

Подготовка данных и выбор алгоритмов

Лекция 2

Формат данных

Переменные



Элементы данных

ID транзакции	Покупатель	Даты	Куплено фруктов	Куплена рыба	Потрачено
1	Пингвин	1 янв.	1	да	5,30 \$
2	Медведь	1 янв.	4	да	9,70 \$
3	Кролик	1 янв.	6	нет	6,50 \$
4	Лошадь	2 янв.	6	нет	5,50 \$
5	Пингвин	2 янв.	2	да	6,00 \$
6	Жираф	3 янв.	5	нет	4,80 \$
7	Кролик	3 янв.	8	нет	7,60 \$
8	Кот	3 янв.	?	да	7,40 \$

Таблица 2. Переформатированный набор данных о покупках за день с дополнительными переменными

Переменные



Дата	Выручка	Число покупателей	Погода	Выходные
1 янв.	21,50 \$	3	солнечно	да
2 янв.	11,50 \$	2	дождливо	нет
3 янв.	19,80 \$	3	солнечно	нет

Типы переменных

- Бинарная
- Категориальная
- Целочисленная
- Непрерывная (количественная)

Типы переменных

Переменные



Элементы данных

ID транзакции	Покупатель	Даты	Куплено фруктов	Куплена рыба	Потрачено
1	Пингвин	1 янв.	1	да	5,30 \$
2	Медведь	1 янв.	4	да	9,70 \$
3	Кролик	1 янв.	6	нет	6,50 \$
4	Лошадь	2 янв.	6	нет	5,50 \$
5	Пингвин	2 янв.	2	да	6,00 \$
6	Жираф	3 янв.	5	нет	4,80 \$
7	Кролик	3 янв.	8	нет	7,60 \$
8	Кот	3 янв.	?	да	7,40 \$

Выбор переменных

← Переменные →

	ID транзакции	Покупатель	Даты	Куплено фруктов	Куплена рыба	Потрачено
↑ Элементы данных ↓	1	Пингвин	1 янв.	1	да	5,30 \$
	2	Медведь	1 янв.	4	да	9,70 \$
	3	Кролик	1 янв.	6	нет	6,50 \$
	4	Лошадь	2 янв.	6	нет	5,50 \$
	5	Пингвин	2 янв.	2	да	6,00 \$
	6	Жираф	3 янв.	5	нет	4,80 \$
	7	Кролик	3 янв.	8	нет	7,60 \$
	8	Кот	3 янв.	?	да	7,40 \$

Конструирование признаков

Переменные

ID транзакции	Покупатель	Даты	Куплено фруктов	Куплена рыба	Потрачено
1	Пингвин	1 янв.	1	да	5,30 \$
2	Медведь	1 янв.	4	да	9,70 \$
3	Кролик	1 янв.	6	нет	6,50 \$
4	Лошадь	2 янв.	6	нет	5,50 \$
5	Пингвин	2 янв.	2	да	6,00 \$
6	Жираф	3 янв.	5	нет	4,80 \$
7	Кролик	3 янв.	8	нет	7,60 \$
8	Кот	3 янв.	?	да	7,40 \$

Неполные данные

Переменные



ID транзакции	Покупатель	Даты	Куплено фруктов	Куплена рыба	Потрачено
1	Пингвин	1 янв.	1	да	5,30 \$
2	Медведь	1 янв.	4	да	9,70 \$
3	Кролик	1 янв.	6	нет	6,50 \$
4	Лошадь	2 янв.	6	нет	5,50 \$
5	Пингвин	2 янв.	2	да	6,00 \$
6	Жираф	3 янв.	5	нет	4,80 \$
7	Кролик	3 янв.	8	нет	7,60 \$
8	Кот	3 янв.	?	да	7,40 \$

Элементы данных



Выбор алгоритма

	Алгоритмы
Обучение без учителя	Метод k -средних Метод главных компонент Ассоциативные правила Анализ социальных сетей
Обучение с учителем	Регрессионный анализ Метод k -ближайших соседей Метод опорных векторов Дерево решений Случайные леса Нейросети
Обучение с подкреплением	Многорукие бандиты

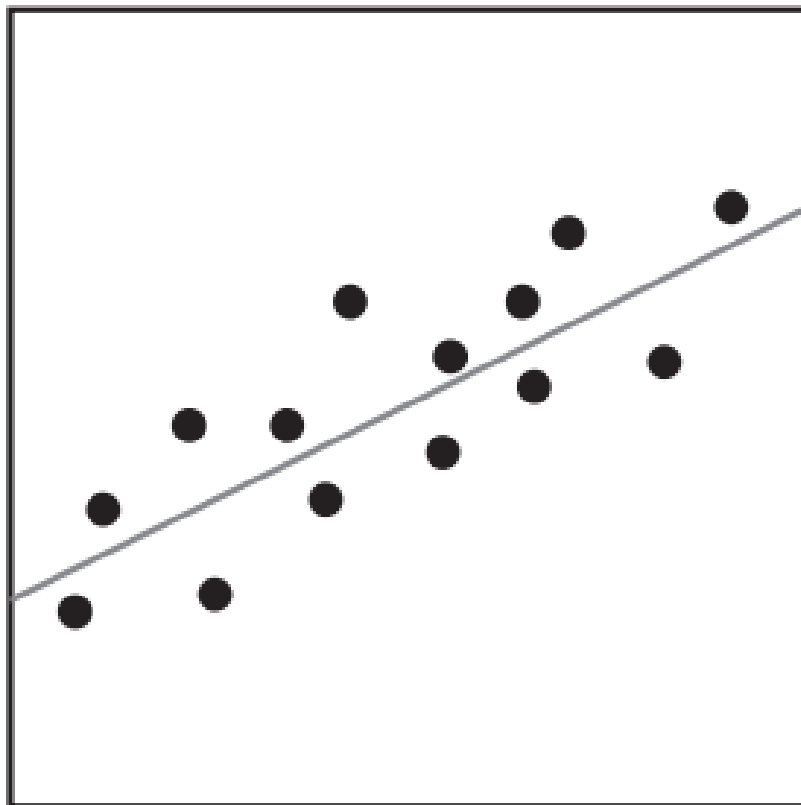
Обучение без учителя

← Переменные →

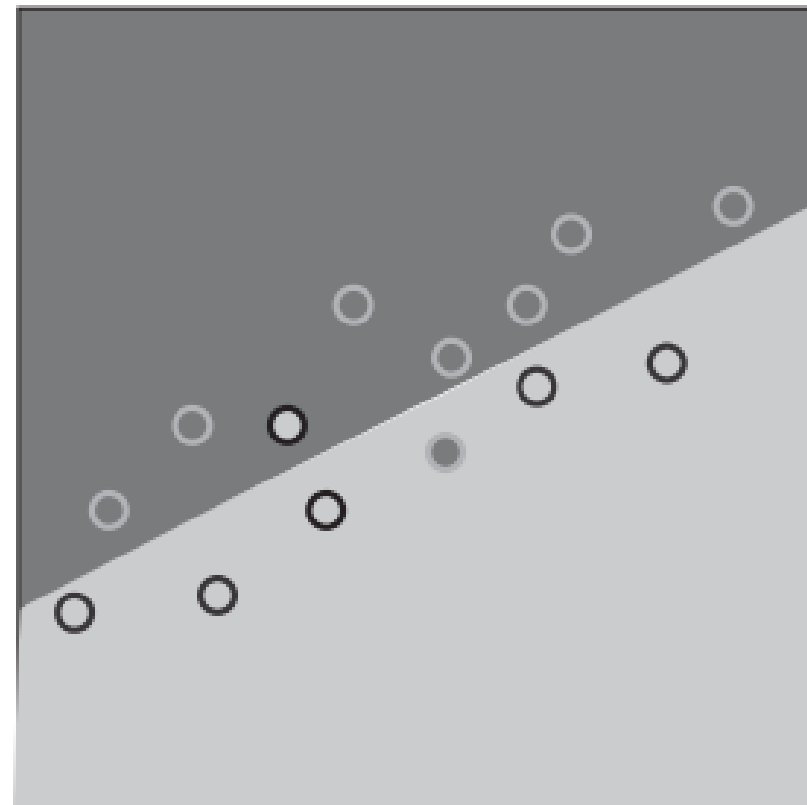
ID транзакции	Покупатель	Даты	Куплено фруктов	Куплена рыба	Потрачено
1	Пингвин	1 янв.	1	да	5,30 \$
2	Медведь	1 янв.	4	да	9,70 \$
3	Кролик	1 янв.	6	нет	6,50 \$
4	Лошадь	2 янв.	6	нет	5,50 \$
5	Пингвин	2 янв.	2	да	6,00 \$
6	Жираф	3 янв.	5	нет	4,80 \$
7	Кролик	3 янв.	8	нет	7,60 \$
8	Кот	3 янв.	?	да	7,40 \$

↑ Элементы данных ↓

Обучение с учителем



а) Регрессия



б) Классификация

Обучение с подкреплением

Өзіңнің GB мен МИНУТТАРЫҢДЫ бер
Передавай свои GB и МИНУТЫ

«Сенің ережелерің»
тарифтік желісі

Тарифная линейка
«Твои правила»

TELE2 4G

Дарим месяц

15 GB ТРАФИКА
120 МИН. НА КРУТКЕ НОМЕРА РЕ

+ **1 МЕСЯЦ БЕСПЛАТНО**

Тариф «ПРОСТО Супер»

1590 ₸

Ещё выгоднее

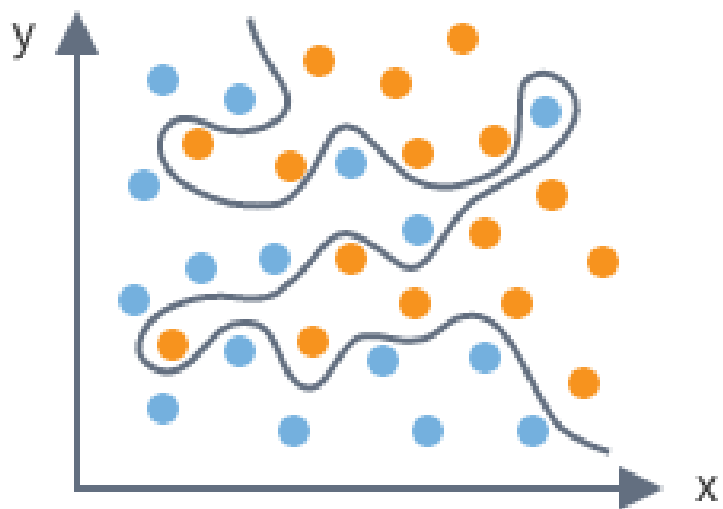
Стартовый комплект

-50%

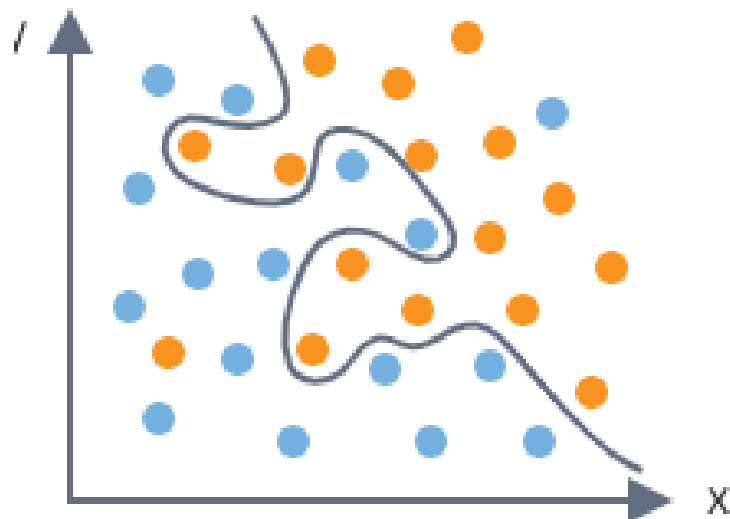
TELE2 4G

Гарантия лучшей цены

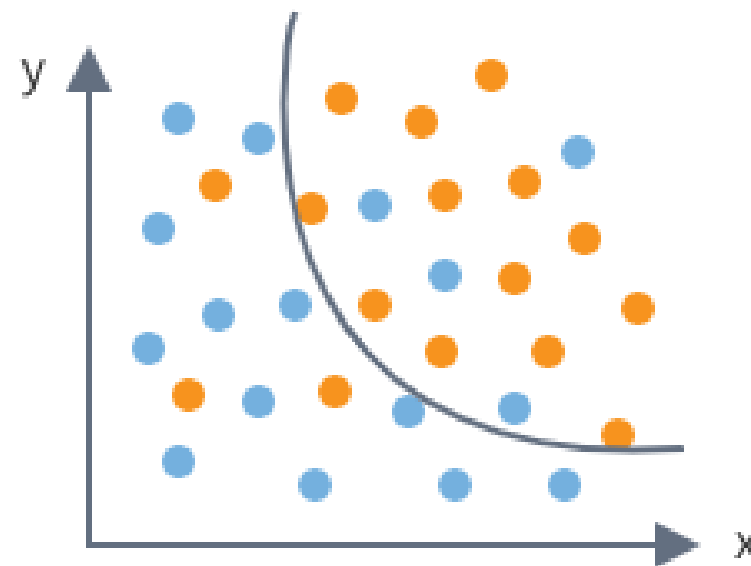
Настройка параметров



а) переобучение



б) идеальное
распределение



с) недообучение