Линейный классификатор и логистическая регрессия

Занятие 1.2

Цели занятия

В конце занятия вы:

- будете знать преимущества и недостатки линейных моделей, а также требования к данным;
- научитесь реализовывать алгоритм градиентного спуска и логистическую регрессию;
- повторите понятие условной вероятности.

О чем поговорим и что сделаем

нетология

План занятия

1. Линейные модели: требования к данным и практика
2. Логистическая регрессия: практическое задание
3. Градиентный спуск: теория и практическое задание
4. Немного про условную вероятность.

Линейные модели

Причины популярности

- Линейные модели подходят для описания многих процессов
- Относительная простота вычислений и интерпретации результатов
- Вклад нескольких факторов часто можно разбить на сумму влияния каждого фактора в отдельности

Примеры использования

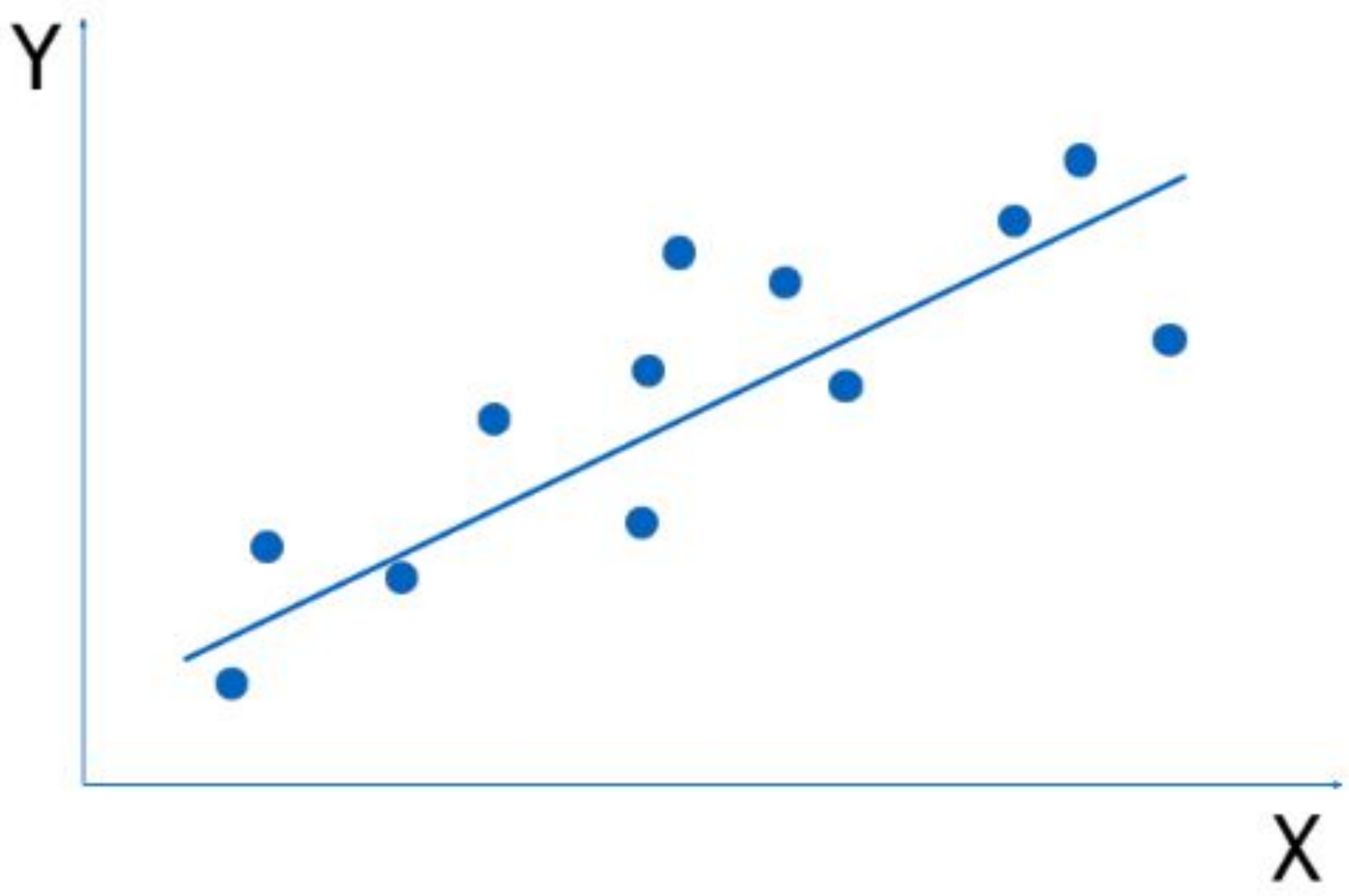
- Прогноз продаж по объему инвентаря, загрузке, площади и другим «линейным» характеристикам
- Построение вероятностных моделей в страховании, кредитном скоринге, инвестиционных проектах
- Предсказание цены товара на основании его характеристик
- Построение трендов

Определение и код

1

нетология

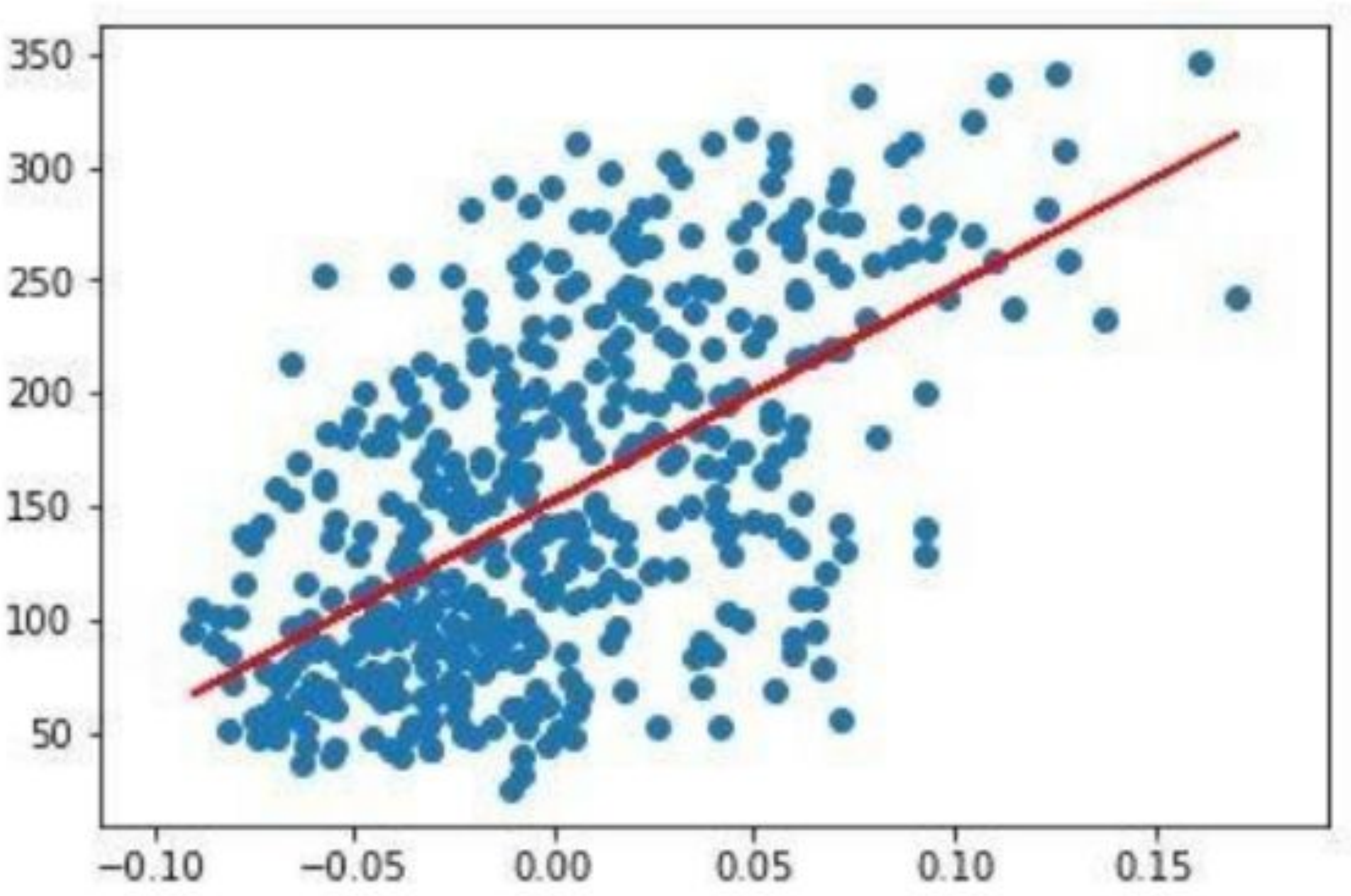
Определение



$$y_i = \sum_{j=1}^{m} w_j X_{ij} + e_i$$

Y - целевая переменная
W - вектор весов модели
X - матрица наблюдений
e - ошибка модели

Пример из кода
LINEAR REGRESSION.IPYNB

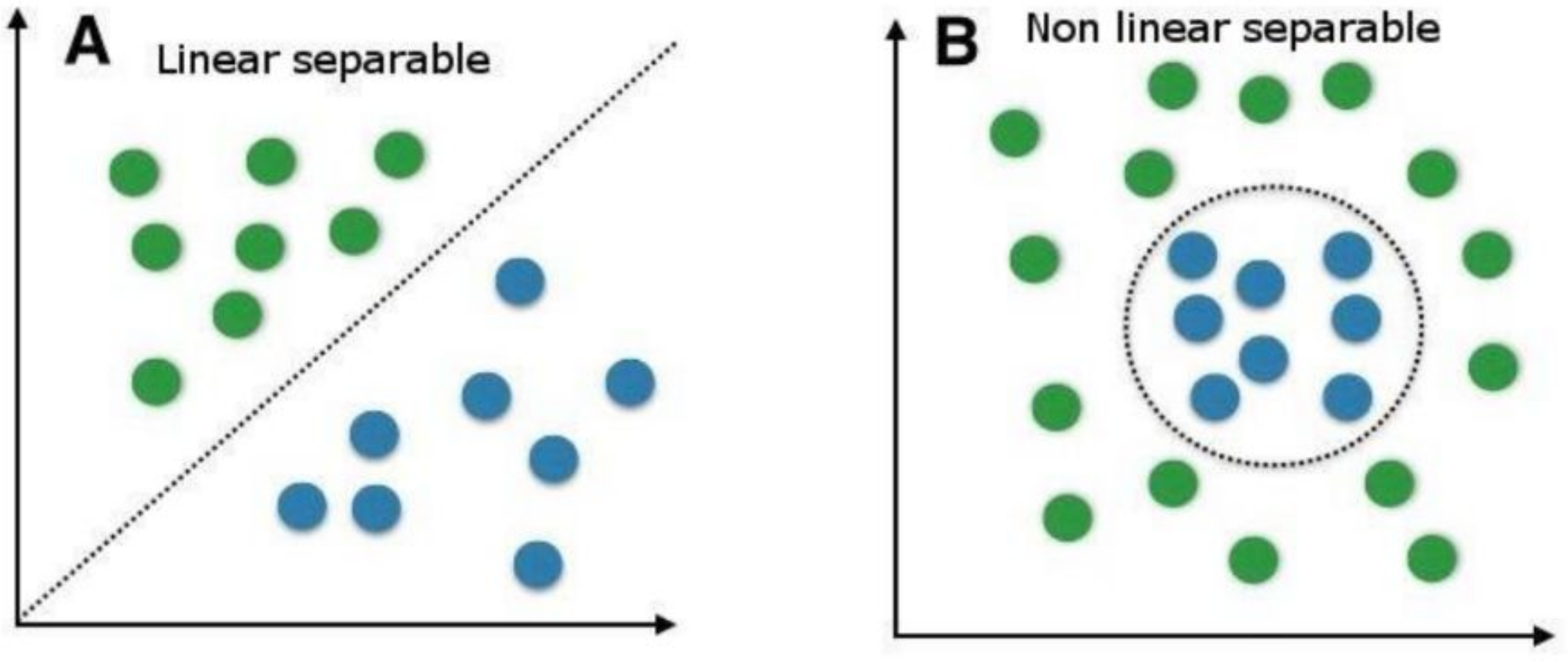


Построение линейной модели

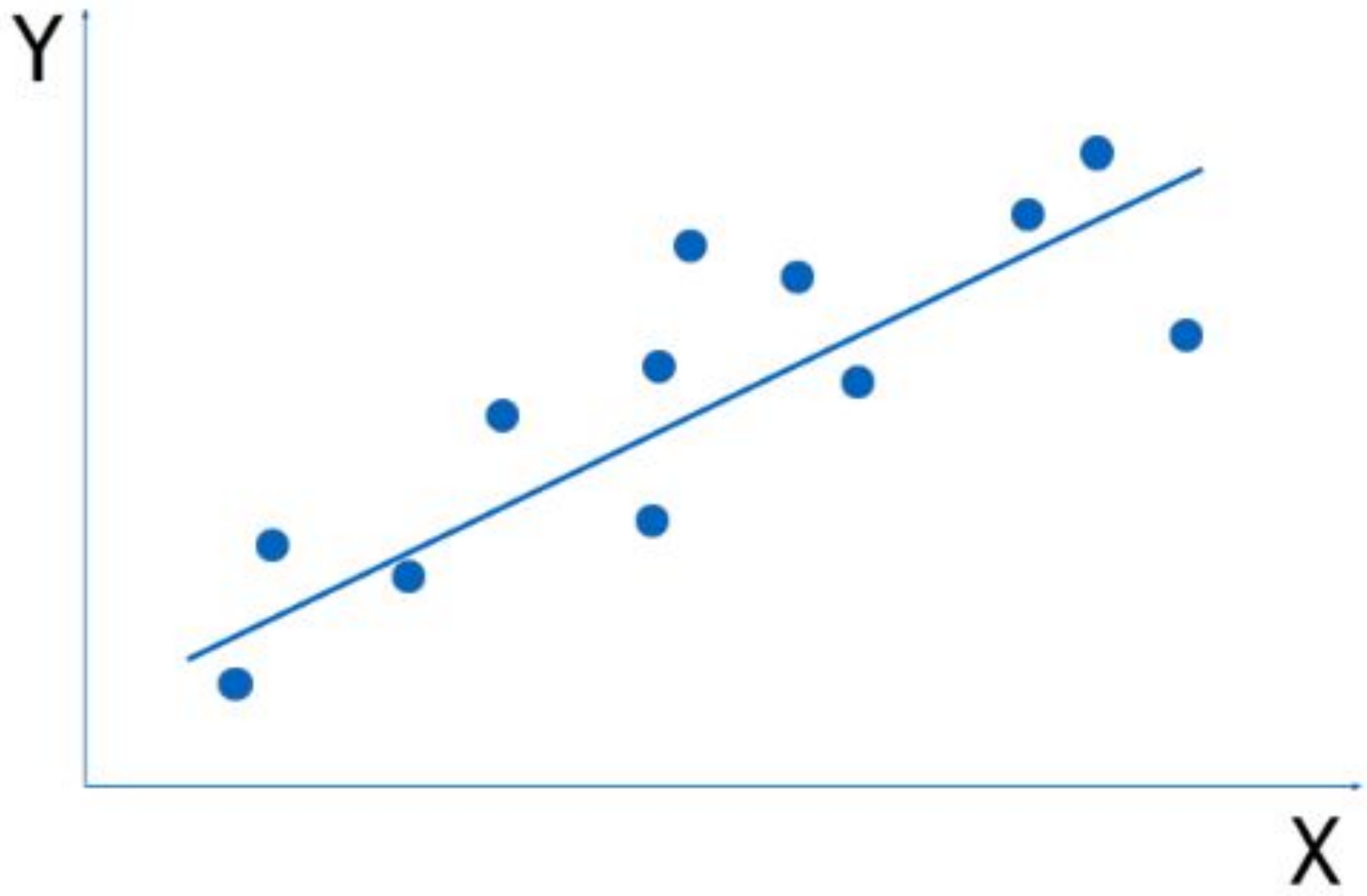
2

нетология

Как строим линейную модель



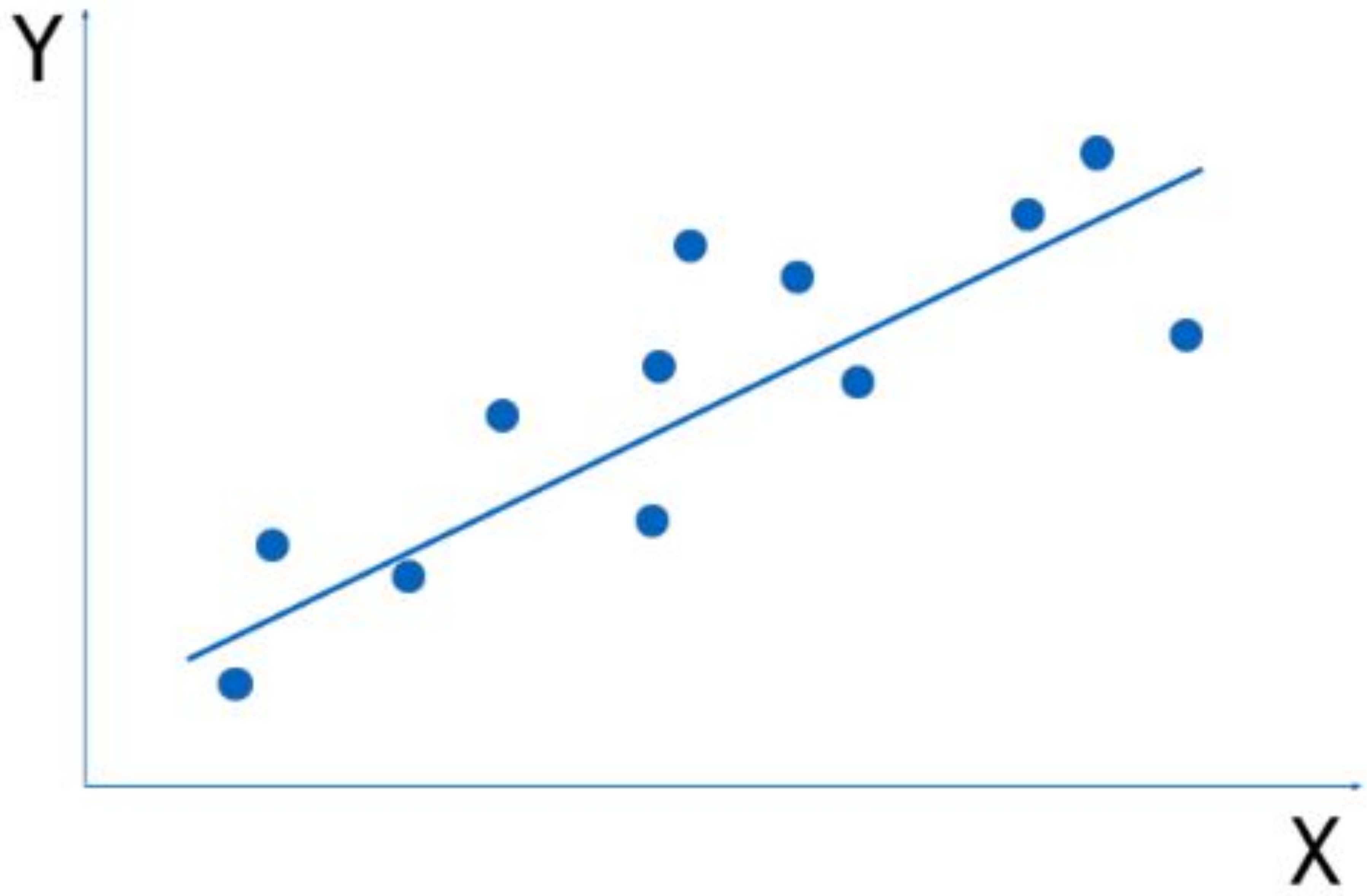
Метод максимального правдоподобия



Как можно получить эту прямую?

p (y | x. a) — вероятность получить *y* при входных данных *x*. *a* — параметр модели

Метод максимального правдоподобия



Как можно получить эту прямую?

p (y | x. a) — вероятность получить *y* при входных данных *x*. *a* — параметр модели

Введем функцию:

$$W(\alpha) = \prod_i p(x_i, \alpha)$$

Метод максимального правдоподобия

Функция максимального правдоподобия:

$$L(\alpha) = \sum_i \log p(x_i, \alpha)$$

Метод максимального правдоподобия

Функция максимального правдоподобия:

$$L(\alpha) = \sum_i \log p(x_i, \alpha)$$

Как подобрать значение a, чтобы максимизировать L (a)?

Необходимо минимизировать среднеквадратичную ошибку между прогнозными и фактическими значениями

Доказательство
https://habrahabr.ru/company/ods/blog/323890/#metod-maksimalnogo-pravdopodobiya

Время кода

REGRESSION_CARS.IPYNB

Практическое задание 1

3

нетология

Время практики

SAT_MODEL.IPYNB

Логистическая регрессия

4

нетология

Прогноз вероятности

Прогнозирует вероятность отнесения наблюдения к определенному классу

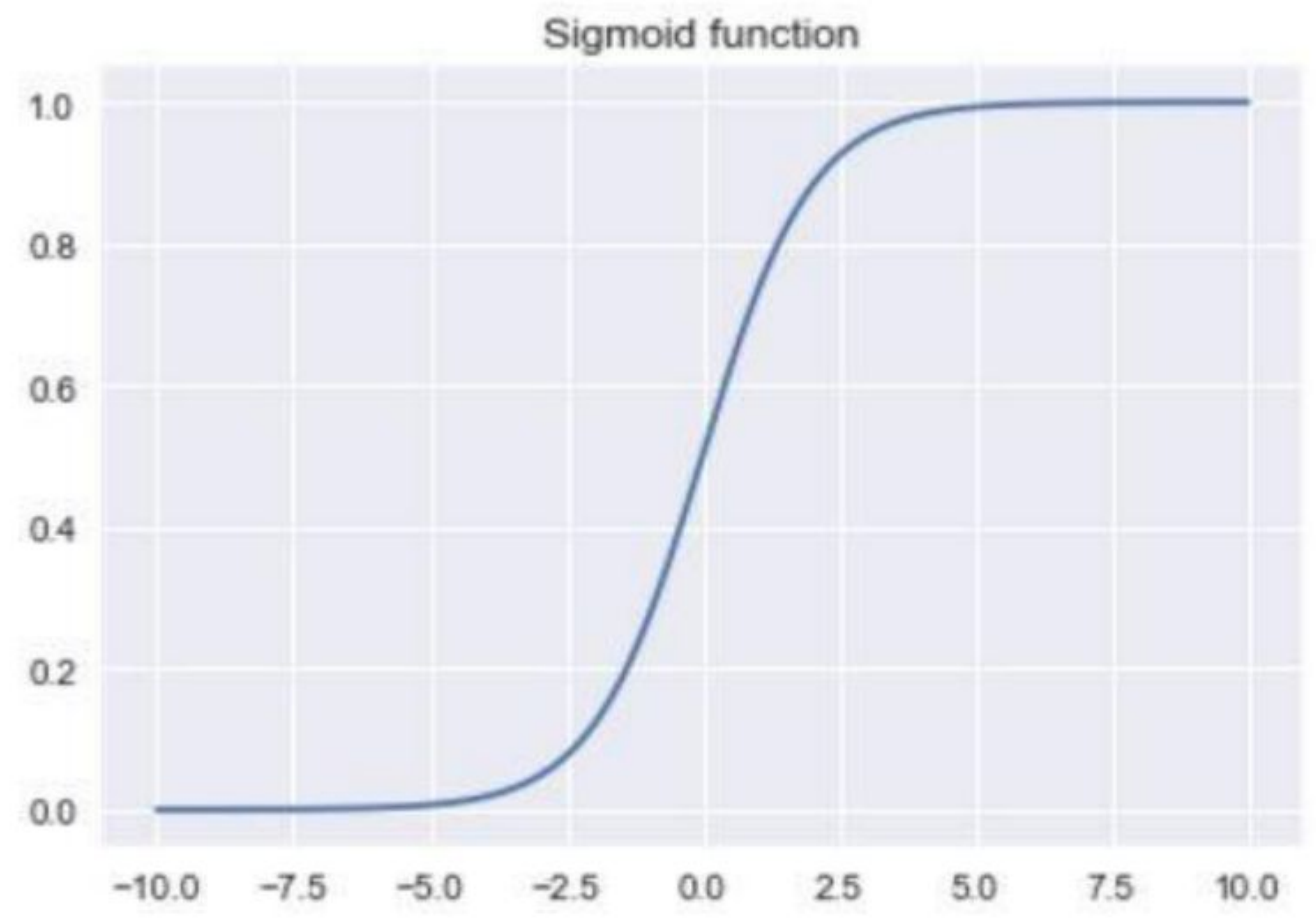
Модель:

$$L = a_0 + a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

Прогноз вероятности

Вероятность:

$$p = \frac{1}{1 + e^{-L}}$$



Основа практики

LOGISTIC_REGRESSION_ATHLETES_CLASSIFIER.IPYNB

Практическое задание 2

4

нетология

Улучшаем точность модели

С новыми признаками

Если классов больше двух

5

нетология

Пример

IRIS_DATASET.IPYNB

Для каких данных это работает?

6

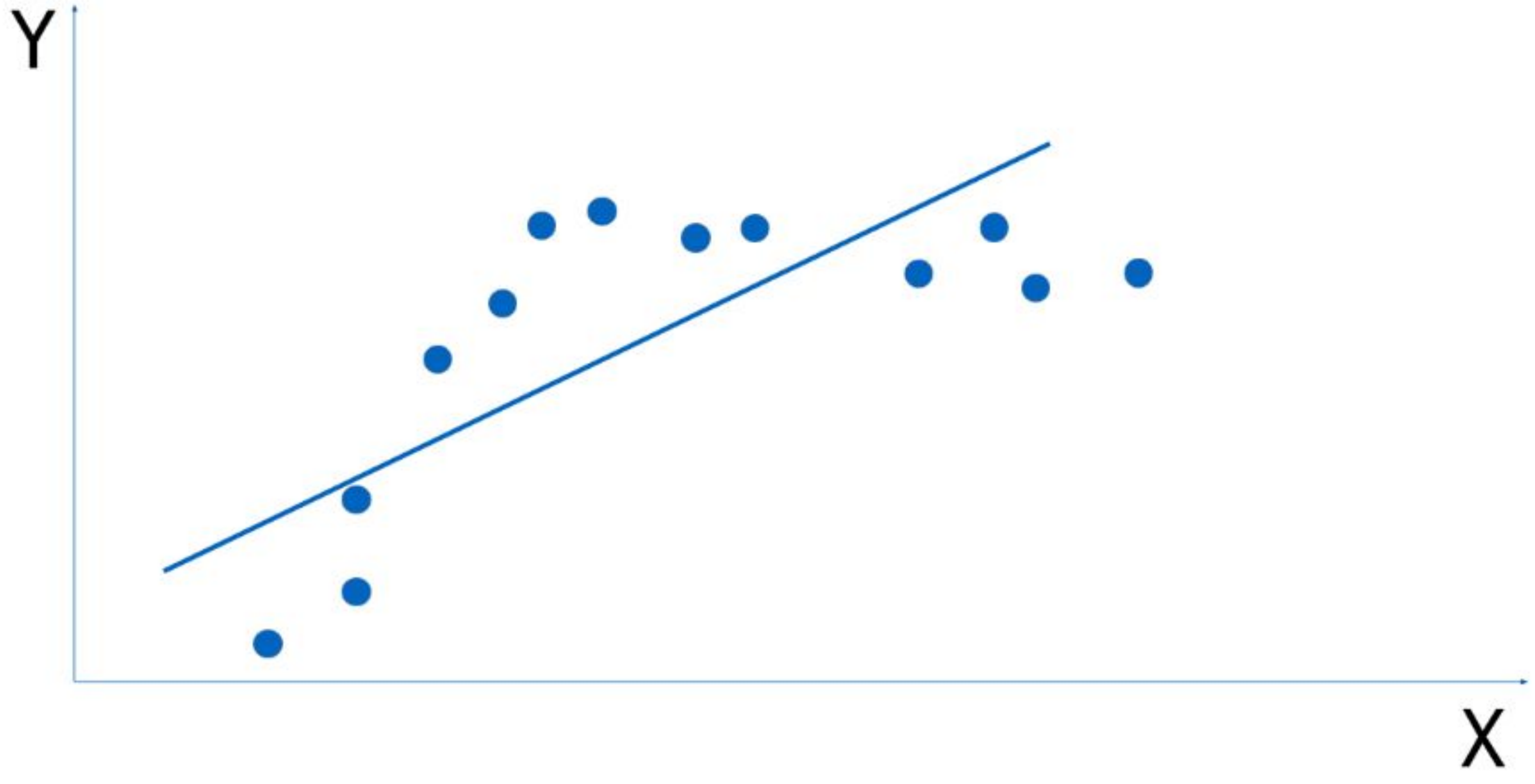
Требования к данным

Линейная зависимость целевой переменной

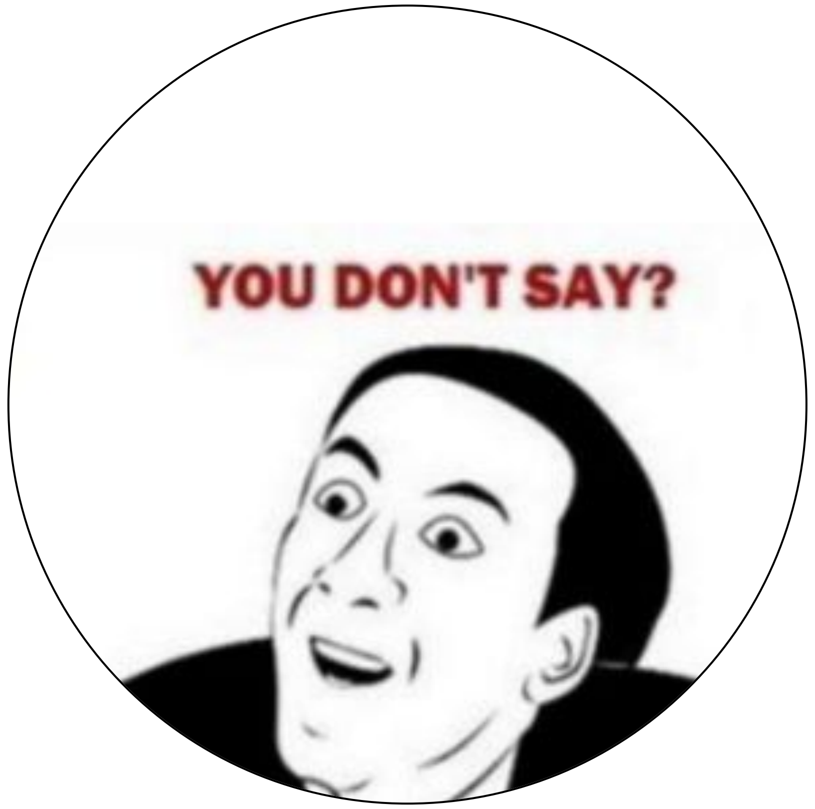
Нормальное распределение остатков

Постоянная изменчивость остатков

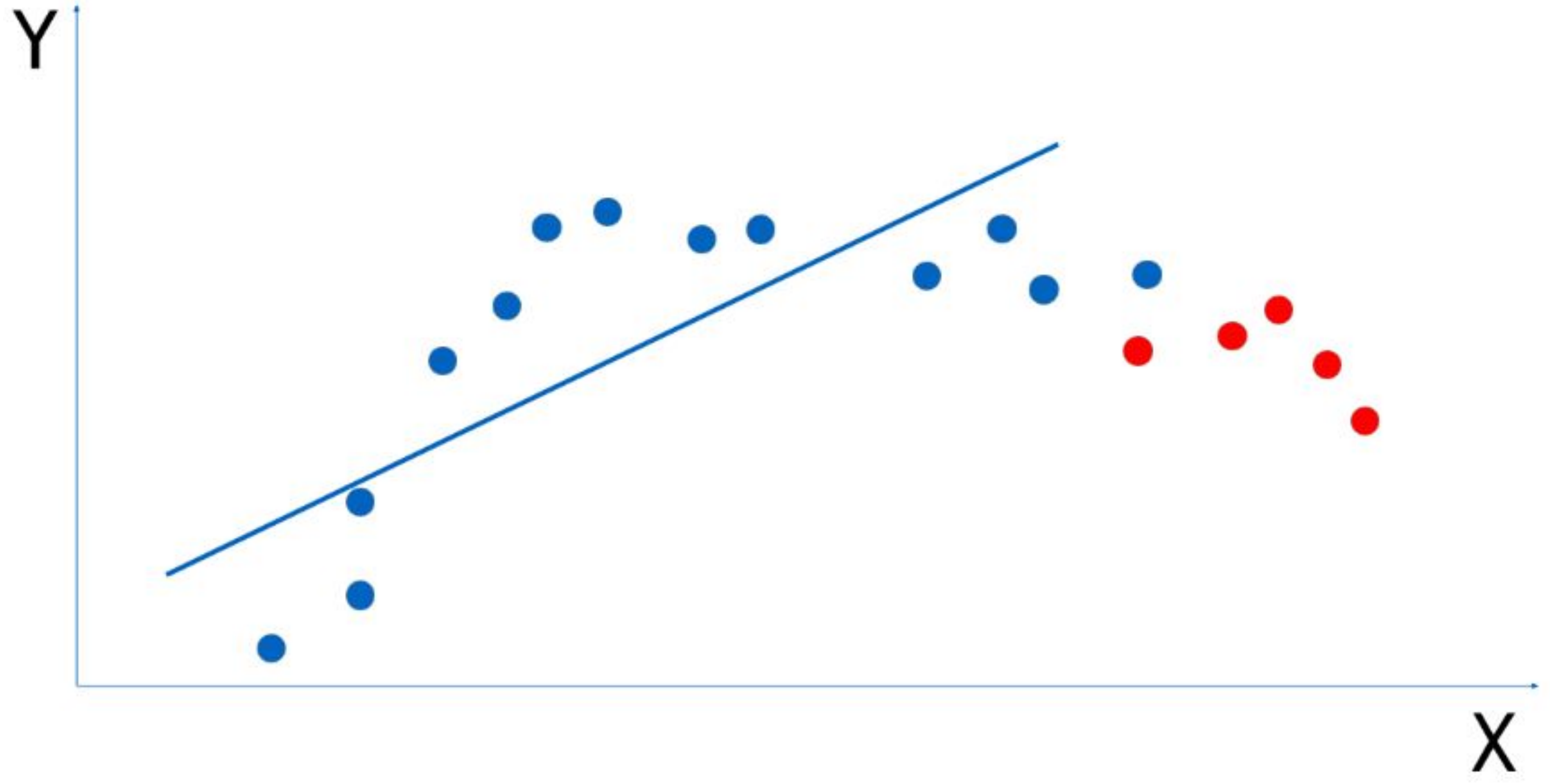
Требования к данным



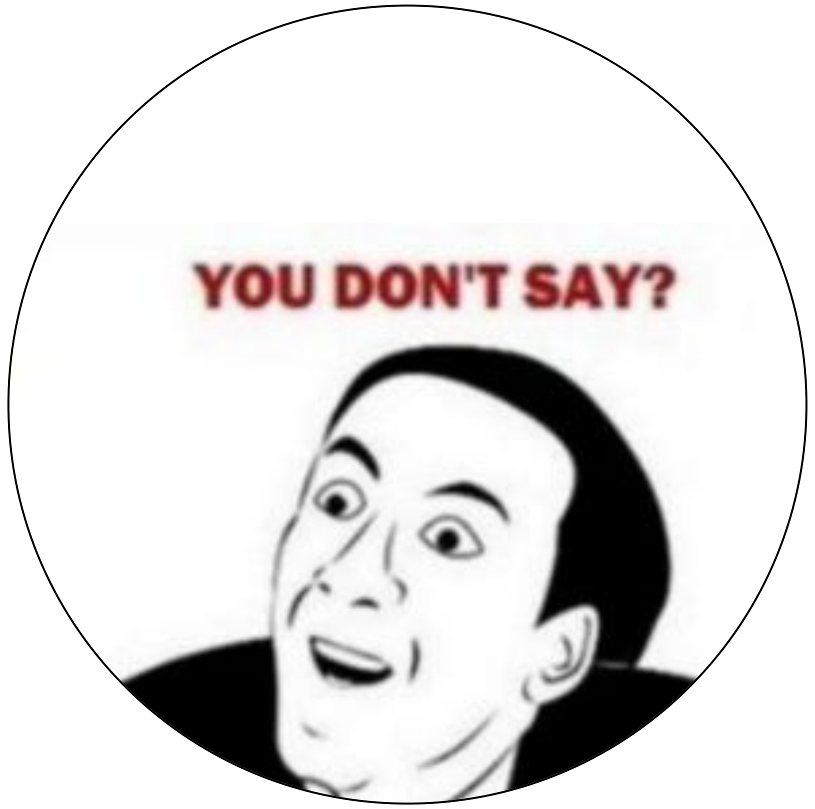
Линейная взаимосвязь X и Y



Требования к данным

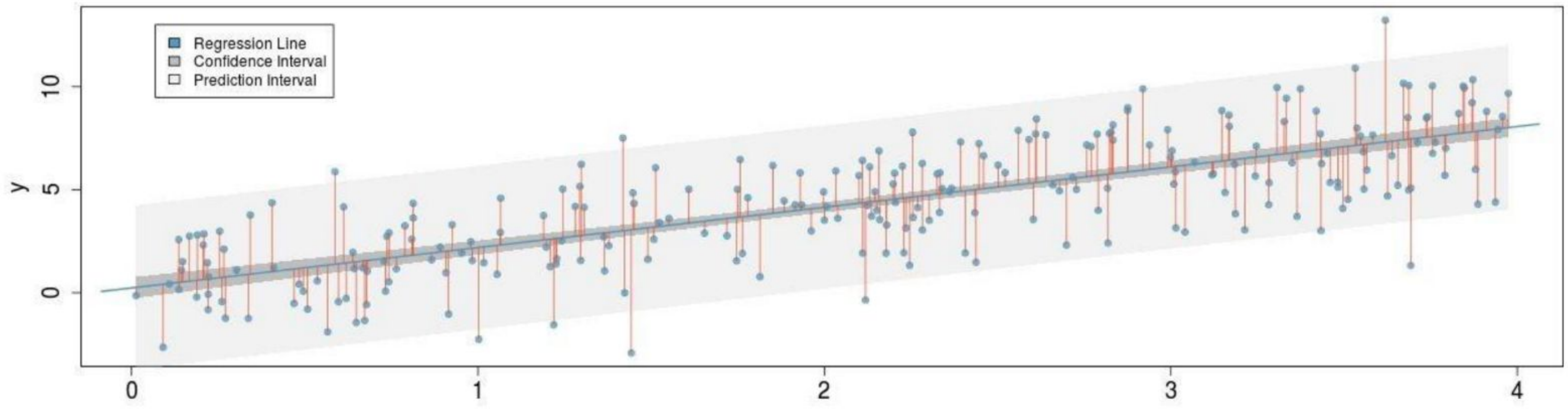


Линейная взаимосвязь X и Y



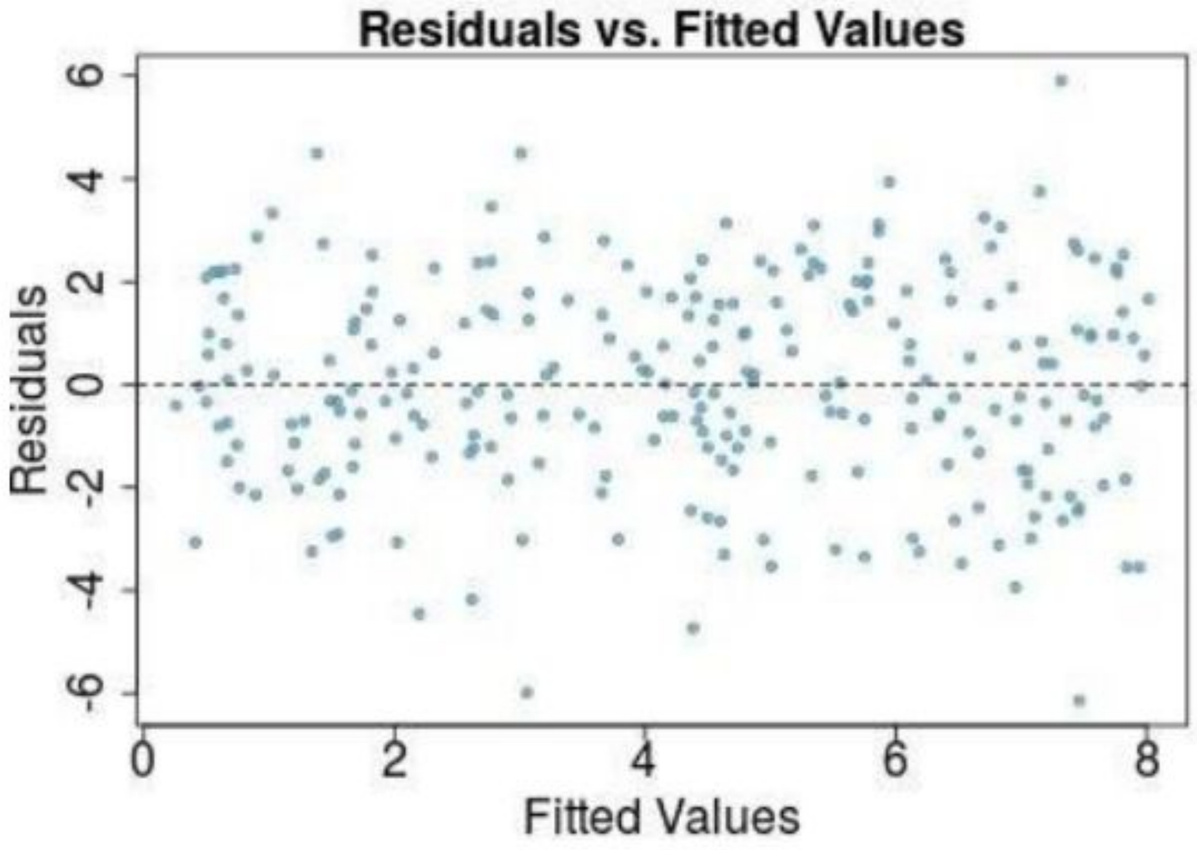
Нормальное распределение остатков

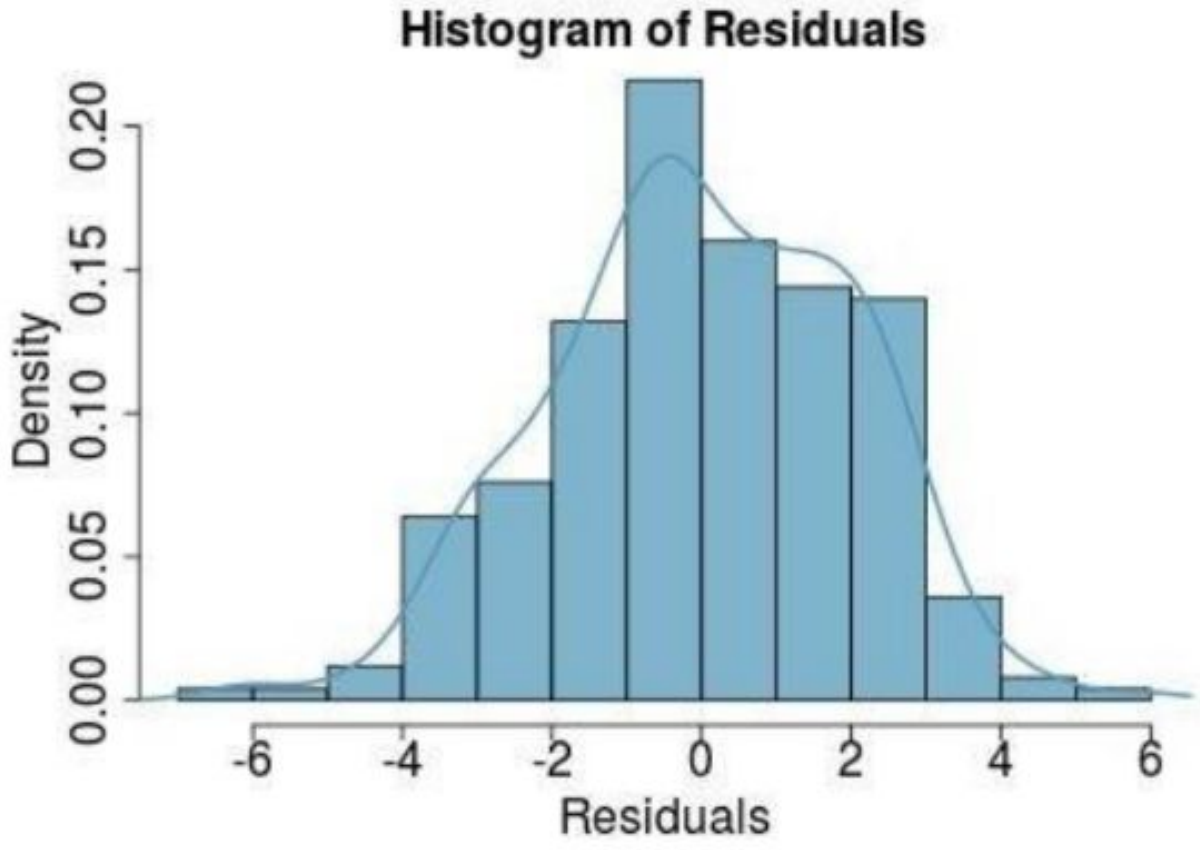
HTTPS:/ GALLERY.SHINYAPPS.IO/SLR_DIAG/

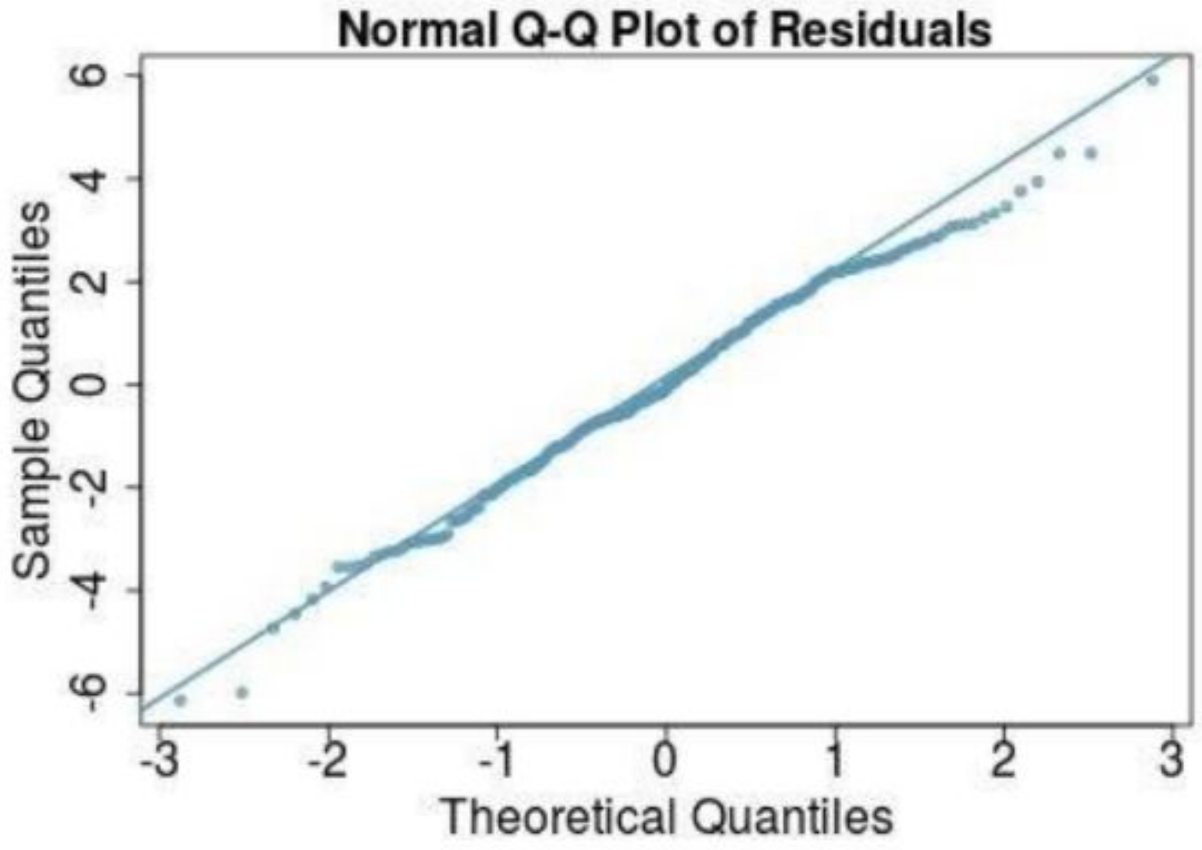


Нормальное распределение остатков

HTTPS:/ GALLERY.SHINYAPPS.IO/SLR_DIAG/



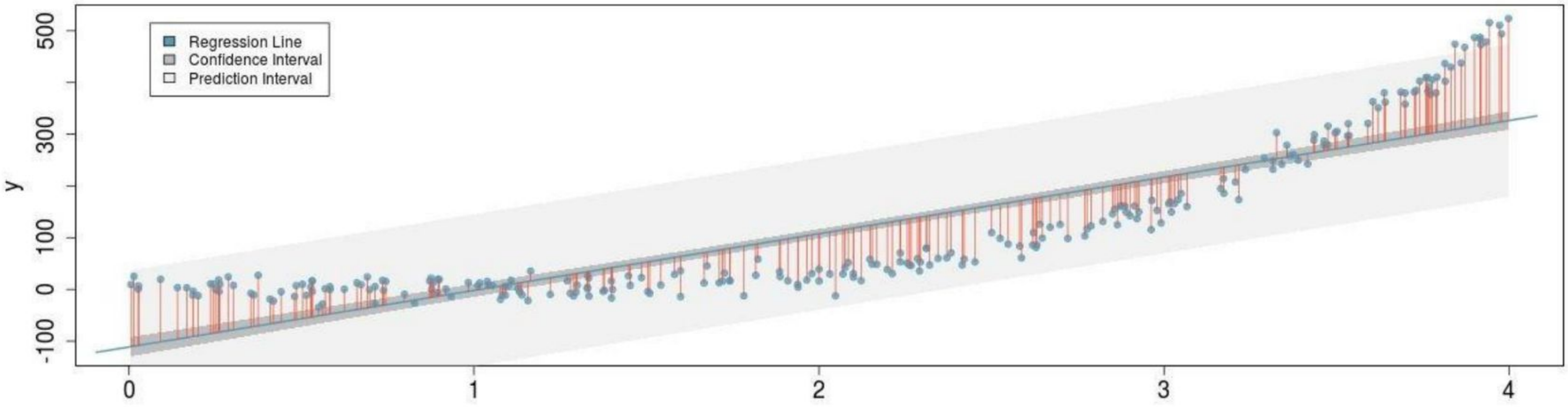




Гомоскедастичность

Постоянная изменчивость остатков

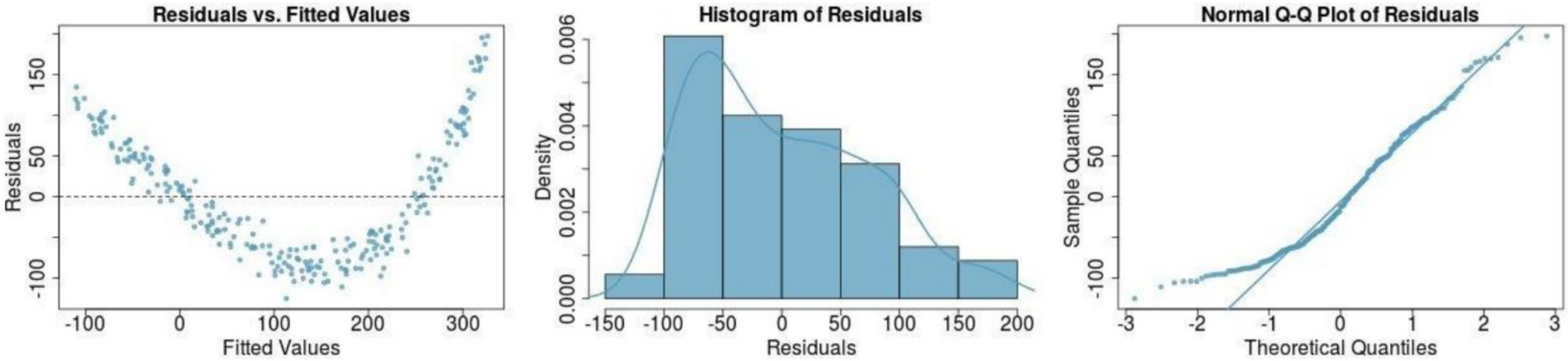
Пример постоянной гетероскедастичной последовательности



Гомоскедастичность

Постоянная изменчивость остатков

Пример постоянной гетероскедастичной последовательности



SVM

7

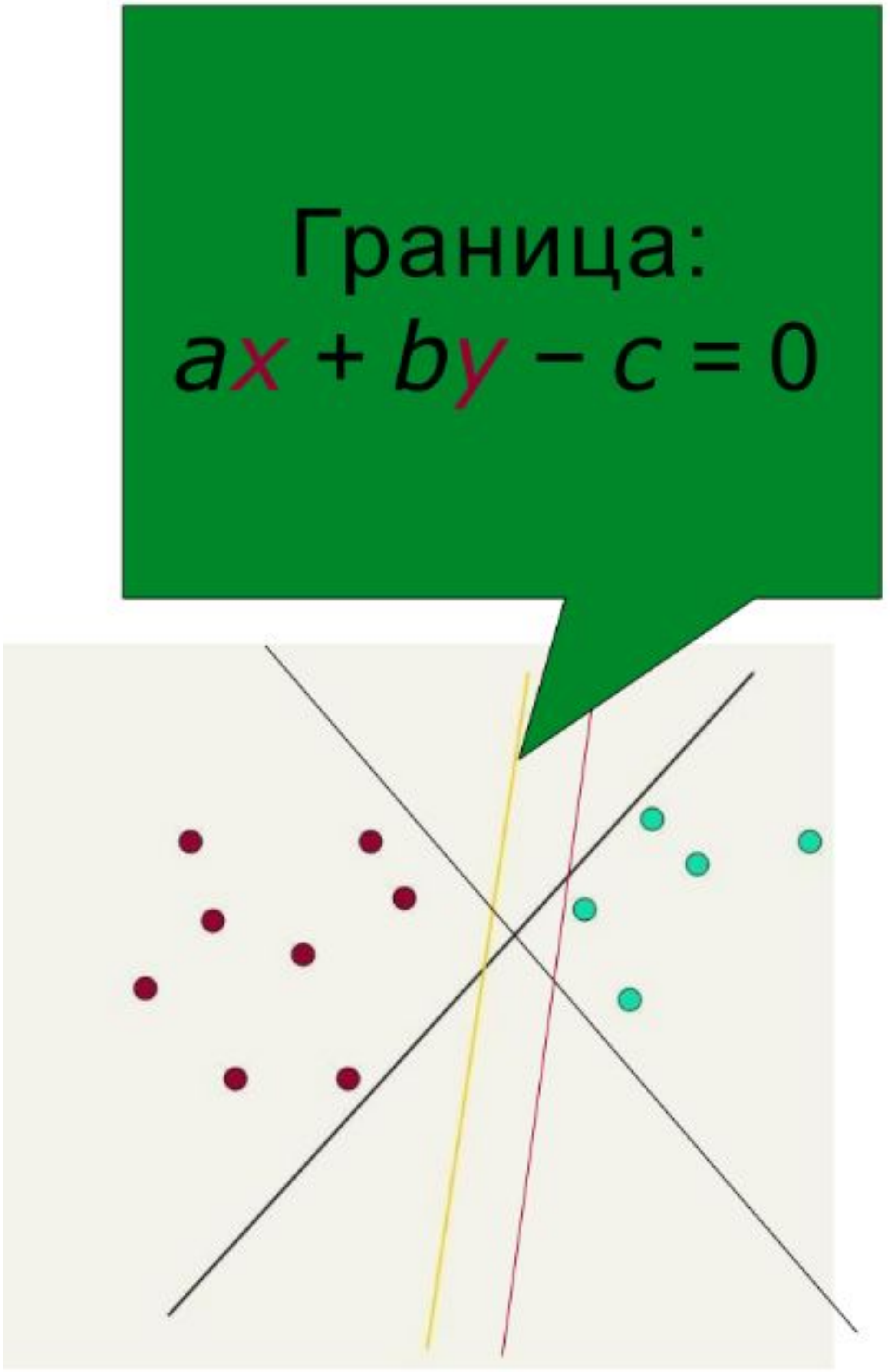
нетология

Множество гиперплоскостей

Множество решений для *a,b,c.*

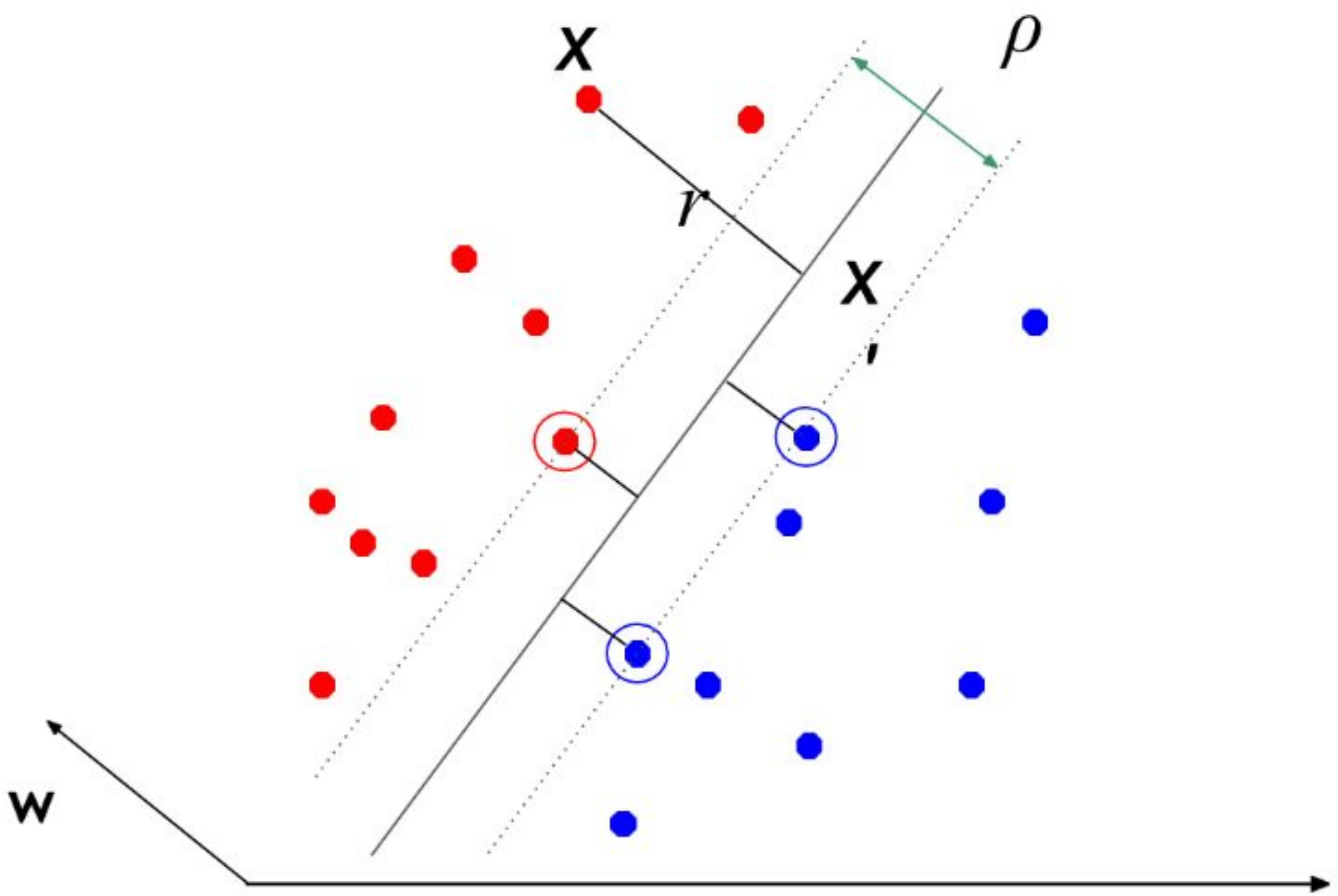
SVM находит оптимальную разделяющую поверхность

Максимизирует «зазор»



Максимальный зазор

- w – нормаль к разделяющей плоскости
- xi - sample
- yi: - класс sample i (+1 or -1) (важно, не 1 и 0)
- Классификатор: f(xi) = sign(wTxi +b)
- Зазор для точки x $r = y\frac{\mathbf{w}^T\mathbf{x}+b}{\|\mathbf{w}\|}$
- Зазор всего датасета – минимум зазора для всех точек



Формула

Итого получаем задачу оптимизации:

Найти $\mathbf{w}$ и b такие что

максимально; и для всех $\{(\mathbf{x_i}, y_i)\}$

$\mathbf{w^T x_i} + b \geq 1$ если $y_i = 1$; $\mathbf{w^T x_i} + b \leq -1$ если $y_i = -1$

Перепишем в более понятном виде:

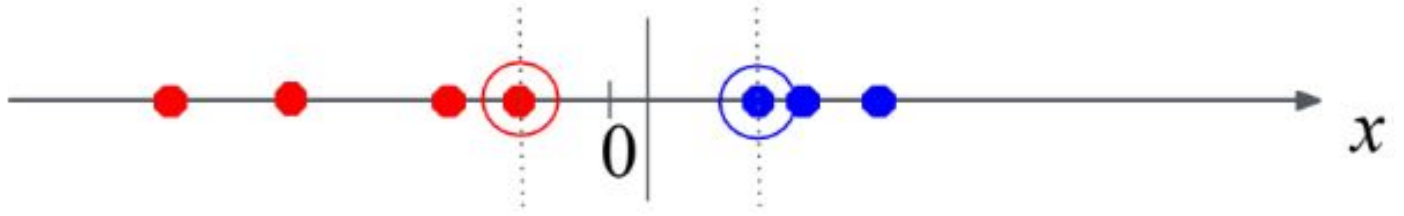
Найти $\mathbf{w}$ и b такие что

$\mathbf{\Phi(w)} = \mathbf{0.5\, w^T w}$ максимально

И для всех $\{(\mathbf{x_i}, y_i)\}$: $y_i(\mathbf{w^T x_i} + b) \geq 1$

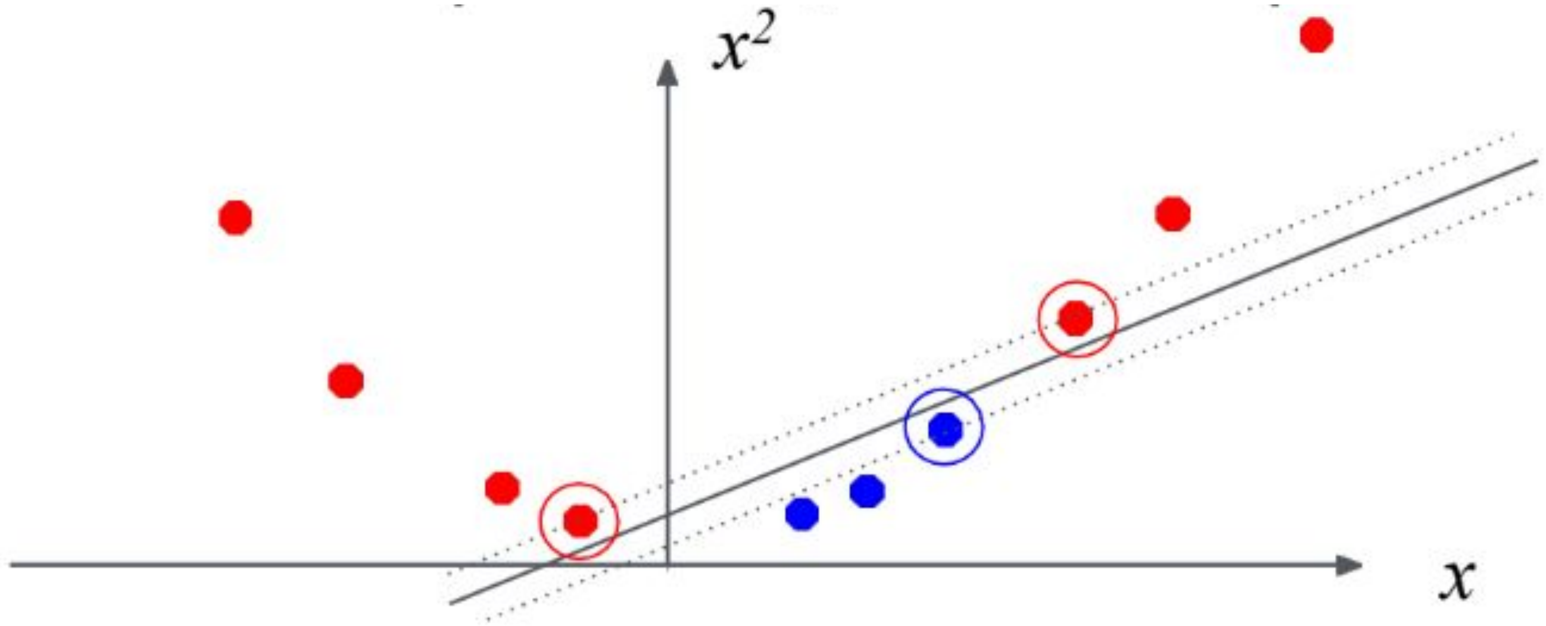
Non-linear SVMs

Линейно разделимые датасеты хорошо классифицируются



Но что делать, если они не линейно разделимы?

Можно попробовать отобразить данные в пр-во более высокой размерности



The «Kernel Trick»

- SVM зависит от скалярного произведения K(xi,xj)=xiTxj

- Если каждая точка отображается в пр-во более высокой размерности при помощи Ф: x ➔ φ(x), тогда скалярное произведение становится:

- K(xi,xj)= φ(xi) Tφ(xj)

- Функция ядра - это функция, соответствующая скалярному произведению в пр-ве более высокой размерности

Kernels

- Примеры
- Линейное
- Полиномное K(x,z) =(1+xTz)d
- RBF

$$K(x_i, x_i) = e^{||x_i - x_j||^2 / 2\boldsymbol{\sigma}^2}$$

Что мы сегодня узнали

8

Итоги занятия

- Вспомнили основы теории вероятностей
- Изучили линейные модели и требования к ним на основе функции правдоподобия
- Реализовали логистическую регрессию
- Изучили алгоритм градиентного спуска и потренировались в его реализации

Полезные материалы

9

нетология

Статья о линейных моделях в ODS
https://habrahabr.ru/company/ods/blog/323890/

Курс «Основы статистики» на Stepik.org
https:/stepik.org/course/Основы-статистики-76

Спасибо за внимание!

fb.com/

mail@mail.com

нетология