НЕТОЛОГИЯ
групп

ЗАНЯТИЕ 1.2

ЛИНЕЙНЫЙ КЛАССИФИКАТОР И ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

Артур Сапрыкин

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 ЧЁМ ПОГОВОРИМ И ЧТО СДЕЛАЕМ

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

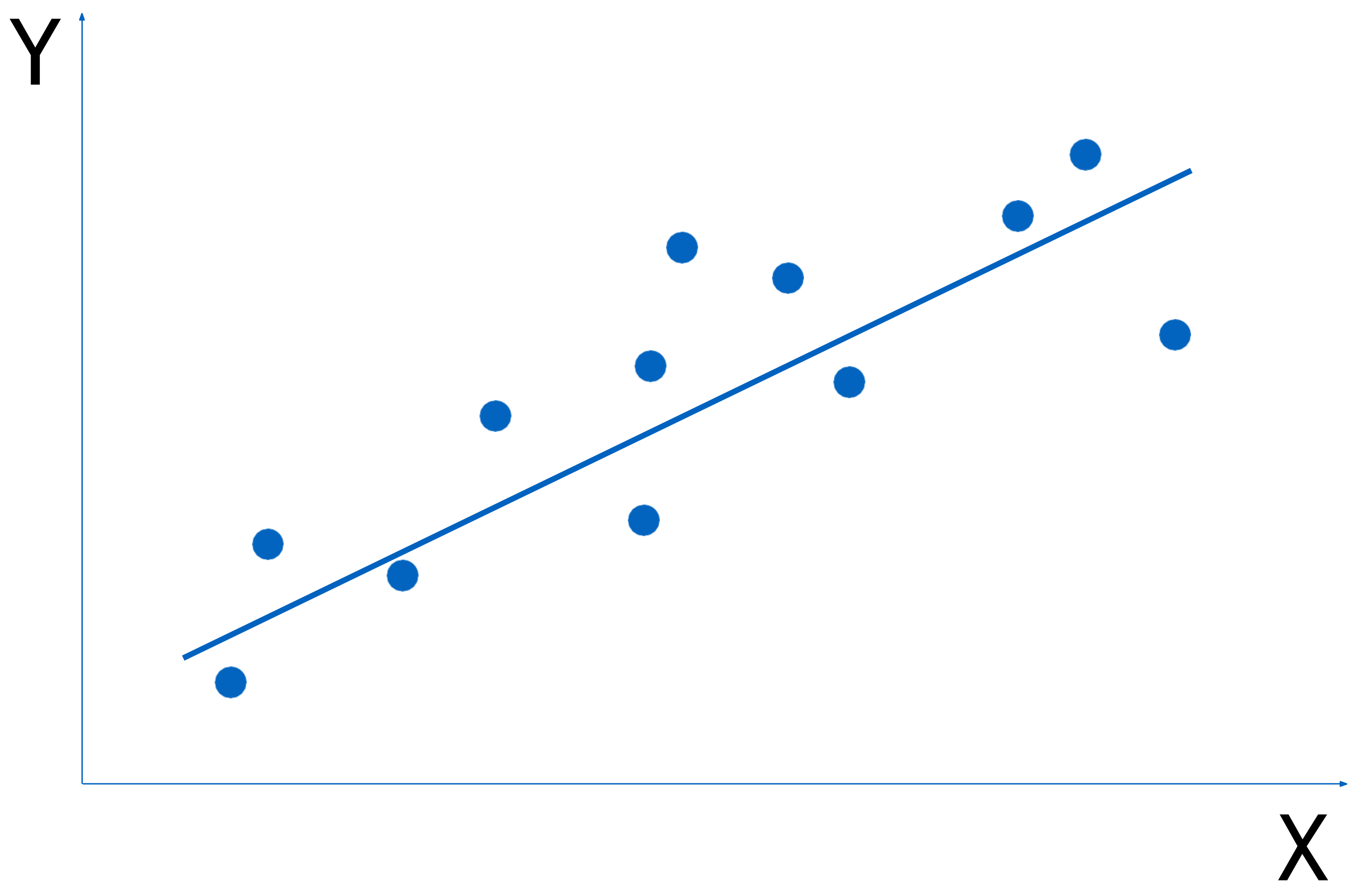
- Линейные модели подходят для описания многих процессов
- Относительная простота вычислений и интерпретации результатов
- Вклад нескольких факторов часто можно разбить на сумму влияния каждого фактора в отдельности

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

ОПРЕДЕЛЕНИЕ



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

Y – целевая переменная

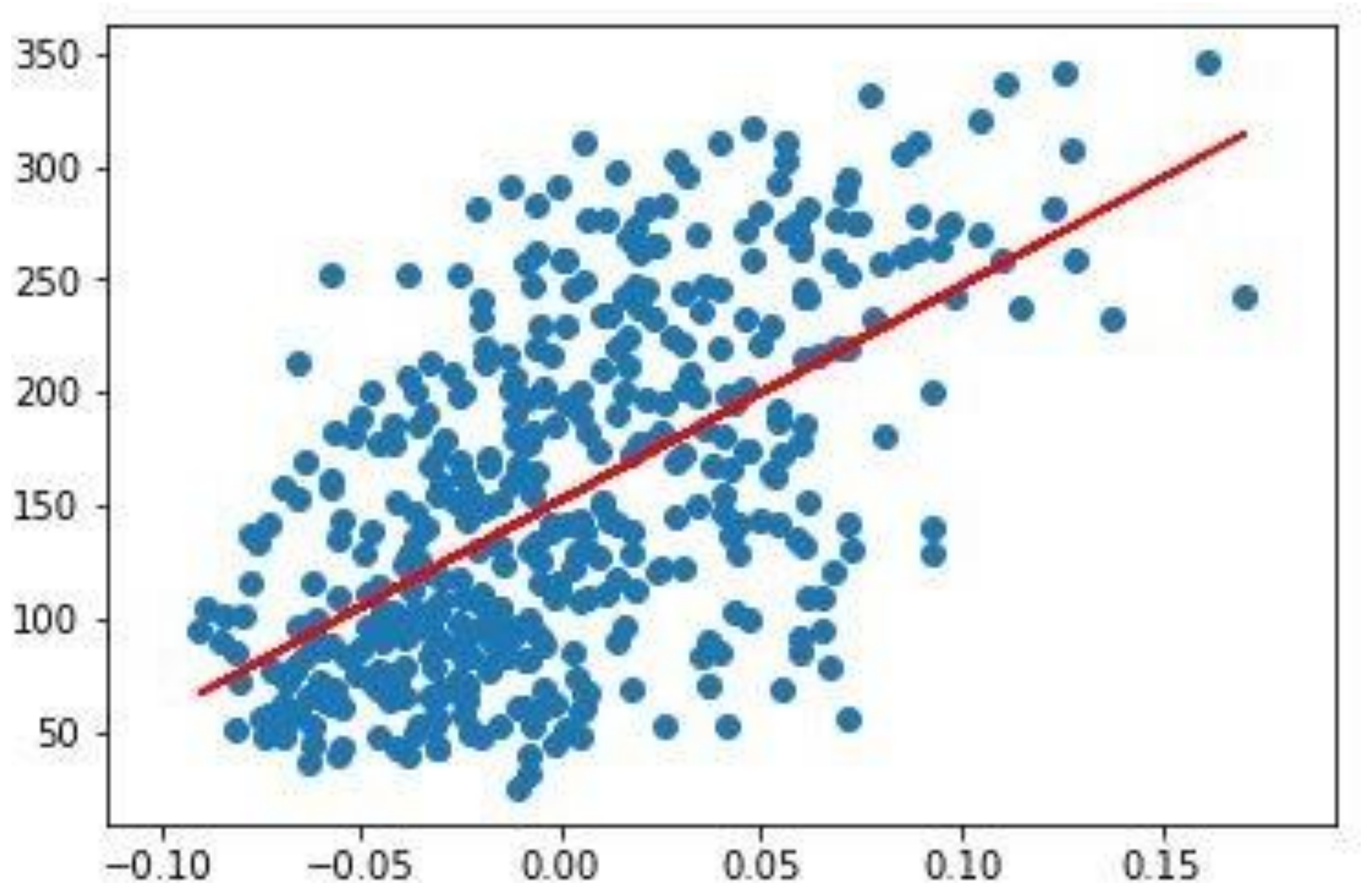
W – вектор весов модели

X – матрица наблюдений

e – ошибка модели

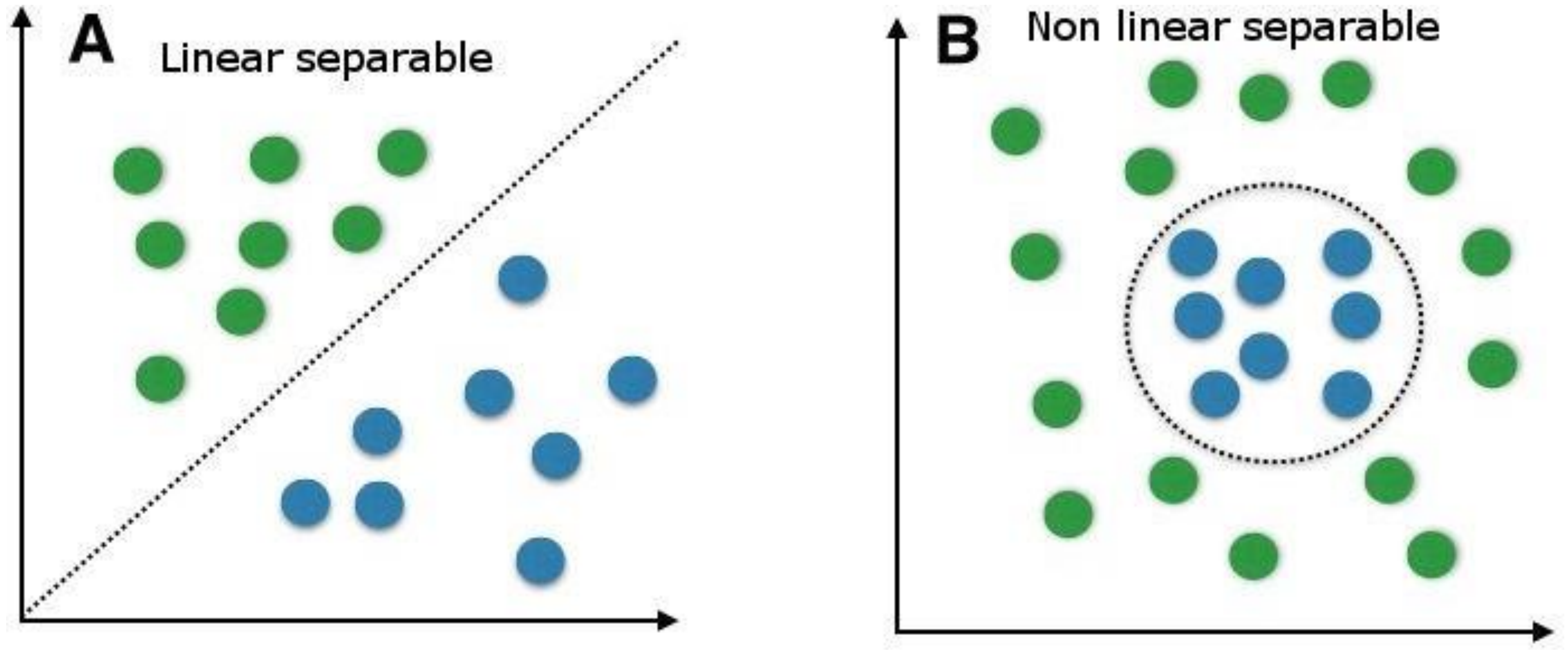
ПРИМЕР ИЗ КОДА

LINEAR REGRESSION.IPYNB

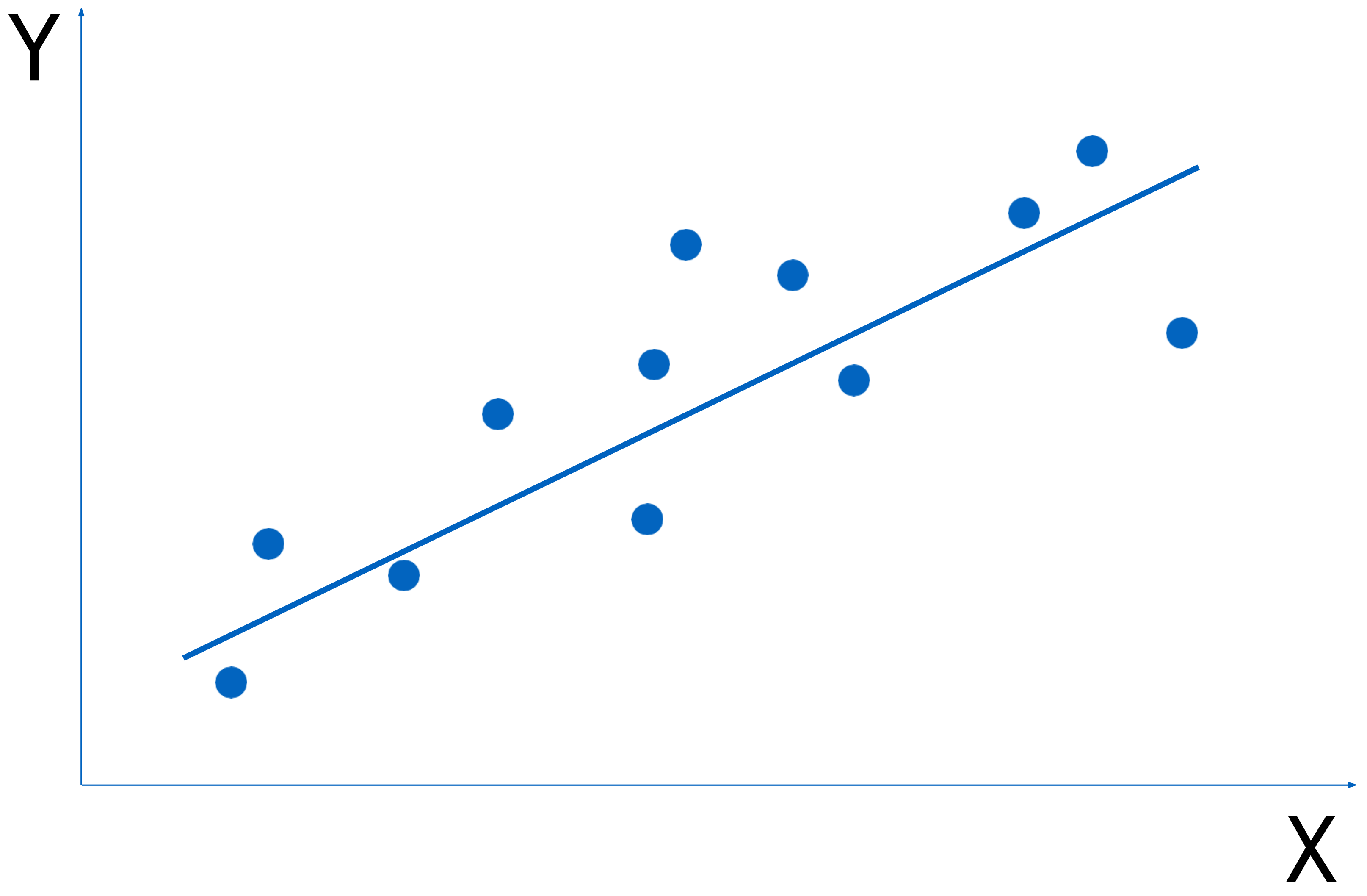


ПОСТРОЕНИЕ ЛИНЕЙНОЙ МОДЕЛИ

КАК СТРОИМ ЛИНЕЙНУЮ МОДЕЛЬ



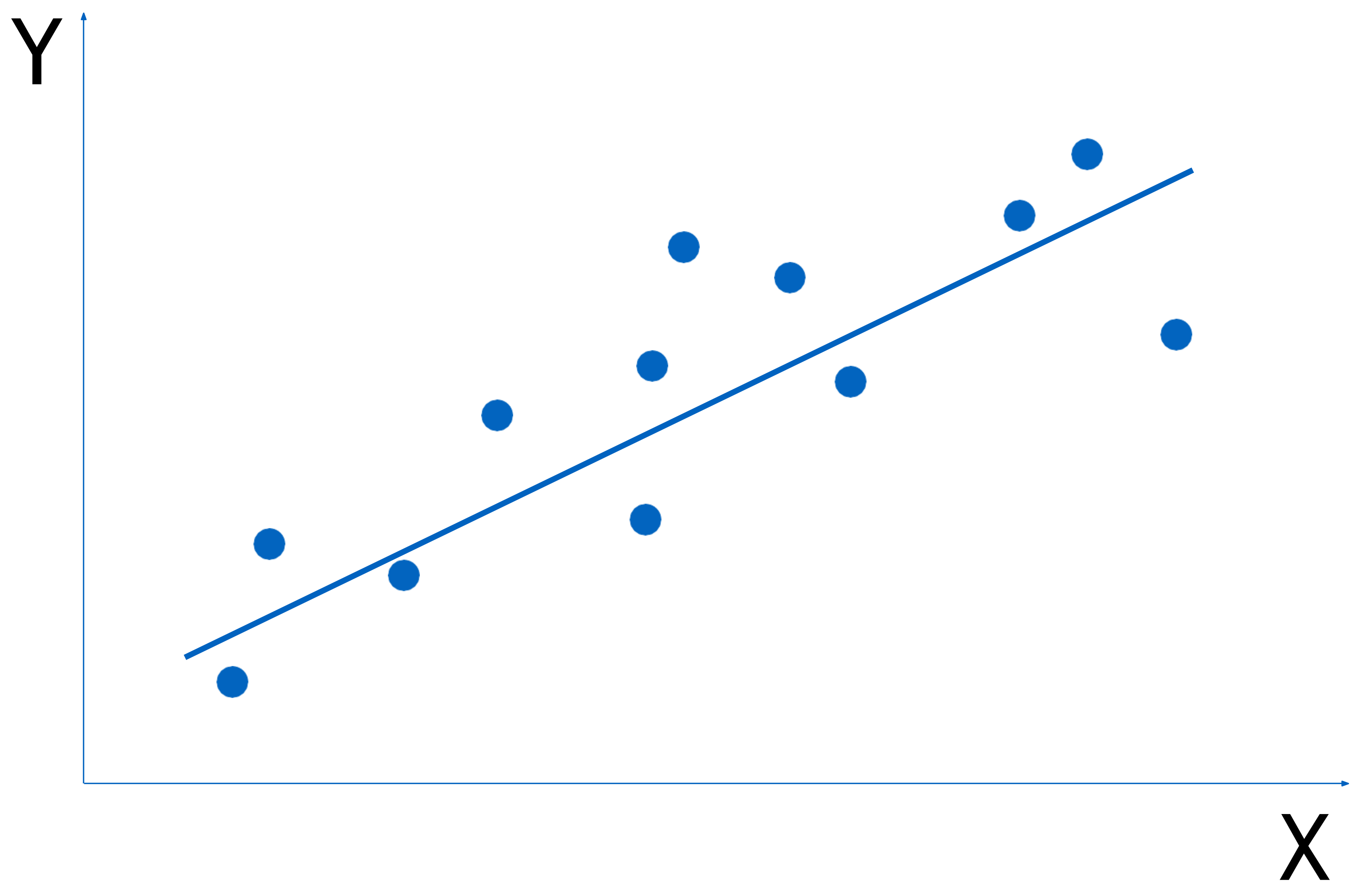
МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ



Как можно получить эту прямую?

p(y | x, α) – вероятность получить y при входных данных x. α – параметр модели

МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ



Как можно получить эту прямую?

p(y | x, α) – вероятность получить y при входных данных x. α – 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 α, чтобы максимизировать L(α)?

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

ВРЕМЯ ПРАКТИКИ

SAT_MODEL.IPYNB

ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

ПРОГНОЗ ВЕРОЯТНОСТИ

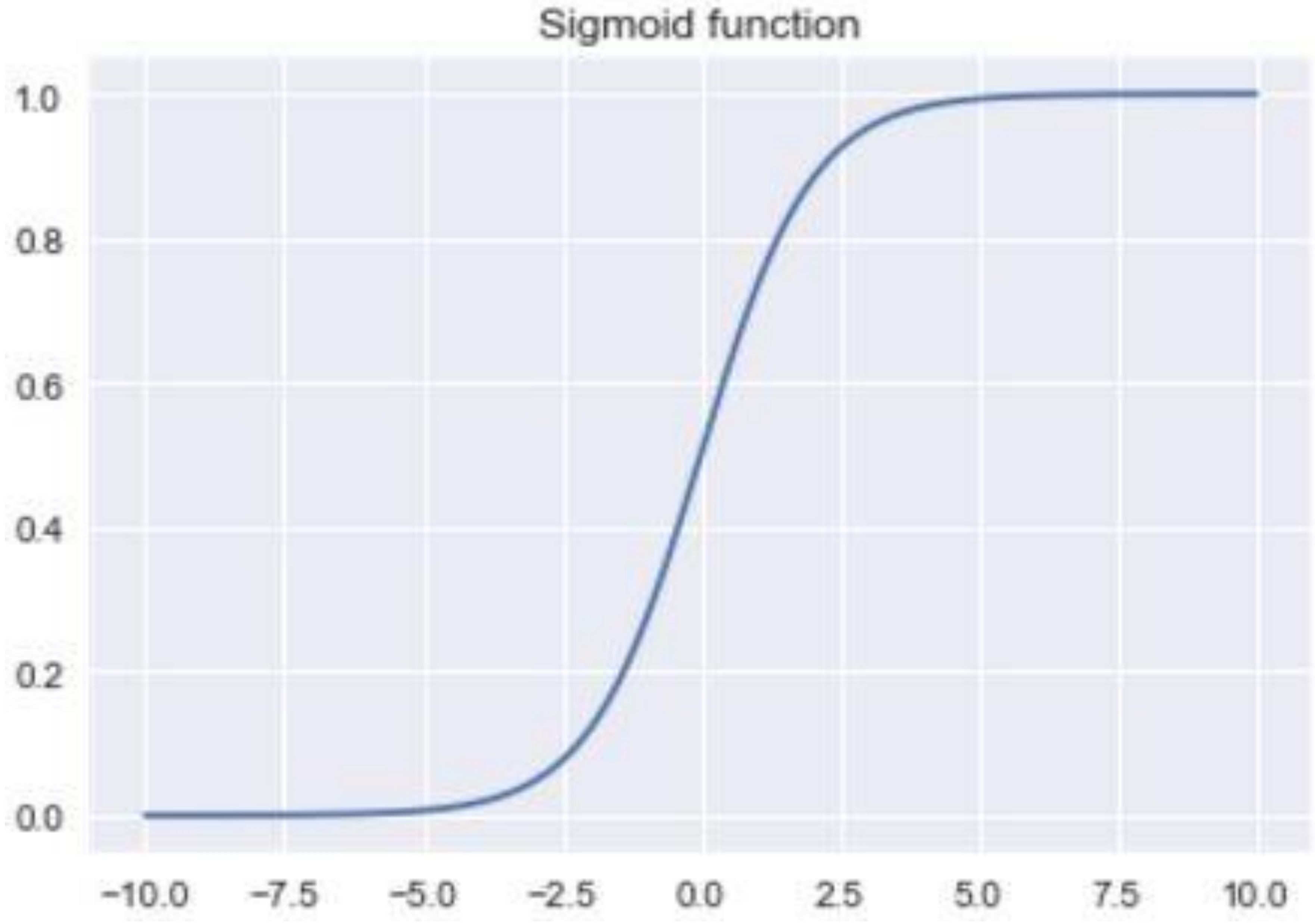
Прогнозирует вероятность отнесения наблюдения к определенному классу

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

ПРОГНОЗ ВЕРОЯТНОСТИ

Вероятность:

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



СНОВА ПРАКТИКА

LOGISTIC_REGRESSION_ATHLETES_CLASSIFIER.IPYNB

ПРАКТИЧЕСКОЕ ЗАДАНИЕ 2

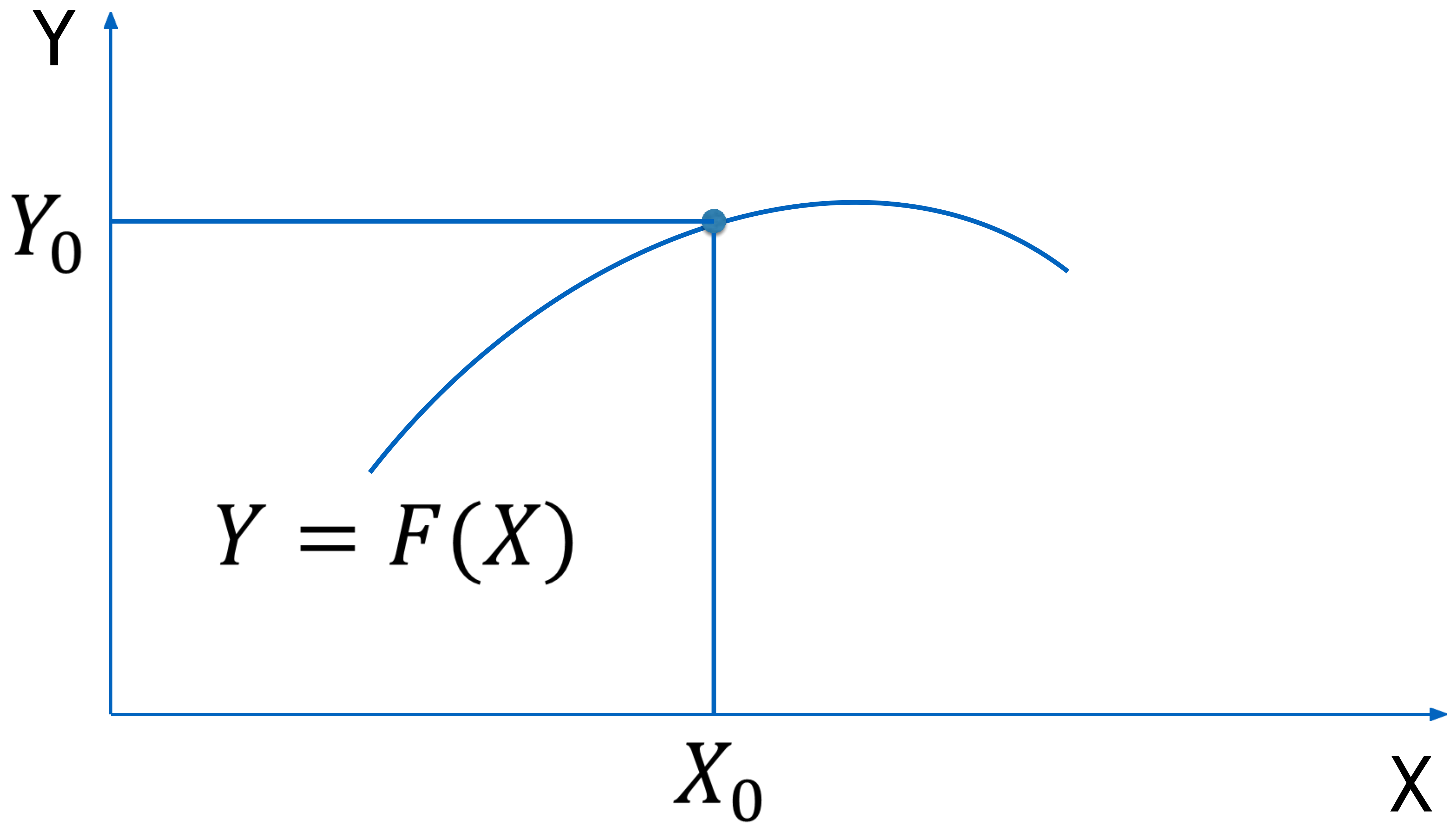
УЛУЧШАЕМ ТОЧНОСТЬ МОДЕЛИ

С НОВЫМИ ПРИЗНАКАМИ

ГРАДИЕНТНЫЙ СПУСК

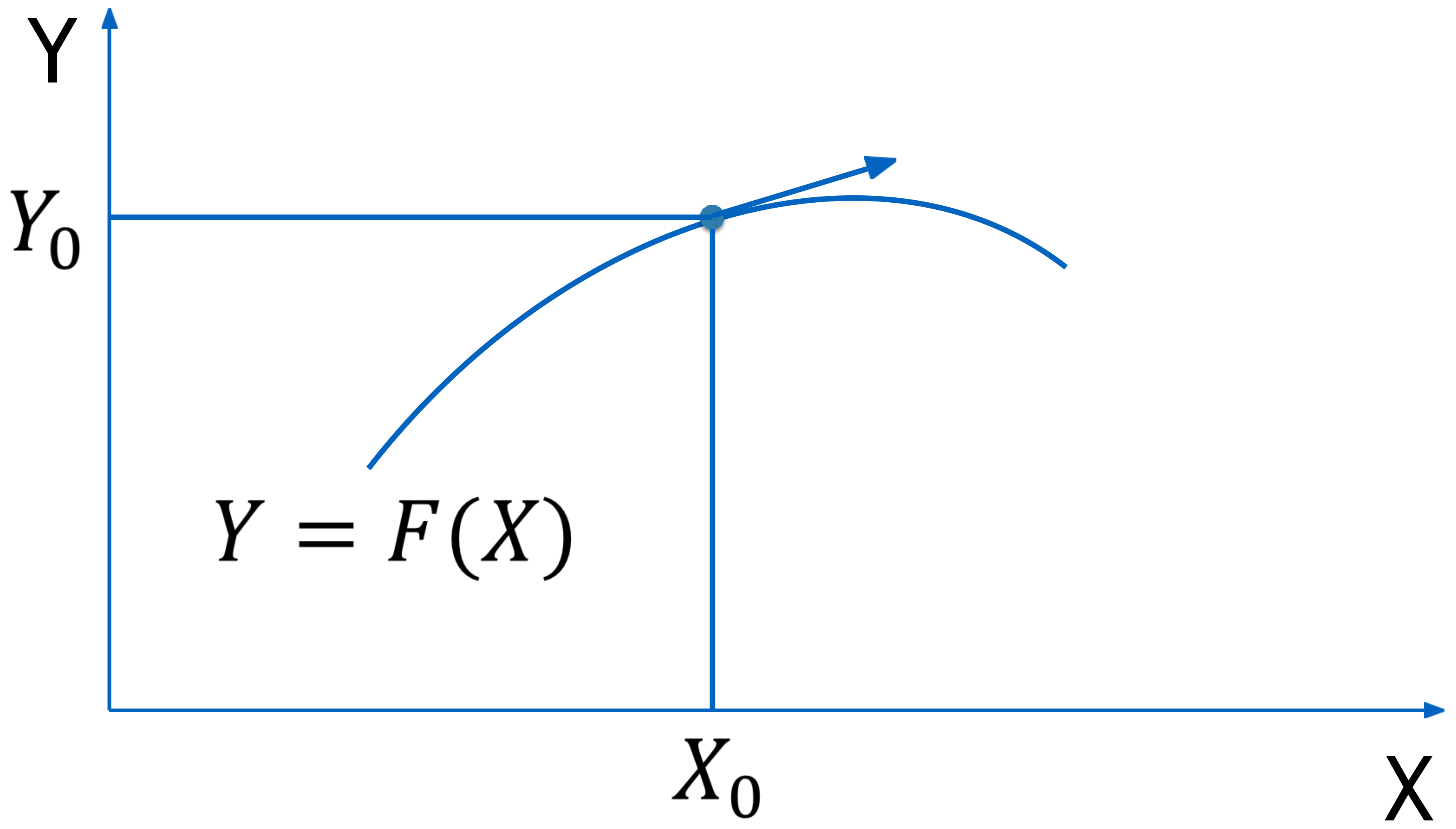
ПРОИЗВОДНАЯ И МИНИМУМ

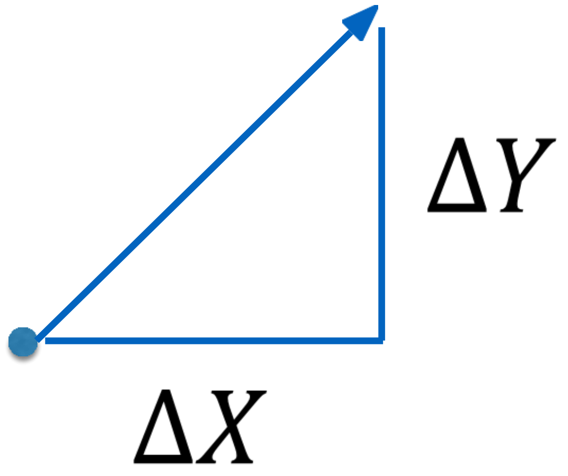
Производная определяет скорость изменения функции в точке



ПРОИЗВОДНАЯ И МИНИМУМ

Производная определяет скорость изменения функции в точке

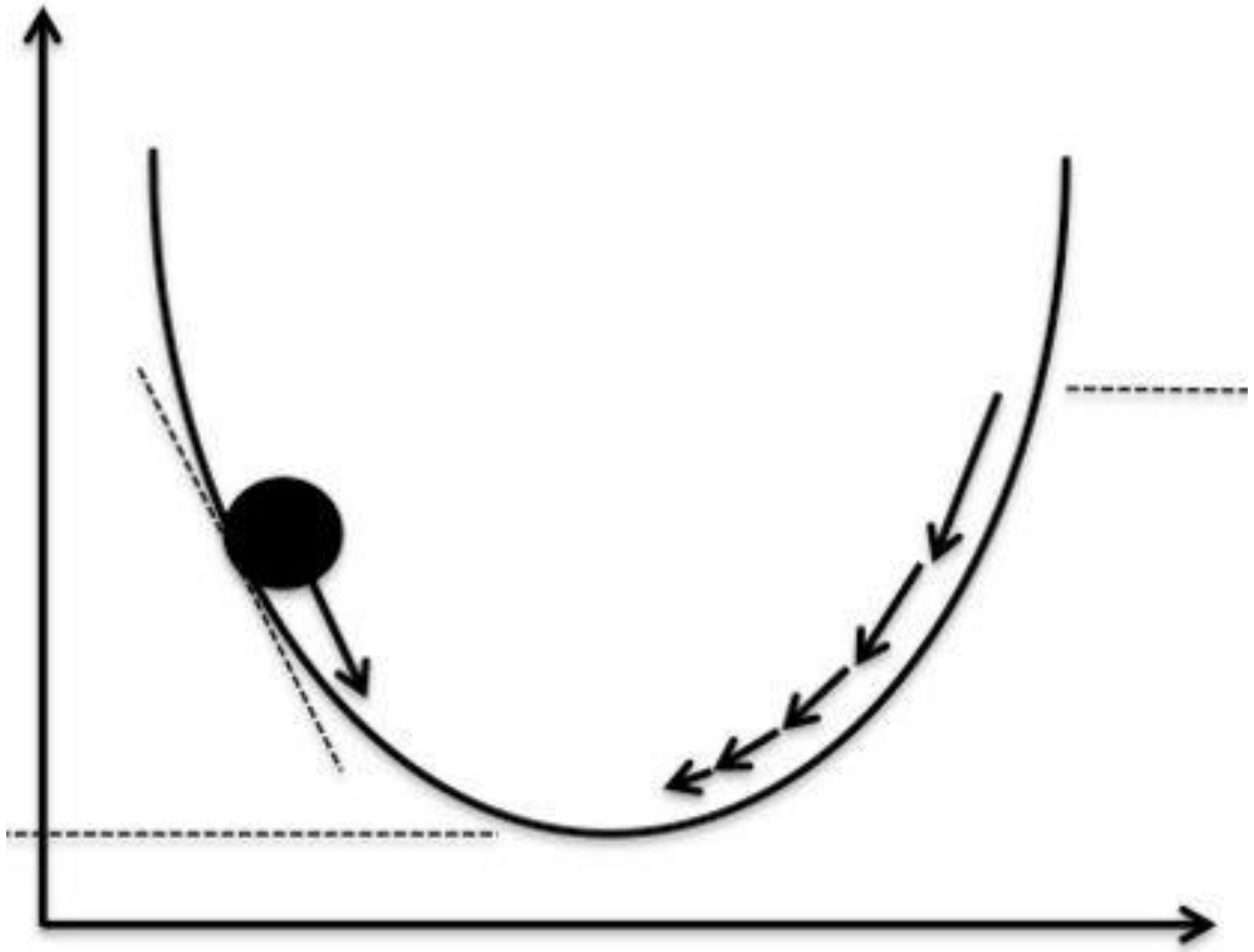




$$F'(X_0) = \lim_{\Delta X \to 0} \frac{\Delta Y}{\Delta X}$$

ИЩЕМ МИНИМУМ

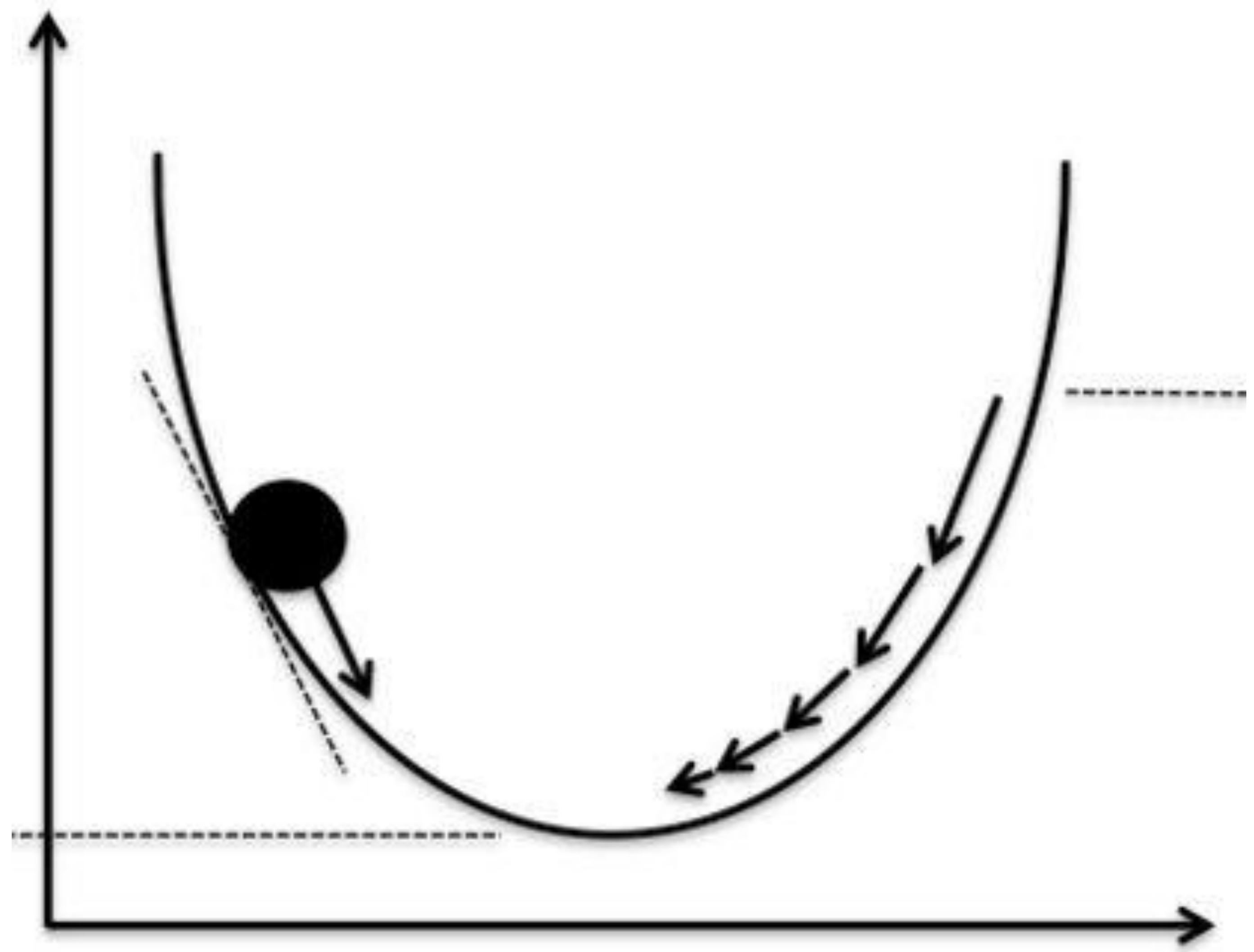
Допустим, необходимо найти минимум суммы среднеквадратичной ошибки для параметров модели



ИЩЕМ МИНИМУМ

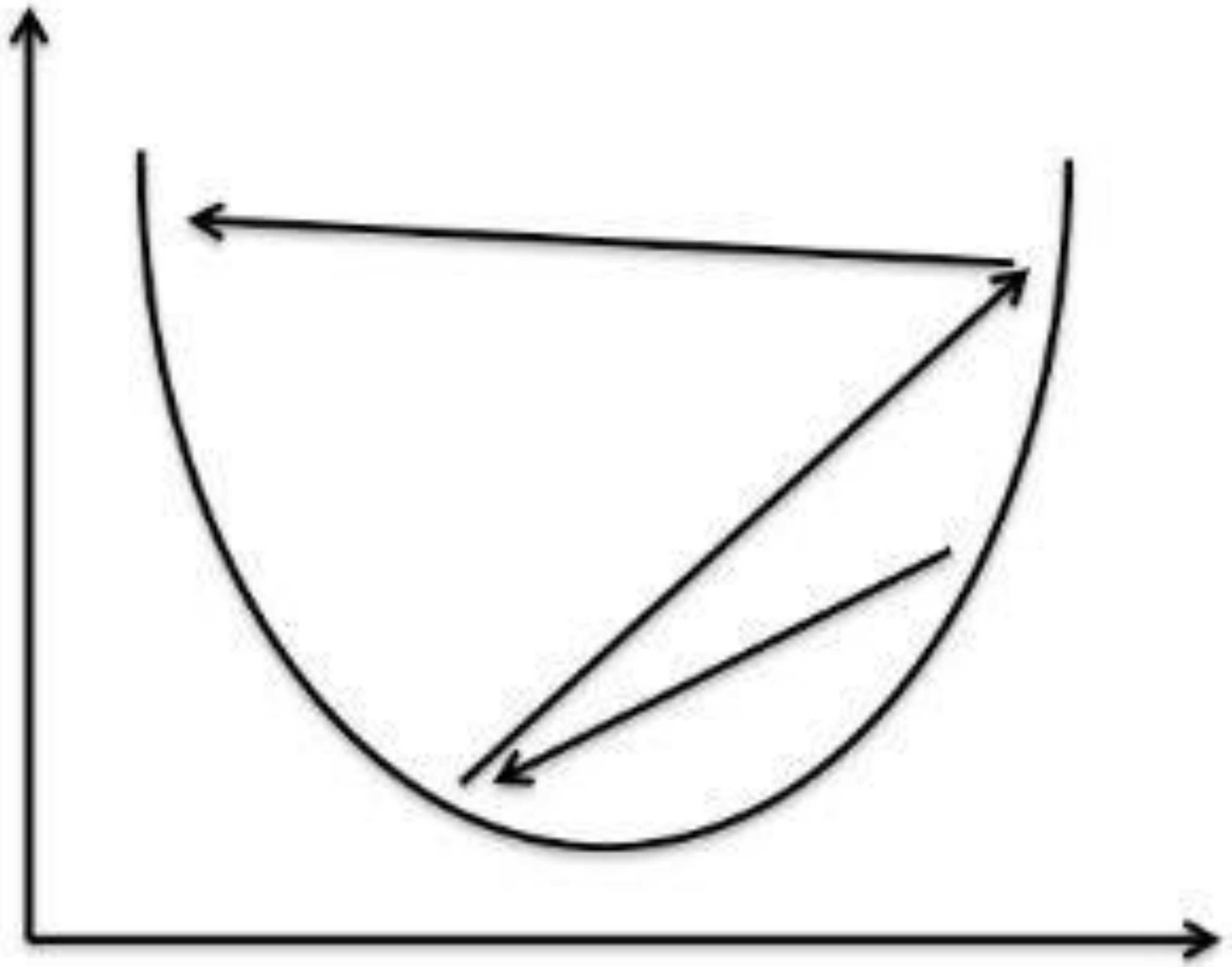
Возьмем произвольную точку на графике и будем пошагово «спускаться» к минимуму

$$x_{i+1} = x_i - \alpha \nabla F(x_i)$$



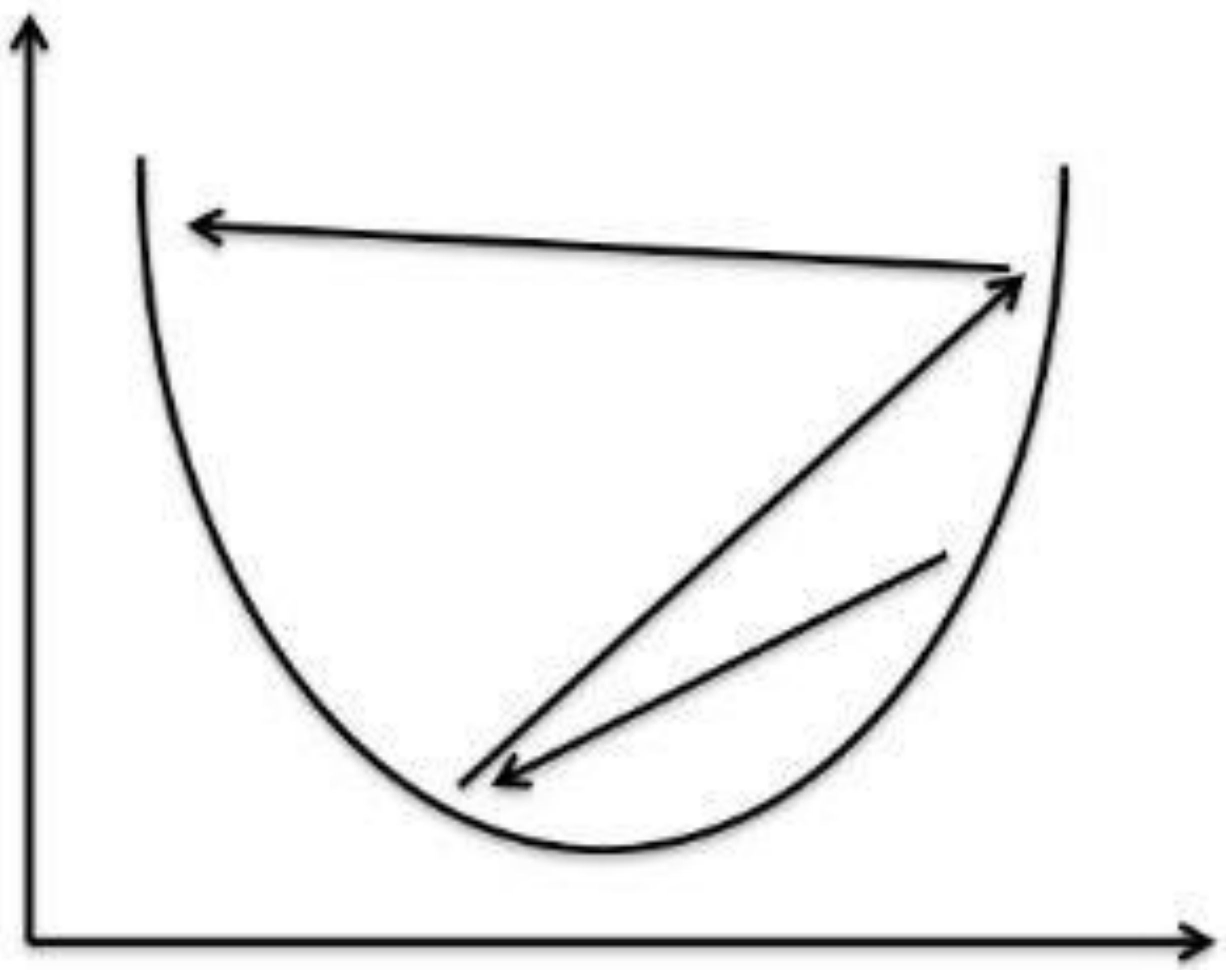
ВОЗМОЖНЫЕ ПРОБЛЕМЫ

Шаг слишком большой

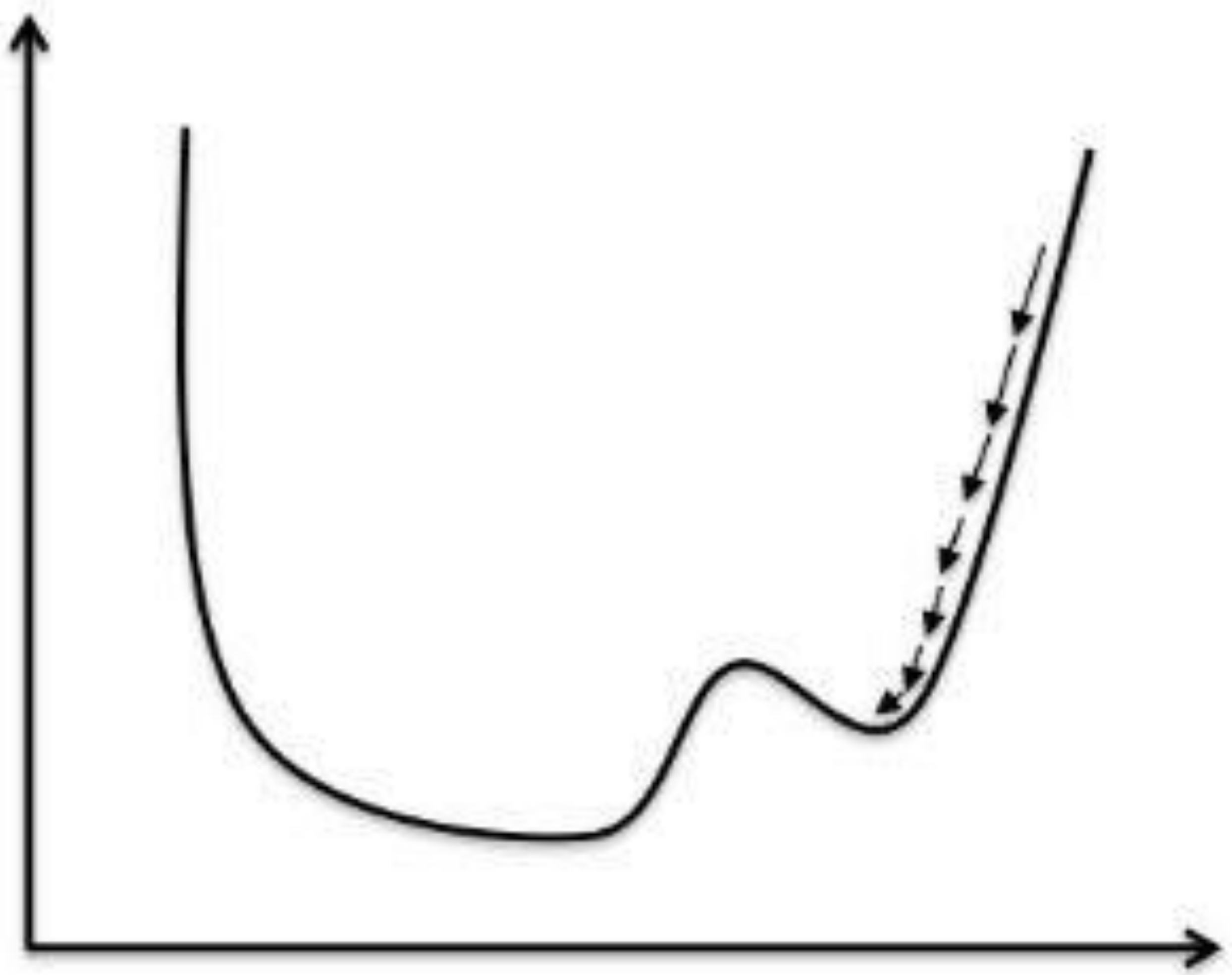


ВОЗМОЖНЫЕ ПРОБЛЕМЫ

Шаг слишком большой

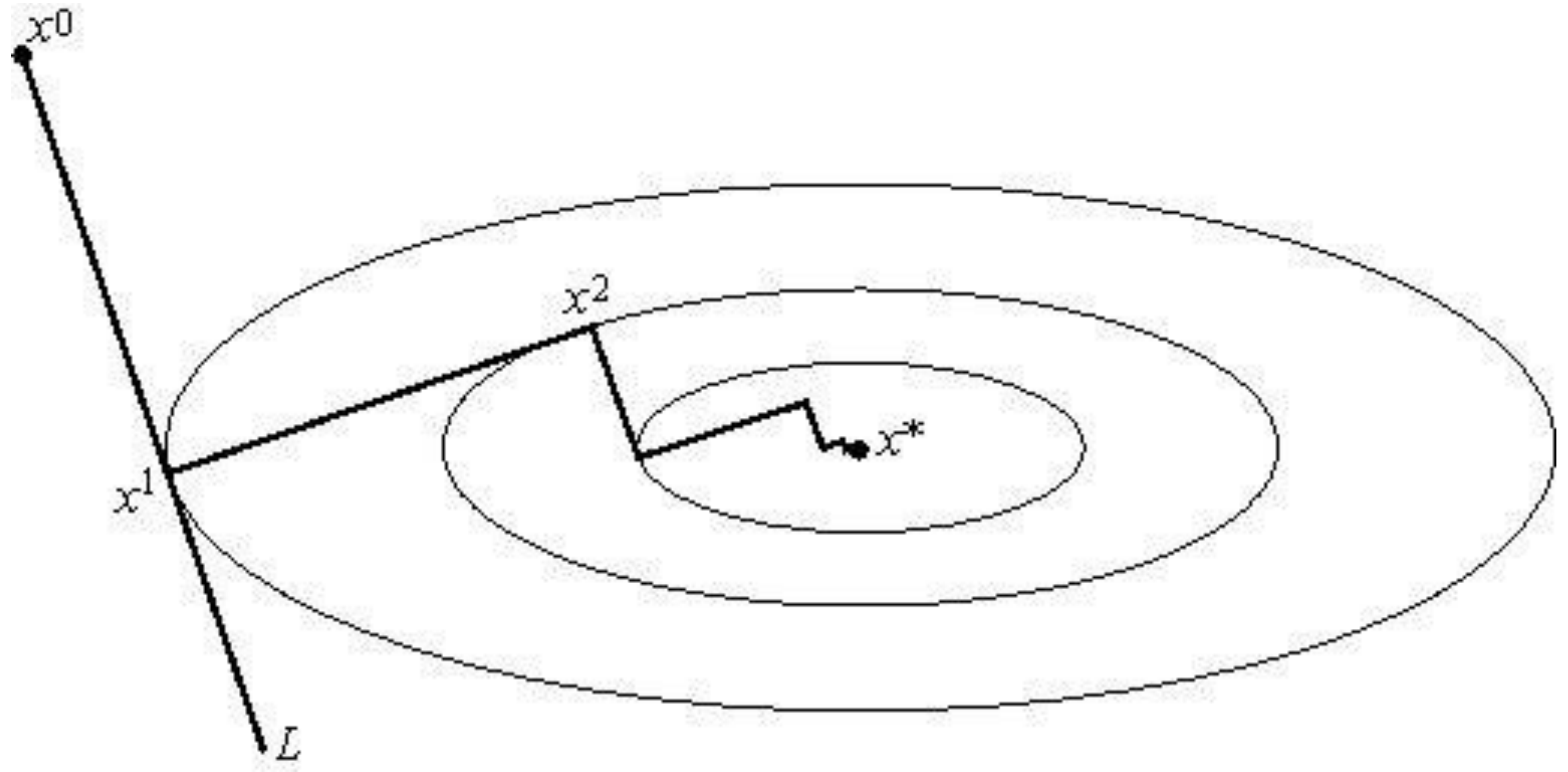


Остаемся в локальном минимуме

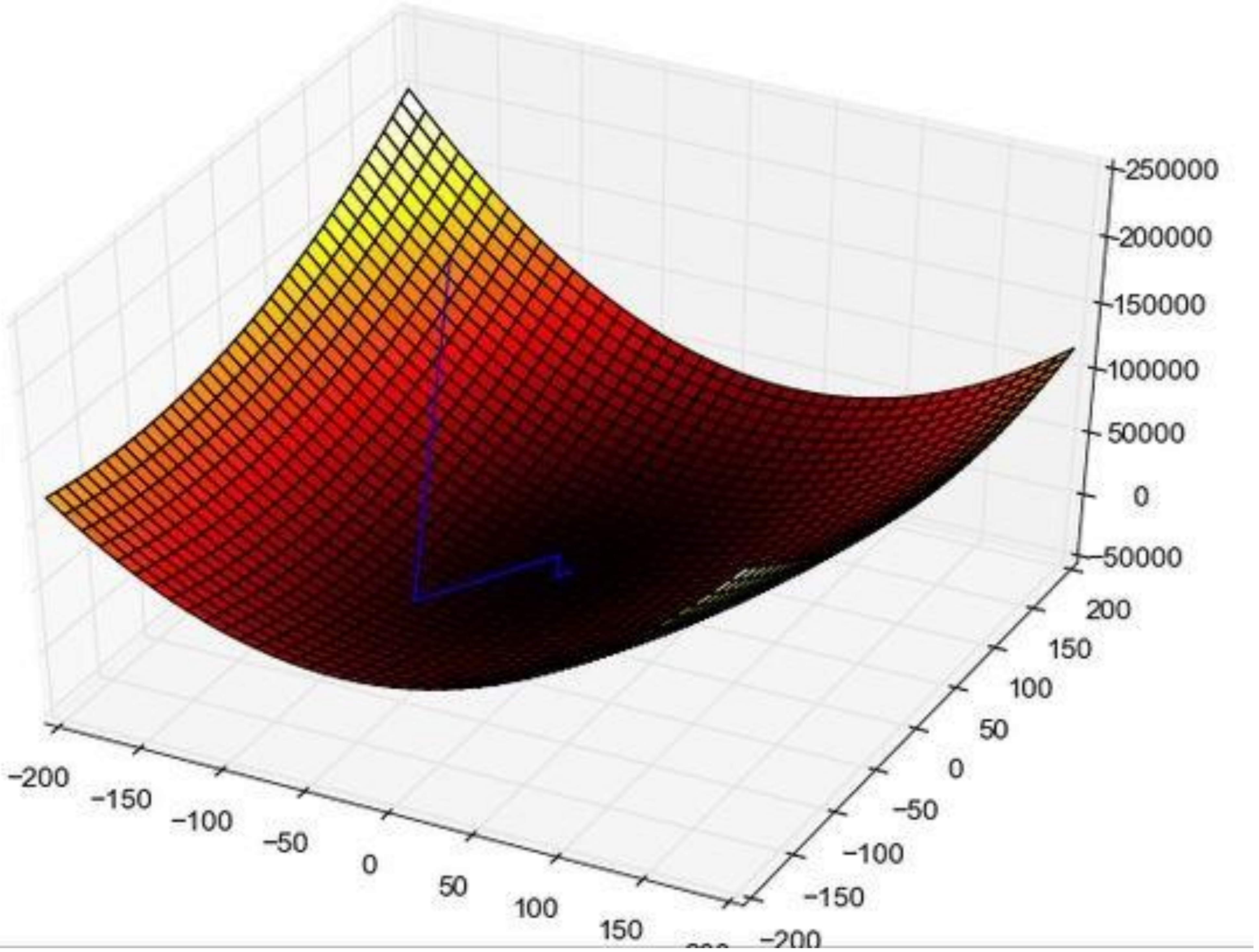


ВАРИАНТЫ ВЫБОРА λ

- Постоянной – метод может расходиться
- С дробным шагом – делим на число каждый шаг
- С наискорейшим спуском – α выбирается так, чтобы следующая итерация была точкой минимума функции f на луче



ПРИМЕР В 3D



250000
200000
150000
100000
50000
0
−50000
200
150
100
50
0
−50
−100
−150
−200
−200
−150
−100
−50
0
50
100
150

РЕАЛИЗУЕМ

GRADIENT_DESCENT.IPYNB

ЕСЛИ КЛАССОВ БОЛЬШЕ ДВУХ

ПРИМЕР

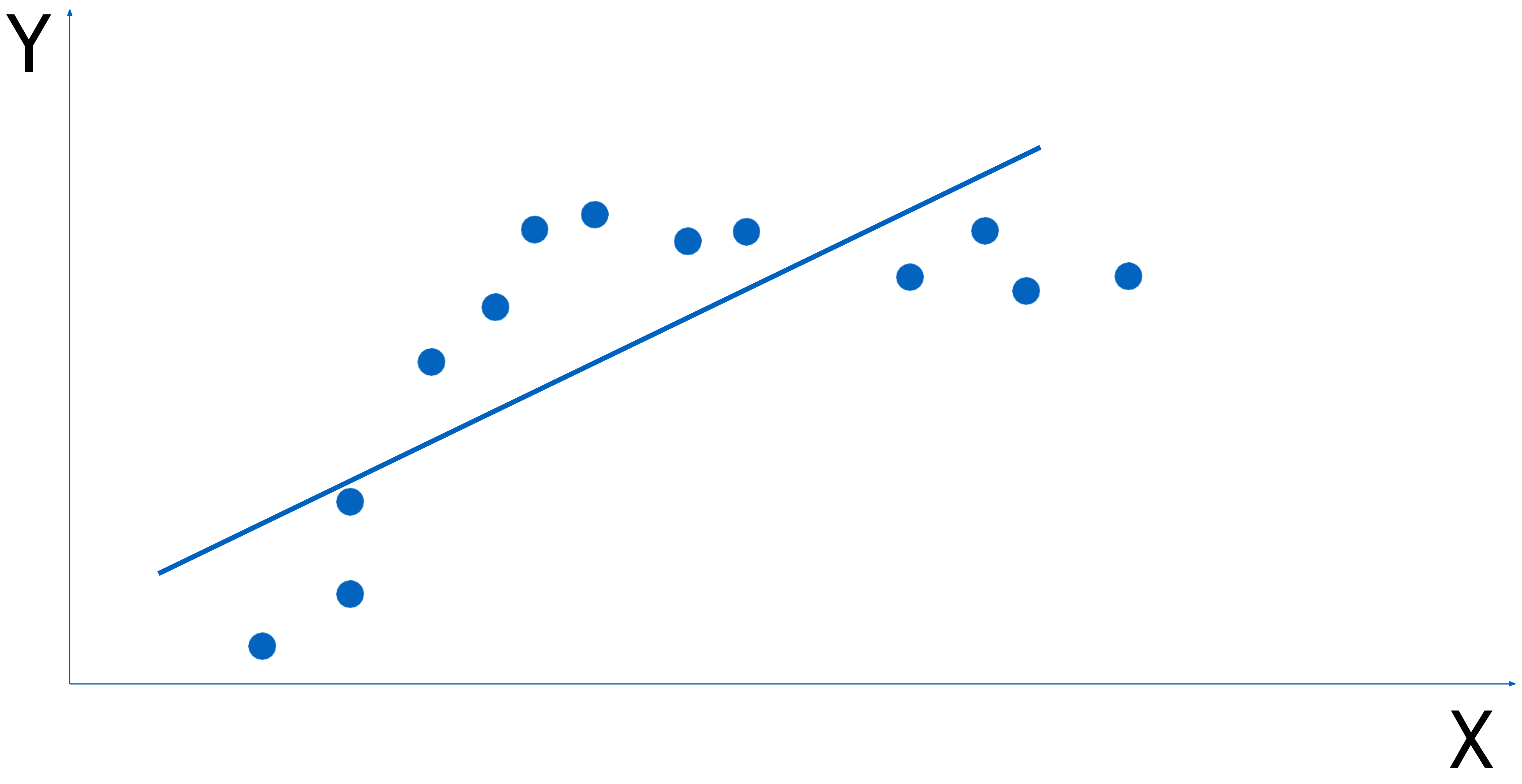
IRIS_DATASET.IPYNB

ДЛЯ КАКИХ ДАННЫХ ЭТО РАБОТАЕТ?

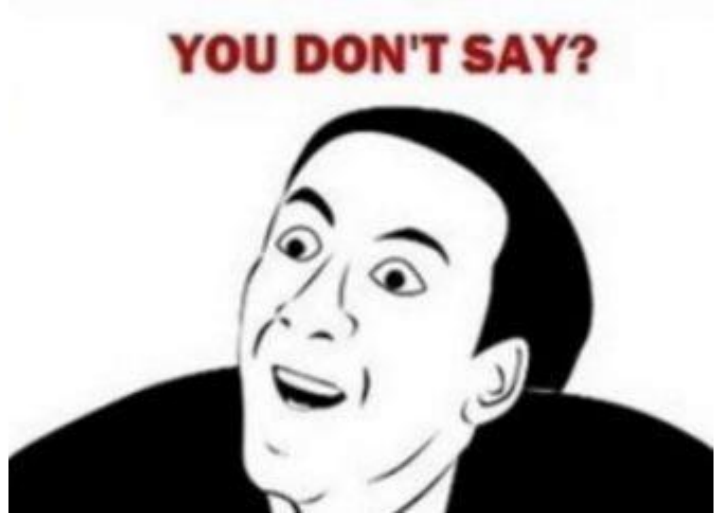
ТРЕБОВАНИЯ К ДАННЫМ

- Линейная зависимость целевой переменной
- Нормальное распределение остатков
- Постоянная изменчивость остатков

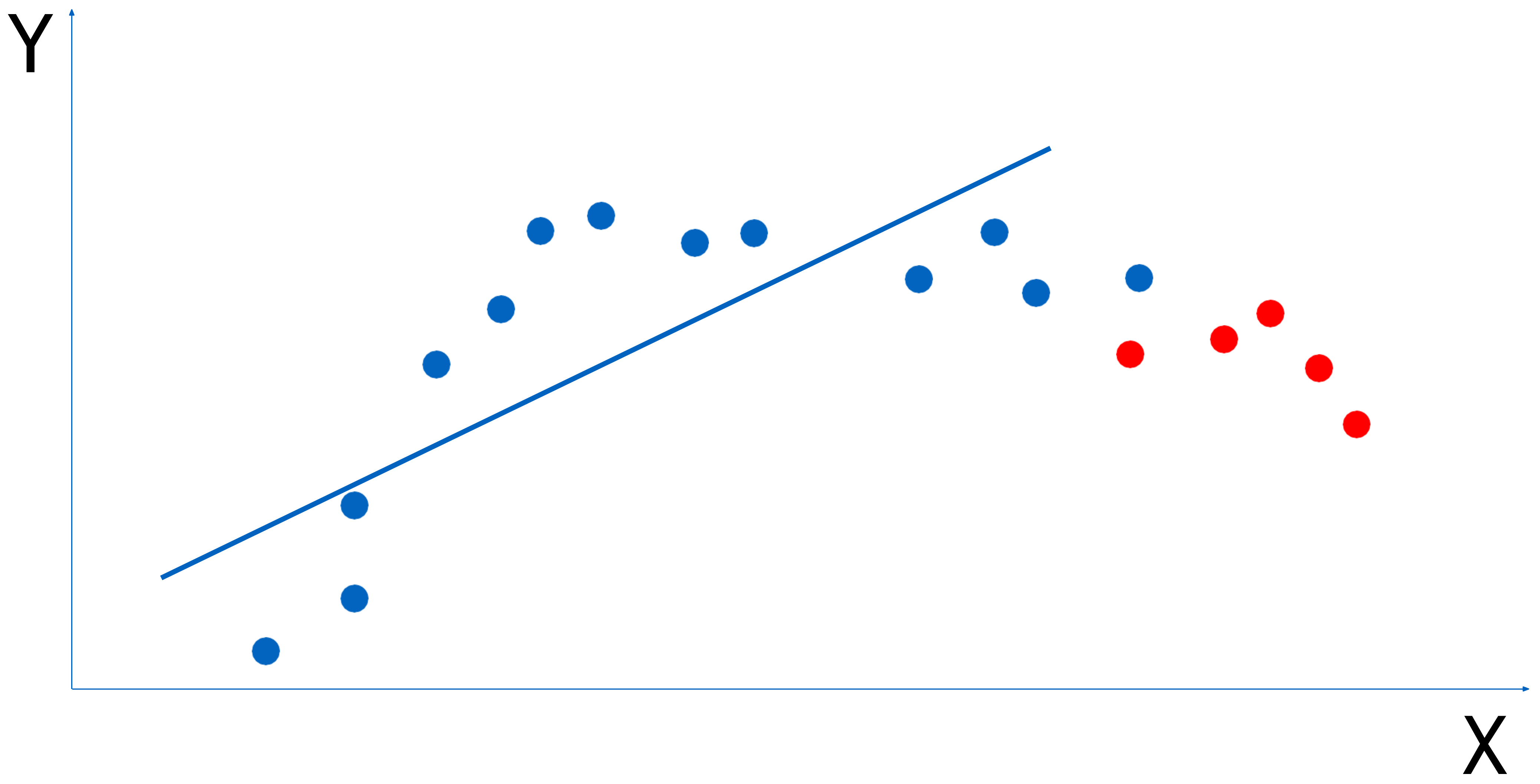
ТРЕБОВАНИЯ К ДАННЫМ



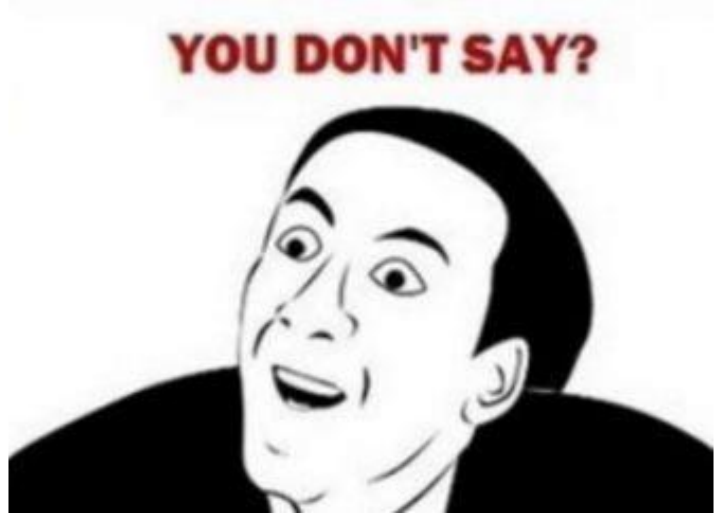
Линейная взаимосвязь X и Y



ТРЕБОВАНИЯ К ДАННЫМ

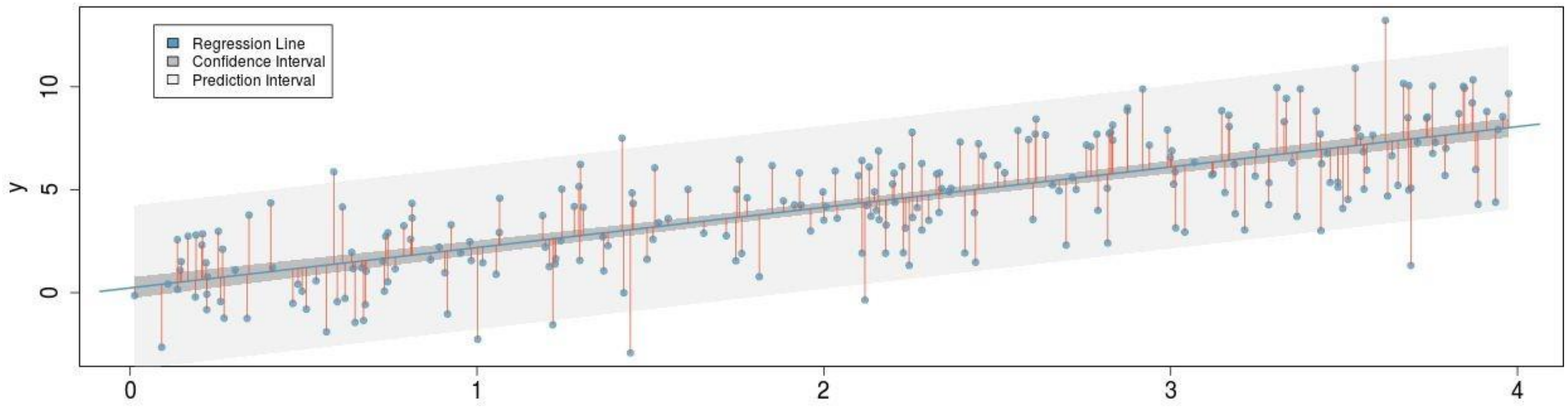


Линейная взаимосвязь X и Y



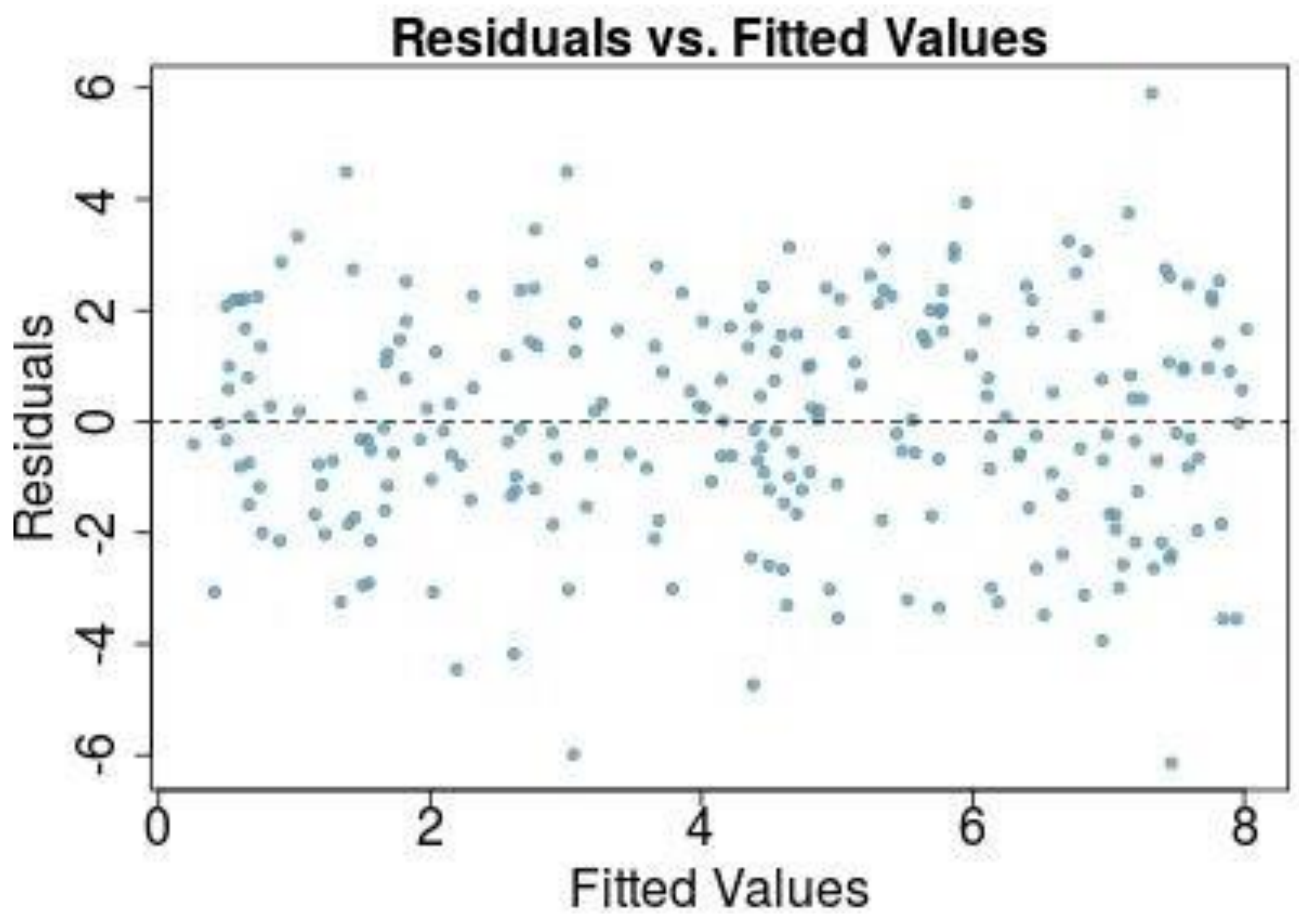
НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ ОСТАТКОВ

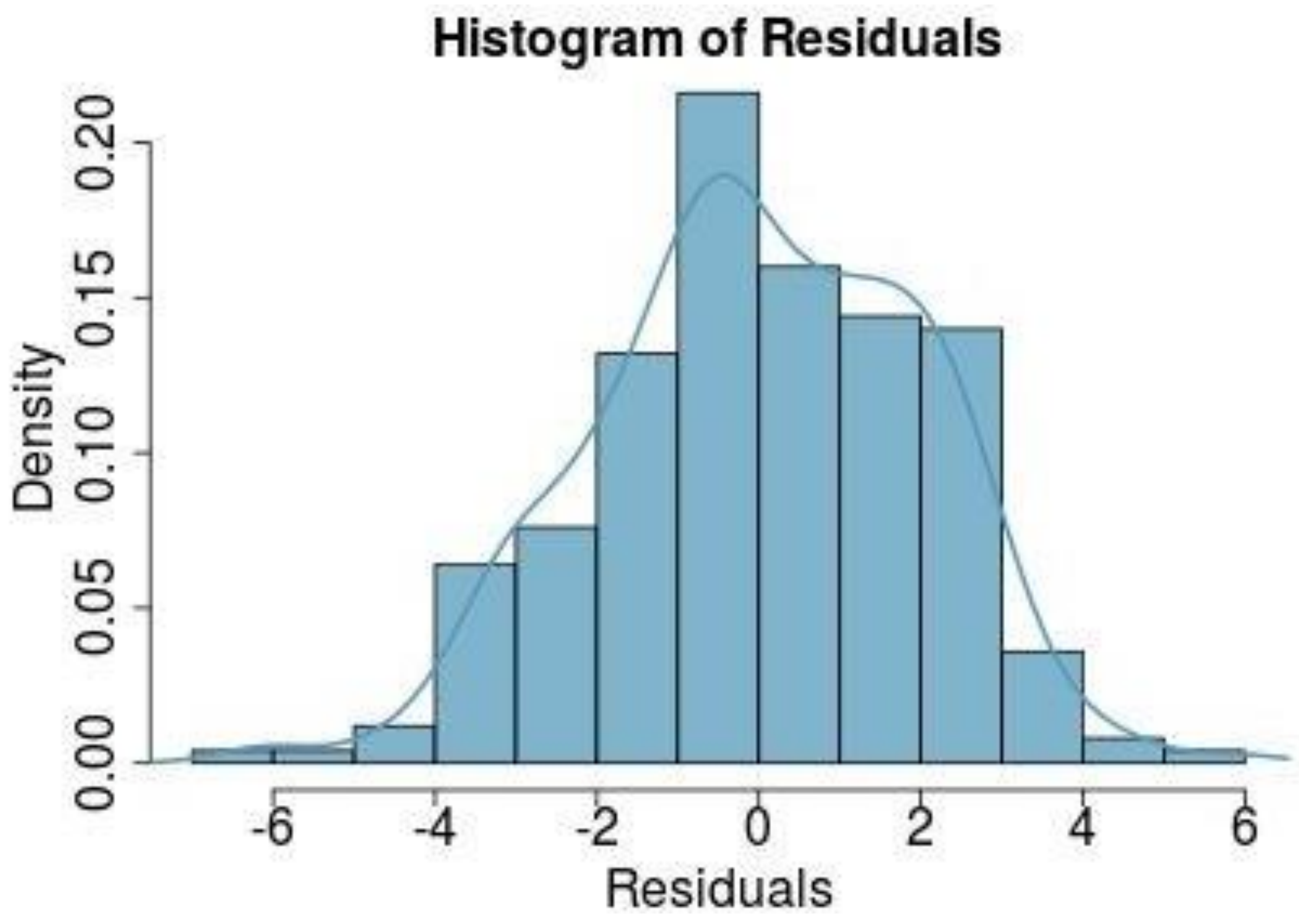
HTTPS:/ GALLERY.SHINYAPPS.IO/SLR_DIAG/

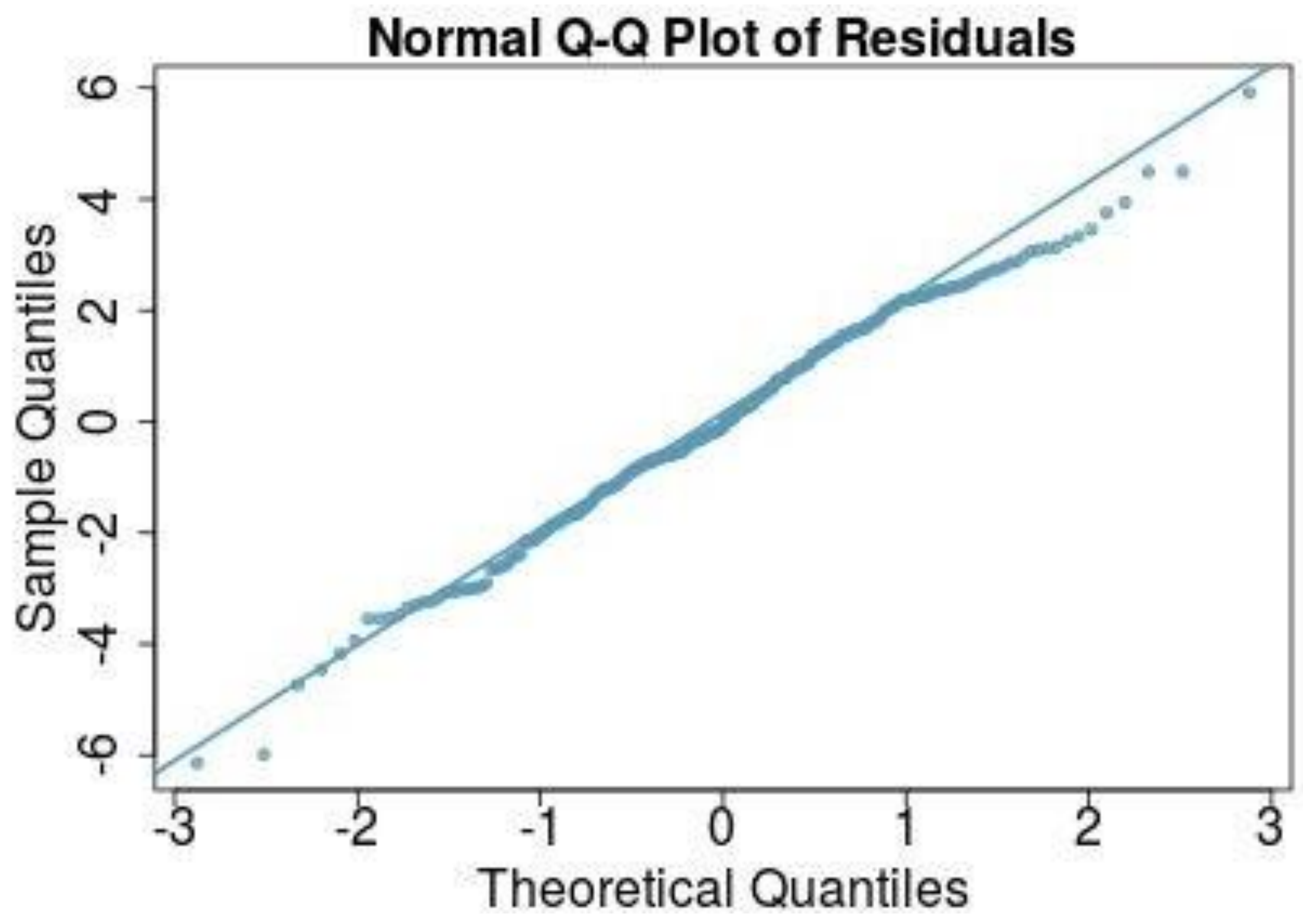


НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ ОСТАТКОВ

HTTPS:/ GALLERY.SHINYAPPS.IO/SLR_DIAG/



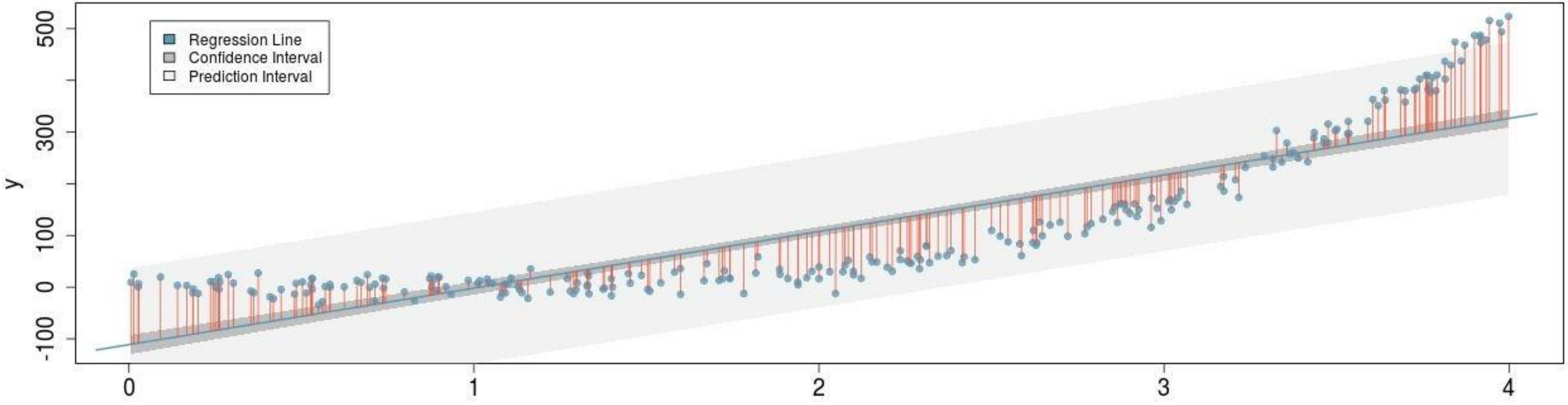




ГОМОСКЕДАСТИЧНОСТЬ

Постоянная изменчивость остатков

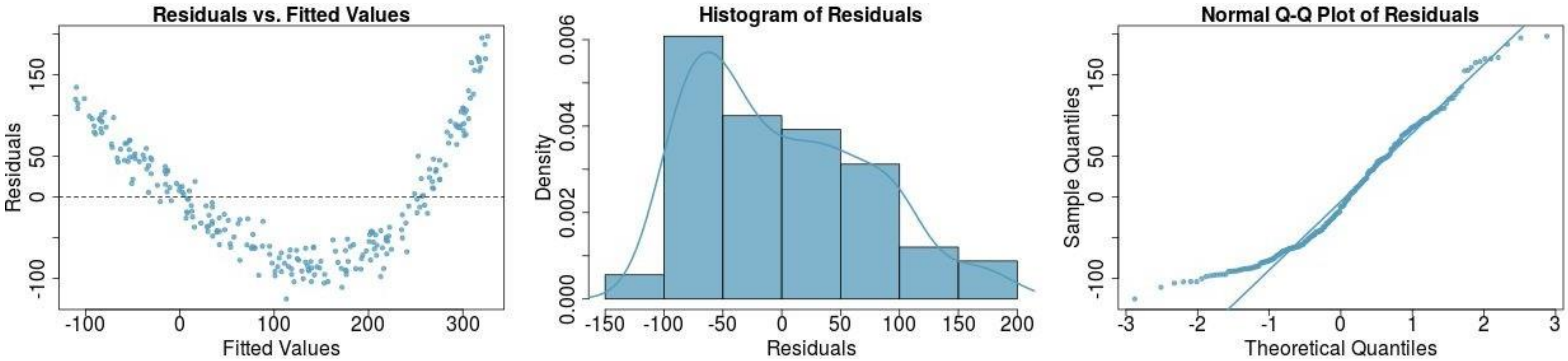
Пример гетероскедастичной последовательности



ГОМОСКЕДАСТИЧНОСТЬ

Постоянная изменчивость остатков

Пример гетероскедастичной последовательности

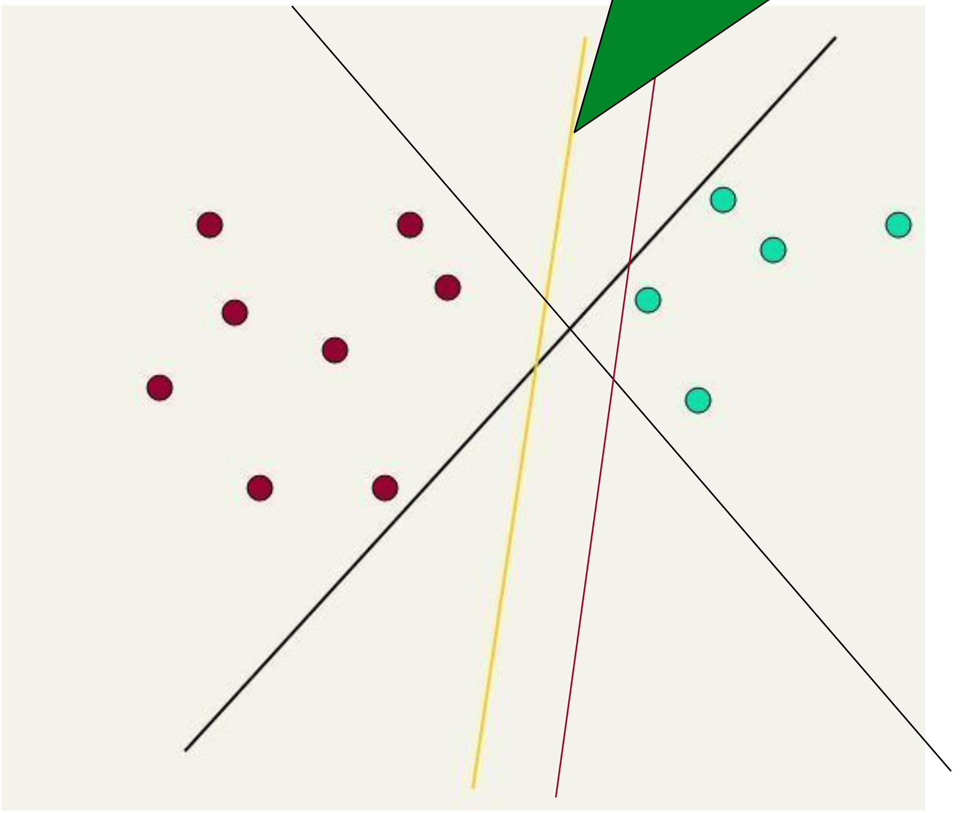


SVM

Множество гиперплоскостей

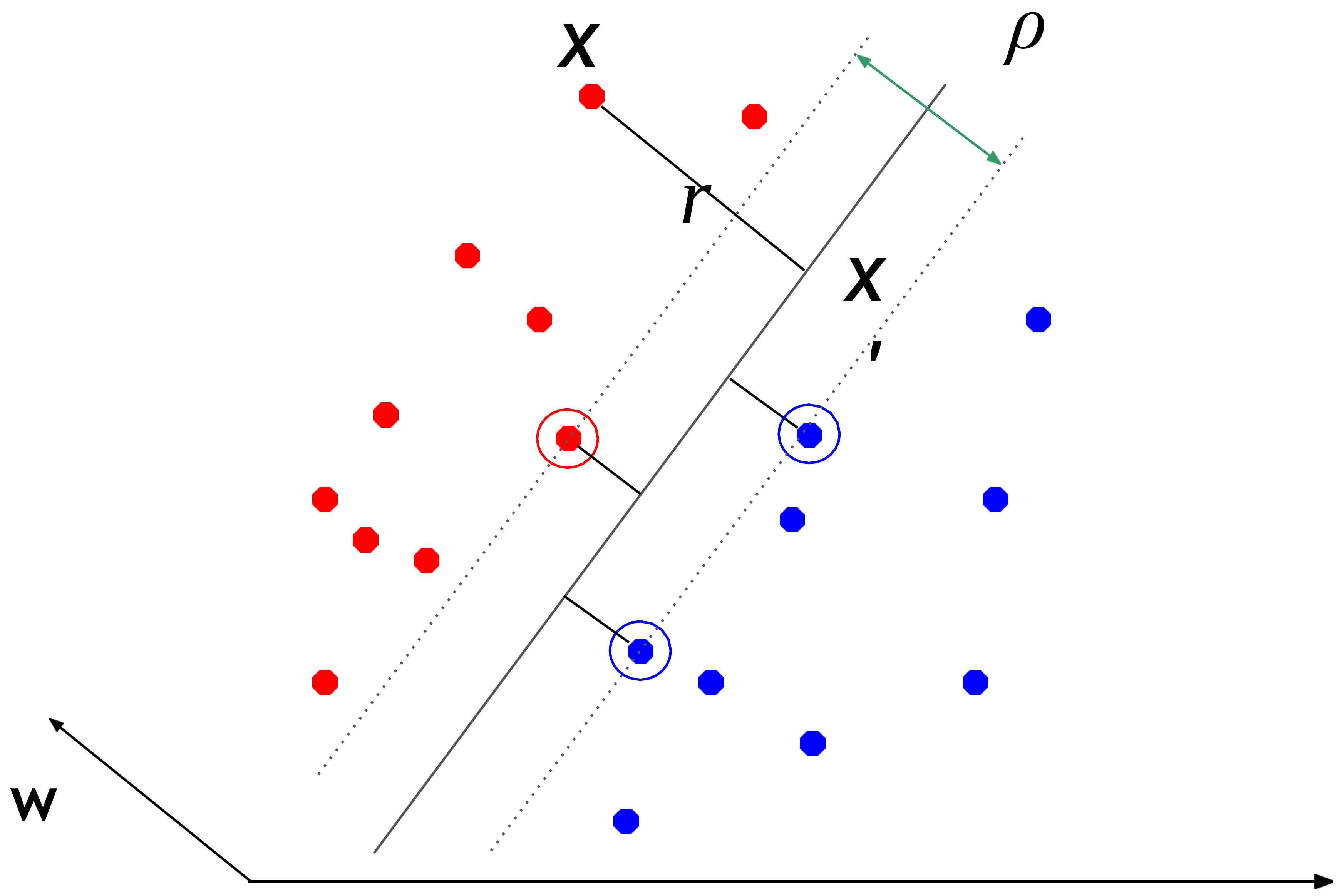
- Множество решений для *a*, *b*, *c*.
- SVM находит оптимальную разделяющую поверхность
- Максимизирует «зазор»

Граница:
$ax + by - c = 0$



Максимальный зазор

- w – нормаль к разделяющей плоскости
- x_i - sample
- y_i: - класс sample *i* (+1or -1) (важно, не 1и 0)
- Классификатор: $f(x_i) = \text{sign}(w^T x_i + b)$
- 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

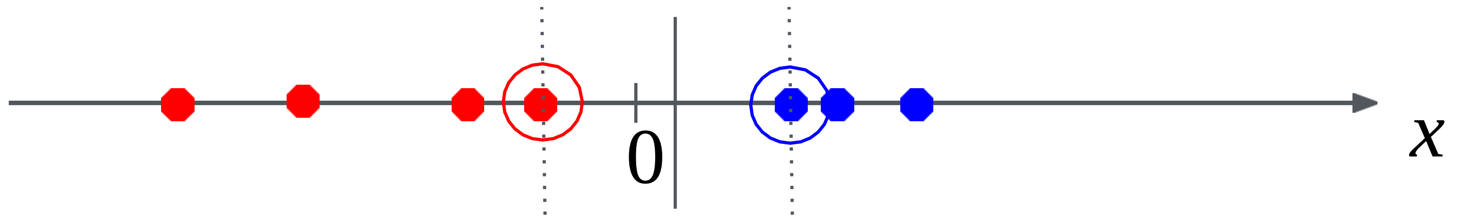
Найти $\mathbf{w}$ и b такие что

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

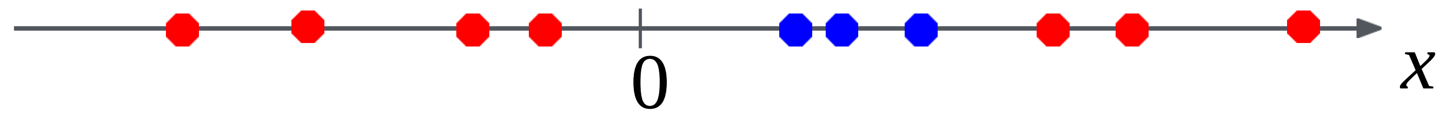
И для всех $\{(\mathbf{x_i}, y_i)\}$: $y_i(\mathbf{w^T x_i} + b) \geq 1$

Non-linear SVMs

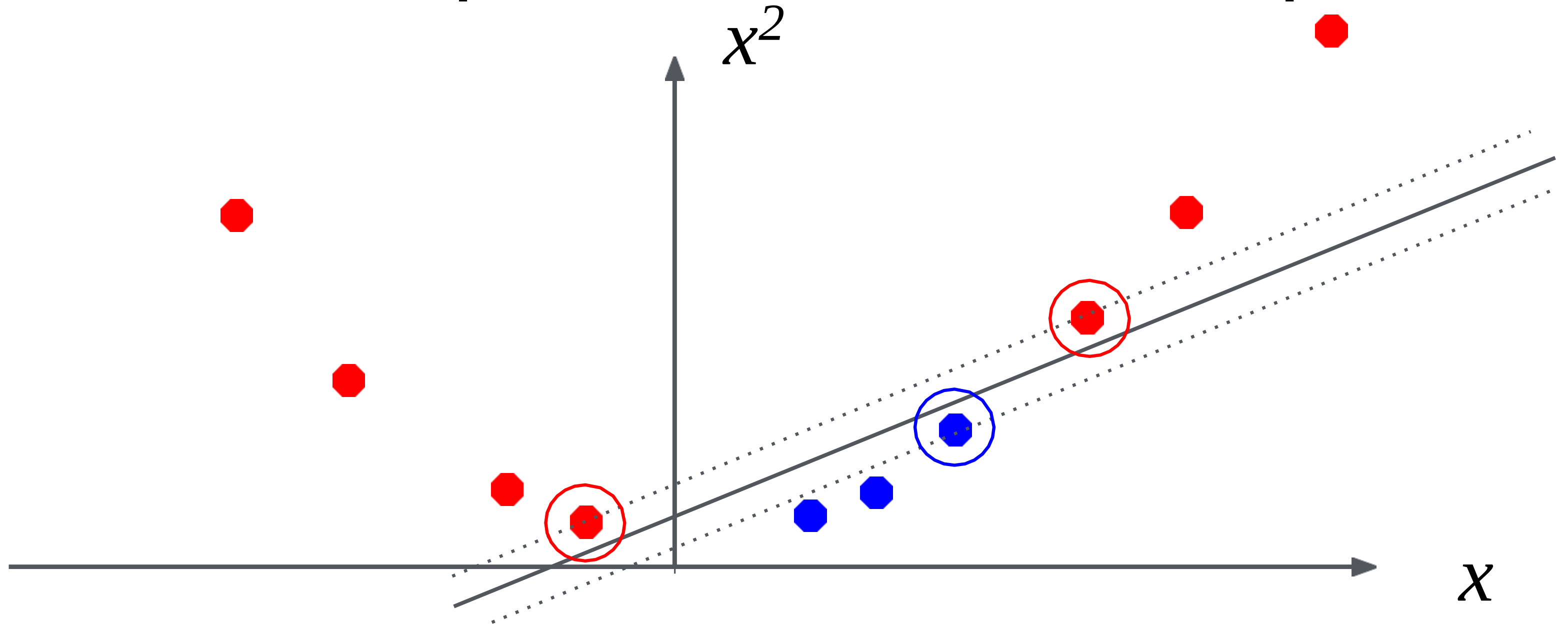
- Линейной разделимые датасеты хорошо классифицируются



- 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 Φ: $x \rightarrow \varphi(x)$, тогда скалярное произведение становится:
- $K(xi,xj)= \varphi(xi)\ T\varphi(xj)$
- Функция ядра – это функция соотв. Скалярному произведению в пр-ве более высокой размерности

Kernels

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

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2}$$

ЧТО МЫ СЕГОДНЯУЗНАЛИ

ЧТО МЫ СЕГОДНЯ УЗНАЛИ

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

ПОЛЕЗНЫЕ МАТЕРИАЛЫ

1. Статья о линейных моделях в ODS
   https://habrahabr.ru/company/ods/blog/323890/
2. Курс «Основы статистики» на Stepik.org
   https://:/ stepik.org/course/Основы-статистики-76

Спасибо за внимание!

Артур Сапрыкин