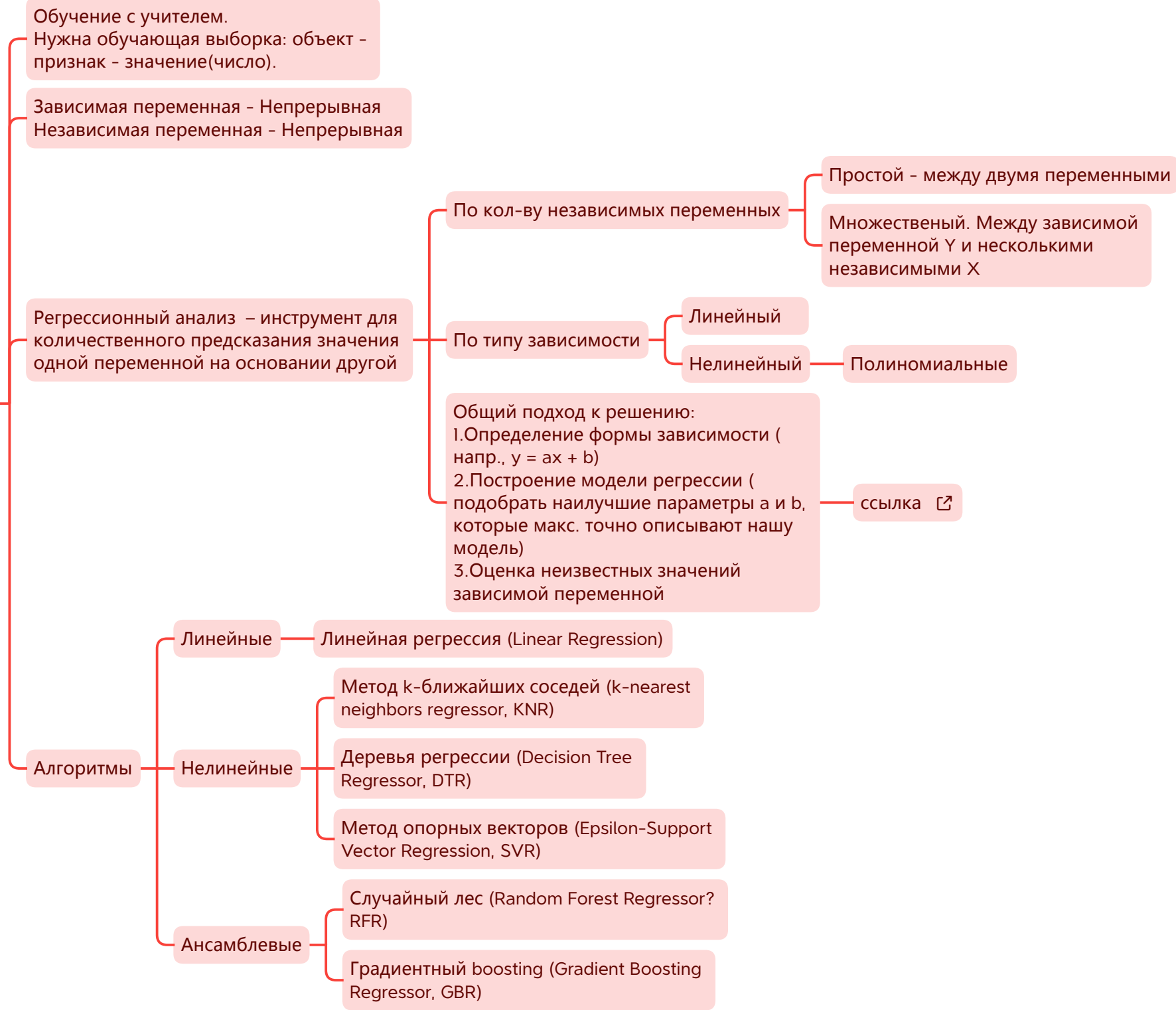
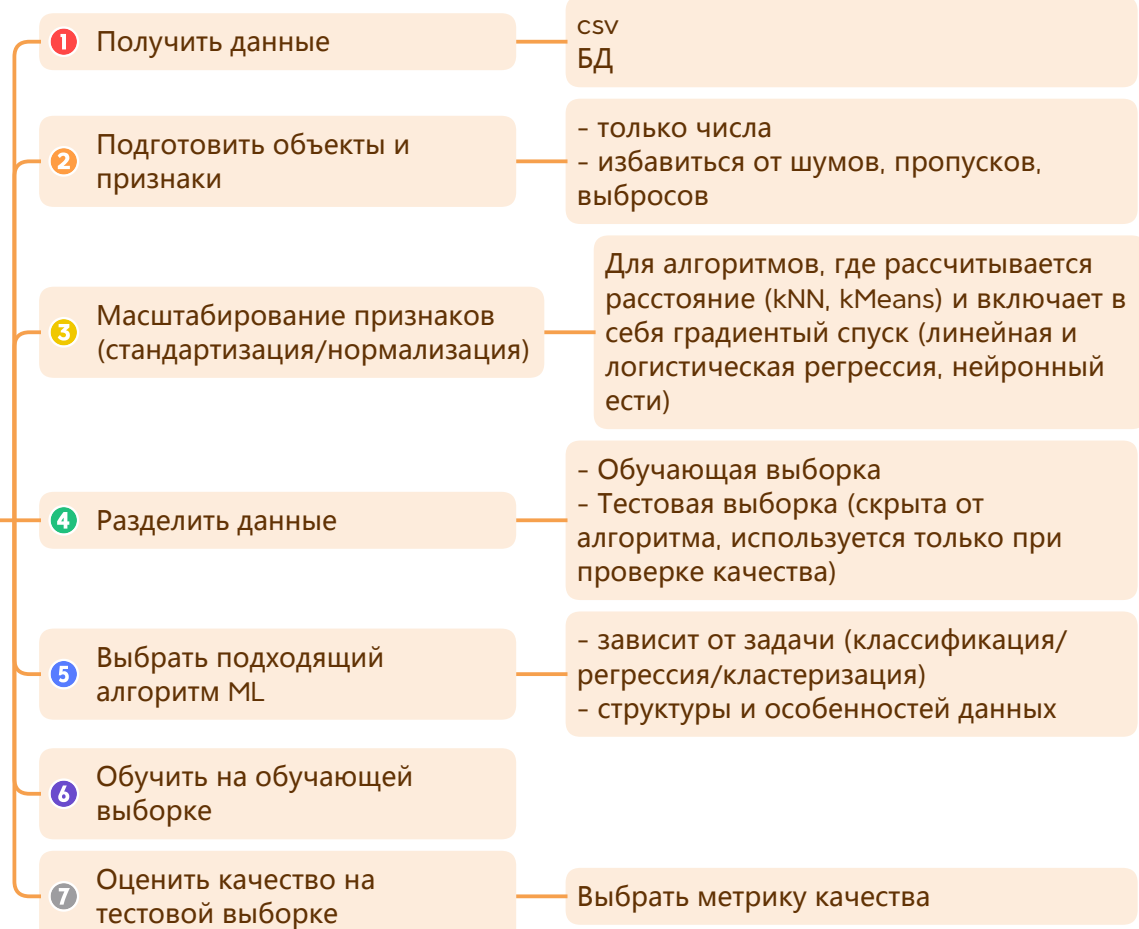


МАШИННОЕ ОБУЧЕНИЕ
MACHINE LEARNING
ML

РЕГРЕССИЯ
предсказание
вещественного значения



Общая схема (для классификации и регрессии)



КЛАССИФИКАЦИЯ
предсказывать метку класса



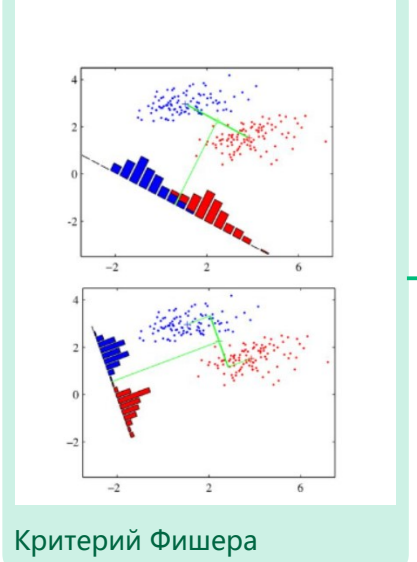
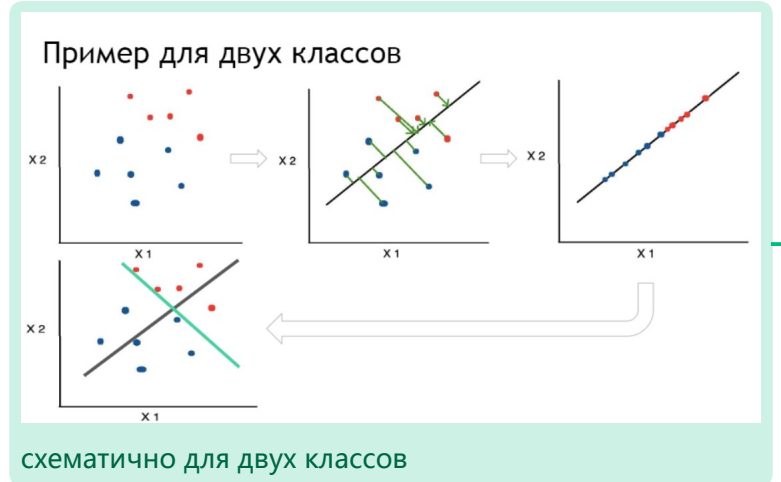
Обучение с учителем.
Нужна обучающая выборка: объект - признаки - класс.

Задача: на основании некоторых признаков (независимых переменных) объект можно причислить к одной из заранее заданных групп

Примеры.

Бинарные:
- фишинг/не фишинг
- нажмет/не нажмет
- перейдет/не перейдет
- определение подозрительных транзакций

Несколько категорий:
- фильтрация входящих сообщений на почте "основные", "соцсети", "промоакции" или "форумы"
- Спам-фильтры
- Определение языка
- Поиск похожих документов
- Распознавание рукописных букв и цифр



Для наблюдений должно выполняться:
- Многомерное нормальное распределение в каждой группе
- Равенство ковариационных матриц для разных групп
- Отсутствие выраженной корреляции внутри групп

Пример по ссылке ↗

Линейный дискриминантный анализ (LDA — Linear Discriminant Analysis).

Линейные

Логистическая регрессия
Метод опорных векторов (SVC)

Метод ближайших соседей (k Nearest Neighbors, или kNN)

Нелинейные

Деревья решений (Decision Tree)
Наивный байесовский алгоритм (Naive Bayes algorithm).
Квадратичный дискриминантный анализ

Случайный лес (Random Forest)

Ансамблевые

Это метрический метод, нужна нормализация нецелевых признаков. Целевой – не трогаем.

1. Задается входной параметр k
2. Находятся k ближайших соседей для исследуемого объекта A
3. Объект A будет отнесен к тому классу, который является наиболее распространенным среди соседей

Неустойчив к выбросам

Плохо работает когда очень много признаков

Часто используется как промежуточный алгоритм при работе с более сложными моделями

КЛАСТЕРИЗАЦИЯ
разделение на кластеры (группы)

