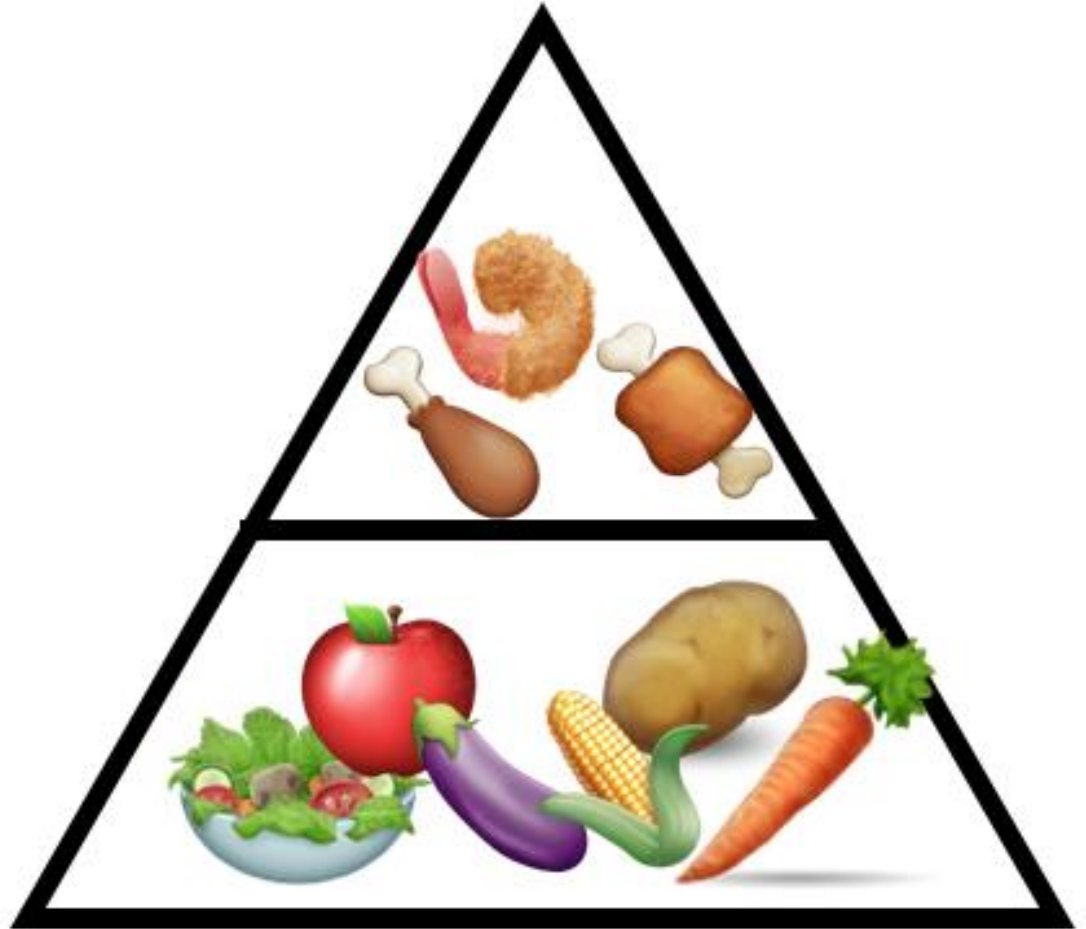


Метод главных компонент

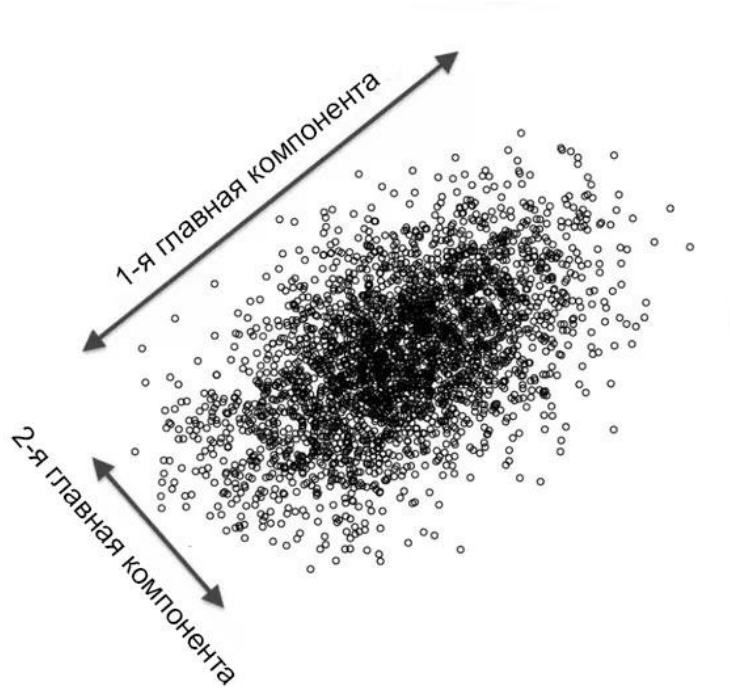
Лекция 5

Изучение пищевой ценности

- Визуализация
- Обнаружение кластеров



Главные компоненты



Витамин С

- Петрушка
- Капуста кале

• Брокколи

• Цветная капуста

• Соя

• Ямс

■ Мясо цесарки

(Витамин С) - (жир)

- Петрушка
- Капуста кале

• Брокколи

• Цветная капуста

• Капуста

• Шпинат

• Ямс

• Сладкая кукуруза

■ Мясо цесарки

• Окунь

• Скумбрия

• Курица

■ Говядина

• Свинина

• Ягнятина

(Витамин С + пищевые волокна) - (жир)

• Петрушка

• Горох

• Корень лотоса

• Лук-резанец

• Цветная капуста

• Соя

• Баклажан

• Сладкая кукуруза

• Грибы

■ Треска

■ Мясо цесарки

• Окунь

• Скумбрия

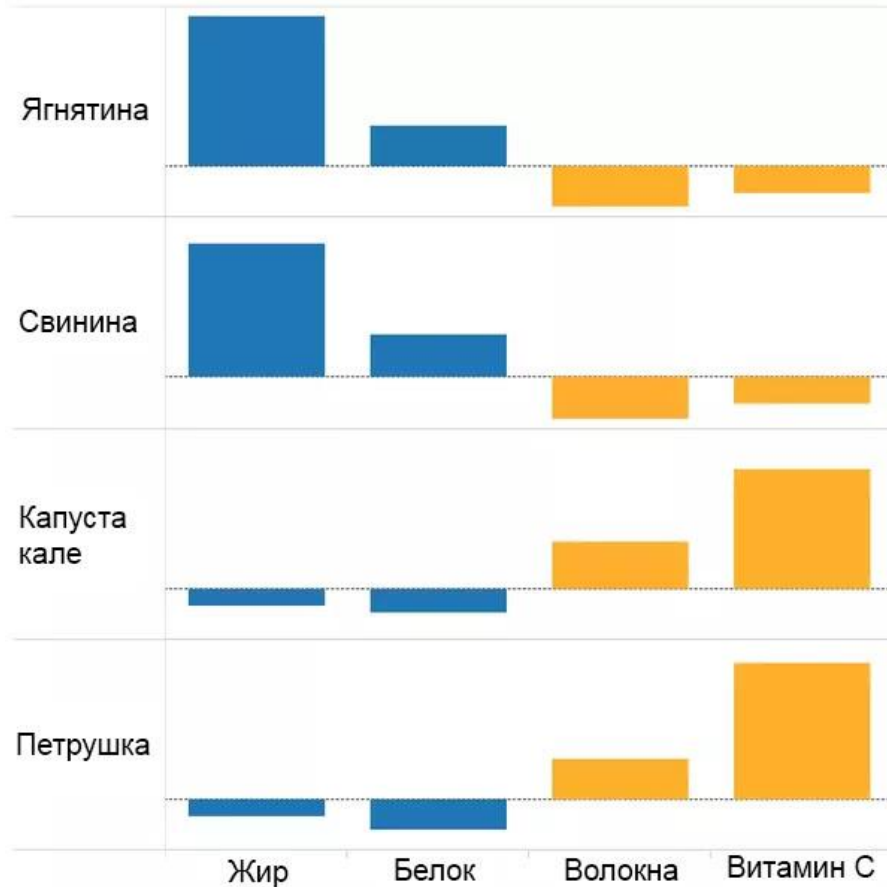
• Курица

■ Говядина

• Свинина

• Ягнятина

Пример: анализ пищевых групп



Чтобы получить компоненту 1 для отдельного продукта, мы можем вычислить следующее:

$$.55(\text{пищевые волокна}) + .44(\text{Витамин С}) - .45(\text{жир}) - .55(\text{белок})$$

	PC1	PC2	PC3	PC4
Жир	-0.45	0.66	0.58	0.18
Белок	-0.55	0.21	-0.46	-0.67
Волокна	0.55	0.19	0.43	-0.69
Витамин С	0.44	0.70	-0.52	0.22

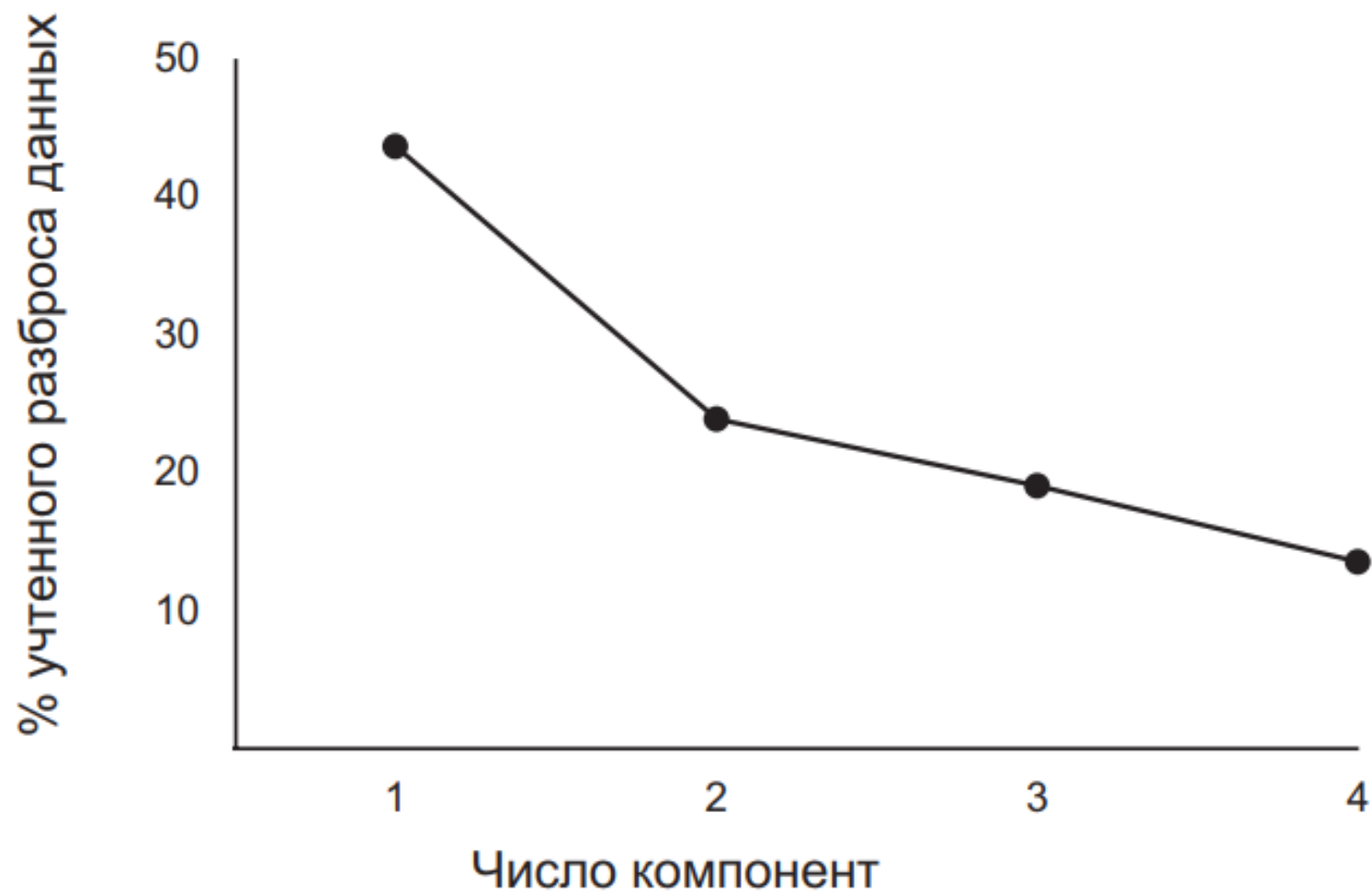
2-я главная компонента



1-я главная компонента

У мясных товаров низкие значения компоненты 1, поэтому они сосредоточены в левой части графика, в противоположной стороне от овощных. Видно также, что среди не овощных товаров низкое содержание жиров у морепродуктов, поэтому значение компоненты 2 для них меньше, и сами они тяготеют к нижней части графика. Схожим образом у тех овощей, которые не являются зеленью, низкие значения компоненты 2, что видно в нижней части графика справа.

График осыпи



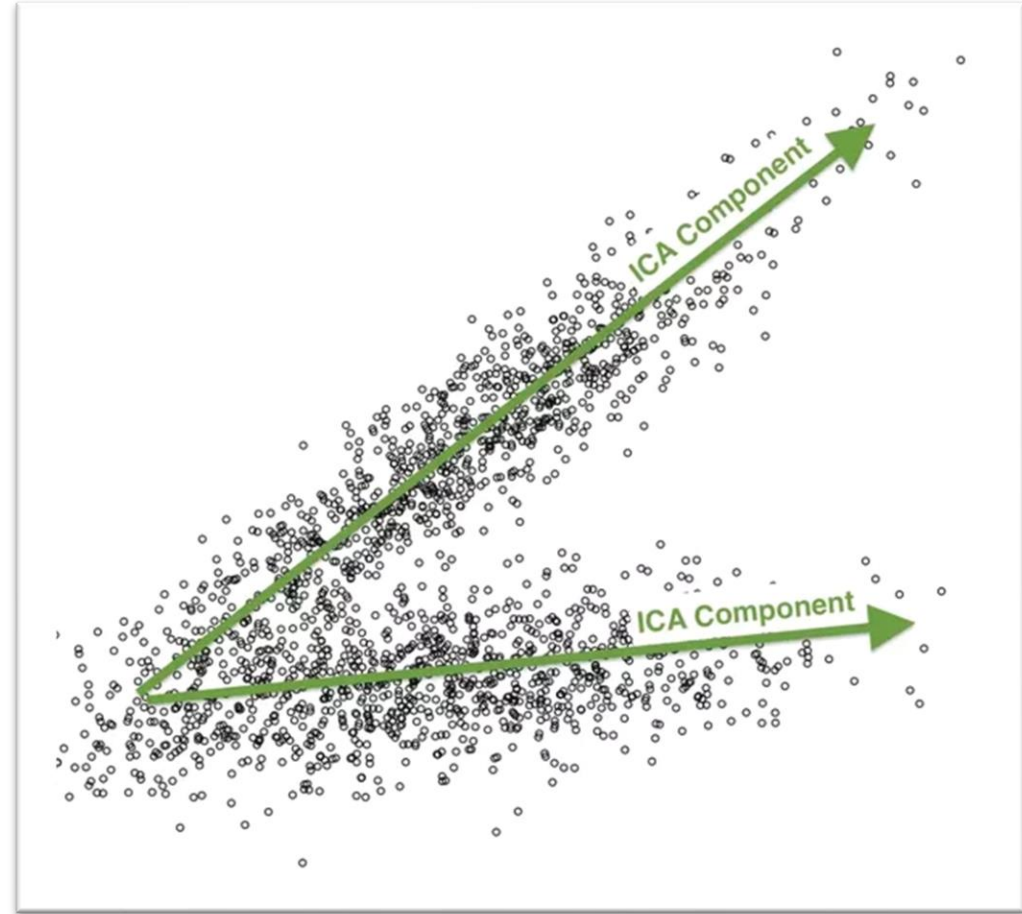
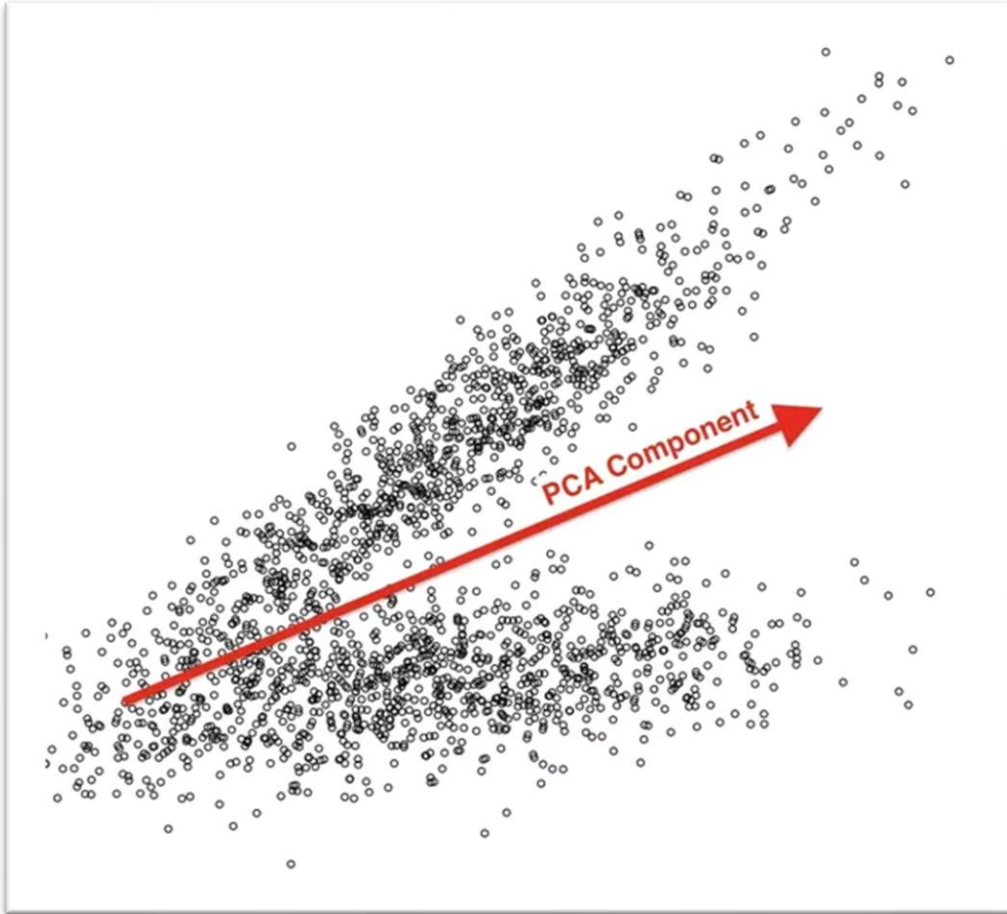
Ограничения

- **Максимизация распределения.** МГК исходит из важного допущения того, что наиболее полезны те измерения, которые дают наибольший разброс. Однако это не всегда так.



- **Интерпретация компонент.** Главная трудность с МГК состоит в том, что необходима интерпретация сгенерированных компонент, и иногда нужно очень постараться, чтобы объяснить, почему переменные должны быть скомбинированы именно выбранным способом.
- **Ортогональные компоненты.** МГК всегда формирует ортогональные главные компоненты, то есть такие, которые размещаются в отношении друг друга под углом 90° . Однако это допущение может оказаться излишним при работе с неортогональными информационными измерениями.

Анализ независимых компонент



Краткие итоги

- Метод главных компонент — это способ уменьшения размерности, который позволяет выразить наши данные через небольшой набор переменных, называемых главными компонентами.
- Каждая главная компонента — это некая сумма изначальных переменных. Лучшие из них могут быть использованы для анализа и визуализации.
- МГК лучше всего работает с теми информационными измерениями, которые дают большой разброс данных и ортогональны друг другу