

## План работы

- Простая линейная регрессия
- Графическая библиотека Vokeh
- Метод стохастического градиентного спуска

## Простая линейная регрессия

- На основании выборки  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$
- Прямая линия

$$y = \beta_0 + \beta_1 x$$

- Прямая выбрана так, чтобы минимизировать расхождение между выборочными данными и точками, лежащими на этой прямой
- Как построить такую прямую (найти  $\beta_0$  и  $\beta_1$ )?

Расхождение между выборочными данными и точками на прямой. Метод наименьших квадратов (ordinary least squares, OLS).

$$\sum_{i=1}^n ((\beta_0 + \beta_1 x_i) - y_i)^2 \rightarrow \min_{\beta_0, \beta_1}$$

- Подберем вручную?
- Продифференцируем и найдем критические точки?
- Используем численный метод нахождения минимума?

## Набор данных

x	y
10.0	8.04
8.0	6.95
13.0	7.58
9.0	8.81
11.0	8.33
14.0	9.96
6.0	7.24
4.0	4.26
12.0	10.84
7.0	4.82
5.0	5.68

## Визуализация в Bokeh

```
from bokeh.plotting import figure, output_file, show

x=[10.0,8.0,13.0,9.0,11.0,14.0,6.0,4.0,12.0,7.0,5.0]
y=[8.04,6.95,7.58,8.81,8.33,9.96,7.24,4.26,10.84,4.82,5.68]

output_file("linreg.html")
p = figure(plot_width=400, plot_height=400)
p.line([1, 10], [1, 15], line_width=2)
p.circle(x, y, fill_color="white", size=8)
show(p)
```

## Градиентный спуск

- Задача оптимизации:

$$Q(w) = \frac{1}{n} \sum_{i=1}^n Q_i(w) \rightarrow \min_w$$

- Стандартный метод градиентного спуска для минимизации  $Q(w)$

$$w := w - \eta \nabla Q(w)$$

В нашем случае

$$w := w - \eta \sum_{i=1}^n \nabla Q_i(w) / n$$

## Стохастический градиентный спуск

1. Выбрать начальный вектор параметров  $w$  и интенсивность обучения  $\eta$
2. Повторять до сходимости:
  - (a) Случайно перемешать выборочные данные
  - (b) Для  $i = 1, 2, \dots, n$ , выполнить

$$w := w - \eta \nabla Q_i(w)$$

## Задание на практику и д/з

1. Запрограммируйте метод градиентного спуска для поиска параметров  $\beta_0$  и  $\beta_1$  уравнения линейной регрессии
2. Запрограммируйте метод стохастического градиентного спуска для поиска параметров  $\beta_0$  и  $\beta_1$  уравнения линейной регрессии
3. Примените полученные методы к набору данных со слайда 4 и измените код на слайде 5 так, чтобы он визуализировал получившееся уравнение



линейной регрессии вместе с выборочными данными.

Сдайте получившуюся программу на листочке.