

Казахский национальный университет им. аль-Фараби

УДК 004.93'12

На правах рукописи

**ГРИШКО МИХАИЛ ВАЛЕРЬЕВИЧ**

**Построение нейросетевых моделей и алгоритмов  
для решения задач с неполной информацией**

6D070300 – Информационные системы

Диссертация на соискание степени  
доктора философии (PhD)

Научный консультант  
доктор физико-математических наук,  
профессор  
А.Е. Дюсембаев

Республика Казахстан  
Алматы, 2022

## СОДЕРЖАНИЕ

<b>ВВЕДЕНИЕ</b> .....	3
<b>1 ОБЗОР ЛИТЕРАТУРЫ</b> .....	7
1.1 Алгебраический подход к решению задач распознавания и классификации (подход Ю.И. Журавлева).....	7
1.2 Обзор методов оценки риска портфеля ценных бумаг.....	16
1.3 Принципы функционирования модели самоорганизующихся карт Байеса.....	23
<b>2 ПОСТРОЕНИЕ ПРОСТРАНСТВЕННОЙ МОДЕЛИ НЕЙРОСЕТИ ДЛЯ ЗАДАЧ РАСПОЗНАВАНИЯ И КЛАССИФИКАЦИИ</b> .....	26
2.1 Постановка задачи распознавания и классификации.....	26
2.2 Построение модели 3-х уровневой пространственной нейросети ( $\mu$ -блок) для модели распознавания $\mathfrak{M}$ .....	29
2.3 Построение корректного алгоритма для $\Omega$ -регулярных задач.....	36
2.4 Построение пространственной нейросети для $\Omega$ -регулярных задач и её особенности.....	40
<b>3 РАЗРАБОТКА МОДЕЛИ БАЙЕСОВОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ЗАДАЧИ ОЦЕНКИ РИСКОВ ПОРТФЕЛЯ ЦЕННЫХ БУМАГ</b> .....	45
3.1 Модель UBSOM оценки и прогнозирования риска портфеля ценных бумаг на основе синтеза многомерного совместного распределения портфельных компонент.....	45
3.2 Вычислительные эксперименты и анализ результатов.....	51
3.2.1 Применение UBSOM к данным по финансовым инструментам фондового рынка США.....	51
3.2.2 Применение UBSOM к данным по финансовым инструментам фондового рынка РФ.....	56
<b>ЗАКЛЮЧЕНИЕ</b> .....	58
<b>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ</b> .....	59

## ВВЕДЕНИЕ

**Актуальность темы диссертационного исследования.** Использование нейронных сетей является одним из основных инструментов современных исследований в области искусственного интеллекта. В этой связи представляет значительный научный интерес построение и исследование новых типов нейронных сетей, а также изучение условий их применимости.

Как известно, актуальным направлением математической теории распознавания образов [1, 2] является операторный подход, предложенный Ю.И. Журавлевым, позволяющий в целом ряде случаев строить корректные алгоритмы [3-10] для задач распознавания и классификации. Среди других подходов к решению задач распознавания получил распространение подход на основе нейросетей, играющий важную роль и в прикладных исследованиях [1, р. 356; 11-15]. И одной из центральных проблем при синтезе нейронных сетей для решения практически любого класса задач, в том числе, задач распознавания и классификации, является проблема повышения точности решения сетью исходной задачи.

Стремительное развитие и усложнение финансовых технологий в течение последнего десятилетия, привлечение широких слоев экономически активного населения Казахстана к сложным финансовым продуктам, таким как инвестирование в финансовые активы вводит в прикладные исследования в области финансовых технологий (fintech) в круг наиболее актуальных прикладных задач современности.

**Цель диссертационной работы.** Построение и исследование нейросетевых моделей и алгоритмов для решения задач с неполной информацией, в частности задач распознавания и классификации, а также задач восстановления зависимостей на основе эмпирических данных.

**Задачи исследования, реализующие цель диссертационной работы:**

- построение модели нейросети для решения задач распознавания с бинарной информацией, воспроизводящей вычисления корректного алгоритма согласно операторной теории Ю.И. Журавлева;
- разработка и исследование подхода к оценке риска портфеля ценных бумаг, основанного на использовании самоорганизующихся карт Байеса.

**Объект исследования.** Операторы вычисления оценок для задач распознавания и классификации, модели оценки риска портфеля ценных бумаг.

**Предмет исследования:**

- способы построения корректного алгоритма и нейросети, воспроизводящей вычисления, осуществляемые корректным алгоритмом;
- модели самоорганизующихся карт Байеса, а также их применения для задачи оценки риска портфеля ценных бумаг.

**Научная новизна.** В работе впервые построена модель нейронной сети, воспроизводящая вычисления корректного алгоритма (по Журавлеву), определены условия корректности задач, для которых может быть построен корректный алгоритм.

Впервые применены самоорганизующиеся карты Байеса для задачи оценки риска портфелей ценных бумаг.

**Теоретическая значимость.** В диссертации определены условия корректности алгебры распознающих алгоритмов с  $\mu$ -операторами вычисления оценок над множеством задач распознавания с бинарной информацией. И эти условия, являясь достаточными условиями корректности, сформулированы в виде ограничений на множество задач распознавания ( $\Omega$ -регулярные задачи), для которых может быть построен корректный алгоритм. Причём каждому оператору исходного семейства соответствует  $\mu$ -блок – пространственная, трёхуровневая, многослойная нейронная сеть, воспроизводящая вычисления, осуществляемые алгоритмом (оператором) исходного семейства.

Более того, учитывая данный результат, для любой  $\Omega$ -регулярной задачи удастся построить 6-уровневую пространственную, многослойную нейросеть, воспроизводящую вычисления, осуществляемые корректным алгоритмом. Одной из особенностей построенной сети является использование диагональных активационных функций, что заметно упрощает вычисления во внутреннем внешнем циклах нейросети.

Кроме того важно, что для каждого контрольного объекта внутренний цикл  $\mu$ -блока состоит из одной итерации, связанной с объектом обучающей выборки с ненулевым весом. Последнее, при подсчете оценки контрольного объекта за класс, позволяет избежать перебора по объектам обучающей выборки, что также дает значительную экономию в вычислениях в сети. При этом здесь возникает естественный вопрос об ослаблении вышеупомянутых условий корректности. Решению этой задачи посвящена работа Дюсембаев А., Гришко М. «Об условиях корректности для алгебры распознающих алгоритмов с исходным множеством  $\mu$ -операторов над множеством задач с бинарной информацией».

В работе введено понятие  $\Omega$ -слабо регулярной задачи, исходя из которого и на основе операторной теории Журавлева с учётом нейросетевой парадигмы и ранее полученных результатов даётся вариант ответа на поставленный вопрос. Интересным аспектом, связанным с полученным результатом, является возможность построения для рассматриваемых задач шестиуровневой пространственной нейросети, каждый  $\mu$ -блок которой является сетью прямого распространения. Таким образом, в работе показана возможность ослабления условий корректности для алгебры распознающих алгоритмов с  $\mu$ -операторами вычисления оценок, что обосновывает повышения эффективности построенной нейросети.

**Практическая значимость работы.** Вместе с тем, касаясь практического аспекта применения нейросетей для решения прикладной задачи, в диссертации рассмотрен подход по использованию байесовских самоорганизующихся карт для более точной оценки рисков портфеля ценных бумаг (метод UBSOM). Этот метод также может помочь преодолеть предположение о нормальности классической модели Марковица и заменить

гауссово распределение смесью гауссианов, которая лучше соответствует реальным распределениям портфеля.

Разработанный подход основан на построении распределения портфеля из совместного распределения компонентов портфеля, которое лучше отражает внутреннюю структуру взаимодействия компонентов в портфеле, чем методы, основанные только на данных о стоимости портфеля.

Как было показано, краткосрочное прогнозирование риска портфеля ценных бумаг не представляет сложностей – простейшие методы оценки риска дают демонстрируют приемлемые результаты ввиду того, что фактическое распределение отклонений стоимости портфеля не может существенно измениться за короткий период времени. Но с увеличением периода прогнозирования точность большинства методов снижается. При этом для долгосрочного прогнозирования разработанная модель UBSOM демонстрирует в численных экспериментах наилучшую точность среди сравнимых методов.

Это связано с тем, что разработанный подход более точно отражает природу распределения приростов стоимости портфеля, по сравнению с альтернативными параметрическими и непараметрическими методами. Суть в том, что мы аппроксимируем распределение в многомерном пространстве, в то время как прямая оценка приростов стоимостей портфеля дает одномерное распределение. Преобразование векторов приростов стоимости финансовых инструментов в одномерную величину прироста стоимости портфеля приводит к потере детальной информации о многомерном распределении, что оказывает влияние на действительный риск портфеля.

Преимущества модели UBSOM были подтверждены результатами вычислительных экспериментов, проведенных на основе данных о стоимости ценных бумаг на фондовых рынках РФ и США.

**Публикации.** Результаты, изложенные в диссертационной работе, обсуждались на научных семинарах кафедры информационных систем КазНУ им. аль-Фараби и Института информационных и вычислительных технологий, докладывались на международных конференциях:

По материалам диссертации соискателем были опубликованы 10 научных работ, в том числе: 1 – в научных изданиях, рекомендуемых ККСОН МОН РК, 4 – в международных научных изданиях, входящих в базу данных Scopus и Web of Science (Clarivate Analytics), 5 – в материалах международных научно-практических конференций.

**Объем и структура работы.** Диссертационная работа состоит из введения, 3 разделов, заключения и списка использованной литературы. Общий объем диссертации 64 страниц, 11 рисунков, 1 таблицы. Список использованной литературы состоит из 65 источников. Во введении обосновывается актуальность выбранной темы диссертационной работы, цели и задачи исследования, краткий обзор полученных результатов, их научная новизна и значимость.

**В первом разделе** излагается современное состояние алгебраического подхода к решению задач распознавания и классификации, излагаются принципы теории Ю.И. Журавлева и теории алгоритмов вычисления оценок. Также в данной главе приводится обзор методов оценки риска портфеля ценных бумаг и принципы функционирования модели самоорганизующихся карт Байеса.

**Второй раздел** посвящен вопросу построения пространственной модели нейросети для задач распознавания и классификации. Здесь подробно разбирается постановка задачи распознавания и классификации, рассматривается поэтапное построение пространственной модели нейросети для задач распознавания и классификации, в частности рассматривается Построение модели 3-х уровневой пространственной нейросети ( $\mu$ -блок) для модели распознавания  $\mathcal{M}$ . В разделе вводится понятие  $\Omega$ -регулярной задачи, и если задача распознавания с бинарной информацией [2, с. 12] является  $\Omega$ -регулярной, то показывается, что для нее может быть построен корректный алгоритм [4, с. 38; 5, с. 729], а также шестиуровневая пространственная нейросеть (далее – сеть), воспроизводящая вычисления, осуществляемые корректным алгоритмом [16, 17].

В работе приводится обоснование такого построения. Предлагаемый подход по построению нейросети не связан с традиционным подходом минимизации функционала [1, р. 50], а основан на операторной теории [2, с. 19; 3, с. 761; 4, с. 37; 5, с. 732; 15, с. 96], разработанной для решения задач распознавания и классификации. Таким образом, цель работы – построение и обоснование нейросетей для решения задач распознавания с бинарной информацией, при этом последние, как известно, образуют свой подкласс в метрическом пространстве задач распознавания  $U$  [2, с. 13; 6, с. 116].

**В третьем разделе** в качестве прикладного исследования в диссертации (раздел 3) предложен новый подход к оценке риска портфеля ценных бумаг, основанный на использовании самоорганизующихся карт Байеса (BSOM). В рамках предложенного подхода построена модель UBSOM, использующая BSOM для аппроксимации Гауссовой смесью многомерного распределения векторов значений компонентов портфеля. В работе, путем проведения численных экспериментов на данных о стоимости ценных бумаг на фондовом рынке США и РФ, сравнивается точность оценки риска портфеля ценных бумаг моделью UBSOM с классической оценкой по Марковицу и более поздней моделью ядерной оценки плотности (KDE).

**В Заключении** представляются основные результаты и выводы диссертационной работы.

# 1 ОБЗОР ЛИТЕРАТУРЫ

## 1.1 Алгебраический подход к решению задач распознавания и классификации (подход Ю.И. Журавлева)

В данной части диссертации (Обзор литературы) будем опираться на книги Журавлева Ю.И. и Дюсембаева А.Е. не оговаривая специально текстуальные совпадения фрагментов с [10, с. 12]. Содержание данного раздела не является результатами диссертации, полученными автором.

С середины 70-х годов актуальным направлением в математической теории распознавания является алгебраическая теория распознающих алгоритмов. Необходимость разработки и создания такой теории, как это можно было заметить, была обусловлена, с одной стороны, отсутствием формального аппарата для полного описания плохо формализуемых областей естествознания (задачи, в которых требуются математические методы решения), а с другой стороны, негарантированной точностью решения данных задач алгоритмами исходного семейства алгоритмов. Основы такой общей алгебраической теории были заложены в ряде работ Ю.И. Журавлева и в дальнейшем получили развитие в работах [18-20].

Подход, предложенный Ю.И. Журавлевым, являлся принципиально новым, поскольку позволял строить корректные (теоретически точные) алгоритмы, не изменяя начальной задачи, т.е. задачи построения алгоритма, дающего точное решение задачи распознавания. До этого как при решении прикладных задач, так и в теоретических исследованиях в основном сужали постановку задачи (выбор конкретного функционала качества), и задача синтеза корректного алгоритма подменялась некоторой экстремальной задачей, решение которой в лучшем случае давало  $\varepsilon$ -оптимальный алгоритм. Например, в [21, 22] для уравнения  $Af(t) = F(x)$ ,  $F(x) \in \{F(x, \alpha)\}$ , задачи интерпретации результатов косвенных экспериментов минимизация функционала

$$I(\alpha) = \int (y - Af(t, \alpha))^2 P(y/x) P(x) dy dx$$

при условии, что задана обучающая информация, по существу означает минимизацию функционала:

$$\int (y - F(x, \alpha))^2 P(y, x) dy dx \rightarrow \min_{\alpha \in \Lambda} \quad (2.1.1)$$

где  $P(y, x)$  – совместная плотность распределения пар  $(y, x)$ , в которой  $x$  – объект, а  $y$  – величина, значения которой определяется числом исходных классов, например,  $y \in \{0, 1\}$ .

Функция  $F(x, \alpha)$  – элемент множества  $\{F(x, \alpha)\}$  решающих правил, каждое  $\alpha$  (набор параметров) определяет конкретную функцию  $F(x, \alpha) \in \{F(x, \alpha)\}$ . Множество значений функции  $F(x, \alpha)$  то же, что и множество значений величины  $y$ .

Для задачи распознавания плотность  $P(y, x)$  предполагается неизвестной, но считается, что задана случайная и независимая выборка

$$(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m).$$

Задача заключается в минимизации функционала (2.1.1) на множестве  $\{F(x, \alpha)\}$ . В общем случае ввиду недостаточно ясной природы функционала (2.1.1) минимизируют функционал эмпирического риска, т.е.

$$I_0(\alpha) = \frac{1}{m} \sum_{i=1}^m (y_i - F(x_i, \alpha))^2 \rightarrow \min_{\alpha} \quad (2.1.2)$$

Здесь важно, чтобы были выполнены условия, при которых оптимальное решение задачи (2.2) является приемлемым в смысле заданной точности и надежности решением задачи (2.1), а именно в смысле оценки

$$P\{\sup | I(\alpha) - I_0(\alpha) | \geq \varepsilon \} < \eta, \quad \varepsilon > 0, \quad \eta \in (0,1)$$

На практике проверка выполнимости этих условий представляет известные затруднения, что побуждает развивать альтернативные подходы для решения задачи распознавания.

Возвращаясь к алгебраической теории, следует заметить, что благодаря тому, что для синтеза корректных алгоритмов требуются простые операции над матрицами, такие алгоритмы могут быть смоделированы на ЭВМ, а аналитический вид корректных алгоритмов удобен в теоретических исследованиях как при построении естественной модели параллельных вычислений, как и при синтезе новых моделей систем аналитических вычислений.

Далее рассмотрим метод опорных подмножеств, применяемый к задачам распознавания и классификации (в литературе класс соответствующих алгоритмов носит название алгоритмы вычисления оценок).

Алгоритмы вычисления оценок (АВО) были предложены академиком РАН Ю.И. Журавлевым в начале 70х годов прошлого века. В их описании были отражены передовые концепции решения задач распознавания. Ниже перечислены некоторые принципы, использованные в модели алгоритмов вычисления оценок [23]:

1. Решение о классификации объекта принимается с помощью анализа оценок близости объекта к классам. За какой класс оценка близости выше – к тому классу и относят объект. Оценки вычисляет распознающий оператор. Классифицирует объекты на основе оценок их близостей к классам решающее правило.

2. При вычислении оценок близости к классам учитывают близость/дальность объекта к эталонным объектам. Близость – схожесть описаний, малое расстояние между значениями признаков. При этом оценка близости объекта к



классу тем выше, чем ближе он к эталонным объектам данного класса и дальше от эталонных объектов других классов.

3. Близость распознаваемого объекта  $S$  к эталонному  $S^t$  определяется на основе расстояний  $\rho_i(a_i(S), a_i(S^t))$ ,  $i = 1, 2, \dots, n$  и формализуется понятием функция близости.

Каждый объект  $x \in X$  – это точка  $n$ -мерного пространства  $R^n$ , при этом координаты  $1, \dots, n$  называются признаками. Пусть множество  $X$  также есть декартово произведение множеств  $E_1, E_2, \dots, E_n$ , где  $E_i$  является областью изменения  $i$ -го признака.

В качестве  $E_i$  могут быть взяты, например, множества:

- $\{0, 1\}$ , состоящее из двух элементов, 0 и 1;
- $\{b_0, b_1, \dots, b_k\}$ , состоящее из  $k + 1$  целых чисел, например  $\{0, 1, \dots, k\}$ ;
- $\Delta$ -промежутки:  $[a_0, a_1], [a_0, a_1], (a_0, a_1), (a_0, a_1)$ , где  $a_0, a_1 \in R_1$ .

Существуют и другие варианты определения  $E_i$ ; обычно предполагается, что  $E_i$  является метрическим или полуметрическим пространством.

При описании объектов в прикладных задачах распознавания признаки имеют, как правило, различную природу. Впервые метод опорных подмножеств был описан Ю.И. Журавлевым и впоследствии, по мере развития этого метода, применялся для решения многих прикладных задач.

Применение данного метода состоит прежде всего в определении  $\omega$ -части объекта  $x$  и в оценке близости  $\omega$ -частей пары объектов  $x', x'' \in R^n$ , один из которых – объект, подлежащий распознаванию, а другой – объект обучения, а затем в вычислении оценок близости объекта  $x$  к классу  $K_j$ ,  $j = 1, 2, \dots, \ell$ . При этом если  $\omega = (i_1, i_2, \dots, i_k) \subset \{1, 2, \dots, n\}$ , а объект  $x = (x_1, x_2, \dots, x_n)$ , то  $\omega$ -часть объекта  $x$  обозначается через  $\omega x = (x_{i_1}, x_{i_2}, \dots, x_{i_k})$ .

Итак, пусть  $Z = \langle I_0, X^q \rangle$  – задача. Рассмотрим следующий класс алгоритмов вычисления оценок:  $\{\tilde{A}(\Omega, \tau_0, \tau_1, \bar{\varepsilon}, \bar{\gamma}, \bar{p})\}$ .

Шаг 1. Фиксируем множество  $\Omega = \{\omega\}$ , каждое  $\omega \in \Omega$  есть подмножество множества признаков  $\{1, 2, \dots, n\}$ , заметим, что существует ряд специальных методов определения системы опорных множеств  $\Omega$  по исходной информации о задаче  $Z$ . В частности, в качестве  $\omega \in \Omega$  могут быть взяты тупиковые тесты матрицы обучения, подмножества множества  $\{1, 2, \dots, n\}$  мощности  $k$  и др. Здесь же, как правило, задается и способ определения веса  $p(\omega)$  опорного множества  $\omega$ . Пусть  $\omega = (i_1, i_2, \dots, i_t)$ ,  $i_j \in \{1, 2, \dots, n\}$ ,  $t \leq n$ , а  $p_1, p_2, \dots, p_n$  – веса признаков  $1, 2, \dots, n$  соответственно,  $p_i \geq 0$ ,  $i = \overline{1, n}$ . Обычно вес опорного множества определяется так:  $p(\omega) = p_{i_1} + p_{i_2} + \dots + p_{i_t}$ .

Шаг 2. Определение функции близости  $F_\varepsilon(\omega x', \omega x'')$  объектов  $x', x''$  по опорному множеству  $\omega$ . Пусть:

$$x' = (x'_1, x'_2, \dots, x'_n), x'' = (x''_1, x''_2, \dots, x''_n),$$

$$\varepsilon = \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, \varepsilon_i \geq 0;$$

$$\omega = (i_1, i_2, \dots, i_t), \omega x' = (x'_{i_1}, x'_{i_2}, \dots, x'_{i_t})$$

$$\omega x'' = (x''_{i_1}, x''_{i_2}, \dots, x''_{i_t}).$$

Пусть  $p_{ik}(x'_{ik}, x''_{ik})$  – функция, оценивающая близость между компонентами  $x'_{ik}, x''_{ik}$  объектов  $x', x''$ .

Тогда  $F_\varepsilon(\omega x', \omega x'')$  определяется, например, так:

$$F_\varepsilon(\omega x', \omega x'') = \begin{cases} 1, & \text{если } p_{i1}(x'_{i1}, x''_{i1}) \leq \varepsilon_{i1} \& p_{i2}(x'_{i2}, x''_{i2}) \leq \varepsilon_{i2} \& \dots \& p_{it}(x'_{it}, x''_{it}) \leq \varepsilon_{it} \\ 0, & \text{если хотя бы одно из неравенств нарушено} \end{cases}$$

Заметим, что существуют различные способы определения функции близости в зависимости от характера решаемых задач, при этом часто предполагается, что  $p_i$  – метрика или полуметрика, заданная на  $E_i \times E_i$ .

Шаг 3. Вычисление оценки  $\Gamma_i(x_i)$  объекта  $x_i \in X^q$  за класс  $K_j$ ,  $j = 1, 2, \dots, \ell$ . Итак, пусть  $X^q = \{x_1, x_2, \dots, x_q\}$ , а  $X^m = \{x'_1, x'_2, \dots, x'_m\}$  – список объектов обучения. Будем считать, что каждому объекту обучения  $x_j \in X^m$  приписан вес  $\gamma_j = \gamma(x_j) > 0$ ,  $j = 1, 2, \dots, m$ , тогда определим:

$$\Gamma_j^1(x_i) = \sum_{x \in \tilde{K}_j} \gamma(x) \cdot \sum_{\omega \in \Omega} p(\omega) \cdot F_\varepsilon(\omega x, \omega x_i), \quad (2.1.3)$$

$$\Gamma_j^0(x_i) = \sum_{x \in C\tilde{K}_j} \gamma(x) \cdot \sum_{\omega \in \Omega} p(\omega) \cdot \bar{F}_\varepsilon(\omega x, \omega x_i) \quad (2.1.4)$$

Окончательно

$$\Gamma_{ij} = \Gamma_j(x_i) = \tau_0 \cdot \Gamma_j^0(x_i) + \tau_1 \cdot \Gamma_j^1(x_i) \quad (2.1.5)$$

где  $\tau_0, \tau_1 \in \{0, 1\}$ ,

$$\tilde{K}_j = K_j \cap X^m, j = 1, 2, \dots, \ell$$

$$C\tilde{K}_j = X^m \setminus K_j,$$

$$\bar{F}_\varepsilon(\omega x', \omega x'') = 1 - F_\varepsilon(\omega x', \omega x'').$$

Как отмечено в [2, с. 21], оценку объекта  $x_i$  за класс  $K_j$  можно трактовать как значение функции принадлежности по L.Zadeh. В нашем случае для этого предварительно требуется осуществить нормировку при вычислении оценки  $\Gamma_{ij}$ .

Итак, оператор  $B$ , задаваемый формулами (2.1.3)-(2.1.5), будучи примененным к задаче  $Z = \langle I_0, X^q \rangle$  вычисляет матрицу  $\|\Gamma_{ij}\|_{q \times \ell}$ , что формально записывается так:

$$B(Z) = \|\Gamma_{ij}\|_{q \times \ell}.$$

В свою очередь решающее правило  $C^*$  по матрице  $\|\Gamma_{ij}\|$  вычисляет матрицу классификации  $\|\beta_{ij}\|_{q \times \ell}$ . Заметим, что именно подкласс  $\{\tilde{A}\}$  использовался впервые для доказательства корректности алгебры  $\mathfrak{A}\{\tilde{A}\}$

относительно множества задач распознавания при условии, что в качестве  $\Omega$  взяты  $\Omega_0 = \{\{1\}, \{2\}, \dots, \{n\}\}$  и  $\Omega_1 = \{1, 2, \dots, n\}$  или  $\Omega = \{\Omega_0 \cap \Omega_1\}$ .

При решении прикладных задач с непересекающимися классами получило распространение использование решающего правила иного типа:

Пусть  $\Gamma_{ij_0} = \max \Gamma_{ij}$  ( $j = 1, 2, \dots$ ), а  $d_1, d_2$  – константы, параметры решающего правила; тогда если:

$$\Gamma_{ij_0} - \Gamma_{ip} > d_1, (p \neq j_0) \quad (2.1.6)$$

а также

$$\Gamma_{ij_0} / \sum_{p \neq j_0} \Gamma_{ip} > d_2 \quad (2.1.7)$$

то объект  $x_i$  заносится в класс  $K_{j_0}$  ( $\beta_{ij_0} = 1$ ) и соответственно  $x_i \notin K_j$  ( $\beta_{ip} = 0$ ),  $p \neq j_0$ . В противном случае, т.е. если хотя бы одно из неравенств нарушено, считается, что вхождение объекта  $x_i$  в любой класс  $K_j$  не определено ( $\beta_{ij} = \Delta$ ).

Таким образом определяется модель алгоритмов, порождаемая параметрами  $\Omega, \tau_0, \tau_1, \bar{\gamma}, \bar{p}, \bar{\varepsilon}, d_1, d_2$  и формулами (2.1.3)-(2.1.5):

$$\mathfrak{M} = \{\mathfrak{M}(\Omega, \tau_0, \tau_1, \bar{\gamma}, \bar{p}, \bar{\varepsilon}, d_1, d_2)\},$$

где часто вместо  $\Omega$  указывается натуральное число  $k$ , показывающее длину опорного множества  $\omega \in \Omega$ . Каждому набору параметров модели соответствует алгоритм распознавания. Как уже отмечалось, при решении задачи распознавания  $Z$  желательно так выбрать параметры алгоритма, чтобы максимально точно разбить множество обучающих объектов на классы. При решении прикладных задач, как правило, используется функционал – доля правильно классифицированных объектов обучения, иначе говоря, если  $A \in \mathfrak{M}$  – алгоритм, то формально качество алгоритма  $A$  на задаче  $Z$  оценивается функционалом  $\varphi_A = m_A / m$ , и требуется определить алгоритм  $A^*$ , доставляющий максимум функционала  $\varphi_A$ :

$$\max_{k, \bar{\gamma}, \bar{p}, \bar{\varepsilon}, d_1, d_2} \varphi_A, A \in \mathfrak{M} \quad (2.1.8)$$

Известно, что вариации параметров  $d_1, d_2$  не оказывают существенного влияния на результат оптимизации в (2.1.8), т.е. вместо (2.1.8) фактически необходимо решать задачу:

$$\max_{k, \bar{\gamma}, \bar{p}, \bar{\varepsilon}} \varphi_A, A \in \mathfrak{M}.$$

При заданных  $k$  и  $\varepsilon$  задача определения весов  $\gamma_1, \gamma_2, \dots, \gamma_m; p_1, p_2, \dots, p_n$  в задаче (2.1.8) сводится с использованием системы вида (2.1.6), (2.1.7) к задаче

выделения максимально совместной подсистемы в системе линейных неравенств, образованных системами типа (2.1.6), (2.1.7). Величина  $k$  определяется, как правило, экспериментально, а параметры  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  определяются в зависимости от границы изменения признаков.

Здесь следует заметить, что в ряде случаев при построении алгоритмов вычисления оценок используются функции близости несколько более общего типа, а именно функция близости  $B_{\varepsilon}^{\varepsilon}(\omega x', \omega x'')$ :

$$B_{\varepsilon}^{\varepsilon}(\omega x', \omega x'') = \begin{cases} 1, \text{ если число невыполненных неравенств} \\ \text{среди } p_{ij}(x'_{ij}, x''_{ij}) \leq \varepsilon_{ij} \text{ не больше } \varepsilon, j = 1, 2, \dots, t \\ \\ 0 \text{ в противном случае} \end{cases}$$

При  $\varepsilon = 0$  получаем исходную функцию близости  $F_{\varepsilon}(\omega x', \omega x'')$ . Для  $B_0^0(\omega x', \omega x'')$  имеем  $B_0^0(\omega x', \omega x'') = 1$ , если  $\omega x' = \omega x''$ , и  $B_0^0(\omega x', \omega x'') = 0$ , если  $\omega x' \neq \omega x''$ .

Далее кратко рассмотрим вопрос о вычислении выражений вида

$$\Gamma_j(x_i) = \sum_{x \in K_j} \gamma(x) \sum_{\omega \in \Omega} p(\omega) \bar{F}_{\varepsilon}(\omega x_i, \omega x), \quad (2.1.9)$$

определяющих оценку объекта  $x_i$  за класс  $K_j$ . Если в качестве  $\omega \in \Omega$  взять наборы длины  $k$  ( $1 \leq k \leq n$ ), то уже в данном случае число слагаемых в сумме будет равно  $|K_j| C_n^k$ . Поэтому актуальной задачей здесь является задача сокращения вычислений в выражении (2.1.9). С этой целью удобно определить вектор  $d$ ; координаты вектора  $d(x', x'')$ :  $d_{ij} = 1$ , если  $p_j(\omega x'_{ij}, \omega x''_{ij}) \leq \varepsilon_{ij}$ , и  $d_{ij} = 0$  в противном случае.

Через  $\|d(\cdot, \cdot)\|$  обозначим норму – суммарное число единичных координат вектора  $d(\cdot, \cdot)$ . Тогда:

1. Для случая системы  $\Omega$ , состоящей из  $k$  элементарных множеств:

$$\Gamma_j(x_i) = \sum_{x \in K_j} \gamma(x) C_{\|d(x_i, x)\|}^k.$$

2. Для системы  $\Omega$ , состоящей из всех непустых подмножеств множества  $\{1, 2, \dots, n\}$ :

$$\Gamma_j(x_i) = \sum_{x \in K_j} \gamma(x) (2^{\|d(x_i, x)\|} - 1)$$

Нетрудно заметить, что в приведенных формулах исчезло суммирование по  $\omega \in \Omega$ . Для функции близости  $B_{\varepsilon}^{\varepsilon}(\cdot, \cdot)$  общего вида выражения для  $\Gamma_j(x_i)$  имеет более сложную форму. В работе [2, с. 23] описан также иной способ упрощения вычисления величин  $\Gamma_j(x_i)$ , основанный на специальном представлении характеристической функции системы опорных множеств.

Рассмотрим операторную теорию алгоритмов распознавания и классификации для случая классического признакового пространства, что означает, что каждый объект  $x$  есть точка  $n$ -мерного признакового пространства. Приведем ряд необходимых определений.

Итак, пусть множество объектов  $X \subset R^n$  покрывается конечным числом классов:

$$K_1, K_2, \dots, K_{\ell} : X = \bigcup_{j=1}^{\ell} K_j,$$

пересекающихся в общем случае. Рассмотрим далее множество одномерных предикатов:

$$P_1, P_2, \dots, P_{\ell} : P_j(x) \equiv "x \in K_j",$$

где  $x \in X, j = 1, 2, \dots, \ell$ .

Информационной строкой  $\tilde{\alpha}_i$  объекта  $x_i \in X$  по системе предикатов  $P_1, P_2, \dots, P_{\ell}$  является строка  $\tilde{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{i\ell})$ ,  $\alpha_{ij} = P_j(x_i), j = 1, 2, \dots, \ell$ .

Пусть  $X^m = \{x'_1, x'_2, \dots, x'_m\}$  – множество допустимых объектов из  $X$ , для которых информационные строки  $\tilde{\alpha}'_1, \tilde{\alpha}'_2, \dots, \tilde{\alpha}'_m$  по системе предикатов  $P_1, P_2, \dots, P_{\ell}$  известны, а  $X^q = \{x_1, x_2, \dots, x_q\}$  – множество допустимых объектов из  $X$ , для которых информационные строки по системе предикатов  $P_1, P_2, \dots, P_{\ell}$  необходимо вычислить.

Определим задачу распознавания  $Z$  как пару  $\langle I^0, X^q \rangle$ , где  $I^0 = \{x'_1, x'_2, \dots, x'_m; \tilde{\alpha}'_1, \tilde{\alpha}'_2, \dots, \tilde{\alpha}'_m\}$  – начальная информация задачи  $Z$ , а  $X^q$  – набор объектов из  $X$ , информационные строки которых необходимо вычислить.

Информационной матрицей задачи  $Z = \langle I^0, X^q \rangle$  называется матрица  $\|\alpha_{ij}\|_{q \times \ell}$ , где  $\alpha_{ij} = P_j(x_i), i = 1, 2, \dots, q; j = 1, 2, \dots, \ell$ .

Пусть  $\{A(\pi_1, \pi_2, \dots, \pi_k)\}$  – некоторая совокупность алгоритмов распознавания, каждый из которых определяется заданием параметров  $\pi_1, \pi_2, \dots, \pi_k$ . Будем считать, что любой из алгоритмов  $A \in \{A(\pi_1, \pi_2, \dots, \pi_k)\}$ , будучи примененным к задаче  $Z$ , вычисляет некоторую матрицу  $\|\beta_{ij}\|_{q \times \ell}$  (что запишем как  $A(Z) = \|\beta_{ij}\|_{q \times \ell}$ ), не обязательно совпадающую с матрицей  $\|\alpha_{ij}\|_{q \times \ell}$ . Элементы матрицы  $\|\beta_{ij}\|_{q \times \ell} : \beta_{ij} \in \{0, 1, \Delta\}$ , причем  $\beta_{ij} = 1$  трактуется так: алгоритм  $A$  относит объект  $x_i \in X^q$  к классу  $K_j$ ;  $\beta_{ij} = 0$ , если  $A$

не относит объект  $x_i \in X^q$  к классу  $K_j$ , и  $\beta_{ij} = \Delta$  означает, что алгоритм  $A$  не классифицирует объект  $x_i \in X^q$ .

Алгоритм  $A^* \in \{A\}$  называется корректным для задачи  $Z = \langle I^0, X^q \rangle$ , если

$$A^*(Z) = \|\alpha_{ij}\|_{q \times \ell},$$

т.е. алгоритм  $A^*$  по задаче  $Z$  вычисляет информационную матрицу этой задачи.

Далее будем рассматривать такие модели (совокупности)  $\{A\}$  алгоритмов распознавания, что каждый алгоритм  $A \in \{A\}$  представим как последовательное выполнение алгоритмов  $B$  и  $C$ , что записывается как  $A = B \circ C$ . Далее алгоритм  $B$  будем называть оператором, а алгоритм  $C$  – решающим правилом, при этом считаем, что оператор  $B$  по задаче  $Z$  вычисляет вещественную матрицу  $\|\Gamma_{ij}\|_{q \times \ell}$ . Содержательно величина  $\Gamma_{ij}$  – это оценка объекта  $x_i$  за класс  $K_j$ ;  $i = 1, 2, \dots, q$ ;  $j = 1, 2, \dots, \ell$ . Решающее правило  $C$  по матрице  $\|\Gamma_{ij}\|_{q \times \ell}$  вычисляет матрицу  $\|\beta_{ij}\|_{q \times \ell}$ ;  $\beta_{ij} \in \{0, 1, \Delta\}$ . Таким образом, множество алгоритмов  $\{A(\pi_1, \pi_2, \dots, \pi_k)\}$  порождает множество операторов  $\{B(\pi_1, \pi_2, \dots, \pi_{k_1})\}$ ,  $k_1 \leq k$ , и множество решающих правил  $\{C\}$ . Заметим здесь же, что в данную схему укладываются многие из известных типов алгоритмов распознавания и классификации.

Зафиксируем константы  $c_{11}, c_{21}, \dots, c_{1\ell}, c_{2\ell}$  и определим решающее правило  $C^*$  следующим образом:

$$C^*(Z) = \|\beta_{ij}\|_{q \times \ell} = \begin{cases} 1, & \text{если } \Gamma_{ij} > c_{2j} \\ 0, & \text{если } \Gamma_{ij} < c_{1j} \\ \Delta, & \text{если } c_{1j} \leq \Gamma_{ij} \leq c_{2j} \end{cases}.$$

Далее в качестве решающего правила  $C$  будем рассматривать правило именно данного типа.

На множестве операторов  $\{B\}$ , а вследствие этого и во множестве алгоритмов  $\{A(\pi_1, \pi_2, \dots, \pi_k)\}$  вводятся операции сложения, умножения и умножения на скаляр, а именно: пусть  $Z$  – задача, а  $B', B''$  – операторы из  $\{B(\pi_1, \pi_2, \dots, \pi_{k_1})\}$  такие, что  $B'(Z) = \|\Gamma'_{ij}\|_{q \times \ell}$ ,  $B''(Z) = \|\Gamma''_{ij}\|_{q \times \ell}$ ; тогда

$$bB = \|b\Gamma'_{ij}\|_{q \times \ell}, B \in \{B\}, \quad (2.1.10)$$

$$B' + B'' = \|\Gamma'_{ij} + \Gamma''_{ij}\|_{q \times \ell}, \quad (2.1.11)$$

$$B' \times B'' = \|\Gamma'_{ij} \cdot \Gamma''_{ij}\|_{q \times \ell} \quad (2.1.12)$$

Замыкание множества  $\{B\}$  относительно операций (2.1.10), (2.1.11) является линейным замыканием множества  $\{B\}$  и обозначается  $L\{B\}$ .

Алгебраическое замыкание множества операторов  $\{B\}$  относительно операций (2.1.10)-(2.1.12) является операторной алгеброй, порождаемой множеством (моделью) операторов  $\{B\}$  и обозначается через  $\mathfrak{A}\{B\}$ . Очевидно, что  $\mathfrak{A}\{B\}$  состоит из операторных многочленов. Максимальная степень многочлена в  $\mathfrak{A}\{B\}$  есть степень алгебраического замыкания  $\mathfrak{A}\{B\}$  и соответствующего замыкания  $\mathfrak{A}\{A\}$ . Пусть  $\mathfrak{A}^k\{B\}$  – совокупность операторных многочленов степени не выше  $k$ , ясно, что

$$\{B\} = \mathfrak{A}^0\{B\} \subseteq L\{B\} = \mathfrak{A}^1\{B\} \subseteq \dots \subseteq \mathfrak{A}^k\{B\}.$$

Алгебра  $\mathfrak{A}\{A\}$ , порождаемая моделью алгоритмов распознавания  $\{A\}$ , называется корректной над множеством задач  $\{Z\}$ , если для любой задачи  $Z \in \{Z\}$  в  $\mathfrak{A}\{A\}$  существует алгоритм  $A^*$  такой, что  $A^*(Z) = \|\alpha_{ij}\|_{q \times \ell}$ .

Пусть  $Z = \langle I^0, X^q \rangle$  – задача. Если множество матриц  $B(Z)$  ( $B$  пробегает все  $\{B\}$ ) размерности  $q \times \ell$  содержит базис пространства матриц размерности  $q \times \ell$ , то алгебра  $\mathfrak{A}\{B\}$  называется полной относительно задачи  $Z$ .

Операторная алгебра  $\mathfrak{A}^k\{B\}$ ,  $k \geq 1$ , называется квазиполной относительно множества задач  $\{Z\}$ , если для любой задачи  $Z$  из  $\{Z\}$  в  $\mathfrak{A}^k\{B\}$  существуют операторы  $B(i, j)$  ( $1 \leq i \leq q, 1 \leq j \leq \ell$ ) такие, что для матрицы  $B(i, j)(Z) = \|\Gamma_{tp}(i, j)\|_{q \times \ell}$  оператора  $B(i, j)$  верны следующие соотношения:

$$\begin{aligned} \Gamma_{tp}(i, j) &\geq 1, \\ |\Gamma_{tp}(i, j)| &\geq 1, (t, p) \neq (i, j). \end{aligned}$$

Совокупность таких операторов  $\{B(i, j)\}_{1,1}^{q,\ell}$  называется квазибазисом алгебры  $\mathfrak{A}\{B\}$ .

Пусть  $Z$  – задача; отметим, что как условие полноты, так и условие квазиполноты операторной алгебры  $\mathfrak{A}\{B\}$  относительно задачи  $Z = \langle I^0, X^q \rangle$  является достаточным условием корректности алгебры  $\mathfrak{A}\{A\}$  над  $Z$ . Однако доказательство квазиполноты технически проще, чем доказательство полноты, и если квазиполнота алгебры  $\mathfrak{A}\{B\}$  для задачи  $Z$  имеет место, то как показано Ю.И. Журавлевым, корректный алгоритм  $A^*$  для  $Z$  представим в виде;

$$A^* = \left( (c_1 + c_2) \sum_{i=1}^q \sum_{j=1}^{\ell} \alpha_{ij} \cdot B^k(i, j) \right) \circ C^*,$$

где  $\|\alpha_{ij}\|_{q \times \ell}$  – информационная матрица задачи  $Z$ , а

$C^*$  – решающее правило с параметрами  $c_{11}, c_{21}, c_{12}, c_{22}, \dots, c_{1\ell}, c_{2\ell}$ , определенное выше.

Для  $k$  имеет место формула (2.1.13):

$$k = \left[ \frac{\ln q + \ln \ell + |\ln(c_1 + c_2)| - \ln c_1}{|\ln a|} \right] + 1, \quad (2.1.13)$$

где  $c_1 = \min_j c_{1j}$ ,  $c_2 = \max_j c_{2j}$ ,

$$a = \max_{(t,p),(i,j)} |\Gamma_{tp}(i,j)| < 1, (t,p) \neq (i,j), 0 < c_1 < c_2.$$

К настоящему времени найдены условия корректности для различных типов алгоритмов распознавания: кусочно линейных, потенциальных функций, вычисления оценок по опорным подмножествам, типа тестовых, типа «Кора», структурных и т.д.

## 1.2 Обзор методов оценки риска портфеля ценных бумаг

Ключевой вопрос для инвестора - как выбрать портфель инвестиций таким образом, чтобы ожидаемая инвестиционная стоимость в течение некоторого времени была максимизирована, в то время как вероятность того, что инвестиционная стоимость упадет ниже ожидаемой стоимости на некоторую предварительно определенную сумму, будет меньше предварительно заданного значения вероятности.

Работы Марковица [24, 25] заложили основы теории портфеля ценных бумаг и описывают проблему принятия инвестиционных решений в условиях неопределенности. По Марковицу, риск портфеля ценных бумаг соответствует дисперсии стоимости портфеля. Более высокая дисперсия стоимости портфеля означает более высокие возможные отклонения от ожидаемой стоимости портфеля. Таким образом, потенциальное падение стоимости портфеля относительно его ожидаемой стоимости увеличивается, и инвестор должен принять соответственно более высокий риск, если решит инвестировать в этот портфель [26].

Теория Марковица имеет несколько допущений, ключевым из которых является ожидание того, что доходность финансовых инструментов, образующих портфель, представляет собой нормальные случайные величины. В действительности большая часть наблюдаемых доходностей финансовых инструментов не распределена нормально – для их распределений характерны асимметрии, наличие «тяжелых хвостов» и выбросов [27, 28]. Это оказывает существенное влияние на обоснованность предположений о риске. Если распределение стоимости финансовых инструментов не согласуется с предположениями теории Марковица, фактический риск портфеля может быть больше или меньше, чем риск, рассчитанный в соответствии с моделью.

В некоторых недавних исследованиях концепция использования дисперсии была расширена, чтобы включить асимметрию в распределении доходности при выборе портфеля [28, p. 148; 29, 30]. Также существует множество методов корректировки, которые касаются вопроса выбора периода данных для оценки дисперсии, эффектов инфляции и т.д. Например, в [29,



р. 69] описаны девять различных способов вычисления стандартного отклонения. Проблема снижения шума исходных данных с целью увеличить устойчивости и повышения прогнозной способности была рассмотрена в ряде исследований, а также в [31].

Для того, чтобы преодолеть предположение о нормальности было предложено много других альтернатив нормального распределения - например, распределение Парето [32], распределение Стьюдента [33], обобщенное гиперболическое распределение [34], и обобщенное logF распределение [35, 36]. Одним особенно интересным подходом является модель гауссовой смеси, которая показывает хорошие результаты в описании эмпирического распределения инструментов, входящих в фондовые индексы.

Например, в работе Behr и Potter [34, р. 49] эта модель была протестирована на индексе S&P 500 в долгосрочном временном периоде 1871-2005, где он показывает 92%-ю точность описания приростов стоимости индекса. Однако определение параметров компонентов гауссовой смеси не является тривиальной задачей. Существуют исследования, в которых гауссова смесь используется для оценки коэффициента Шарпа, который представляет собой коэффициент, показывающий соотношение между премией за риск и риском отклонения актива [37]. Ruppert [38] предлагает использовать для нахождения параметров метод ядерной оценки плотности. Аппроксимация гауссовой смесью распределения стоимости портфеля также можно построить с использованием алгоритма максимального правдоподобия и максимизации ожидания (EM), но такие методы очень чувствительны к выбросам в данных [39].

Допустим, что инвестор имеет некоторый начальный капитал  $x$ , который может быть распределен между финансовыми инструментами  $A_1, \dots, A_N$ . Цены данных финансовых инструментов в момент  $t = 0$  составляют  $S_0(A_1), \dots, S_0(A_N)$ .

Пусть

$$X_0(b) = b_1 S_0(A_1) + \dots + b_N S_0(A_N), \quad (2.2.1)$$

где  $b_i \geq 0, i = 1, \dots, N$ .

Тогда

$$B(x) = \{b = (b_1, \dots, b_N): b_i \geq 0, X_0(b) = x\}, x > 0 \quad (2.2.2)$$

есть портфель финансовых инструментов, где  $b_i$  – доля финансового инструмента  $A_i$  с ценой  $S_0(A_i)$ .

Если инвестор сформировал портфель  $b = (b_1, \dots, b_N)$ , тогда в момент времени  $t = t_1$  его начальный капитал  $X_0(b) = x$  будет составлять

$$X_1(b) = b_1 S_1(A_1) + \dots + b_N S_1(A_N), \quad (2.2.3)$$

и, как ожидается, получит некоторый прирост при допустимом риске (т.е. принимая возможность снижения стоимости ниже ожидаемых значений).

Сделаем допущение, что стоимость финансовых инструментов подчиняется нормальному распределению в любой момент времени.

Используя данное допущение Макровиц [24, р. 81] рассматривает два параметра стоимости портфеля  $X_1(b)$ :

1.  $EX_1(b)$  – математическое ожидание, которое после вычитания величины начального капитала даст величину прибыли портфеля.

2.  $DX_1(b)$  – дисперсия, при этом соответствующее стандартное отклонение рассматривается как эквивалент риска портфеля.

Если цены финансовых инструментов подчиняются нормальному распределению, то и стоимость портфеля, составленного из данных финансовых инструментов, также подчиняется нормальному распределению и для любого значения вероятности возможно вычислить максимальное ожидаемое значение отклонения стоимости портфеля как произведение стандартного отклонения и стоимости портфеля.

Определим вес каждого финансового инструмента в портфеле  $d = (d_1, \dots, d_N)$  как;

$$d_i = \frac{b_i S_0(A_i)}{x}. \quad (2.2.4)$$

Тогда изменение стоимости портфеля составляет

$$q_P(d) = d_1 q(A_1) + \dots + d_N q(A_N), \quad (2.2.5)$$

где

$$q(A_i) = \frac{\Delta S_1(A_i)}{S_0(A_i)} \quad (2.2.6)$$

Для  $X_1(b)$  мы получим

$$X_1(b) = x(1 + q_P(d)). \quad (2.2.7)$$

Дисперсия  $q_P(d)$  имеет следующий вид

$$Dq_P(d) = \sum_{i=1}^N d_i^2 Dq(A_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^N d_i d_j V_{ij} \quad (2.2.8)$$

где  $V_{ij}$  – ковариация между ценами финансовых инструментов  $A_i$  и  $A_j$ ,

$Dq(A_i)$  – дисперсия цен  $A_i$ .

Это показывает, что в соответствии с подходом Марковица риск портфеля зависит от дисперсий портфельных компонент и ковариации между значениями цен данных финансовых инструментов.

Значительное усиление результатов Марковица достигается за счет внедрения в модель асимметрии распределения [28, р. 148; 29, р. 206; 30, р. 109]. Основные положения данной модели могут быть сформулированы в соответствии с [30, р. 109] следующим образом:

$$\begin{aligned} \text{Mean} &= d^T \tilde{Q}, \\ \text{Variance} &= d^T V d, \\ \text{Skewness} &= E[d^T \hat{Q}]^3, \end{aligned} \quad (2.2.9)$$

где  $d^T = (d_1, d_2, \dots, d_N)$  – транспонированный вектор  $d$  портфельных компонент, где  $d_i$  – доля капитала, инвестированная в  $i$ -й рисковый финансовый инструмент;

$Q = (q(A_1), \dots, q(A_N))$  – вектор приростов стоимости финансовых инструментов;

$\bar{r}$  – средняя величина прироста стоимости безрисковых активов, таких как банковский депозит или государственная ценная бумага;

$\tilde{Q} = (q(A_1) - \bar{r}, \dots, q(A_N) - \bar{r})$  –  $N$ -мерный вектор ожидаемых превышений приростов стоимости;

$$V = \begin{pmatrix} \text{Cov}(q(A_1), q(A_1)) & \dots & \text{Cov}(q(A_1), q(A_N)) \\ \dots & \dots & \dots \\ \text{Cov}(q(A_N), q(A_1)) & \dots & \text{Cov}(q(A_N), q(A_N)) \end{pmatrix} \quad \text{– ковариационная}$$

матрица для  $Q$ ;

$\bar{Q} = \text{avg}(q(A_1), \dots, q(A_N))$  – среднее значение приростов стоимости финансовых инструментов;

$\hat{Q} = (q(A_1) - \bar{Q}, \dots, q(A_N) - \bar{Q})$  – вектор, содержащий средние приросты стоимости по каждому финансовому инструменту.

Необходимо отметить, что данная модель применима при допущении, что существуют  $N$  финансовых инструментов и  $(N + 1)$ -й финансовый инструмент является безрисковым.

Таким образом задача выбора портфеля трансформируется в задачу многокритериальной оптимизации, которую можно записать как

$$\begin{aligned} &\max\{d^T \tilde{Q}\} \\ &\min\{E[d^T \hat{Q}]^3\} \end{aligned} \quad (2.2.10)$$

при условии  $d^T V d = 1$ .

Имеется несколько более поздних альтернатив, преодолевающих допущение о нормальности распределения цен финансовых инструментов. Один из вариантов заключается в рассмотрении обобщенного гиперболического распределения с вектором параметров  $\Psi = (\alpha, \beta, \delta, \lambda, \mu)$ , функция плотности распределения которой имеет следующий вид [32, р. 334]:

$$f_{GH}(x; \Psi) = k\{\delta^2 + (x - \mu)^2\}^{\frac{1}{2}(\lambda - \frac{1}{2})} K_{\lambda - \frac{1}{2}}\left(\alpha\sqrt{\delta^2 + (x - \mu)^2}\right) e^{\beta(x - \mu)}$$

$$0 \leq |\beta| < \alpha, \delta > 0,$$

$$\text{где } k = \frac{(\alpha^2 - \beta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi}\alpha^{\lambda - \frac{1}{2}}\delta^{\lambda}K_{\lambda}(\delta\sqrt{\alpha^2 - \beta^2})}$$

$$K_{\lambda}(t) = \frac{1}{2} \int_0^{\infty} u^{\lambda-1} e^{-\frac{1}{2}t(u+u^{-1})} du, t > 0.$$

Другой альтернативой является использование обобщенного логистического распределения [34, р. 49; 35, р. 414; 39, р. 340] с вектором параметров  $\Psi = (a, b, \mu, \sigma)$ :

$$f_{LF}(x; \Psi) = \frac{a^a b^b}{B(a, b)} e^{a \frac{x-\mu}{\sigma}} \left(b + a e^{\frac{x-\mu}{\sigma}}\right)^{-(a+b)},$$

где  $a, b > 0$  и  $B(a, b)$  имеют следующий вид:

$$B(a, b) = B(b, a) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

где  $\Gamma(x)$  обозначает гамма функцию.

Следующий вариант заключается в использовании - распределения Стьюдента в качестве функции плотности вероятности, которое было обобщено Хансеном для учета эксцесса и асимметрии [33, р. 340]:

$$f(z; v, \lambda) = \begin{cases} bc \left(1 + \frac{1}{v-2} \left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(v+1)/2}, & \text{if } z < -a/b \\ bc \left(1 + \frac{1}{v-2} \left(\frac{bz+a}{1+\lambda}\right)^2\right)^{-(v+1)/2}, & \text{if } z \geq -a/b \end{cases}$$

где  $2 < v < \infty$  обозначает степень свободы параметра;

$-1 < \lambda < 1$  – параметр асимметрии;

а параметры  $a, b, c$  имеют следующие ограничения:

$$a = 4\lambda c \left(\frac{v-2}{v-1}\right),$$

$$b = \sqrt{1 + 3\lambda^2 - a^2},$$

$$c = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi(v-2)\Gamma\left(\frac{v}{2}\right)}}.$$

Также имеется ряд непараметрических подходов для оценки плотности вероятности, среди которых необходимо отметить метод ядерной оценки

плотности (Kernel Density Estimation, KDE) [40].

Пусть  $X_1, \dots, X_n$  выборка из генеральной совокупности  $X$ , где  $n$  – объем выборки. Тогда KDE имеет следующий вид:

$$f(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right),$$

где  $h$  – ширина полосы;

$k$  – ядерная функция, которая должна удовлетворять условиям:

$$\begin{aligned} \int_{-\infty}^{\infty} |k(y)| dy &< +\infty; \\ \lim_{n \rightarrow \infty} yk(y) &= 0; \\ k(y) &= k(-y); \\ \int_{-\infty}^{\infty} k(y) dy &= 1. \end{aligned}$$

Существует множество типов ядерных функций, таких как однородная, треугольная, Епанечникова, биквадратное, триквадратное, Гауссово, логистическое и другие [40, р. 169]. Метод KDE очень чувствителен к параметру  $h$  - большое значение  $h$  влечет слишком сглаженную оценку, в то же время малая величина  $h$  приводит к адаптации к имеющемуся в данных «шуму». Имеется несколько методов выбора оптимальной ширины  $h$ , большая часть из которых основана на минимизации среднеквадратической ошибки:

$$MSE = E \left[ \sum_{i=1}^n (f(\hat{x}_i) - f(x_i))^2 \right]$$

или интегральной среднеквадратической ошибки (для непрерывных распределений)

$$MISE = E \left[ \int_x (f(\hat{x}) - f(x))^2 dx \right].$$

Как известно, нейронные сети имеют широкий перечень применений в области финансов, и, в частности, в портфельной теории. Имеются десятки исследований, которые исследуют применение нейронных сетей для прогнозирования цен финансовых инструментов [41, 42]. Задача выбора и

оптимизации портфеля также была рассмотрена некоторыми исследователями [43, 44]. Например, в [43, р. 1178] задача выбора портфеля, определенная в (2.2.10) сформулирована следующим образом:

$$\min \left\{ \lambda \left[ \sum_{i=1}^N \sum_{j=1}^N x_i V_{ij} x_j \right] + (1 - \lambda) \left[ - \sum_{i=1}^N \mu_i x_i \right] \right\}$$

при условии

$$\sum_{i=1}^N x_i = 1, \quad 0 \leq x_i \leq 1, \quad i = 1, \dots, N,$$

где  $N$  – количество финансовых инструментов в портфеле,

$V_{ij}$  – ковариация между приростами цены финансовых инструментов  $i$  и  $j$ ,

$\mu_i$  – среднее значение прироста цены  $i$ -го финансового инструмента,

$\lambda$  – допустимый уровень риска, являющийся компромиссом между приростом цены финансового инструмента и дисперсией,

$x_i$  – искомая переменная, представляющая вес  $i$ -го финансового инструмента в портфеле.

В такой форме задача становится сходной с функцией энергии в сети Холлфилда и оптимизация портфеля достигается через процесс обучения сети. В [43, р. 1184] было показано, что нейронные сети Холлфилда могут достигать лучших результатов в портфельной оптимизации чем другие методы оптимизации, такие как генетические алгоритмы, поиск с запретами и алгоритм имитации отжига.

Использование нейронных сетей для прогнозирования цены финансовых инструментов позволяет перейти от классического определения задачи оптимизации портфеля к формулированию риска как ошибки прогнозирования [44, р. 28]. Тогда риск портфеля определяется как

$$\hat{D} = \hat{\sigma}_p^2 = \sum_{i=1}^M X_i^2 \sigma_{\varepsilon i}^2 + \sum_{i=1}^M \sum_{\substack{j=1 \\ i \neq j}}^M X_i X_j \gamma_{\varepsilon ij},$$

где  $\sigma_{\varepsilon}^2 = \frac{1}{n-1} \sum_{t=1}^n \varepsilon_t^2$ ;

$$\hat{\gamma}_{ij} = \gamma_{\varepsilon ij} = \frac{1}{n-1} \sum_{t=1}^n \varepsilon_{it} \varepsilon_{jt};$$

$$\varepsilon_t = P_t - \hat{P}_t;$$

$Q_t$  – фактический прирост цены финансового инструмента в момент времени  $t$ ;

$\hat{Q}_t$  – прогнозируемое значение цены финансового инструмента в момент времени  $t$ .

Данный подход был протестирован на финансовых инструментах, включенных в индекс IBONESPA и показал более высокую точность по сравнению с классической моделью Марковица.

### **1.3 Принципы функционирования модели самоорганизующихся карт Байеса**

Самоорганизующиеся карты (Self-organized maps, SOM) относятся к классу нейронных сетей с обучением без учителя и используются для отображения данных из пространств с высокой размерностью в пространства с более низкой [45].

SOM состоят из набора нейронов или узлов, которые задаются двумя векторами: прототип-вектором и вектором координат. Прототип-вектор  $m$  имеет ту же размерность, что и входные данные. Вектор координат  $r$  представляет собой позицию нейрона в отображаемом пространстве. Вектора координат SOM нейронов обычно привязаны к регулярной сетке в отображаемом пространстве. Прототип-векторы инициализируются случайными значениями или в соответствии с некоторыми процедурами (например кластеризации), основанными на обработке исходных данных [46].

В ходе стадии обучения входные векторы итеративно и последовательно представляются на вход SOM. В ходе winner-take-all процесса нейрон победитель притягивает представленный входной вектор. Здесь, как правило, используются сравнения на базе вычисления евклидова расстояния и нейрон с наиболее близким прототип-вектором становится победителем, однако могут использоваться и иные методы измерения расстояния [47].

Прототип-вектор нейрона-победителя, а также всех тех нейронов, которые имеют достаточно близкий к нейрону-победителю вектор координат, сдвигается в направлении представленного входного вектора. Достаточность близости вектора координат определяется величиной лимита, уменьшающейся с каждой итерацией обучения. Процесс обучения продолжается до тех пор, пока величина лимита не уменьшается до такой степени, что только нейрон-победитель получает сдвиг в сторону входного вектора, либо пока не будет достигнуто заданное значение лимита.

Результатом обучения является проекция данных из исходного пространства в пространство векторов координат таким образом, что векторы из окрестности в пространстве векторов координат соответствуют векторам из окрестности в исходном пространстве (при этом обратное не всегда верно). Отображенные данные представляют собой аппроксимацию распределения данных в исходном пространстве и, в случае уменьшения размерности пространства отображенных данных, могут рассматриваться как сжатое представление исходных данных.

Использование SOM производится путем применения winner-take-all процедуры к входным векторам [46, р. 105]. В качестве альтернативы, SOM

может использоваться для аппроксимации распределения данных для анализа и оценки данных [48].

Самоорганизующиеся карты Байеса (Bayesian self-organizing map, BSOM) являются расширением SOM, предложенным Ином и Элисоном [49, 50] для построения оценки функции плотности вероятности на основе эмпирических данных. В данной модели функция плотности вероятности аппроксимируется гауссовой смесью распределений и каждый нейрон представляет собой гауссиан в смеси.

Концепция гауссовой смеси распределений заключается в следующем. Пусть  $x$  - элемент из  $d$ -мерного множества входных векторов  $X$ . Допустим, что элементы множества сгенерированы случайно и подчиняются нормальному распределению, при этом имеется  $K$  нормальных распределений с различными параметрами, из которых мы можем выбрать любое для генерации каждого отдельного входного вектора. Обозначим данные  $K$  нормальных распределений как  $(w_1, \dots, w_K)$ , а их векторы параметров через  $\Theta = (\theta_1, \dots, \theta_K)$ , где  $\theta_i = \{m_i, \Sigma_i\}$ ,  $i = 1, \dots, K$  включает вектор средних значений и ковариационную матрицу гауссиана.

Тогда функция плотности вероятности нормального распределения  $w_i$  имеет вид:

$$p(x|w_i, \theta_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(x - m_i)^T \Sigma_i^{-1} (x - m_i)\right), \quad (2.2.11)$$

где  $m_i$  – вектор средних значений  $i$ -го гауссиана,

$\Sigma_i$  – ковариационная матрица  $i$ -го гауссиана.

Функцией плотности вероятности гауссовой смеси является:

$$p(x|\Theta) = \sum_{i=1}^K p(x|w_i, \theta_i) P(w_i) \quad (2.2.12)$$

где  $P(w_i)$  или  $P_i$  – априорная вероятность принадлежности элемента к гауссиану  $w_i$ , при этом  $\sum_{i=1}^K P(w_i) = 1$ .

Каждый нейрон в BSOM представляет один гауссиан из смеси. В дополнение к стандартным для SOM векторам, нейроны BSOM также имеют соответствующие ковариационные матрицы и значения априорной вероятности, представляющие вклад гауссиана в параметры гауссовой смеси, которую представляет вся BSOM. Таким образом BSOM задает параметры распределения гауссовой смеси, определяемой уравнением (2.2.12). Для обучения нейроны BSOM инициализируются однородным распределением и диагональной ковариационной.

В ходе обучения векторы  $x(t)$  представляются на вход BSOM. Нейрон-победитель  $v$  определяется путем определения апостериорной вероятности для представленного входного вектора:



$$v = \arg \max_i \left\{ P(w_i, x(t), \theta_i) = \frac{p[x|w_i, \theta_i]P(w_i)}{\sum_{i=1}^K p[x|w_i, \theta_i]P(w_i)} \right\}, \quad (2.2.13)$$

где  $P(w_i, x(t), \theta_i)$  – апостериорная вероятность  $i$ -го нейрона в соответствии с уравнением (2.2.11),

$t$  – порядковый номер входного вектора.

Окрестность  $\eta_v$  нейрона-победителя определяется также, как и в стандартной модели SOM на основе расстояния между векторами координат в отображаемом пространстве.

Уравнения обновления параметров нейронов BSOM имеют следующий вид:

$$m_i(t+1) = m_i(t) + \alpha(t)P[w_i, x(t), \theta_i][x(t) - m_i(t)], \quad i \in \eta_v, \quad (2.2.14)$$

$$\Sigma_i(t+1) = \Sigma_i(t) + \alpha(t)P[w_i, x(t), \theta_i][x(t) - m_i(t)][x(t) - m_i(t)]^T - \Sigma_i(t), \quad i \in \eta_v, \quad (2.2.15)$$

$$P_i(t+1) = P_i(t) + \alpha(t)[P[w_i, x(t), \theta_i] - P_i(t)], \quad i \in \eta_v, \quad (2.2.16)$$

где  $\alpha(t)$  – скорость обучения на шаге  $t$ , задаваемая как:

$$\alpha(t) = \frac{\alpha_0}{1 + t/\tau}, \quad (2.2.17)$$

где  $\alpha_0$  и  $\tau$  – гиперпараметры модели.

Алгоритм BSOM имеет близкое родство с методом Generative Topographic Mapping (GTM), который был предложен Бишопом и Свенсеном [51] и другими похожими методами. Метод GTM основан на построении функции правдоподобия для гауссовой смеси. Поиск параметров модели GTM производится при помощи EM-алгоритма (expectation-maximization algorithm). Как было показано в [50, p. 234] скорость сходимости BSOM значительно превосходит EM-алгоритм в задачах аппроксимации гауссовой смеси распределений.

## 2 ПОСТРОЕНИЕ ПРОСТРАНСТВЕННОЙ МОДЕЛИ НЕЙРОСЕТИ ДЛЯ ЗАДАЧ РАСПОЗНАВАНИЯ И КЛАССИФИКАЦИИ

### 2.1 Постановка задачи распознавания и классификации

Итак, пусть  $X = \{x | x = (x_1, x_2, \dots, x_n), x_i \in \{0,1\}, i = 1, 2, \dots, n\}$  пространство исходных объектов, а  $C'_1, C'_2, \dots, C'_\ell$  классы целиком покрывающие пространство исходных объектов  $X$ . Пусть  $Q_1, Q_2, \dots, Q_\ell$  система одноместных 2- $x$ -значных предикатов над  $X$  таких, что  $Q_j(x) \equiv \ll x \in C_j \gg, x \in X, j = 1, 2, \dots, \ell$ .

Задача распознавания  $u \in U$  это упорядоченная пара:

$$u = (I_0, X^q)$$

где  $I_0 = \langle X^m, \alpha \rangle$  – начальная информация задачи  $u$ ;

$X^m = \{x^1, x^2, \dots, x^m\}$  – обучающая выборка задачи  $u$ ;

$\alpha$  – матрица классификации выборки  $X^m$ ;

$X^q = \{x^1, x^2, \dots, x^q\}$  – выборка контрольных объектов задачи  $u$ .

В качестве исходного уравнения возьмем операторное уравнение, с которым обычно связывают решение прямой или обратной задачи:

$$\mathcal{A}u = f, \mathcal{A} \in H, u \in U, f \in F \quad (3.1.1)$$

где  $\mathcal{A}$  – распознающий алгоритм(оператор), действующий из метрического пространства  $U$  в метрическое пространство  $F$ ;

$U$  – пространство задач распознавания;

$f = \|f_{ij}\|_{q \times \ell}$  – матрица классификации задачи  $u$  (выборки  $X^q$ ) [2, с. 15];

Для задачи  $u$  начальная информация  $I_0$ , как видим, содержит список обучающих объектов  $X^m = \{x^1, x^2, \dots, x^m\}$  и их информационную матрицу  $\alpha = \|\alpha_{ij}\|_{m \times \ell}$ , показывающую принадлежность объектов выборки  $X^m$  классам  $C'_1, C'_2, \dots, C'_\ell$ . При этом  $\alpha_{ij} = Q_j(x^i), j = 1, 2, \dots, \ell; i = 1, 2, \dots, m$ ;

Другой набор  $X^q = \{x^1, x^2, \dots, x^q\}$ , как мы уже знаем, выборка контрольных (рабочих) объектов, которые подлежат распознаванию. Элементы информационной матрицы для  $X^q$ , т.е. элементы матрицы классификации  $f$  задачи  $u$ , также определяются как  $f_{ij} = Q_j(x^i), j = 1, 2, \dots, \ell; i = 1, 2, \dots, q$ .

В качестве  $H$ , как правило, берется параметрическое семейство алгоритмов  $\{\mathcal{A}_p\} = \{\mathcal{A}_p | p = (p_1, p_2, \dots, p_c) \in P\}$ , среди элементов которого при помощи той или иной оптимизационной процедуры стараются выделить алгоритм, гарантирующий приемлемое решение задачи распознавания или классификации. Здесь обычно исходят из неразрешимости обратной задачи для уравнения (3.1.1) и в случае нормированного пространства  $F$  задача синтеза алгоритма  $\mathcal{A}$ , обеспечивающего точное решение задачи  $u \in U$ , подменяется экстремальной задачей типа:

$$\min_{p \in P} \|\mathcal{A}_p u - f\|^2 \quad (3.1.2)$$

Её решение, если оно существует, определяет алгоритм  $\mathcal{A}_{p^*}$ , который и является приближенным решением обратной задачи для (3.1.1), при этом условия для построения  $\mathcal{A}_{p^*}$ , с практической точки зрения, могут представляться достаточно ограничительными, а их проверка вызывать определенные затруднения. В нашем случае используя теорему о декомпозиции распознающего оператора [2, с. 24.; 4, с. 40] уравнение (3.1) заменим системой:

$$\begin{aligned} Au &= \varphi, \quad A \in \mathfrak{A}\{A_p\}, \quad u \in U \\ R^* \varphi &= f, \quad \varphi \in \Phi, \quad f \in F, \end{aligned}$$

где  $\mathfrak{A}$  – алгебра над  $\{A_p\}$ ;

$A$  – оператор, вычисляющий матрицу оценок  $\varphi = \|\varphi_{ij}\|_{q \times \ell}$ ;

$\Phi$  – метрическое пространство матриц размерности  $q \times \ell$ ;

$R^*$  – решающее правило [2, с. 24], которое по матрице  $\varphi$  определяет принадлежность объектов выборки  $X^q$  классам  $C'_1, C'_2, \dots, C'_\ell$ .

Таким образом, задача построения алгоритма  $\mathcal{A}^*$  сводится к задаче построения  $A^* \in \mathfrak{A}\{A_p\}$ , такого, что  $\mathcal{A}^* = A^* \circ R^*$  и представляет собой специальный вариант классической задачи о восстановлении оператора по информации о решении и правой части операторного уравнения. Операторный подход позволяет за счет пополнения семейства  $\{A_p\}$  в целом ряде случаев [2, с. 36; 3, с. 762; 4, с. 37; 5, с. 737; 6, с. 54; 7, с. 1280; 8, с. 176; 9, с. 1140; 10, с.12; 52, 53] определить условия корректности алгебры  $\mathfrak{A}$  над множеством задач распознавания и классификации, что подразумевает описание множества задач  $U$ , таких, что для решения каждой задачи  $u \in U$  в  $\mathfrak{A}\{A_p\}$  существует алгоритм  $\mathcal{A}^*$ , дающий точное решение задачи  $u$ , т.е.  $\mathcal{A}^* u = f$ ,  $u \in U$ ,  $\mathcal{A}^* \in \mathfrak{A}\{A_p\}$ .

При условии, что исходное множество алгоритмов – это семейство алгоритмов вычисления оценок, а  $u$  это регулярная задача, в [4, с. 41] найдено явное выражение для  $\mathcal{A}^*$  в виде операторного многочлена (оператор Журавлева), вид (1.3) которого носит универсальный характер:

$$\mathcal{A}^* = ((\theta_1 + \theta_2) \sum_{i=1}^q \sum_{j=1}^{\ell} f_{ij} \cdot B^k(i, j)) \circ R^* \quad (3.1.3)$$

$$\begin{aligned} k &= [(\ln q + \ln \ell + \ln(\theta_1 + \theta_2) - \ln \theta_1) / |\ln a|] + 1 \\ a &= \max_{i,j} \max_{(s,h) \neq (i,j)} |\Gamma_{sh}(i, j)|, \end{aligned} \quad (3.1.4)$$

Заметим, что имеется возможность понизить  $k$  степень замыкания [7, с. 1286; 8, с. 176; 9, с. 1142], однако в данной работе мы будем ориентироваться на формулу (3.1.4). В (3.1.3) пороговое решающее правило  $R^*(\theta_1, \theta_2)$  с

параметрами  $\theta_1 = \min \theta_{1j}$ ,  $\theta_2 = \max \theta_{2j}$ , по матрице  $\varphi = \|\varphi_{ij}\|_{q \times \ell}$  и при  $k$ , подсчитанном согласно (3.1.4), даёт матрицу  $\beta = \|\beta_{ij}\|_{q \times \ell}$ , совпадающую с матрицей  $f$  задачи  $u$ . В более общем случае для алгоритма распознавания  $\mathcal{A}$  решающее правило  $R^*(\theta_{1j}, \theta_{2j})$  с параметрами  $\theta_{1j}, \theta_{2j}$  по  $\varphi$  вычисляет матрицу  $\beta = \|\beta_{ij}\|_{q \times \ell}$ , которая может содержать элементы отличные от соответствующих элементов матрицы  $f = \|f_{ij}\|_{q \times \ell}$ .

Наша цель здесь состоит в том, чтобы, опираясь на операторный подход [2, с. 41; 3, с. 761; 4, с. 37; 5, с. 728; 6, с. 31] и нейросетевую парадигму, показать возможность построения для задачи  $u \in U$  корректного алгоритма и нейросети, воспроизводящей вычисления осуществляемые корректным алгоритмом. В нашем случае речь идет о нейросети, последовательно вычисляющей строки матрицы  $\beta$ , т.е. матрицы, совпадающей с матрицей  $f$  задачи  $u$ . Для построения нейросети нам будет удобно переписать выражение (3.1.3) для  $\mathcal{A}^*$  в виде:

$$\mathcal{A}^* = \sum_{f_{ij}=1}^{i,j} (\theta_1 + \theta_2) \cdot B^k(i, j) \circ R^* \quad (3.1.5)$$

Элемент  $\beta_{ij}$  матрицы  $\beta$  определяется так:

- если вычисленная оператором  $A$  оценка  $\varphi_{ij} > \theta_{2j}$ , то  $\beta_{ij} = 1$ , т.е. распознающий алгоритм  $\mathcal{A}$  относит объект  $x^i$  к классу  $C_j$ ;
- если  $\varphi_{ij} < \theta_{1j}$ , то  $\beta_{ij} = 0$ , что трактуется как  $x^i \notin C_j$ ;
- и, наконец,  $\mathcal{A}$  не определяет вхождение (не вхождение)  $x^i$  в класс  $C_j$ , если  $\theta_{1j} \leq \varphi_{ij} \leq \theta_{2j}$ , что записывается как  $\beta_{ij} = \Delta$ .

Параметры  $\theta_{1j}, \theta_{2j}$  таковы:  $0 < \theta_{1j} \leq \theta_{2j}$  [2],  $j = \overline{1, \ell}$ . Как видим, правило  $R^*$  практически обобщает известное правило, задающее акты возбуждения и торможения биологического нейрона. Пусть  $A', A'', A \in \mathfrak{M}$  операторы с матрицами  $\varphi', \varphi'', \varphi$  соответственно, операции алгебры  $\mathfrak{A}$  таковы [2, с. 43]:

- а)  $const \cdot Au = \|const \cdot \varphi_{ij}\|$ ;
- б)  $(A' + A'')u = A'u + A''u = \|\varphi'_{ij} + \varphi''_{ij}\|$ ;
- в)  $(A' \cdot A'')u = A'u \cdot A''u = \|\varphi'_{ij} \cdot \varphi''_{ij}\|$

Здесь для нас имеется дополнительное условие, связанное с влиянием нейросетевой парадигмы, которое состоит в том, что ввиду ограниченности синаптической силы биологического нейрона как матрицы-операнды операций а) – в), так и результирующие матрицы в а) – в) не должны иметь элементов с большими модулями. Зададим эти значения промежутком  $(-1, 1]$ , а в качестве операции в) будем иметь в виду операцию возведение в степень, т.е.  $A^k$ . По тем же причинам ограничим и величины  $0 < \theta_{1j}, \theta_{2j} < 1$ , добавив условие:  $\theta_1 \leq (1/2)(1 - \theta_2)$ .

В [10, с. 12; 52, с. 713] условия разрешимости обратной задачи для (3.1.1) найдены и для задачи сегментации программ – дискретной экстремальной

задачи, не имеющей в смысле [2, с. 44] классического описания. Универсальность подхода [2, с. 12; 3, с. 761; 4, с. 37; 5, с. 728; 6, с. 31] показывает, что подобные условия могут быть найдены и для других типов родственных задач и алгоритмов. В связи с чем обратим внимание на алгоритмы обучения искусственных нейронных сетей, которые, как известно, преимущественно следуют стратегии, заключенной в (3.1.2), позволяющей в общем случае получать лишь приближенные решения задач распознавания и классификации.

Проблемы обучения сетей, связанные со стратегией (3.1.2), а с другой стороны, результаты о корректности моделей, построенных на основе [2, с. 19.; 3, с. 761; 4, с. 37; 5, с. 728; 6, с. 31; 7, с. 1280; 8, с. 176; 9, с. 1140], вызывают определенный интерес к моделям, цель построения которых состоит в том, чтобы показать принципиальную возможность построения на основе [2, с. 19; 3, с. 761; 4, с. 37; 5, с. 729; 6, с. 80] нейросетей для получения точных решений задач распознавания и классификации.

Преследуя эту цель, в качестве исходного семейства распознающих операторов возьмем семейство операторов вычисления оценок такое, что каждому оператору этого семейства соответствует элементарная пространственная нейросеть ( $\mu$ -блок), воспроизводящая те же вычисления, что и исходный оператор. Такие операторы и назовем  $\mu$ -операторами, при этом предполагая, что соблюдается и упомянутая особенность нейросетевой парадигмы. Говоря о нейронах внутренних слоёв сети и, в том числе и  $\mu$ -блока, будем иметь в виду модель нейрона с полностью или частично заданной диагональной активационной функцией, которая может иметь ненулевые значения лишь на промежутке  $(-1, 1]$ .

В случае полностью заданной диагональной активационной функции будем использовать также обозначение  $R$ . У нас для подобного типа активационной функции значение, выработанное сумматором нейрона, и является значением активационной функции. Для частично заданной диагональной активационной функции будем использовать и обозначение  $R^\setminus$ . Заметим, что  $\mu$ -блок может рассматриваться также и как самостоятельный распознающий алгоритм, например, с активационной функцией  $R^*(\theta_{1j}, \theta_{2j})$ ,  $j = 1, 2, \dots, \ell$ , однако этот случай потребует, разумеется, наличия алгоритма обучения  $\mu$ -блока.

Кроме того полагаем, что нейроны, вовлекаемые в процессе построения нейросети обладают особенностями, необходимыми для такого построения. Говоря о пространственных нейросетях добавим, что этот класс нейросетей вызывает [13, р. 555; 14, р. 436] и, по-видимому, должен вызывать вполне определенный интерес, как «эволюционное» развитие концепции плоских структур. Пусть далее  $C_j = X^m \cap C'_j$ ,  $j = 1, 2, \dots, \ell$ .

## **2.2 Построение модели 3-х уровневой пространственной нейросети ( $\mu$ -блок) для модели распознавания $\mathfrak{M}$**

Рассмотрим модель  $\mathfrak{M} = \{A(\mu, \Omega, \varepsilon^n, p^n, \gamma^m)\}$  операторов вычисления оценок по опорным множествам, использованную в [53, с. 1491] для доказательства корректности алгоритмов «тестового» типа. Зафиксируем произвольный оператор  $A \in \mathfrak{M}$ , вычисляющий по  $u = (I_0, X^q)$  матрицу оценок  $\Gamma = \|\Gamma_{ij}\|_{q \times \ell}$  и задаваемый набором параметров:

$$\begin{aligned} \mu &\in \{0,1\}, \Omega = \{e_a\}, \varepsilon^n = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n), \\ p^n &= (p_1, p_2, \dots, p_n), \gamma^m = (\gamma_1, \gamma_2, \dots, \gamma_m), \end{aligned}$$

где  $p_1, p_2, \dots, p_n$  – веса признаков, они же синаптические веса нейронов 1-го уровня сети ( $\mu$ -блока), и здесь обратим внимание на рисунках 3.2.1, 3.2.2;

$\gamma_1, \gamma_2, \dots, \gamma_m$  – веса обучающих объектов, которые в каждом классе могут иметь свои значения. Они же являются синаптическими весами  $w_1, w_2, \dots, w_m$  нейронов 2-го уровня сети (рисунки 3.2.1, 3.2.2);

$\varepsilon^n = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$  – параметры функции близости. Они же являются параметрами активационной функции нейрона 1-го уровня;

Детали, связанные с параметрами для алгоритмов вычисления оценок обсуждаются, например, в [54]. В нашем случае  $\varepsilon_1 = \varepsilon_2 = \dots = \varepsilon_n = \varepsilon$ , где  $\varepsilon$  целое из  $[0, n]$ , параметр  $\mu \in \{0,1\}$ ,  $X = \{x\}$  – исходное пространство объектов – бинарных векторов размерности  $n$ .

Пусть  $u = (I_0, X^q)$  произвольная задача из  $U$ . Для подкласса алгоритмов вычисления оценок, использованного, например, в [53, с. 1491], каждый алгоритм  $\mathcal{A}$  подкласса представим в виде суперпозиции двух операторов  $\mathcal{A} = A \circ R^*$  и будем считать, что заданы:

1. Система опорных множеств  $\Omega = \{e_a\}$ ,  $e_a \subset \{1, 2, \dots, n\}$ , «веса» признаков  $p_1, p_2, \dots, p_n$ ,  $p_i \geq 0$ ,  $i = 1, 2, \dots, n$ . Вес опорного множества  $e = (t_1, t_2, \dots, t_{|e|})$ , есть  $p(e) = p_{t_1} + p_{t_2} + \dots + p_{t_{|e|}}$ . Веса  $\gamma_1, \gamma_2, \dots, \gamma_m$  обучающих объектов  $x^1, x^2, \dots, x^m$  из  $X^m$ , где  $\gamma_i \geq 0$ ,  $\gamma_i = \gamma(x^i)$ ,  $i = 1, 2, \dots, m$ .

2. Функция близости  $F_e^\varepsilon(x', x'')$  объектов  $x', x''$  по опорному множеству  $e$ . Функция  $F_e^\varepsilon(x', x'')$  задаётся следующим образом:

$$F_e^\varepsilon(x', x'') = 1, \text{ если } \rho_e(x', x'') \geq n - \varepsilon; \text{ и}$$

$$F_e^\varepsilon(x', x'') = 0 \text{ в противном случае.}$$

Здесь функция  $\rho_e(x', x'')$  это число совпавших координат у векторов  $x'_e = (x'_{t_1}, x'_{t_2}, \dots, x'_{t_{|e|}})$ ,  $x''_e = (x''_{t_1}, x''_{t_2}, \dots, x''_{t_{|e|}})$ ;  $\varepsilon$  – целое из отрезка  $[0, n]$ . Функция антиблизости:

$$\bar{F}_e^\varepsilon(x', x'') = 1 - F_e^\varepsilon(x', x'').$$

3. Правило вычисления оценки  $\Gamma_{ij}$  для объекта  $x^i \in X^q$  за класс  $C_j$ :

$$\Gamma_{ij}^1 = \Gamma_j^1(x^i) = \sum_{x \in C_j} \gamma(x) \sum_{e \in \Omega} F_e^\varepsilon(x, x^i) \cdot p(e),$$

$$\Gamma_{ij}^0 = \Gamma_j^0(x^i) = \sum_{x \in \bar{C}_j} \gamma(x) \sum_{e \in \Omega} \tilde{F}_e^\varepsilon(x, x^i) \cdot p(e),$$

$$\Gamma_{ij} = (1 - \mu) \cdot \Gamma_{ij}^0 + \mu \cdot \Gamma_{ij}^1, \quad \mu \in \{0,1\},$$

где  $\bar{C}_j$  это дополнение к классу  $C_j$  в  $X^m$ ,  $\mu \in \{0,1\}$ .

4. Параметры  $\theta_{1j}, \theta_{2j}$ ,  $j = 1, 2, \dots, \ell$  решающего правила  $R^*$ , по которым вычисляется матрица  $\beta = \|\beta_{ij}\|_{q \times \ell}$ .

Будем исходить из предположения, что информационный образ  $x_1, x_2, \dots, x_n$  объекта, подаваемого на вход нейрона 1-го слоя  $\mu$ -блока, складывается не из однородной массы однотипных входных сигналов, а из групп их различных комбинаций. В связи с чем, считаем аргументы выделенными множества комбинаций входов нейрона 1-го слоя, а именно:  $e_1, e_2, \dots, e_{|\Omega|}$ ;  $\Omega = \{e_a\}$ .

Пусть  $e_a \subset \{1, 2, \dots, n\}$  и  $e = (t_1, t_2, \dots, t_{|e|})$ ,  $p_e = (p_{t_1}, p_{t_2}, \dots, p_{t_{|e|}})$ ,  $ex = x_e = (x_{t_1}, x_{t_2}, \dots, x_{t_{|e|}})$ ,  $ab = (x_e, p_e)$ . Далее при заданном  $p = (p_1, p_2, \dots, p_n)$  будем вычислять произведение  $b$  лишь для обучающих объектов, т.е. объектов из  $X^m$ . Заметим, что выбор  $e_a = (t_1, t_2, \dots, t_{|e_a|}) \in \Omega$  определяет и нулевые веса среди  $p_1, p_2, \dots, p_n$  у  $n$  входов для каждого из нейронов  $e_a$ -столбца первого уровня (рисунок 3.2.1)  $\mu$ -блока. Речь здесь идёт о тех из весов, номера которых не входят в  $e_a$ , при этом  $a = 1, 2, \dots, |\Omega|$ .

Представим элементарный блок ( $\mu$ -блок), вычисляющий матрицу оценок  $\Gamma = \|\Gamma_{ij}^\mu\|_{q \times \ell}$  выборки  $X^q = \{x^1, x^2, \dots, x^q\}$ , соответствующий оператору  $A \in \mathfrak{M}$ , как 3-х уровневую пространственную многослойную сеть, где на рисунке 3.2.1 и рисунке 3.2.2 показано лишь вычисление оценки  $\Gamma_{ij}^\mu$ .

«Средний» фрагмент для  $e_a$ -го и  $j$ -го столбцов этой сети изображен на рисунке 3.2.1. Будем считать, что расслоение сети  $\mu$ -блока (рисунок 3.2.2) в плоскости  $XOY$  осуществляется по уровням, а вдоль оси  $Z$  по слоям. Кроме того, говоря о нейронах одного уровня в плоскости  $XOY$  для самой нейросети (см. п. 3.), мы также будем использовать термин расслоение.

Число нейронов  $\mu$ -блока равно  $2 \cdot |\Omega| \cdot \ell + \ell$ . Число нейронов 1-го уровня в плоскости  $XOY$  равно  $|\Omega|$ , каждый из них допускает расслоение по оси  $Z$  с числом нейронов  $\ell$  в каждом расслоении. Выход активационной функции нейрона 1-го уровня:  $F_\mu^\varepsilon(ex', ex) = \mu b$ , если  $\rho_e(x', x) \geq n - \varepsilon$  и  $F_\mu^\varepsilon(ex', ex) = (1 - \mu)b$ , если  $\rho_e(x', x) < n - \varepsilon$ ; где  $\rho_e(x', x)$  функция, как уже говорилось, число совпавших координат у векторов  $ex', ex$ , вычисляемая для  $x' \in X^m$ ,  $x \in X^q$ .

Рассмотрим одноименную с  $C_j$  переменную и введём обозначение  $C_j^\mu = \mu \cdot C_j \vee \bar{\mu} \cdot \bar{C}_j$ , полагая  $C_j^\mu$  при  $\mu = 1$  это  $C_j$ , или при  $\mu = 0$  это  $\bar{C}_j$ , где  $\bar{C}_j$  это дополнение к классу  $C_j$  в  $X^m$ . Пусть, теперь  $x'$  обозначает произвольный объект из  $C_j^\mu \cup \{\#\}$ . Пусть далее при  $\mu = 1$  «вес»  $x'$  из  $C_j$  это  $w_j' = w(x') \geq 0$ . Для себя

отметим, что когда речь идет об операторах вычисления оценок, то эта же величина имеет другое, традиционное обозначение  $\gamma(x')$ . Присоединим (рисунок 3.2.2) ко входу сети ( $\mu$ -блока) в конец списка объектов любого  $C_j^\mu$  формальный объект-значение  $\#$ , вес которого  $w_\# = 0$  и, игнорируя вычисления на сумматоре, положим:

$$F_\mu^\varepsilon(e\#, ex) = \# \quad , \quad \forall x \in X^q \text{ и } \# \cdot 0 = 0 \cdot \# = 0.$$

Число нейронов 2-го уровня  $\mu$ -блока в плоскости  $XOY$  равно  $\ell$ , и каждый из них допускает расслоение по оси  $Z$  с числом нейронов  $|\Omega|$  в каждом расслоении. Выход активационной функции нейрона 2-го уровня при условии, что в (3.2.1)  $C_j^\mu \neq \emptyset$  есть

$$f_e^\mu(x', \xi_{i,j}^{(t,e)}) = \begin{cases} \sum_{\substack{x \in C_j^\mu: \\ F_\mu^\varepsilon(ex, ex') \neq 0}} w_j(x) \cdot (p_e, x_e), \text{ если } x' = \#, \\ \Delta \text{ пока } x' \neq \# \quad (\text{обновление из } C_j^\mu) \end{cases} \quad (3.2.1)$$

где  $\xi_{i,j}^{(t,e)}$  это значение, вычисляемое на сумматоре нейрона,  $t = 1, 2, \dots, |C_j^\mu|$ , при этом на рисунке 3.2.1, рисунке 3.2.2 используется более простое обозначение  $\xi_{ij}^{(t)}$ . Равенство  $f_e^\mu(x', \xi_{i,j}^{(t,e)}) = \Delta$  означает, что значение функции  $f_e^\mu$  при заданных аргументах функции не определено (не вычисляется), что подразумевает лишь обновление объекта из  $C_j^\mu$  для продолжения внутреннего цикла по  $x'$ . При  $C_j^\mu = \emptyset$  полагаем  $f_e^\mu(\#, \xi_{i,j}^{(0,e)}) = 0 \quad \forall e, \mu, i, j$ .

Каждый нейрон 1-го уровня, как и его дубликаты (по оси  $Z$ ), имеет  $n$  входов и один выход, связанный с нейроном 2-го уровня в соответствующем расслоении (рисунок 3.2.1), например, нейрон 1-го уровня  $e_a, j$  связан с нейроном 2-го уровня  $j, e_a$ , где  $j$  – номер класса,  $e_a \in \Omega$ ,  $e_a = (i_1, i_2, \dots, i_{|e_a|})$ .



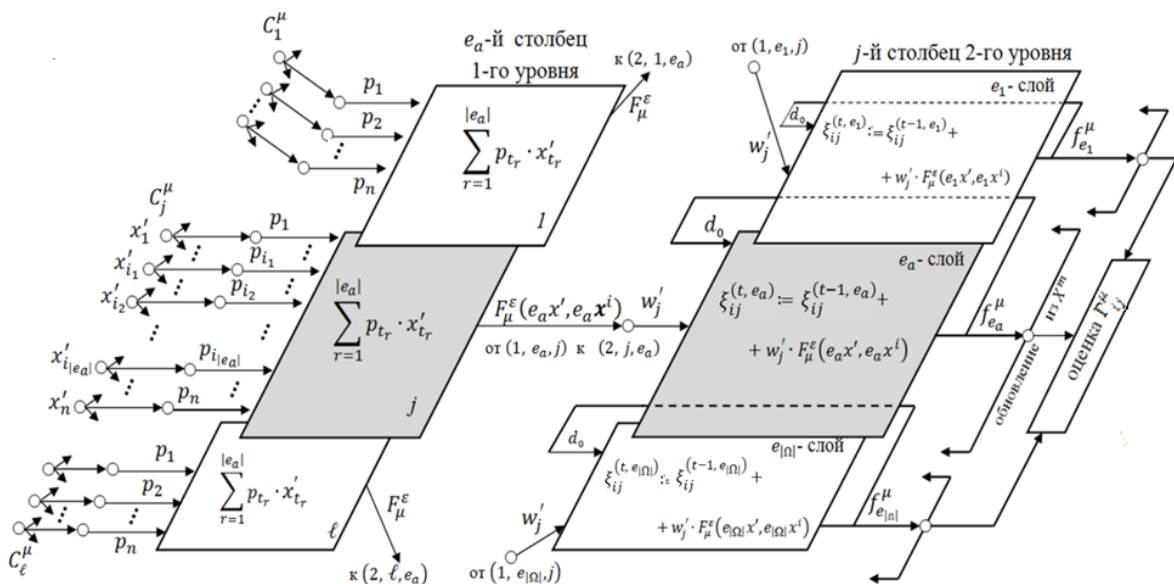


Рисунок 3.2.1 – Средний фрагмент  $\mu$ -блока для  $e_a$ -го и  $j$ -го столбцов

Каждый нейрон в расслоении 2-го уровня имеет еще один вход от своего же выхода – для аккумуляции  $e$ -части оценки  $\Gamma_{ij}^\mu$ , и его синаптический вес  $d_0 = 1$  пока  $x' \neq \#$  ( $x' \in C_j^\mu$ ). При появлении  $x' = \#$ , точнее непосредственно после этого, синаптический вес  $d_0$  станет «0» (очистка сумматора) и восстановит значение равное «1», после того, как на вход  $\mu$ -блока придет первый элемент из  $C_j^\mu$  для нового  $x^i \in X^q$ . Для каждого  $x^i \in X^q$  на вход  $\mu$ -блока последовательно подаются объекты  $x'$  из  $C_j^\mu$  пока  $x' \neq \#$ .

Случай  $x' = \#$  на входе сети (вспомним,  $F_\mu^\varepsilon(e\#, ex) = \#$  и  $\# \cdot 0 = 0 \cdot \# = 0$ ) дает «0»-й вклад в окончательную оценку на сумматоре нейрона 2-го уровня

$$\begin{aligned} \xi_{ij}^{(t+1, e_a)} &= \xi_{ij}^{(t, e_a)} + w_j' \cdot F_\mu^\varepsilon(e_a x', e_a x^i) = \xi_{ij}^{(t, e_a)} + 0 \cdot \# = \\ &= \sum_{\substack{x' \in C_j^\mu: \\ F_\mu^\varepsilon(e_a x', e_a x^i) \neq 0}} w_j(x') \cdot (p_{e_a}, x'_{e_a}), \quad t = 0, 1, \dots, |C_j^\mu|. \end{aligned}$$

и, если  $x' = \#$  имеет место для всех  $j = 1, 2, \dots, \ell$  (т.е. все  $C_j^\mu$  исчерпаны), то это приводит и к смене контрольного объекта, т.е. объекта из  $X^q$ . Процесс полностью завершается, когда исчерпаются объекты выборки  $X^q$ .

Каждый нейрон  $\mu$ -блока сохраняет вычисленное на «сумматоре» значение, скажем  $s$ , лишь непродолжительное время. И очистка сумматора любого нейрона  $\mu$ -блока, как и всех нейронов сети выполняется автоматически. Небольшой вопрос возникает при последней итерации у нейронов с обратной связью (уровни 2, 5 сети). В нашей модели для уровня 2, нейрон, точнее его сумматор, сохраняет  $s (s \neq 0)$  до момента передачи  $s$  в качестве выходного значения своей активационной функции, или в качестве промежуточного

значения до передачи  $s$  по соединению обратной связи к своему же входу. В последнем случае, синхронно с другим входом нейрона новые аргументы подаются на сумматор для следующей итерации.

При этом структура  $\mu$ -блока позволяет осуществлять вычисление самих оценок  $\Gamma_{i1}^\mu, \Gamma_{i2}^\mu, \dots, \Gamma_{il}^\mu$  (3-й уровень, рисунок 3.2.2.) во внутреннем цикле, одновременно и независимо, в параллельном режиме. Здесь при «вычислении» значения активационной функции для нейрона 1-го уровня выполняется проверка  $\rho(ex', ex) \geq n - \varepsilon$  ( $x' \neq \#$ ), а для нейрона 2-го уровня проверка  $x' = \#$ .

Обратим внимание, что для нейронов 1-го уровня активационная функция – это диагональная активационная функция  $F_\mu^\varepsilon$ , а для нейронов 2-го уровня активационная функция – это частично заданная диагональная активационная функция (2.1). Таким образом, для нейрона 2-го уровня результат вычисления  $s = \xi_{i,j}^{(t,e)}$  окажется на выходе нейрона, если внутренний цикл завершен ( $x' = \#$ ). В противном случае ( $x' \neq \#$ ), значение  $s$  передается на свой же вход для следующей итерации с обновленным  $x'$ , пока  $x' \neq \#$ , к этому моменту должна произойти и очистка сумматора нейрона. И, в целом, считаем, что для всех нейронов сети к моменту активации входов нейрона значение  $s$  на сумматоре это 0. Что касается активационной функции нейронов 3-го уровня  $\mu$ -блока, то далее будем считать, что их активационная функция это  $R$ .

Говоря сейчас о всей сети, скажем, что вычисления для объекта  $x^i \in X^q$  носят асинхронный характер и окончательный для  $x^i$  результат, это  $i$ -я строка матрицы  $f = \|f_{ij}\|_{q \times \ell}$ . Её элементы будут вычисляться, хотя и независимо, но в выходной строке (см. рисунок 3.4.1), будут появляться не одновременно, что, в первую очередь, объясняется тем, что длины последовательностей в  $C_j^\mu$  при разных  $j = 1, 2, \dots, \ell$  могут быть различны.

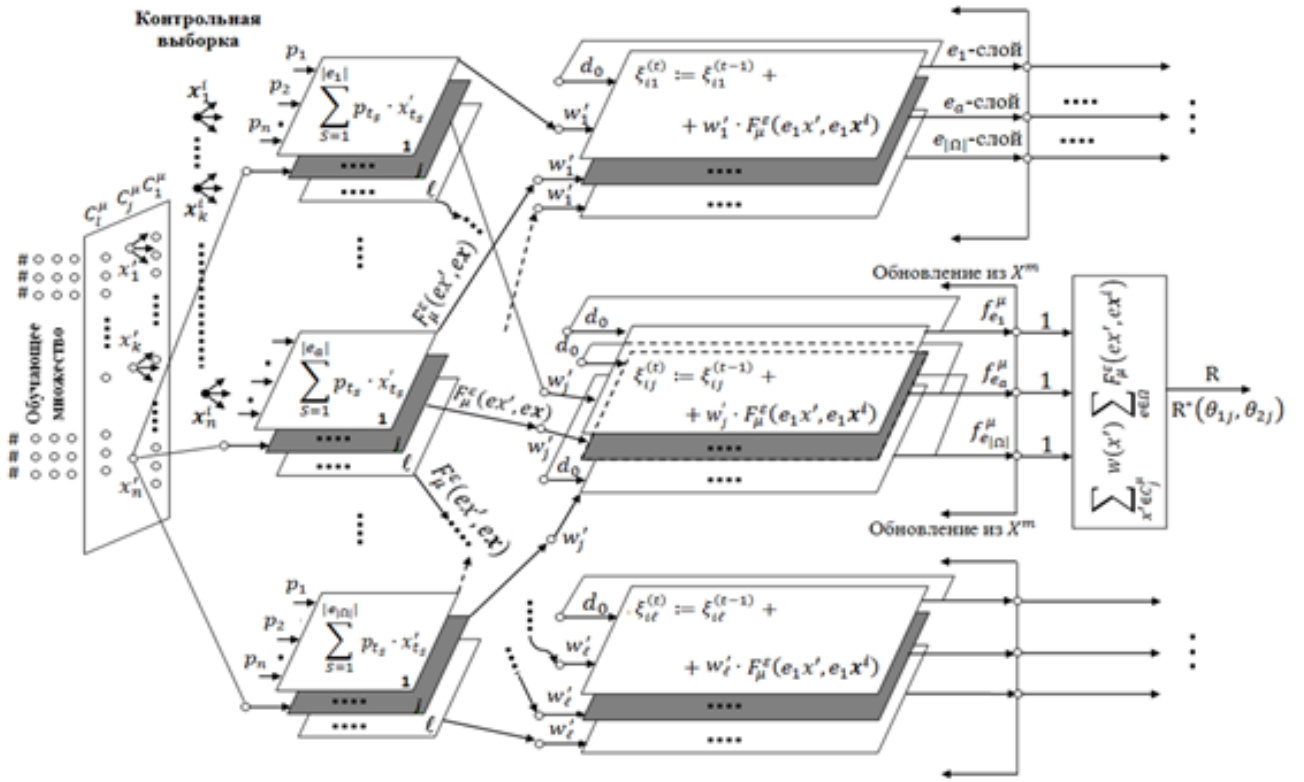


Рисунок 3.2.2 –  $\mu$ -блок

Как видно из рисунка 3.2.2. для нейрона 2-го уровня вычисление оценки  $\xi_{ij}^{(t,e)}$  осуществляется при  $t = 1, 2, \dots, |C_j^\mu|$  и последнее вычисленное значение и есть  $e$ -часть оценки  $\Gamma_{ij}^\mu$ , начальное же значение  $\xi_{ij}^{(0,e)} = 0$  для всех  $i, j, e$ . При исчерпании  $C_j^\mu$  все выходы от нейронов в расслоении нейрона 2-го уровня поступают на единичные входы соответствующего нейрона 3-го уровня (всего таких нейронов  $\ell$ ), где на «сумматоре» и образуется оценка  $\Gamma_{ij}^\mu$ , оценка объекта  $x^i \in X^q$  за класс  $C_j$ .

Формула (3.2.2), (3.2.3), (3.2.4) для вычисления  $\Gamma_{ij}^\mu$  такова:

$$\Gamma_{ij}^\mu = \mu \sum_{x' \in C_j} w(x') \sum_{e \in \Omega} F_\mu^e(ex', ex^i) + (1 - \mu) \sum_{x' \in \bar{C}_j} w(x') \sum_{e \in \Omega} F_\mu^e(ex', ex^i) = \quad (3.2.2)$$

$$= \mu \sum_{x' \in C_j} \gamma(x') \sum_{e \in \Omega} F_e^e(x', x^i) \cdot (p_e, x_e') + (1 - \mu) \sum_{x' \in \bar{C}_j} \gamma(x') \sum_{e \in \Omega} \bar{F}_e^e(x', x^i) \cdot (p_e, x_e') = \quad (3.2.3)$$

$$= \sum_{x' \in C_j^\mu} w(x') \sum_{e \in \Omega} F_\mu^e(ex', ex^i), \quad j = 1, 2, \dots, \ell \quad (3.2.4)$$

Заметим, что здесь за счет дублирования  $\mu$ -блоков можно пойти по пути дальнейшего распараллеливания вычислительного процесса (оценки  $\Gamma_{i1}^\mu, \Gamma_{i2}^\mu, \dots, \Gamma_{i\ell}^\mu$ ).

Оценка (3.2.4) для  $\Gamma_{ij}^\mu$  на выходе нейрона 3-го уровня практически совпадает с оценкой для классической модели оператора вычисления оценок.

Незначительное отличие (3.2.4) от классического случая алгоритма вычисления оценок состоит в том, что вместо произведения  $(p_e, \bar{1}_e)$ , как, например, в схеме, использованной в [53, с. 1491], в правой части формул (3.2.4) неявно, и, соответственно, для (3.2.3) явно присутствует множитель  $b = (p_e, x_e)$ , где  $e = (t_1, t_2, \dots, t_{|e|})$ ,  $x_e = (x_{t_1}, x_{t_2}, \dots, x_{t_{|e|}})$ ,  $\bar{1}_e = (1_{t_1}, 1_{t_2}, \dots, 1_{t_{|e|}})$ .

Будем считать, что операторы вычисления оценок семейства  $\mathfrak{M} = \{A(\mu, \Omega, \varepsilon^n, p^n, \gamma^m)\}$  вычисляют оценку  $\Gamma_{ij}^\mu$  способом, который передаёт формула (3.2.3). Кроме того, скажем, что, далее, для нас важно, что, как выход нейрона 1-го и 2-го уровней  $\mu$ -блока, так и результирующая оценка  $\Gamma_{ij}^\mu$  в (3.2.2)-(3.2.4), должны быть ограничены промежутком  $(-1, 1]$ . Отрицательные значения оценки  $\Gamma_{ij}^\mu$  могут возникнуть, например, в случае задания отрицательных, тормозящих  $(p_i \leq 0, i \in \{1, 2, \dots, n\})$  весов среди весов  $p_1, p_2, \dots, p_n$  входов нейронов 1-го уровня. В нашем же случае с отрицательными весами и оценками нейронов мы можем столкнуться при построении 4-го уровня сети (рисунок 3.4.1) для операторов квази-базиса.

Отметим, что для нейронов 1-го уровня в качестве активационной функции можно использовать и функцию:

$$F_{e_a}^{\bar{\varepsilon}}(x', x) = \begin{cases} b, & \text{если } |x'_{t_1} p_{t_1} - x_{t_1} p_{t_1}| \leq \varepsilon_{t_1} \& \dots \& |x'_{t_{|e_a|}} p_{t_{|e_a|}} - x_{t_{|e_a|}} p_{t_{|e_a|}}| \leq \varepsilon_{t_{|e_a|}} \\ 0, & \text{если по крайней мере одно из условий нарушено} \end{cases}$$

где  $\bar{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ ,

$$e_a = (t_{i_1}, t_{i_2}, \dots, t_{i_{|e_a|}}), t_{i_j} \in \{1, 2, \dots, n\}.$$

Для нейрона 3-го уровня, активационная функция это  $R$ , однако активационной функцией может быть и решающее правило  $R^*(\theta_{1j}, \theta_{2j})$ .

### 2.3 Построение корректного алгоритма для $\Omega$ -регулярных задач

Итак, рассмотрим семейство операторов вычисления оценок  $\mathfrak{M} = \{A(\mu, \Omega, \varepsilon^n, p^n, \gamma^m)\}$ , где для каждого оператора  $A$  из  $\mathfrak{M}$  веса обучающих объектов  $\gamma_1, \gamma_2, \dots, \gamma_m; \gamma_i \geq 0, i = 1, 2, \dots, m$  могут зависеть от класса, которому принадлежат объекты из  $X^m = \{x^1, x^2, \dots, x^m\}$ , и определяются при задании оператора  $A$ . Остальные параметры алгоритма имеют известный смысл, о котором также уже упоминалось.

Пусть теперь  $A, B (A \neq B)$  это множества, определим  $\nabla(A, B) = B \setminus A$ , если  $A \subsetneq B$  и  $\nabla(A, B) = A \setminus B$ , в противном случае. Пусть для каждой задачи  $u = (I_0, X^q) \in U$ , как и в [2, с. 13; 4, с. 41] выполнено:

- 1)  $C_{j'} \neq C_{j''} (j' \neq j'', 1 \leq j', j'' \leq \ell)$  и
- 2)  $X^m \cap X^q = \emptyset$ .

Зафиксируем  $\Omega = \{e_a\}$  произвольную, непустую систему подмножеств, где  $e_a \subset \{1, 2, \dots, n\}$ . Пусть  $\Omega = \{e_a\}$ ,  $e_a \subset \{1, 2, \dots, n\}$  и  $e, e' \in \Omega$ , обозначим

$$D_e = \bigcup_{e' \neq e} e'$$

и, если  $\Omega = e_a$ , то полагаем  $D_e = \phi$ .

Пусть далее  $q > 1$ ,  $\ell > 1$ ,  $m > 1$ ,  $n > 1$ . Задача  $u = (I_0, X^q)$  является  $\Omega$ -регулярной, если

I. для каждой пары классов  $C_j, C_t$ , ( $1 \leq j, t \leq \ell, t \neq j$ ) найдутся такие  $x \in \nabla(C_j, C_t)$ ,  $e \in \Omega$ ,  $i_\lambda \in e$ ,  $i_\lambda \notin D_e$  и  $x_{i_\lambda} \neq 0$ :

II. для каждой пары  $x', x'' \in X^q$  ( $x' \neq x''$ ) найдутся такие  $x \in X^m$ ,  $e \in \Omega$ ,  $i_r \in e$ : а)  $i_r \notin D_e$ ,  $x_{i_r} \neq 0$       б)  $\rho(ex', ex) \neq \rho(ex'', ex)$

Заметим, что поскольку  $\Omega$  фиксирована, то в качестве системы опорных множеств любого рассматриваемого ниже оператора вычисления оценок. будет использована именно  $\Omega$ . Пусть далее  $B(i, j)$  оператор квази-базиса  $i = 1, 2, \dots, q$ ;  $j = 1, 2, \dots, \ell$  с матрицей  $\Gamma = \|\Gamma_{rs}(i, j)\|_{q \times \ell}$ , для которой выполняются соотношения:  $\Gamma_{ij}(i, j) = 1$ ,  $|\Gamma_{rs}(i, j)| < 1$ ,  $(r, s) \neq (i, j)$ .

Ниже займёмся построением  $B(i, j)$ ,  $i = 1, 2, \dots, q$ ;  $j = 1, 2, \dots, \ell$ , помня об упомянутой нейросетевой парадигме. Говоря об этом, имеем в виду, во-первых, ассоциацию оператора вычислительных оценок с  $\mu$ -блоком, который воспроизводит те же вычисления, что и ассоциированный с этим блоком оператор вычисления оценок ( $\mu$ -оператор вычисления оценок). Во-вторых, каждый нейрон  $\mu$ -блока не должен вырабатывать большое суммарное значение синаптической силы. И кроме того, матрица любого оператора алгебры  $\vartheta$ , используемого при построении оператора  $A^*$  ( $\mathcal{A}^* = A^* \circ R^*$ ), включая матрицу самого оператора  $A^*$ , не должны иметь элементов с большими модулями. Здесь имеются в виду те операторы, которые воспроизводят операции  $const \cdot B$ ,  $B' + B''$ ,  $B^k$  алгебры  $\vartheta$ , вовлекаемые в процесс в ходе построения  $A^*$ . Для нас упомянутые значения, как договорились, должны быть заключены в промежутке  $(-1, 1]$ .

Имеет место [55]

**Теорема 1.** Пусть задана непустая система  $\Omega = \{e_a\}$  и пусть  $u = (I_0, X^q) \in U$  произвольная  $\Omega$ -регулярная задача, а  $f = \|f_{ij}\|_{q \times \ell}$  это матрица классификации задачи  $u$ . Тогда алгоритм (оператор) (1.3) с  $k$  выбранным согласно (1.4) является корректным для задачи  $u$ .

**Доказательство.** Пусть  $u = (I_0, X^q)$  – произвольная  $\Omega$ -регулярная задача, покажем, что в алгебре  $\mathfrak{A}\{\mathcal{A}\}$  для  $u$  существует алгоритм, дающий точное решение задачи  $u$  (для заданной контрольной выборки). Проведём доказательство теоремы используя  $\mu$ -операторы, которые участвуют в построении операторов квази-базиса,  $\{B(i, j)\}_{1,1}^{q,\ell}$  алгебры  $\mathfrak{A}\{\mathcal{A}\}$ . Заметим, что должный выбор параметров  $\mu$ -оператора вычисления оценок позволит «сэкономить» при подсчете оценок по формуле (2.3).

Итак, пусть  $q > 1, \ell > 1, m > 1, n > 1$ . Каждый из операторов  $B(i, j), i = 1, 2, \dots, q; j = 1, 2, \dots, \ell$  квазибазиса, согласно схемы [3, с. 762], есть сумма:  $B(i, j) = B^j + B_i^j$ , где

$$B^j = \sum_{t \in \{1, 2, \dots, \ell\} \setminus \{j\}} B_{jt}, \quad B_i^j = \sum_{\tau \in \{1, 2, \dots, \ell\} \setminus \{i\}} B_{i\tau}^j \quad (3.3.1)$$

В свою очередь в (3.3.1) каждый из операторов  $B_{jt}, j = 1, 2, \dots, \ell; t = 1, 2, \dots, j-1, j+1, \dots, \ell (j = 1, t \geq 2)$  является  $\mu$ -оператором вычисления оценок. Каждый из операторов  $B_{i\tau}^j, j = 1, 2, \dots, \ell; i = 1, 2, \dots, q; \tau = 1, 2, \dots, i-1, i+1, \dots, q (i = 1, \tau \geq 2)$  либо  $\mu$ -оператор вычисления оценок, либо разность двух  $\mu$ -операторов вычисления оценок. При построении  $B_{jt}, j = 1, 2, \dots, \ell; t = 1, 2, \dots, j-1, j+1, \dots, \ell$ , матрица которого есть  $\|\Gamma'_{kh}\|_{q \times \ell}$  примем во внимание условие I определения  $\Omega$ -регулярной задачи и рассмотрим два случая:

а.1)  $C_j \subsetneq C_t$ , тогда имеется объект  $x_v \in C_t \setminus C_j$ . Положим  $\mu = 0, \gamma(x_v) = \sigma > 0, \gamma_i = 0, i = 1, 2, \dots, v-1, v+1, \dots, m; p_{i_\lambda} = 1/(\ell-1), p_j = 0; j = 1, 2, \dots, i_\lambda-1, i_\lambda+1, \dots, n; \varepsilon = 0$ .

а.2)  $C_j \not\subset C_t$ , тогда имеется объект  $x_v \in C_j \setminus C_t$ . Положим  $\mu = 1, \varepsilon = n$  значения остальных параметров те же, что и в случае а.1)

В а.1), а.2) номер  $i_\lambda$  это тот номер признака, о котором идёт речь в п.1. определения  $\Omega$ -регулярной задачи. Подсчитаем, используя формулу (3.2.3), оценки элементов матрицы  $\|\Gamma'_{kh}\|_{q \times \ell}$  для случаев а.1) и а.2):

$$\Gamma'_{kj} = \frac{\sigma}{\ell-1}, \Gamma'_{kt} = 0, k = 1, 2, \dots, q;$$

и для всех элементов матрицы  $\|\Gamma'_{kh}\|_{q \times \ell}$  верно:

$$\Gamma'_{kh} \leq \frac{\sigma}{\ell-1}, k = 1, 2, \dots, q; h = 1, 2, \dots, \ell.$$

Оценивая далее элементы матрицы  $\|\Gamma_{kh}\|_{q \times \ell}$  оператора  $B^j$ , учтем оценки для  $\|\Gamma'_{kh}\|_{q \times \ell}$  и получим:

$$\Gamma_{kj} = \sigma, \Gamma_{kh} < \sigma; k = 1, 2, \dots, q; h = 1, 2, \dots, j-1, j+1, \dots, \ell.$$

При наших построениях параметр  $\sigma$  в найденных оценках не должен превосходить 1.

Перейдём к операторам  $B_{i\tau}^j, j = 1, 2, \dots, \ell; i = 1, 2, \dots, q; \tau = 1, 2, \dots, i-1, i+1, \dots, q$  (при  $i = 1, \tau \geq 2$ ). Кстати говоря, возможно, что некоторые из них также являются  $\mu$ -операторами вычисления оценок и обратимся к условию II определения  $\Omega$ -регулярной задачи.

Зафиксируем произвольную пару  $x^i, x^\tau \in X^q$  объектов из контроля ( $\tau \neq i$ ). По условию для этой пары найдётся объект  $x \in X^m$ , найдутся множество  $e \in \Omega$  и признак  $i_r \in e$ , такие, что верны условия II. а), б) определения  $\Omega$ -регулярной задачи, причем условие б) означает, что имеет место одно из двух неравенств:

$$\text{b1.) } \rho(ex, ex^i) > \rho(ex, ex^\tau)$$

$$\text{b2.) } \rho(ex, ex^i) < \rho(ex, ex^\tau)$$

Случай б1.) и условие  $x \in C_j$  дают следующую последовательность значений параметров:

$$\mu = 1, \gamma(x) = \sigma > 0, \gamma(x') = 0, \forall x' \in X^m \setminus \{x\};$$

$$p_{i_r} = \frac{1}{q-1}, p_j = 0, j = 1, 2, \dots, i_r - 1, i_r + 1, \dots, n, \varepsilon = n - \rho(ex, ex^i).$$

Тогда для элементов матрицы  $\|\Gamma_{uh}(i, j, \tau)\|_{q \times \ell}$   $\mu$ -оператора вычисления оценок  $B^j_{i\tau}$  имеют место оценки:

$$\Gamma_{ij}(i, j, \tau) = \sigma/(q-1), \quad \Gamma_{\tau j}(i, j, \tau) = 0, \quad (3.3.2)$$

$$\Gamma_{uh}(i, j, \tau) \leq \sigma/(q-1), \quad u = 1, 2, \dots, q; i = 1, 2, \dots, \ell; \quad (3.3.3)$$

Пусть далее выполнено всё то же условие б1.), но теперь  $x \in \bar{C}_j$ , определим  $B^j_{i\tau} = B^j_{1i\tau} - B^j_{2i\tau}$  как разность двух  $\mu$ -операторов вычисления оценок. Для  $B^j_{1i\tau}$  определим  $\mu = 0, \varepsilon = 0$ , остальные параметры, т.е.  $\gamma_1, \gamma_2, \dots, \gamma_m$  и  $p_1, p_2, \dots, p_n$ , такие как в предыдущем случае. Для  $B^j_{2i\tau}$  параметры зададим так же, как для  $B^j_{1i\tau}$ , кроме параметра  $\varepsilon: \varepsilon = n - \rho(ex, ex^i)$ . Выбор параметров даёт следующие оценки для матриц  $\|\Gamma_{uh}(1)\|_{q \times \ell}, \|\Gamma_{uh}(2)\|_{q \times \ell}, \|\Gamma_{uh}(i, j, \tau)\|_{q \times \ell}$  операторов  $B^j_{1i\tau}, B^j_{2i\tau}, B^j_{i\tau}$ :

$$\Gamma_{ij}(1) = \sigma/(q-1); \quad \Gamma_{\tau j}(1) = \sigma/(q-1); \quad (3.3.4)$$

$$\Gamma_{ij}(2) = 0, \quad \Gamma_{\tau j}(2) = \sigma/(q-1); \quad (3.3.5)$$

$$\Gamma_{ij}(i, j, \tau) = \sigma/(q-1), \quad \Gamma_{\tau j}(i, j, \tau) = 0; \quad (3.3.6)$$

$$\begin{aligned} \Gamma_{uh}(1) &\leq \sigma/(q-1), & \Gamma_{uh}(2) &\leq \sigma/(q-1), \\ |\Gamma_{uh}(i, j, \tau)| &\leq \sigma/(q-1); \\ u &= 1, 2, \dots, q; i = 1, 2, \dots, \ell; \end{aligned} \quad (3.3.7)$$

Далее, при условиях б2.) и  $x \in C_j$  определим оператор  $B^j_{i\tau} = B^j_{1i\tau} - B^j_{2i\tau}$  и зададим параметры  $\mu$ -оператора вычисления оценок.  $B^j_{1i\tau}, B^j_{2i\tau}$ . Для  $B^j_{1i\tau}$  выберем параметры как в случае б1.) и  $x \in C_j$ . Для  $B^j_{2i\tau}$  параметр  $\varepsilon = n - \rho(ex, ex^\tau)$ , остальные параметры те же, что и для  $B^j_{1i\tau}$ . Нетрудно видеть, что

для элементов матриц  $\|\Gamma_{uh}(1)\|_{q \times \ell}, \|\Gamma_{uh}(2)\|_{q \times \ell}, \|\Gamma_{uh}\|_{q \times \ell}$  операторов  $B^j_{1i\tau}, B^j_{2i\tau}, B^j_{i\tau}$  имеют место оценки (3.3.4)-(3.3.7) соответственно.

Пусть теперь при условии b.2.) объект  $x \in \bar{C}_j$ . Определим  $B^j_{i\tau}$  как  $\mu$ -оператор вычисления оценок, при  $\mu = 0$ , с параметрами, как и в предыдущем случае (оператор  $B^j_{2i\tau}$ ), в том числе  $\varepsilon = n - \rho(ex, ex^T)$ . Нетрудно показать, что для элементов матрицы  $\|\Gamma_{uh}(i, j, \tau)\|_{q \times \ell}$  оператора  $B^j_{i\tau}$  имеют место оценки (3.3.2)-(3.3.3). Обратившись к (3.3.2)-(3.3.3), (3.3.6)-(3.3.7) видим, что для матрицы  $\|\Gamma_{uv}(i, j)\|_{q \times \ell}$  оператора  $B^j_i$  верно:

$$\Gamma_{ij}(i, j) = \sigma, \quad |\Gamma_{uj}(i, j)| < \sigma, \quad u = 1, 2, \dots, i - 1, i + 1, \dots, q; \quad (3.3.8)$$

$$|\Gamma_{uv}(i, j)| \leq \sigma, \quad u = 1, 2, \dots, q; \quad v = 1, 2, \dots, \ell. \quad (3.3.9)$$

И, наконец, учитывая оценки для матрицы  $\|\Gamma_{kh}\|_{q \times \ell}$  оператора  $B^j$  и оценки (3.3.8)-(3.3.9), имеем для матрицы  $\Gamma$  оператора квазибазиса  $B(i, j)$ :

$$\begin{aligned} \Gamma_{ij}(i, j) &= 2\sigma, \\ |\Gamma_{rs}(i, j)| &< 2\sigma, \quad (r, s) \neq (i, j); \quad 1 \leq r \leq q, \quad 1 \leq s \leq \ell; \end{aligned}$$

Задание должным образом  $\sigma$  даёт нужные соотношения для элементов матриц квазибазиса. Приняв во внимание Теорему 1 из [4, с. 42], значение  $\sigma = 1/2$ , ограничения на  $\theta_1, \theta_2$ :  $0 < \theta_1 \leq \theta_2 < 1, \theta_1 \leq (1/2)(1 - \theta_2)$ , получаем, что алгоритм  $\mathcal{A}^*$  вида (1.3) с  $k$ , выбранным согласно (3.1.4), является корректным для  $\Omega$ -регулярной задачи  $u \in U$ . И, если в матрице  $f$  некоторое  $f_{rs} = 1$  ( $r \in \{1, 2, \dots, q\}, s \in \{1, 2, \dots, \ell\}$ ), то для элемента  $\varphi_{rs}$  матрицы  $\varphi = \|\varphi_{rs}\|_{q \times \ell}$  оператора  $A^*$  выполняется:

$$\theta_2 < \varphi_{rs} = (\theta_1 + \theta_2) \sum_{i=1}^q \sum_{j=1}^{\ell} f_{ij} \cdot \Gamma_{rs}^k(i, j) \leq \theta_1 + \theta_2 + a^k \cdot q \cdot \ell \cdot (\theta_1 + \theta_2) \leq 1,$$

а в случае, если  $f_{rs} = 0$ , имеем:  $|\varphi_{rs}| < \theta_1$ . Теорема доказана.

Теорема даёт основание заключить, что выбор системы  $\Omega$ , оказывает влияние на множество задач, для которых в алгебре  $\mathcal{Y}$  строится корректный алгоритм.

## 2.4 Построение пространственной нейросети для $\Omega$ -регулярных задач и её особенности

Теорема 1 показывает, какова должна быть структура нейросети (рисунок 3.4.1), воспроизводящей вычисления, осуществляемые корректным алгоритмом. Используя результат вычислений на 2-м уровне  $\mu$ -блока (см. 3.2, а



также [55, р. 166]) на сумматоре нейрона 3-го уровня  $\mu$ -блока (всего таких нейронов  $\ell$ ) при  $\mu = 0$  или  $\mu = 1$ , как мы видели, образуется оценка:

$$\Gamma_{rj}^{\mu} = \sum_{x' \in C_j^{\mu}} w(x') \sum_{e \in \Omega} F_{\mu}^{\varepsilon}(ex', ex^r), \quad j = 1, 2, \dots, \ell,$$

совпадающая (см. формулу (3.2.3) с оценкой алгоритма вычисления оценок [2, с. 19; 6, с. 145] для объекта  $x^r \in X^q$  за класс  $C_j^{\mu}$ ,  $r = 1, 2, \dots, q$ . И кроме того, в нашем случае, для  $\Omega$ -регулярной задачи, при соответствующем выборе параметра  $\sigma > 0$ , выполняется  $\Gamma_{rj}^{\mu} \in (-1, 1]$ ,  $j = 1, 2, \dots, \ell$  (см. теорема 1).

Согласно Теореме 1 (см. доказательство выше), при надлежащем выборе  $\sigma$ , это же условие выполняется и для элементов матриц для всех операторов, участвующих в построении оператора кази-базиса  $B(i, j)$ ,  $1 \leq i \leq q; 1 \leq j \leq \ell$ , в том числе и для элементов самой матрицы квазибазиса  $\|\Gamma_{rh}(i, j)\|_{q \times \ell}$ .

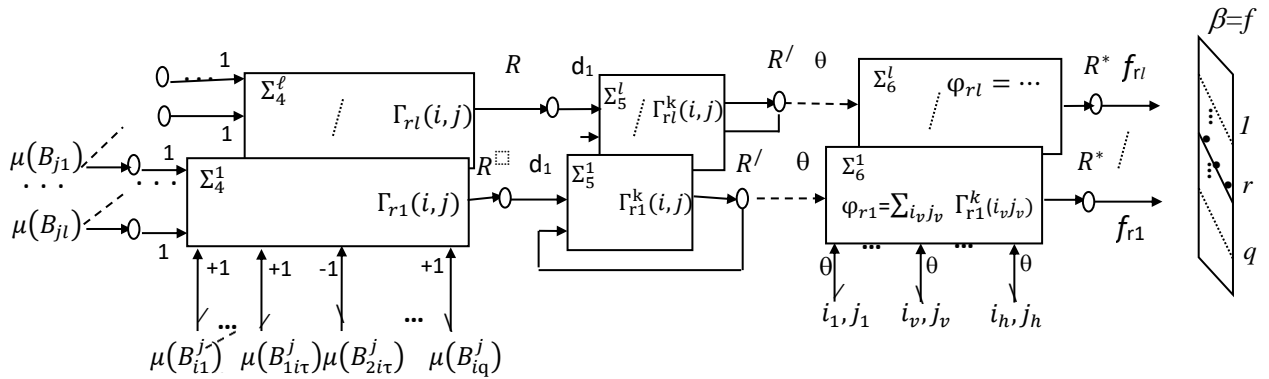


Рисунок 3.4.1 – Структура нейросети, воспроизводящая вычисления, осуществляемые корректным алгоритмом

Теперь, учитывая 3-и уровня  $\mu$ -блока, формулу (3.1.5) и доказательство Теоремы 1, дополним конструкцию сети тремя недостающими уровнями, где  $\theta = \theta_1 + \theta_2$ .

Преследуя цель воспроизвести сеть вычисления осуществляемые корректным алгоритмом, первоначально заметим, что для нейронов 3-го, 4-го, уровней сети, активационной функцией является, активационная функция типа  $R$ . Для нейронов 5-го уровня, активационной функцией является функция  $R'$  (см. рисунок 3.4.1 и формулу (3.1.5)). Для последнего, 6-го уровня сети активационная функция это  $R^*(\theta_1, \theta_2)$ , при этом величины  $\theta_1, \theta_2$  удовлетворяют введенным ограничениям (см.п.1), а также и дополнительному условию:  $\theta_1 \leq (1 - \theta_2)/2$ . Нейроны 4-го, 5-го и 6-го уровней образуют расслоения (в ХОУ), с числом нейронов  $\ell$  в каждом расслоении (рисунок 3.4.1).

Вычислительный процесс в сети составляют два цикла: внешний, по объектам выборки  $X^q$  и внутренний, по объектам выборки  $X^m$ . При этом, в

отличии от классической схемы построения операторов квазибазиса [4, с. 40; 6, с. 211] (см. также п. 3.2), здесь, для любого  $x^r, r = 1, 2, \dots, q$ ; внутренний цикл по каждому  $\mu$ -блоку состоит лишь из одной итерации, определяемой одним объектом обучения, а именно, тем, вес которого равен  $\sigma > 0$  (см. доказательство теоремы 1), что, без сомнения, даст ощутимую экономию в вычислениях.

Очередной шаг с  $x^r \in X^q$  завершается, когда для  $x^r$  исчерпаются внутренние уровни сети и активационная функция  $R^*$ , по вычисленной строке матрицы  $\varphi$ , дает строку матрицы  $\beta$  для  $x^r$ . Если и внешний цикл завершен, а это произойдет при исчерпании  $X^q$ , то, в итоге,  $R^*$  по  $\varphi$  даст матрицу  $\beta = \|\beta_{ij}\|_{q \times \ell}$  с  $\beta_{ij}$ , вообще говоря, из  $\{1, 0, \Delta\}$ . В нашем же случае, вычисления в сети должны выполняться в соответствии с правилами построения  $\mathcal{A}^*$  и конечная цель здесь, получить матрицу  $\beta$ , совпадающую с матрицей классификации  $f$  задачи  $u$ . Итак, имеет место

*Теорема 2.* Пусть задана непустая система  $\Omega = \{e_a\}$  и пусть  $u = (I_0, X^q) \in U$  произвольная  $\Omega$ -регулярная задача, а  $f = \|f_{ij}\|_{q \times \ell}$  это матрица классификации задачи  $u$ . Тогда для задачи  $u$  существует  $6$ -и уровневая пространственная сеть с не более чем  $\ell \cdot (\ell - 1) + 2\ell \cdot q(q - 1)$   $\mu$ -блоками, которая обеспечивает вычисление сетью матрицы  $\beta$ , совпадающей с матрицей  $f$  задачи  $u$ .

*Доказательство.* Опираясь на доказательство Теоремы 1, где, по существу, явно обосновано построение при заданных  $\Omega$  и  $u = (I_0, X^q) \in U$  первых трех уровней сети ( $\mu$ -блоки), рассмотрим этап построения сети для операторов квазибазиса  $\{B(i, j), 1 \leq i \leq q; 1 \leq j \leq \ell\}$ , именно, для тех  $(i, j): f_{ij} \neq 0$ . Для этого, учтя вышеприведенные рассуждения, обратимся к рисунку 3.4.1, и формуле (3.1.5.) из п. 3.1.

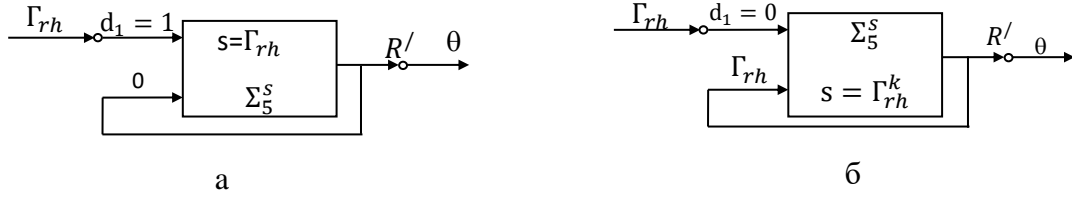
Как мы знаем, оператор  $B(i, j) = \sum_{t \neq j} B_{jt} + \sum_{\tau \neq i} B_{i\tau}^j$ , где  $B_{jt} (t \neq j)$  это оператор вычисления оценок, а  $B_{i\tau}^j (\tau \neq i)$  либо оператор вычисления оценок либо разность двух операторов вычисления оценок, т.е. в последнем случае имеем:  $B_{i\tau}^j = B_{1i\tau}^j - B_{2i\tau}^j$ . Не трудно (см. п. 3.2, 3.3) задать  $\mu$ -блок для каждого из операторов вычисления оценок  $B_{1i\tau}^j, B_{2i\tau}^j, B_{i\tau}^j, B_{jt}$ , пусть это будут:  $\mu(B_{1i\tau}^j), \mu(B_{2i\tau}^j), \mu(B_{i\tau}^j), \mu(B_{jt})$ , каждый, из которых имеет  $\ell$  выходов.

Каждый такой выход связан с соответствующим входом нейрона в расслоении нейрона 4-го уровня (рисунок 3.4.1). Нейроны для операторов вычисления оценок  $B_{jt}$  и  $B_{i\tau}^j$ , и операторов вычисления оценок  $B_{1i\tau}^j, B_{2i\tau}^j$  (если имеются, см. рисунок 3.4.1) непосредственно соединяются с фрагментом сети, соответствующим оператору  $B(i, j)$ . Каждый нейрон (всего их  $\ell$ ) фрагмента имеет не более  $2q + \ell - 3$  входов с синаптическими весами равными  $+1$  или  $-1$ .

На 5-м уровне сети реализуется операция возведения в степень, каждого оператора квазибазиса:  $B^k(i, j)$ , для которого  $f_{ij} \neq 0$ . Нейрон этой операции имеет два входа, синаптический вес одного из них это  $d_1$ . И, именно, этот вход

связан с выходом, соответствующего нейрона в расслоении 4-го уровня (см. формулу (3.1.5) и рисунки 3.4.1, 3.4.2).

Пусть  $k_0 = [(\ln q + \ln \ell + \ln(\theta_1 + \theta_2) - \ln \theta_1) / |\ln a|] + 1$ .



а – первоначальное состояние при  $d_1 = 1$ ; б – состояние при  $d_1 = 0$

Рисунок 3.4.2 – Нейрон 5-го уровня, осуществляющий операцию  $B^k$

Частично заданную диагональную активационную функцию  $R'$ , которая выдает нужное значение определим так:

$$R'(s) = \Gamma_{rh}^k, \text{ если } k = k_0 \text{ и } R'(s) = \Delta, \text{ если } k < k_0,$$

и, в случае завершения процесса, до передачи по линии обратной связи, вычисленного значения  $\Gamma_{rh}^{k_0}$ , вес второго входа должен принять значение 0 (очистка сумматора). К ситуации на рисунке 3.4.2а с  $d_1 = 1$  возвращаемся при обновлении контрольного объекта. Выход функции  $R'(s)$  не определен ( $R'(s) = \Delta$ ), в случае, если величина  $k$  не достигла значения  $k_0$ , что подразумевает лишь передачу на свой же вход последнего вычисленного значения  $s$  для следующей итерации (рисунок 3.4.2б). При этом как об этом уже говорилось, сумматор нейрона сохраняет вычисленное значение лишь непродолжительное время.

Первоначально, как видим,  $d_1 = 1$ , чтобы передать на сумматор  $\Sigma_5^s$ , вычисляемую ранее оценку  $\Gamma_{rh}$ ,  $h = 1, 2, \dots, \ell$ , синаптический вес второго входа равен 0 (рисунок 3.4.2а). Далее  $d_1$  меняется на 0, а  $\Gamma_{rh}$  станет синаптическим весом второго входа (рисунок 3.4.2б). К нему и ведёт обратная связь от выхода нейрона, так посредством активационной функции по заданным  $k_0 = [(\ln q + \ln \ell + \ln \theta - \ln \theta_1) / |\ln a|] + 1$  и  $a = \max_{i,j} \max_{(r,h) \neq (i,j)} |\Gamma_{rh}(i,j)|$  нейрон вычисляет элемент матрицы оператора  $B^k(i,j)$ . Выход данного нейрона «усиливается» синаптическим весом  $\theta = \theta_1 + \theta_2$  (при  $f_{i_v j_v} = 1, v = 1, 2, \dots, h$ ) входа нейрона последнего, 6-го уровня, а сама сеть пошагово (для  $x^r \in X^q, r = 1, 2, \dots, q$ ) вычисляет строки матрицы  $f = \|\|f_{rh}\|\|_{q \times \ell}$ , осуществляет те же вычисления, что и корректный для  $u = (I_0, X^q)$  алгоритм  $\mathcal{A}^*$ .

При этом имея ввиду  $\sigma = 1/2$  и  $k = k_0$ , а также ограничения:  $0 < \theta_1 \leq \theta_2 < 1, \theta_1 \leq (1/2)(1 - \theta_2)$ , получаем, что окончательные оценки  $\varphi_{rh}$  ( $\varphi =$

$\|\varphi_{rh}\|_{q \times \ell}$ ), вычисляемые сумматором нейронов 6-го уровня, также заключены в промежутке  $(-1, 1]$ . Таким образом, Теорема 2 доказана.

Заметим, что те же рассуждения, как и в случае Теоремы 2, а также наличие дополнительной информации, доказывают

*Теорема 3. Пусть задана непустая система  $\Omega = \{e_a\}$  и пусть  $u = (I_0, X^q) \in U$  произвольная  $\Omega$ -регулярная задача, а  $f = \|f_{ij}\|_{q \times \ell}$  это матрица классификации задачи  $u$ . Тогда для задачи  $u$  может быть построена  $b$ -и уровневая пространственная сеть с не более чем  $\ell \cdot (\ell - 1) + 2\ell \cdot q(q - 1)$   $\mu$ -блоками, такая, что задание матрицы  $f = \|f_{ij}\|_{q \times \ell}$  задачи  $u$ , а также параметра  $k$ , обеспечивает вычисление сетью матрицы  $\beta$ , совпадающей с матрицей  $f$ .*

Доказательство теоремы 1 основано на теореме 2 из [4, с. 43], сама же теорема 1 лежит в основе доказательства теорем 2, 3. Заметим, что в качестве  $\Omega$  могут выступать  $\Omega_0 = \{\{1\}, \{2\}, \dots, \{n\}\}, \Omega_1 = \{\{1, 2, \dots, n\}\}$  [2, с. 19; 4, с. 43], подмножества  $\Omega_0$  мощности  $k \geq 1$ , базы множества признаков [10, с. 33; 52, с. 713], представительные наборы [6, с. 223] для задач с непересекающимися классами, подмножества тупиковых тестов и др.

Структура сети и значения её параметров, например,  $\mu, \tilde{p}^n, \tilde{\gamma}^m, \varepsilon$  определяются задачей  $u$ , часть из параметров задается в ходе вычислений и может динамически изменяться. Одна из отличительных особенностей, описанной сети, это наличие в её внутренних слоях диагональных активационных функций, заметно упрощающих промежуточные вычисления во внутреннем и внешних циклах.

### 3 РАЗРАБОТКА МОДЕЛИ БАЙЕСОВОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ЗАДАЧИ ОЦЕНКИ РИСКОВ ПОРТФЕЛЯ ЦЕННЫХ БУМАГ

#### 3.1 Модель UBSOM оценки и прогнозирования риска портфеля ценных бумаг на основе синтеза многомерного совместного распределения портфельных компонент

Итак, пусть имеется капитал  $x$ , распределенный среди ценных бумаг  $A = (A_1, A_2, \dots, A_N)$ . Стоимость портфеля ценных бумаг в момент  $t$  определяется как:

$$X_t(b) = b_1 S_t(A_1) + \dots + b_N S_t(A_N),$$

где  $b = (b_1, b_2, \dots, b_N)$ ,  $b_i \geq 0$  – доли ценных бумаг в портфеле;

$S_t(A_i)$  – стоимость ценной бумаги  $A_i$  в момент  $t$ ,  $t = 0, 1, \dots, n$ .

Портфелем ценных бумаг со стартовым капиталом  $x$ ,  $x > 0$  будем называть вектор:

$$B_x = \{(b_1, b_2, \dots, b_N) \mid b_i \geq 0, X_0(b) = x\}$$

Пусть  $d = (d_1, d_2, \dots, d_N)$  вектор удельных весов ценных бумаг в портфеле  $B_x$ , где  $d_i = b_i S_0(A_i) / x$ ,  $i = 1, 2, \dots, N$ . Прирост стоимости портфеля в момент  $t$  определяется как:

$$Z_t = d \cdot q_t = d_1 q_t(A_1) + \dots + d_N q_t(A_N),$$

где  $q_t = (q_t(A_1), \dots, q_t(A_N))$ ;

$$q_t(A_i) = S_t(A_i) - S_{t-1}(A_i) / S_{t-1}(A_i);$$

$$q_t(A_i) \geq -1, t = 1, 2, \dots, n \text{ и}$$

$$q(A_i) = (q_1(A_i), q_2(A_i), \dots, q_n(A_i)), i = 1, 2, \dots, N.$$

Тогда согласно классической модели Марковица [24, р. 81] под риском портфеля ценных бумаг понимается дисперсия  $D$  отклонений стоимости портфеля:

$$Dq(B_x) = \sum_{i=1}^N d_i^2 Dq(A_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^N d_i d_j V_{ij}, \quad (4.1.1)$$

где  $q(B_x) = (q(A_1), \dots, q(A_N))$ ,  $Dq(A_i)$  – дисперсия отклонения стоимости бумаги  $A_i$ ;

$V_{ij}$  – ковариация между стоимостью бумаг  $A_i$  и  $A_j$ , и пусть  $q = q(B_x)$ .

Данная модель имеет несколько существенных допущений, ключевым из которых является предположение о нормальности доходностей финансовых инструментов, входящих в портфель. Однако в действительности большая часть наблюдаемых значений доходностей финансовых инструментов не соответствует нормальному распределению – для них характерны асимметрия, выбросы значений, а также "тяжелые хвосты" в распределении [25, р. 112]. Для

уточнения модели Марковица был предложен ряд альтернативных распределений, таких как распределение Парето, логнормальное распределение [34, р. 49] и др.

Одной из альтернатив является построение функций распределения значений портфеля на основе эмпирических данных. Здесь используются смеси нормальных распределений (гауссовы смеси), которые [27, р. 35] зачастую дают приемлемые результаты при моделировании фактических распределений фондовых индексов. Для нахождения параметров компонент гауссовой смеси применяются методы ядерной оценки плотности (KDE метод), а также EM-алгоритмы в разных вариантах. В то же время данные методы в определенных ситуациях показывают избыточную чувствительность к качеству исходных данных, в том числе и к шумовым "выбросам"[34, р. 49].

С другой стороны известно, что метод BSOM для задачи построения гауссовых смесей демонстрирует [34, р. 49] лучшую вычислительную эффективность по сравнению, например, с EM-алгоритмом. Представляет интерес и применение метода BSOM для задач финансового моделирования. Тогда, в нашем случае, закономерным шагом будет построение модели, основанной на BSOM ( $N$ -мерный случай), для задачи оценки и долгосрочного прогноза риска портфеля. Назовем такую модель для определенности UBSOM.

Метод BSOM был предложен в работах [41, р. 211; 56] для построения оценок функций плотности вероятности. В модели BSOM неизвестная функция плотности вероятности аппроксимируется гауссовой смесью, где отдельный нейрон соответствует одной компоненте в смеси. Пусть в смеси имеется  $K$  компонент  $(w_1, w_2, \dots, w_K)$  с параметрами, записанными в комбинациях  $\Theta = (\theta_1, \theta_2, \dots, \theta_K)$ , где комбинация  $\theta_i = \{m_i, \Sigma_i\}$  включает  $N$ -мерный оценочный вектор средних значений  $m_i$  и ковариационную матрицу  $\Sigma_i$  размерности  $N \times N$ . Плотность условной вероятности для  $w_i$ -й компоненты смеси это [57]:

$$p(Q|w_i, \theta_i) = \frac{1}{(2\pi)^{c/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(Q - m_i)^T \Sigma_i^{-1} (Q - m_i)\right) \quad (4.1.2)$$

где  $Q = (X_1, \dots, X_N)$  – вектор отклонений компонент портфеля, который в момент  $t$  есть  $q_t = (q_t(A_1), \dots, q_t(A_N))$ .

Плотность совместного распределения  $P(q(B_x))$  гауссовой смеси вычисляется как:

$$p(Q|\Theta) = \sum_{i=1}^K p(Q|w_i, \theta_i) P(w_i)$$

где  $P_i = P(w_i)$  – априорная вероятность принадлежности многомерной случайной величины  $Q = (X_1, \dots, X_N)$  к  $w_i$ -й компоненте, здесь  $\sum_{i=1}^K P(w_i) = 1$ . При этом вычисление величин  $m_i, \Sigma_i, P_i$  осуществляется по схеме (4.1.3).

Каждый нейрон в модели BSOM представляет одну компоненту  $w_i$ , ( $1 \leq i \leq K$ ) из смеси и инициализируется равномерным распределением с диагональной ковариационной матрицей. Считаем, что в течение всего периода обучения на вход сети подаются векторы  $q_t = (q_t(A_1), \dots, q_t(A_N))$ ,  $t = 1, 2, \dots, n$ . Номер  $v$  выигравшего нейрона определяется как:

$$v = \arg \max_i \left\{ P(w_i | q_t, \theta_i) = \frac{p(q_t | w_i, \theta_i) P(w_i)}{\sum_{i=1}^K (p(q_t | w_i, \theta_i) P(w_i))} \right\},$$

где  $t$  – номер входного вектора;

$\arg$  – выделяет номер выигравшего нейрона, который соответствует компоненте  $w_v$ , имеющей наибольшую апостериорную вероятность  $P(w_i | q_t, \theta_i)$  среди  $w_i$ ,  $i = 1, 2, \dots, K$ .

Окрестность  $\eta_v$  для выигравшего нейрона определяется по стандартному для самоорганизующихся карт алгоритму. Уравнения для изменения параметров нейронов в модели UBSOM те же, что и в [56, р. 623]:

$$\begin{aligned} m_i(t+1) &= m_i(t) + \alpha_t P(w_i | q_t, \theta_i) (q_t - m_i(t)), \forall i \in \eta_v; \\ \Sigma_i(t+1) &= \Sigma_i(t) + \alpha_t P(w_i | q_t, \theta_i) \left( (q_t - m_i(t))(q_t - m_i(t))^T - \Sigma_i(t) \right), \forall i \in \eta_v; \\ P_i(t+1) &= P_i(t) + \alpha_t (P(w_i | q_t, \theta_i) - P_i(t)), \forall i \in \eta_v; \end{aligned} \quad (4.1.3)$$

где  $\alpha(t) = \frac{\alpha_0}{1+t/\tau}$  – скорость обучения на шаге  $t$ .

Параметры скорости обучения  $\alpha_0$  и  $\tau$  могут быть заданы согласно процедуре, описанной в [56, р. 623]. Начальные значения  $m_i(0)$  и  $\Sigma_i(0)$  выбираются по методу главных компонент (PCA), начальные значения вероятностей  $P_i(0)$  принимаются равными для всех нейронов и задаются согласно равномерному распределению.

В разработанной модели [58] метод BSOM применяется для построения многомерного совместного распределения портфельных компонент  $p(Q | \Theta)$  на основе имеющихся эмпирических данных. Следующий шаг: "погружение" функций плотности распределения в  $(N+1)$ -мерное евклидово пространство ( $B$ -пространство портфеля), где первым  $N$  измерениям соответствуют отклонения  $q_t(A_i)$  портфельных компонент, а  $(N+1)$ -му измерению соответствуют значения функции плотности вероятности в точках  $Q$ . Пусть  $z_t$  – отклонение стоимости портфеля в текущий период от значения в предыдущий:  $z_t = \Delta Z_t / Z_{t-1}$ ,  $t > 1$ . Тогда для каждого  $z \geq -1$  можно вычислить значение поверхностного интеграла который геометрически является проекцией функции плотности распределения, полученной посредством BSOM, на гиперплоскость  $H_z: d \cdot Q = z$ .

$$p(z) = \iint_{d \cdot Q = z} P(q(B_x)) dQ,$$

Таким образом оценка риска портфеля в модели UBSOM заключается в вычислении интеграла  $p(z)$  для всех заданных  $z$  отклонений портфеля. Проекция функции плотности совместного распределения  $p(Q|\Theta)$  на гиперплоскость  $H_z$  для случая  $N = 2$  изображена на рисунке 4.1.1, где осям  $X$ ,  $Y$  соответствуют значения отклонений компонент портфеля от своих значений в предыдущий момент. Ось  $OZ$  соответствует вероятности отклонения стоимости портфеля при заданных отклонениях компонент. Поверхности  $S$  соответствует функция плотности совместного распределения. И, наконец, плоскость на рисунке 4.1.1 это портфельная гиперплоскость  $H_z$  при  $z = 0$ .

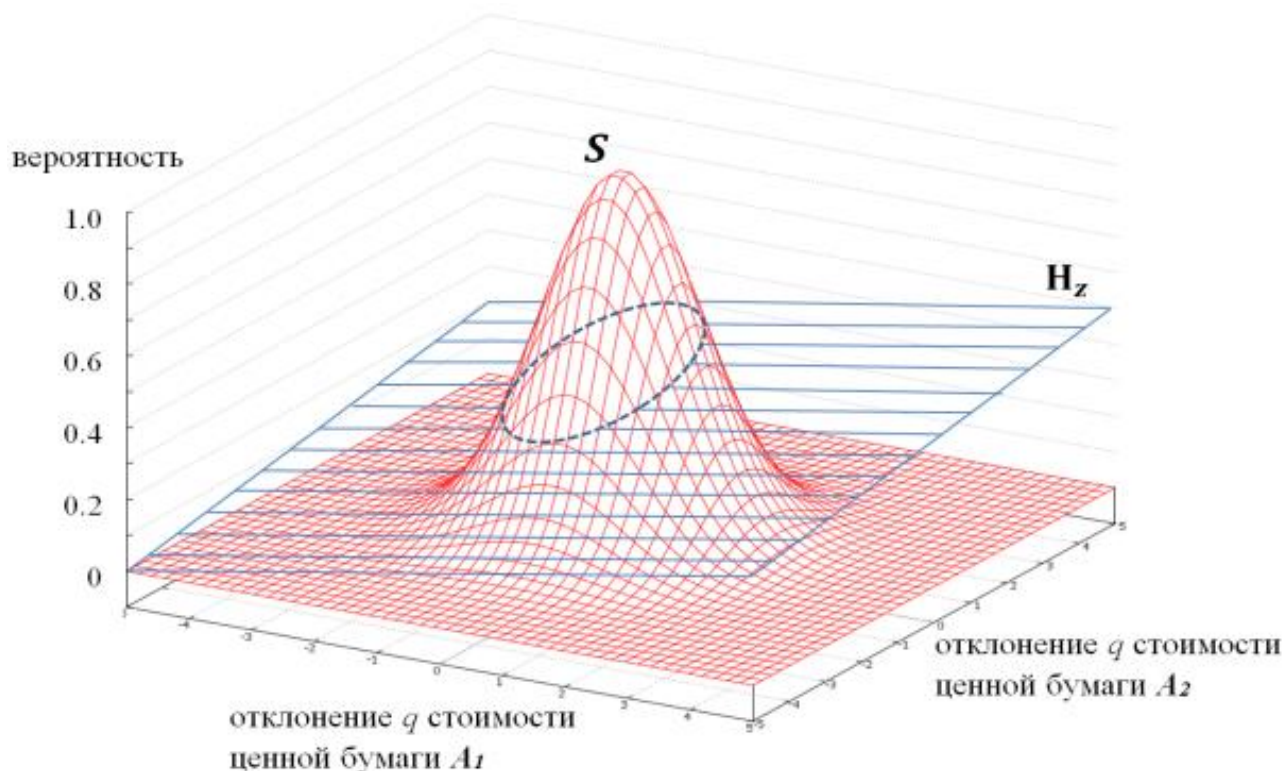


Рисунок 4.1.1 – Графическое представление функции совместного распределения двухкомпонентного портфеля ценных бумаг и гиперплоскости  $H_z$

Преимущество предлагаемого подхода и модели UBSOM по сравнению с одномерными оценками распределения отклонений стоимости портфеля  $B_x$  заключается в том, что одномерные модели, как и те, что рассматриваются в работе, оценивают только амплитудную компоненту проекции вектора портфельных отклонений  $q_t = (q_t(A_1), \dots, q_t(A_N))$ ,  $t = 1, 2, \dots, n$  на вектор удельных весов  $d$ . В отличие от одномерных моделей у нас оценка многомерного совместного распределения портфельных компонент  $q_t$  позволяет учесть и взаиморасположение векторов  $q_t$  и  $d$ .

Для иллюстрации давайте рассмотрим два подхода к непараметрической аппроксимации распределения приростов стоимости портфеля, используя гауссову смесь для одномерного случая с приростами стоимости портфеля и многомерного случая с приростами стоимости финансовых инструментов,



входящих в состав портфеля. Их плотности распределения задаются следующими уравнениями:

$$p_1(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \frac{1}{n} \cdot \sum_{i=1}^n e^{-\frac{(x-b \cdot q_i(A))^2}{2\sigma^2}}$$

$$p_2(x) = \int_{b \cdot q(A)=x; q(A) \geq -1} \frac{1}{(2\pi \cdot \sigma^2)^{\frac{n}{2}}} \cdot \frac{1}{n} \cdot \sum_{i=1}^n e^{-\frac{\|q(A)-q_i(A)\|^2}{2\sigma^2}} dq(A)$$

где  $\sigma$  – некоторое стандартное отклонение используемое для всех гауссианов.

Теперь, для уравнения (4.1.5) мы имеем:

$$b \cdot (q(A) - q_i(A)) = x - b \cdot q_i(A) \quad (4.1.6)$$

и если  $\alpha(q(A), b)$  является углом между векторами  $b$  и  $q(A)$ , то тогда выражение (4.1.6) дает нам длину проекции вектора  $q(A) - q_i(A)$  на вектор  $b$ , т.е. получаем:

$$\|q(A) - q_i(A)\|^2 = (x - b \cdot q_i(A))^2 \cdot (1 + \operatorname{tg}(\alpha(q(A), b))^2) \quad (4.1.7)$$

см. рисунок 4.1.2а.

Теперь, мы можем переписать уравнение (4.1.5) в виде:

$$p_2(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \frac{1}{n} \cdot \sum_{i=1}^n e^{-\frac{(x-b \cdot q_i(A))^2}{2\sigma^2}} \cdot \left[ \int_{b \cdot q(A)=x; q(A) \geq -1} \frac{1}{(2\pi \cdot \sigma^2)^{\frac{n}{2}-1}} \cdot e^{-\frac{(x-b \cdot q_i(A))^2 \cdot \operatorname{tg}(\alpha(q(A), b))^2}{2\sigma^2}} dq(A) \right]. \quad (4.1.8)$$

Обозначим для  $i$ -го компонента:

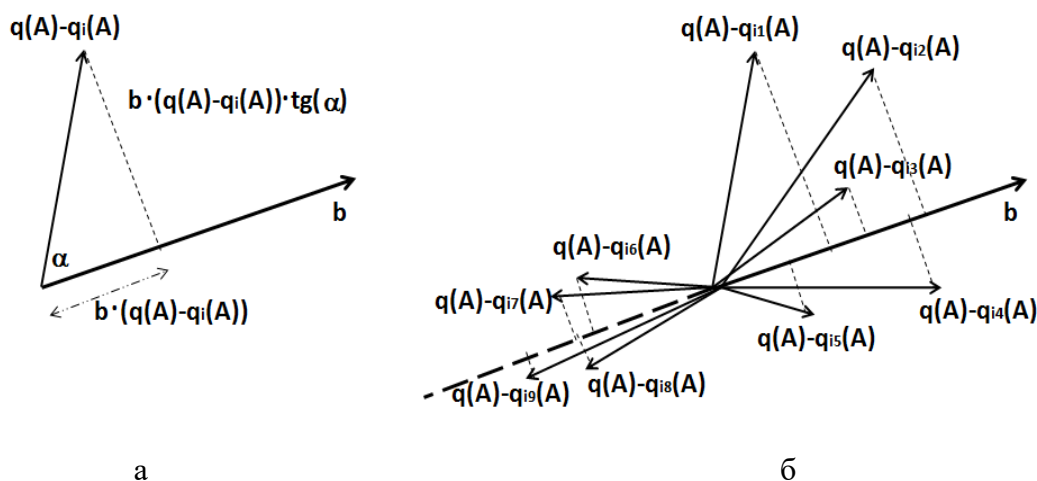
$$w_i(x) = \int_{b \cdot q(A)=x; q(A) \geq -1} \frac{1}{(2\pi \cdot \sigma^2)^{\frac{n}{2}-1}} \cdot e^{-\frac{(x-b \cdot q_i(A))^2 \cdot \operatorname{tg}(\alpha(q(A), b))^2}{2\sigma^2}} dq(A) \quad (4.1.9)$$

и получим:

$$p_2(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \frac{1}{n} \cdot \sum_{i=1}^n w_i(x) \cdot e^{-\frac{(x-b \cdot q_i(A))^2}{2\sigma^2}} \quad (4.1.10)$$

Выражение в (4.1.10) сходно с (4.1.4), однако разница заключается в появлении весового коэффициента  $w_i(x)$  при суммировании. Данное различие между  $p_1(x)$  и  $p_2(x)$  показывает, что для  $p_1(x)$  мы принимаем во внимание только проекции  $q(A) - q_i(A)$  на  $b$ , в то время как в  $p_2(x)$  мы также принимаем во внимание угол между  $q(A) - q_i(A)$  и  $b$ . Другими словами,  $p_1(x)$  учитывает только амплитудную информацию о приростах стоимости портфеля, в то время как  $p_2(x)$  учитывает не только данную амплитудную информацию, но и фазовую компоненту, представленную взаимным расположением векторов  $q_i(A)$ .

Учет фазовой компоненты при оценке риска улучшает точность оценки, основанной на аппроксимации многомерной функции плотности вероятности (рисунок 4.1.2б). В общем случае можно было бы предположить, что аппроксимация многомерной функции плотности вероятности будет менее точной, чем аппроксимация одномерной функции [59-61]. Однако, вклад коррекции, производимой фазовой информацией, улучшает оценку риска, вычисленного путем использования аппроксимации многомерной функции плотности вероятности. Если данное улучшение достаточно велико, оценка риска на основе многомерной аппроксимации будет лучше, чем оценка риска, основанная на одномерной аппроксимации функции плотности вероятности.



а – декомпозиция вектора прироста стоимости портфеля на амплитудную ( $b \cdot (q(A) - q_i(A))$ ) и фазовую ( $b \cdot (q(A) - q_i(A)) \cdot \text{tg}(\alpha)$ ) компоненты; б – гипотетическая ситуация, иллюстрирующая важность фазовых компонент для оценки риска – изменения векторов с положительной проекцией на  $b$  значительно больше, чем изменения векторов с отрицательной проекцией, однако это может не в полной мере быть отражено в суммарной проекции векторов

Рисунок 4.1.2 – Учет фазовой компоненты при оценке риска

Точность аппроксимации повышается при увеличении количества входных данных, при этом повышение происходит быстрее для пространств с меньшей размерностью, по сравнению с пространствами с высокой размерностью [60, р. 6; 61, р. 1209]. Таким образом, корректирующий эффект от использования приростов стоимости финансовых инструментов будет улучшать оценку риска

портфеля в случаях небольшого количества входных данных либо присутствия зашумленных данных.

## 3.2 Вычислительные эксперименты и анализ результатов

3.2.1 Применение UBSOM к данным по финансовым инструментам фондового рынка США

Для проведения вычислительных экспериментов в данном исследовании были использованы данные, предоставленные FINAM Group (Limassol, Cyprus) и содержащие ежедневные цены закрытия торгов за период с 24 января 2008 года до 01 января 2022 года по следующим финансовым инструментам: AAPL - Apple Inc; BAC - Bank of America; GE - General Electric; XOM - Exxon Mobil; PFE - Pfizer Inc.

Данные финансовые инструменты в равных долях (по 20%) были использованы для образования портфеля. Данные были разбиты на непересекающиеся наборы данных, содержащих 100 последовательных элементов в каждом.

Для каждого финансового инструмента во всех наборах мы рассчитали приросты стоимости  $q_t(A_i) = (S_t(A_i) - S_{t-1}(A_i))/S_{t-1}(A_i)$ , при этом значение  $S_0(A_i)$  было присвоено первому (раннему) значению цены финансового инструмента в наборе. Значение стоимости портфеля было рассчитано по формуле (4.2.1):

$$q_{P,t} = \sum_{i=1}^5 0.2 \cdot q_t(A_i) \quad (4.2.1)$$

Для каждого набора данных мы также использовали данные из следующего набора (т.е. значения цен финансовых инструментов за следующие 100 дней) для оценки фактического распределения приростов стоимости портфеля. Фактическое распределение оценивалось путем построения гистограммы, которая использовалась для определения точности оценок риска для рассматриваемого набора данных.

В случае использования более длинных сроков оценки (мы рассмотрели наборы данных размером в 200, 300 и 400 дней) были использованы соответственно более длительные наборы данных для оценки фактического распределения.

Для сравнения точности оценки риска были использованы следующие 5 методов:

1) модель UBSOM. Риск оценивался согласно описанной выше модели на основе аппроксимации многомерной функции плотности вероятности приростов стоимости финансовых инструментов, образующих портфель;

2) оценка по гистограмме. Строится гистограмма приростов стоимости портфеля в рассматриваемом наборе данных, риск оценивается по данной гистограмме;

3) модель Марковица. Расчет дисперсии и ковариаций, а также оценка риска в соответствии с уравнением (4.1.1);

4) аппроксимация нормальным распределением. Аппроксимация распределения приростов стоимости портфеля одним гауссианом;

5) ядерная оценка плотности (KDE). Аппроксимация функции плотности распределения приростов стоимости портфеля комбинацией гауссианов с использованием реализации метода статистическим пакетом NumXL (Spider Financial, Chicago, US).

Модель BSOM была реализована в соответствии с алгоритмом, описанным Утсуги (<http://staff.aist.go.jp/utsugi-a/Lab/BSOM1/sources.html>), который был доработан для работы с многомерными входными данными и настроен для использования 5 гауссианов в смеси. Количество гауссианов было задано на основе эвристического анализа чувствительности и точности BSOM при различном количестве гауссианов в смеси. Рисунок 4.2.1 иллюстрирует пример расчета функции плотности вероятности приростов стоимости портфеля упомянутыми выше методами.

Учитывая, что фактическое распределение не является нормальным, для сравнения оценок риска, полученных упомянутыми выше методами, вычислим значение  $x_0$ , такое что:

$$\int_{x_0}^{+\infty} p(x)dq(A) = 0.975$$

где  $p(x)$  – вычисленное аппроксимированное значение функции плотности вероятности приростов стоимости портфеля.

Другой, вычислительно более легкий способ заключается в следующем. Возьмем симметричный интервал отклонений стоимости портфеля около 0 и вычислим интеграл по этому интервалу упомянутыми выше методами. Например, взяв интервал  $[-0.025, 0.025]$  значения интеграла показывают величину вероятности изменения стоимости портфеля, вычисленную каждым методом, между 97,5 и 102,5% от начального значения стоимости портфеля.

Вычислим ту же величину по гистограмме, построенной по набору данных, следующим по времени за рассматриваемым, и будем использовать ее как референсное значение фактического риска портфеля. Чем ближе оценка риска портфеля, полученная каким-либо методом к фактическому риску портфеля, тем выше его точность. В соответствии с нашим теоретическим анализом мы ожидаем, что точность оценки рисков, полученная BSOM, будет выше чем у остальных рассмотренных методов. При увеличении временного разрыва между оцениваемым и референсным наборами данных точность оценки риска падает в виду изменения параметров распределения и накопления шума в данных. Также мы ожидаем, что оценки, полученные BSOM окажутся более точными при увеличении упомянутого выше временного разрыва.

Мы измерили среднеквадратичную ошибку и дисперсию ошибки для каждой оценки риска, полученной упомянутыми выше методами, по 14 наборам данных. Точность оценок также была измерена для различных

временных интервалов от 10 до 400 дней. Для этих целей оценки сравнивались с фактическим распределением приростов стоимости портфеля в следующем временном периоде (референсным значением риска). Результаты сравнения показаны на рисунке 4.2.1.

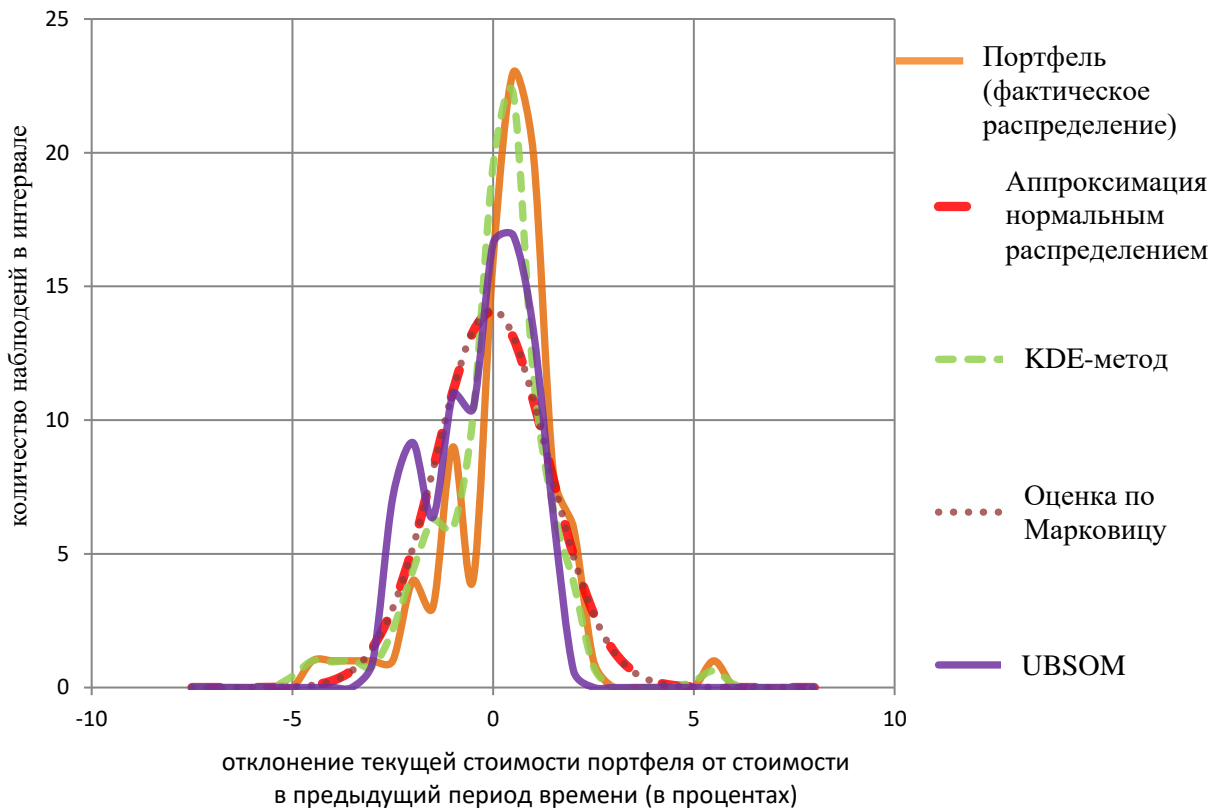
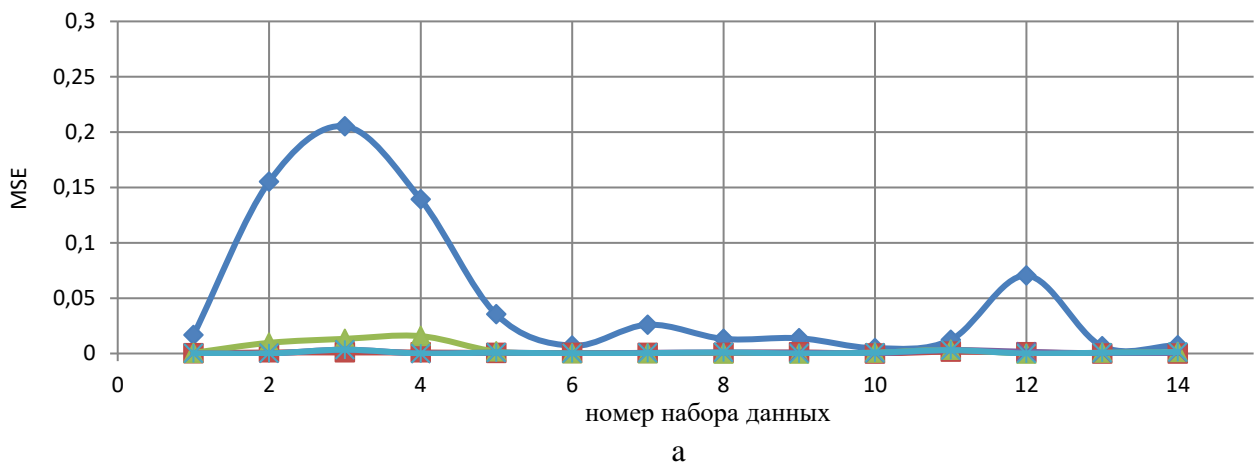
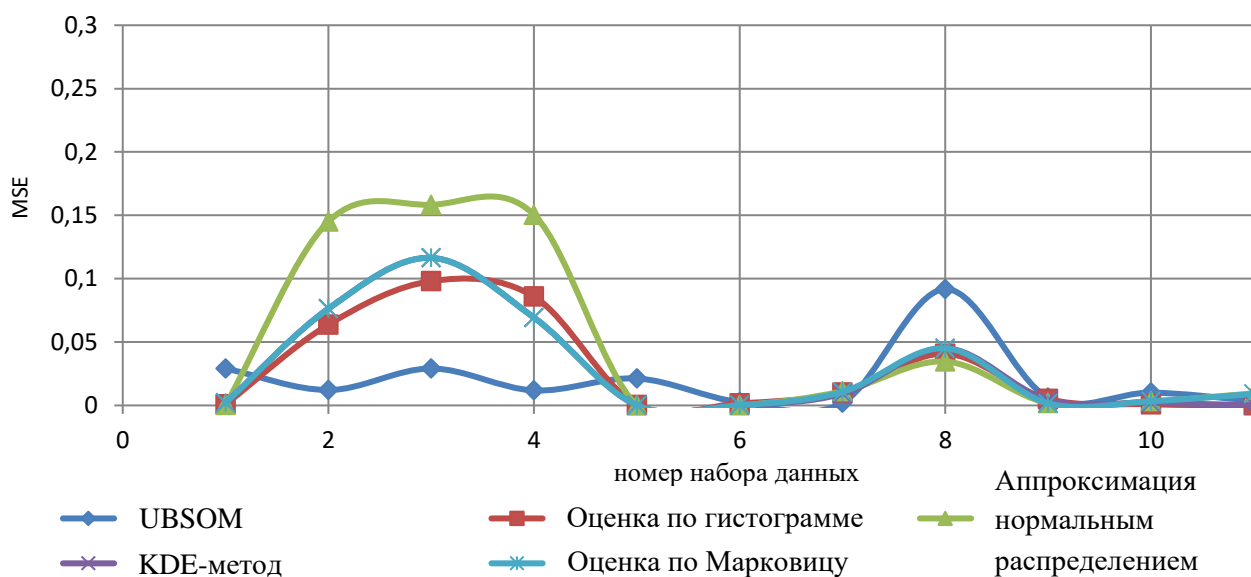
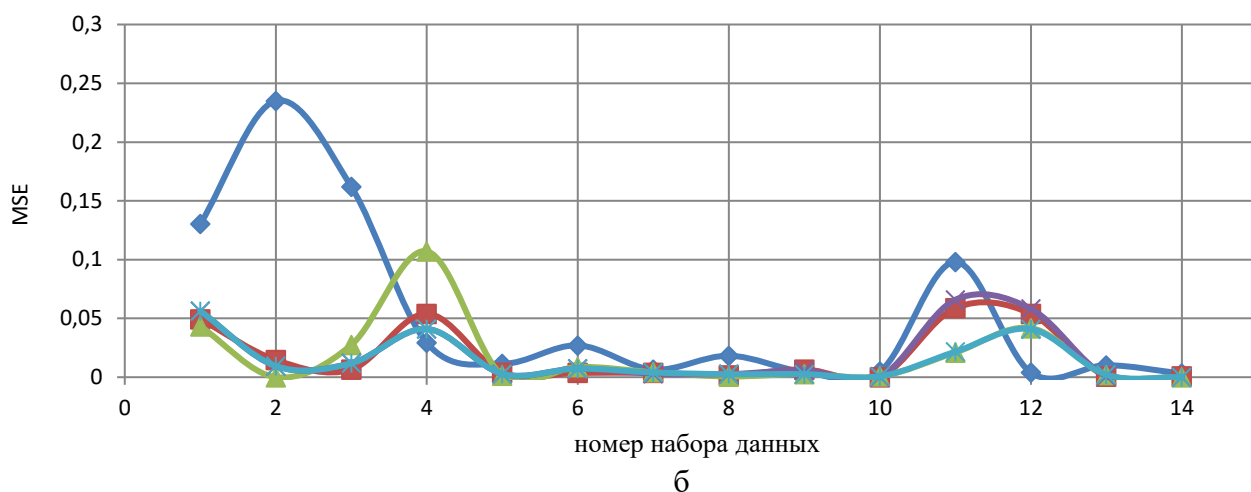


Рисунок 4.2.1 – Оценка функции плотности вероятности приростов стоимости портфеля различными методами (использованы наборы данных за период с 1 января 2010 года до 26 мая 2010 года)



а – 10 дней

Рисунок 4.2.2 – Среднеквадратическая ошибка (вертикальная ось) оценок риска по каждому набору данных (горизонтальная ось) при временном интервале, составляющем, лист 1



UBSOM
  Оценка по гистограмме
  KDE-метод
  Оценка по Марковицу
  Аппроксимация нормальным распределением

в

б – 100 дней; в – 400 дней

Рисунок 4.2.2, лист 2

Рисунок 4.2.2 показывает, что оценка риска на коротком временном интервале является самым плохим случаем для BSOM, а наилучший результат дает оценка по гистограмме. Это имеет простое объяснение – распределение в референсном периоде не успеваешь измениться за короткое время и практически идентично распределению оцениваемого периода (рисунок 4.2.2а). Однако, при увеличении временного интервала и, соответственно, временного разрыва между оцениваемым и референсным набором данных, точность оценок всех методов становится примерно одинаковой (рисунок 4.2.2б).

Увеличение ошибки, которое мы можем наблюдать в левой части графиков на рисунке 4.2.2 также имеет объяснение, т.к. эти наборы данных соответствуют периоду 2008-2009 годов, когда как мы знаем происходили значительные изменения на финансовых рынках. Дальнейшее увеличение временного интервала демонстрирует увеличение ошибки и ее дисперсии для всех методов за исключением BSOM.

Средняя среднеквадратическая ошибка по всем наборам данных в различных временных периодах показана на рисунке 4.2.3.

Рисунок 4.2.3 показывает, что точность модели BSOM увеличивается с увеличением временного интервала в отличие от других методов, которые показывают наилучшие результаты на коротком периоде. Стандартное отклонение среднеквадратических ошибок, показанных на рисунке 4.2.2 приведено в таблице 1.

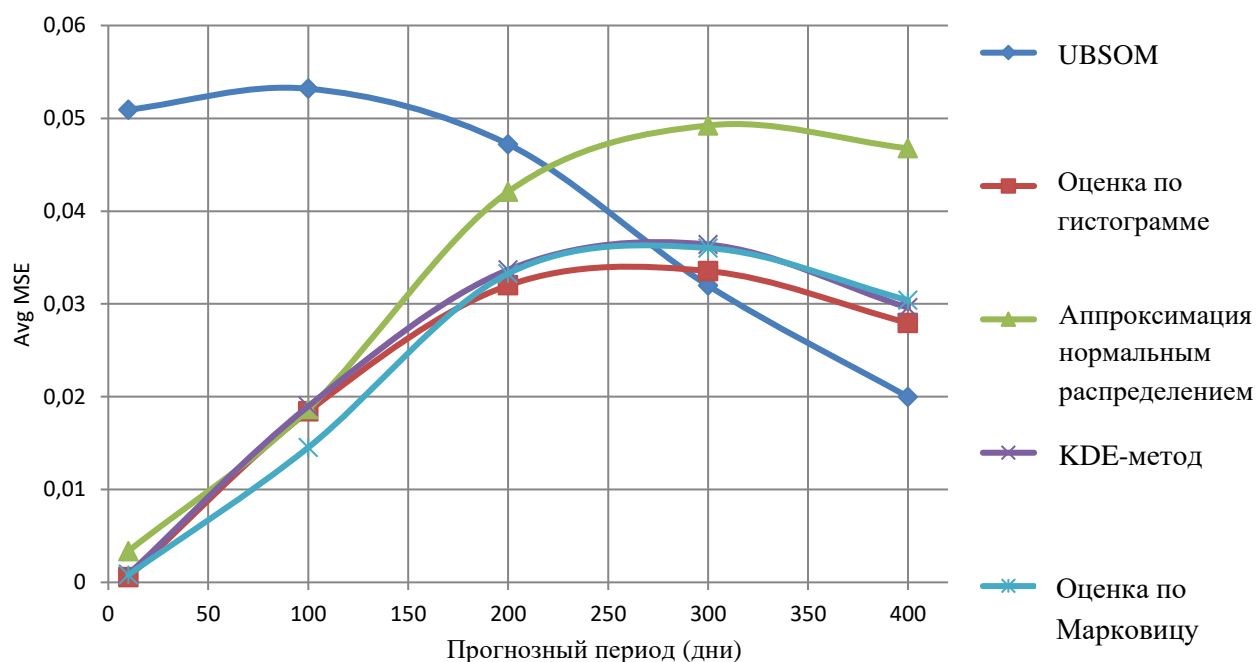


Рисунок 4.2.3 – Средняя среднеквадратическая ошибка методов при временном интервале от 15 до 400 дней

Таблица 1 – Средняя среднеквадратическая ошибка и ее стандартное отклонение\*

Наименование метода	Прогнозный период (дней)				
	15	100	200	300	400
UBSOM	0,050930 (0.066267)	0,053187 (0.073780)	0,047204 (0.072735)	0,031981 (0.048099)	0,019855 (0.024545)
Оценка по гистограмме	0,000568 (0.000483)	0,018416 (0.023723)	0,031998 (0.043879)	0,033542 (0.041883)	0,027956 (0.037743)
Аппроксимация нормальным распределением	0,003366 (0.005348)	0,018660 (0.029768)	0,042085 (0.059821)	0,049231 (0.063616)	0,046766 (0.067911)
Ядерная оценка плотности	0,000851 (0.001092)	0,018967 (0.024630)	0,033672 (0.047376)	0,036400 (0.047514)	0,029596 (0.040870)
Модель Марковица	0,000800 (0.001106)	0,014524 (0.018319)	0,033269 (0.047371)	0,036018 (0.046550)	0,030403 (0.040339)

\* – Стандартное отклонение показано в скобках

### 3.2.2 Применение UBSOM к данным по финансовым инструментам фондового рынка РФ

Для проверки предлагаемого подхода в качестве тестовых данных были использованы цены закрытия с 9 января 2008 по 01 января 2021 простых акций следующих компаний (ОАО): "Газпром"; "Лукойл"; "Магнит"; "Сбербанк России"; "Мобильные ТелеСистемы", объединенные в портфель с равными долями каждой акции. Исходные данные были разделены на наборы по 100 дней, цены акций преобразовывались в значения отклонений по формуле  $q_t(A_i) = (S_t(A_i) - S_{t-1}(A_i))/S_{t-1}(A_i)$ . В качестве тестовой задачи было выбрано прогнозирование риска данного портфеля для периодов прогнозирования 15, 100, 200, 300, 400 дней.

Расчет и сравнение результатов осуществлялось по следующим методам: модель UBSOM, оценка по гистограмме, оценка по Марковицу формула (1), аппроксимация нормальной кривой, KDE-методом (реализован в программном пакете NumXL (SpiderFinancial, Chicago, US)). Процедура оценки точности прогноза заключалась в следующем: фиксируется 100-дневный набор данных; на этих данных каждым методом строится оценка риска; задаются границы отклонения портфельных значений, например  $[-1\%; +1\%]$ ; производился расчет риска портфеля каждым методом, после чего оценивается вероятность нахождения портфельных значений в заданном интервале отклонений. Полученные оценки сравниваются с фактическими данными, полученными при помощи гистограммы для наборов, расположенных в следующем временном периоде по отношению к заданному. Пример построения оценок вышеназванными методами приведен на рисунке 4.2.4.

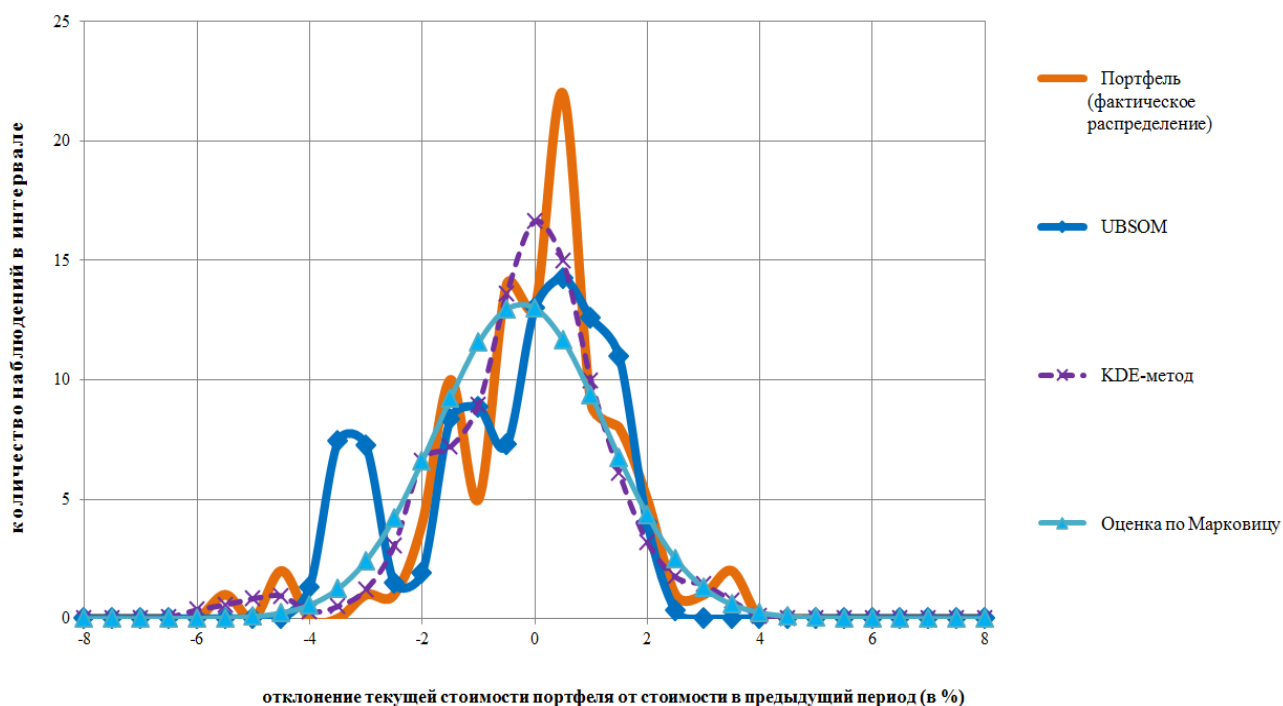


Рисунок 4.2.4 – Пример оценки функции плотности распределения для периода с 1 января 2010 до 26 мая 2010



На рисунке 4.2.5, представлены результаты сравнения упомянутых методов для различных сроков прогнозирования. Для краткосрочного периода лучший прогноз дает простейшая оценка по гистограмме, т.к. обычно распределение портфеля не успевает значительно измениться в течение короткого промежутка времени. С увеличением прогнозного периода все методы, за исключением модели UBSOM, демонстрируют падение точности. Модель UBSOM напротив показывает повышение точности на более длительных периодах прогноза (300-400 дней).

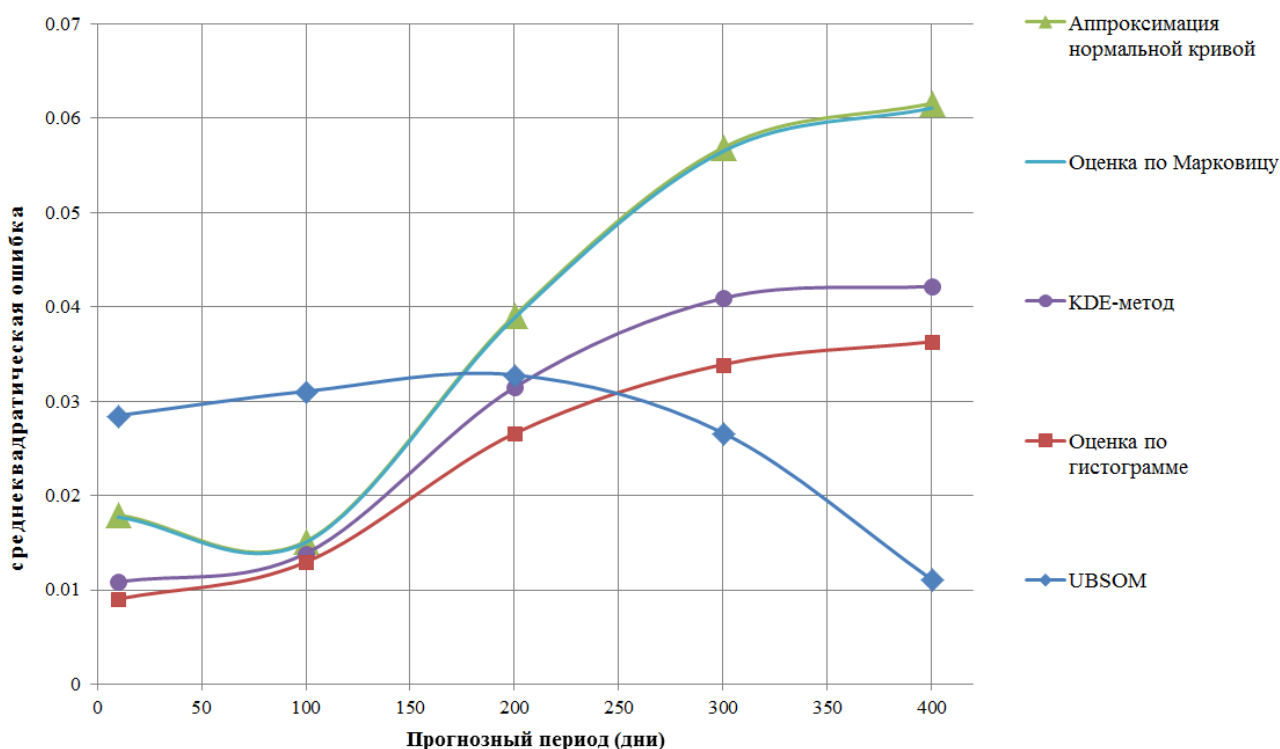


Рисунок 4.2.5 – Значения среднеквадратической ошибки для рассматриваемых методов по всем наборам данных для различных временных периодов прогнозирования (от 15 до 400 дней)

Модель UBSOM оценки и прогнозирования риска портфеля ценных бумаг на основе синтеза многомерного совместного распределения портфельных компонент демонстрирует более высокую точность прогноза риска по сравнению с классическими моделями, что подтверждают проведенные эксперименты. Результат достигается за счет использования значений компонент портфеля в явном виде и более точного учета их взаимных корреляций по сравнению с другими методами, использующими только значения стоимостей портфеля.

## ЗАКЛЮЧЕНИЕ

В диссертации определены условия корректности алгебры распознающих алгоритмов с  $\mu$ -операторами вычисления оценок над множеством задач распознавания с бинарной информацией. И эти условия, являясь достаточными условиями корректности, сформулированы в виде ограничений на множество задач распознавания ( $\Omega$ -регулярные задачи), для которых может быть построен корректный алгоритм [62-64]. Причём каждому оператору исходного семейства соответствует  $\mu$ -блок – пространственная, трёхуровневая, многослойная нейронная сеть [11, р. 870; 12, р. 37], воспроизводящая вычисления, осуществляемые алгоритмом (оператором) исходного семейства.

Более того, учитывая данный результат, для любой  $\Omega$ -регулярной задачи удастся построить 6-уровневую пространственную, многослойную нейросеть, воспроизводящую вычисления, осуществляемые корректным алгоритмом. Одной из особенностей построенной сети является использование диагональных активационных функций, что заметно упрощает вычисления во внутреннем внешнем циклах нейросети.

Кроме того важно, что для каждого контрольного объекта внутренний цикл  $\mu$ -блока состоит из одной итерации, связанной с объектом обучающей выборки с ненулевым весом. Последнее, при подсчете оценки контрольного объекта за класс, позволяет избежать перебора по объектам обучающей выборки, что также дает значительную экономию в вычислениях в сети. При этом здесь возникает естественный вопрос об ослаблении вышеупомянутых условий корректности. Решению этой задачи посвящена работа Дюсембаев А., Гришко М. «Об условиях корректности для алгебры распознающих алгоритмов с исходным множеством  $\mu$ -операторов над множеством задач с бинарной информацией» [64, р. 421].

В приведенной выше работе введено понятие  $\Omega$ -слабо регулярной задачи, исходя из которого и на основе операторной теории Журавлева с учётом нейросетевой парадигмы и ранее полученных результатов даётся вариант ответа на поставленный вопрос. Интересным аспектом, связанным с полученным результатом, является возможность построения для рассматриваемых задач шестиуровневой пространственной нейросети, каждый  $\mu$ -блок которой является сетью прямого распространения. Таким образом, в работе [64, р. 421] показана возможность ослабления условий корректности для алгебры распознающих алгоритмов с  $\mu$ -операторами вычисления оценок, что обосновывает повышения эффективности построенной нейросети.

Вместе с тем, касаясь практического аспекта применения нейросетей для решения прикладной задачи, в диссертации рассмотрен подход по использованию байесовских самоорганизующихся карт для более точной оценки рисков портфеля ценных бумаг (метод UBSOM). Этот метод также может помочь преодолеть предположение о нормальности классической модели Марковица и заменить гауссово распределение смесью гауссианов, которая лучше соответствует реальным распределениям портфеля.

Разработанный подход основан на построении распределения портфеля из совместного распределения компонентов портфеля, которое лучше отражает внутреннюю структуру взаимодействия компонентов в портфеле, чем методы, основанные только на данных о стоимости портфеля.

Как было показано, краткосрочное прогнозирование риска портфеля ценных бумаг не представляет сложностей – простейшие методы оценки риска дают демонстрируют приемлемые результаты ввиду того, что фактическое распределение отклонений стоимости портфеля не может существенно измениться за короткий период времени [65]. Но с увеличением периода прогнозирования точность большинства методов снижается. При этом для долгосрочного прогнозирования разработанная модель UBSOM демонстрирует в численных экспериментах наилучшую точность среди сравнимых методов.

Это связано с тем, что разработанный подход более точно отражает природу распределения приростов стоимости портфеля, по сравнению с альтернативными параметрическими и непараметрическими методами. Суть в том, что мы аппроксимируем распределение в многомерном пространстве, в то время как прямая оценка приростов стоимостей портфеля дает одномерное распределение. Преобразование векторов приростов стоимости финансовых инструментов в одномерную величину прироста стоимости портфеля приводит к потере детальной информации о многомерном распределении, что оказывает влияние на действительный риск портфеля.

Преимущества модели UBSOM были подтверждены результатами вычислительных экспериментов, проведенных на основе данных о стоимости ценных бумаг на фондовых рынках РФ и США.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Haykin S. *Neural Networks and Learning Machines*. – Ed. 3rd. – New Jersey: PrenticeHall, 2009. – 906 p.
- 2 Журавлев Ю.И. Об алгебраическом подходе к задачам распознавания и классификации // *Проблемы кибернетики*. – 1978. – №33. – С. 3-68.
- 3 Журавлев Ю.И. Алгебры над множествами некорректных (эвристических) алгоритмов // *Докл. Академии наук*. – 1977. – Т. 235, №4. – С. 761-763.
- 4 Журавлев Ю.И. Корректные алгебры над множеством некорректных (эвристических) алгоритмов // *Кибернетика*. – 1978. – №2. – С. 37-45.
- 5 Журавлев Ю.И., Исаев И.В. Построение алгоритмов распознавания корректных для заданной контрольной выборки // *Журнал вычислительной математике и математической физики*. – 1979. – Т. 19, №3. – С. 726-738.
- 6 Журавлев Ю.И. *Избранные научные труды*. – М.: Магистр, 1998. – 416 с.
- 7 Матросов В.Л. Корректные алгебры ограниченной ёмкости над множеством алгоритмов вычисления оценок // *Журнал вычислительной математике и математической физики*. – 1981. – Т. 21, №5 – С. 1276-1291.
- 8 Рудаков К.В. Об алгебраической теории универсальных и локальных ограничений для задач распознавания и классификации // *Распознавание, классификация, прогноз: сб. ст.* – М.: Наука, 1989. – Вып. 1. – С. 176-201.
- 9 Дьяконов А.Г. Алгебра над алгоритмами вычисления оценок: минимальная степень корректного алгоритма // *Журнал вычислительной математике и математической физики*. – 2005. – Т. 45, №6. – С. 1134-1145.
- 10 Дюсембаев А.Е. *Математические модели сегментации программ*. – М.: Физматлит, 2001. – 207 с.
- 11 Bauer A., Gornitz N., Biegler F. et al. Efficient Algorithms for Exact Inference in Sequence Labeling SVMs // *IEEE TNNLS*. – 2014. – Vol. 25, №5. – P. 870-881.
- 12 Xiao L. Neural Network Method for Solving Linear Fractional programming // *Procced. internat. conf. «Computational Intellegence and Security»*. – Nanning, 2010. – P. 37-41.
- 13 Matsugu M., Mori K., Mitari Y. et al. Subject independent facial expression recognition with robust face detection using a convolutional neural network // *Neural Networks*. – 2003. – №16. – P. 555-559.
- 14 LeCun Y., Bengio Y., Hinton G. Deep learning // *Nature*. – 2015. – Vol. 521, №7553. – P. 436-444.
- 15 Глызин С.Д., Колесов А.Ю., Розов Н.Х. Двухчастотные автоколебания в нейронной сети ФитцХью–Нагумо // *Журнал вычислительной математике и математической физики*. – 2017. – Т. 57, №1. – С. 94-110.
- 16 Дюсембаев А.Е. Об одном подходе к решению задач распознавания на основе нейросетей // *Докл. Академии наук*. – 2017. – Т. 473, №2. – С. 127-130.

- 17 Дюсембаев А.Е., Калиаждаров Д.Р. О точных решениях задач распознавания на основе нейросетевого подхода // Докл. Академии наук. – 2015. – Т. 461, №3. – С. 268-271.
- 18 Кочетков Д.В. Инвариантные решающие функции. Общий вид и условия корректности. – М.: ВЦ АН СССР, 1987. – 47 с.
- 19 Матросов В.Л. Корректные алгебры ограничений емкости над множеством алгоритмов вычисления оценок // Журнал вычислительной математики и математической физики. – 1981. – Т. 21, №5. – С. 1276-1291.
- 20 Матросов В.Л. О неполноте модели алгоритмов вычисления оценок. АН СССР // Журнал вычислительной математики и математической физики. – 1983. – Т. 23. – С. 453-466.
- 21 Алгоритмы и программы восстановления зависимостей / под ред. В.Н. Вапника. – М.: Наука, 1984. – 816 с.
- 22 Вапник В.Н., Червоненкис А.Я. Теория распознавания образов. – М.: Наука, 1974. – 416 с.
- 23 Дьяконов А.Г. Алгебра над алгоритмами вычисления оценок: учеб. пос. – М.: Изд. отдел Факультета вычислительной математики и кибернетики МГУ, 2006. – 70 с.
- 24 Markowitz H. Portfolio selection // Journal of Finance. – 1952. – Vol. 7. – P. 77-91.
- 25 Markowitz H. Mean-Variance Analysis in Portfolio Choice and Capital Markets. – Cambridge, MA: Blackwell, 1990. – 400 p.
- 26 Huang X. A risk index model for portfolio selection with returns subject to experts' estimations // Fuzzy Optimization and Decision Making. – 2012. – Vol. 11. – P. 451-463.
- 27 Mandelbrot B., Hudson R.L. The (Mis)Behaviour of Markets: A Fractal View of Risk, Ruin, and Reward. – London: Profile Books, 2004. – 353 p.
- 28 Chunchinda P., Dandapani K., Hamid S. et al. Portfolio selection and skewness: evidence from international stock market // Journal of Banking and Finance. – 1997. – Vol. 21, №2. – P. 143-167.
- 29 Lee C.F., Lee A.C., Lee J.C. Handbook of Quantitative Finance and Risk Management. – NY.: Springer Academic Publishers, 2010. – 1716 p.
- 30 Okhrin Y., Schmid W. Comparison of different estimation techniques for portfolio selection // AStA Advances in Statistical Analysis. – 2007. – Vol. 91(2). – P. 109-127.
- 31 Kondor I., Pafka S., Karádi R. et al. Portfolio Selection in a Noisy Environment Using Absolute Deviation as a Risk Measure // Practical Fruits of Econophysics: proceed. of the 3rd Nikkei Econophysics sympos. – Tokyo: Springer, 2006. – P. 220-225.
- 32 Doganoglu T., Hartz C., Mittnik S. Portfolio optimization when risk factors are conditionally varying and heavy tailed // Computational Economics. – 2007. – Vol. 29. – P. 333-354.

- 33 Lu X.F., Lai K.K., Liang L. Portfolio value-at-risk estimation in energy futures markets with time-varying copula-GARCH model // *Annals of Operations Research*. – 2011. – Vol. 219, №1. – P. 333-357.
- 34 Behr A., Pötter U. Alternatives to the normal model of stock returns: Gaussian mixture, generalized logF and generalised hyperbolic models // *Annals of Finance*. – 2009. – Vol. 5. – P. 49-68.
- 35 Wu X. The Empirical Study of Portfolio Risk for Chinese Stock Market Based on Web // *Advances in Computer Science, Environment, Ecoinformatics, and Education: proced. internat. conf.* – Berlin: Springer, 2011. – P. 413-417.
- 36 Kamdem J.S. VaR and ES for linear portfolios with mixture of generalized Laplace distributions risk factors // *Annals of Finance*. – 2012. – Vol. 8. – P. 123-150.
- 37 Sharpe W.F. Mutual Fund Performance // *Journal of Business*. – 1966. – Vol. 39. – P. 119-138.
- 38 Ruppert D. *Statistics and Data Analysis for Financial Engineering*. – NY.: Springer Texts in Statistics, 2011. – 719 p.
- 39 Peel D., McLachlan G.J. Robust mixture modeling using the t distribution // *Statistics and Computing*. – 2000. – Vol. 10. – P. 339-348.
- 40 Chen R., Wang Z. The Comparison of Beta Distribution Estimation and Gauss Kernel Density Estimation in the Recovery Rates of Municipal Bonds // *Proced. 5th internat. conf. on Business Intelligence and Financial Engineering*. – Lanzhou, 2012. – P. 168-171.
- 41 Li Y., Ma W. Applications of Artificial Neural Networks in Financial Economics: A Survey // *Computational Intelligence and Design internat. sympos. (ISCID, 2010)*. – Hangzhou, 2010. – Vol. 1. – P. 211-214.
- 42 Lin W.-Y., Hu Y.-H., Tsai C.-F. Machine Learning in Financial Crisis Prediction: A Survey // *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions*. – 2012. – Vol. 42, Issue 4. – P. 421-436.
- 43 Fernandez A., Gomez S. Portfolio selection using neural networks // *Computers & Operations Research*. – 2007. – Vol. 34. – P. 1177-1191.
- 44 Steiner M., Wittkemper H.-G. Portfolio optimization with a neural network implementation of the coherent market hypothesis // *European Journal of Operational Research*. – 1997. – Vol. 100. – P. 27-40.
- 45 Yin H. Learning Nonlinear Principal Manifolds by Self-Organising Maps // In book: *Principal Manifolds for Data Visualization and Dimension Reduction*. – Berlin: Springer, 2008. – P. 68-95.
- 46 Kohonen T. *Self-Organizing Maps*. – Ed. 3rd. – Berlin: Springer, 2001. – 502 p.
- 47 Kaski S., Sinkkonen J., Peltonen J. Learning metrics for self-organizing maps // *Neural Networks: proced. internat. joint conf. (IJCNN '01)*. – Washington, 2001. – Vol. 2. – P. 914-919.
- 48 Yin H., Allinson N.M. Self-organizing mixture networks for probability density estimation // *Neural Networks, IEEE Transactions*. – 2001. – Vol. 12, Issue 2. – P. 405-411.

- 49 Yin H., Allinson N.M. Bayesian learning for self-organising maps // *Electronics Letters*. – 1997. – Vol. 33, Issue 4. – P. 304-305.
- 50 Yin H., Allinson N.M. Bayesian self-organising map for Gaussian mixtures // *Vision, Image and Signal Processing, IEE Proceedings*. – 2001. – Vol. 148, №4. – P. 234-240.
- 51 Bishop C.M., Svensen M. The Generative Topographic Mapping // *Neural Computation*. – 1998. – Vol. 10. – P. 215-234.
- 52 Дюсембаев А.Е. Об одном подходе к задаче сегментации программ // *Докл. Академии наук*. – 1993. – Т. 323, №6. – С. 711-714.
- 53 Дюсембаев А.Е. О корректности алгебраического замыкания тестовых алгоритмов распознавания // *Журнал вычислительной математике и математической физики*. – 1982. – Т. 22, №6. – С. 1491-1499.
- 54 Журавлев Ю.И. Алгоритмы распознавания, основанные на вычислении оценок. Содержательный смысл параметров задающих алгоритмы // *Тр. междунар. симпоз. по практическому применению методов распознавания образов*. – М.: ВЦ АН СССР, 1973. – С. 116-124.
- 55 Dyusembaev A.E., Grishko M.V. Conditions of the correctness for algebra of estimates calculation algorithms with  $\mu$ -operators over a setoff binary-data recognition problems // *Journal of Pattern Recognition and Image Analysis*. – 2017. – Vol. 27, №2. – P. 166-174.
- 56 Utsugi A. Hyperparameter selection for self-organizing maps // *Neural Computation*. – 1997. – Vol. 9. – P. 623-635.
- 57 Utsugi A. Topology selection for self-organizing maps // *Network: Computation in Neural Systems*. – 1997. – Vol. 7. – P. 727-740.
- 58 Grishko M.V., Dyusembaev A.E., Andras P. Securities portfolio risk estimation and forecasting by the use of Bayesian self-organizing maps // *Proceed. of 8th internat. conf. on Computational and Financial Econometrics (CFE 2014)*. – Pisa, 2014. – P. 140.
- 59 Barron A.R. Universal approximation bounds for superpositions of a sigmoidal function // *IEEE Transactions on Information Theory*. – 1993. – Vol. 39. – P. 930-945.
- 60 Gnecco G. A comparison between fixed-basis and variable-basis schemes for function approximation and functional optimization // *Journal of Applied Mathematics*. – 2012. – Vol. 2012. – P. 1-17.
- 61 Kainen P.C., Kurkova V., Sanguineti M. Dependence of Computational Models on Input Dimension: Tractability of Approximation and Optimization Tasks // *IEEE Transactions on Information Theory*. – 2012. – Vol. 58. – P. 1203-1214.
- 62 Dyusembaev A.E., Grishko M.V., Kaliazhdarov D. The Conditions of Solvability of the Inverse Problem of Operator Equation for a Pattern Recognition Neurooperator Model // *Australian Journal of Intelligent Information Processing Systems* – 2014. – Vol. 14. № 1 – P. 15-20.
- 63 Dyusembaev A.E., Grishko M.V. Construction of a Correct Algorithm and Spatial Neural Network for Recognition Problems with Binary Data // *Comput. Math. and Math. Phys.* – 2018. – Vol. 58. – P. 1673-1686.

64 Dyusembaev A.E., Grishko M.V. On Correctness Conditions for Algebra of Recognition Algorithms with  $\mu$ -Operators over Pattern Problems with Binary Data // Dokl. Math. – 2018. – Vol. 98. – P. 421-424.

65 Grishko M.V., Murzakhmetov A.N. Making investment decisions based on econometric analysis in the conditions of Kazakhstan stock market // Bulletin of National academy of sciences of the Republic of Kazakhstan – 2015. – № 2 – P. 250-256.