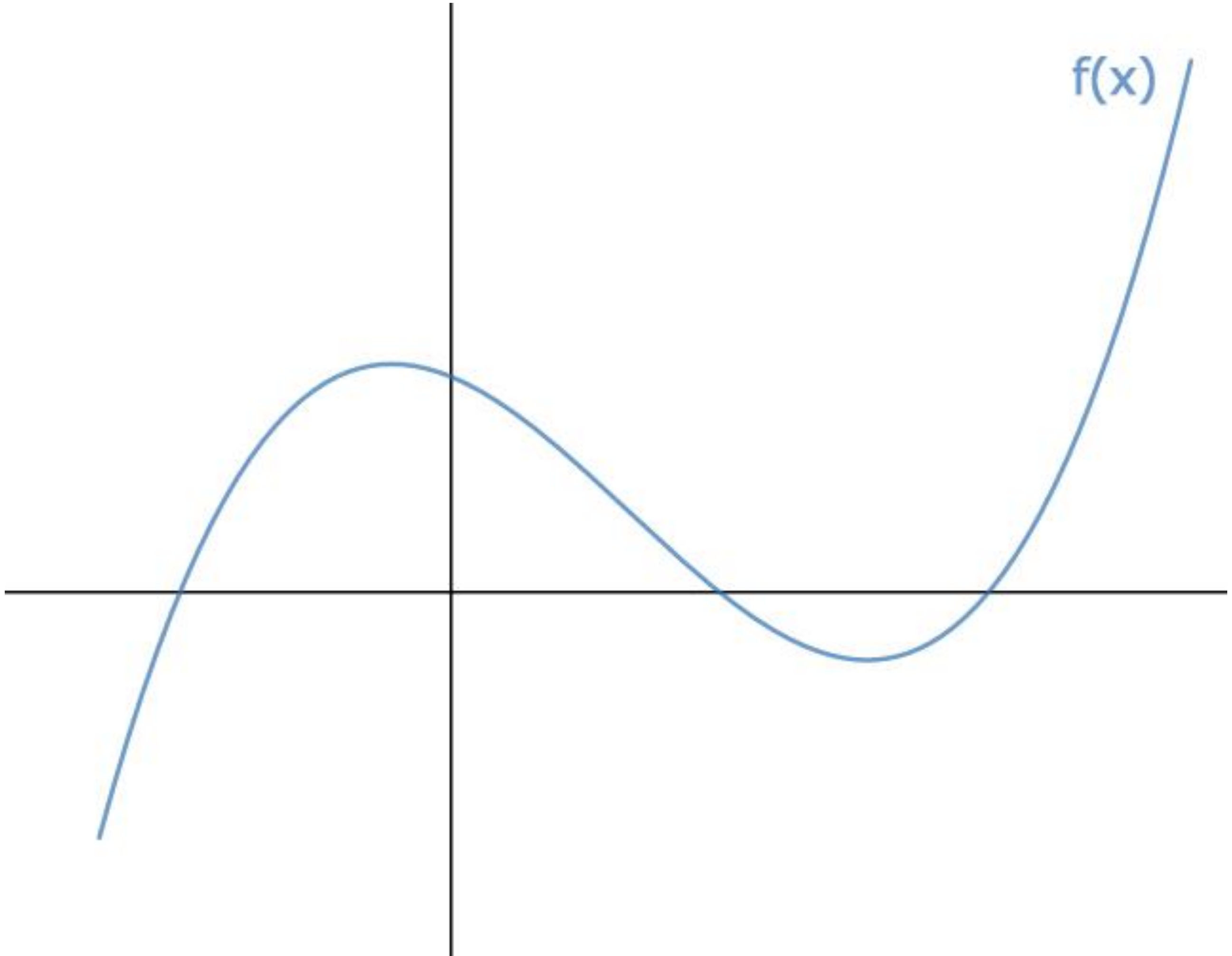
Машинное обучение

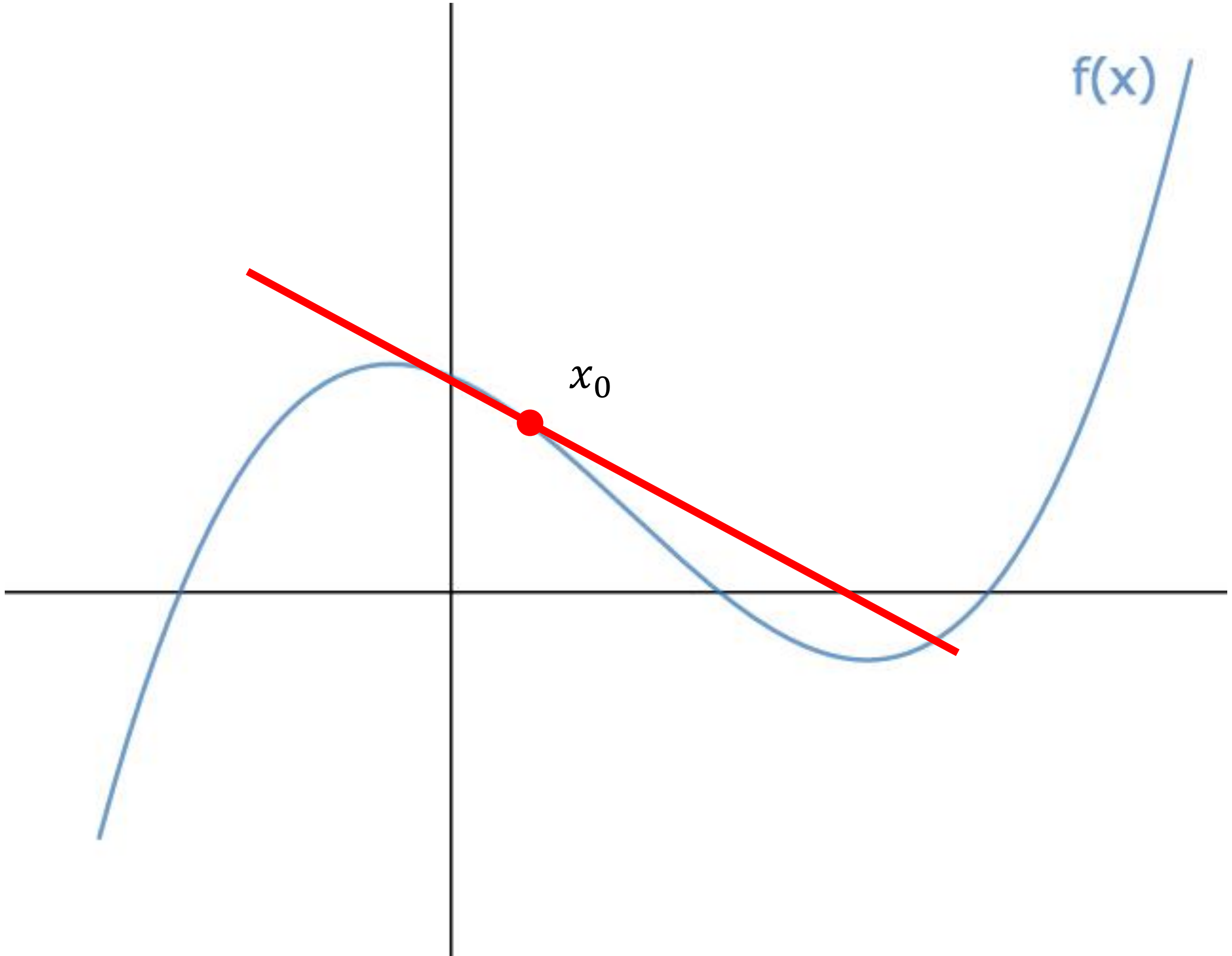
Производная

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f^{'}(x_0)$$



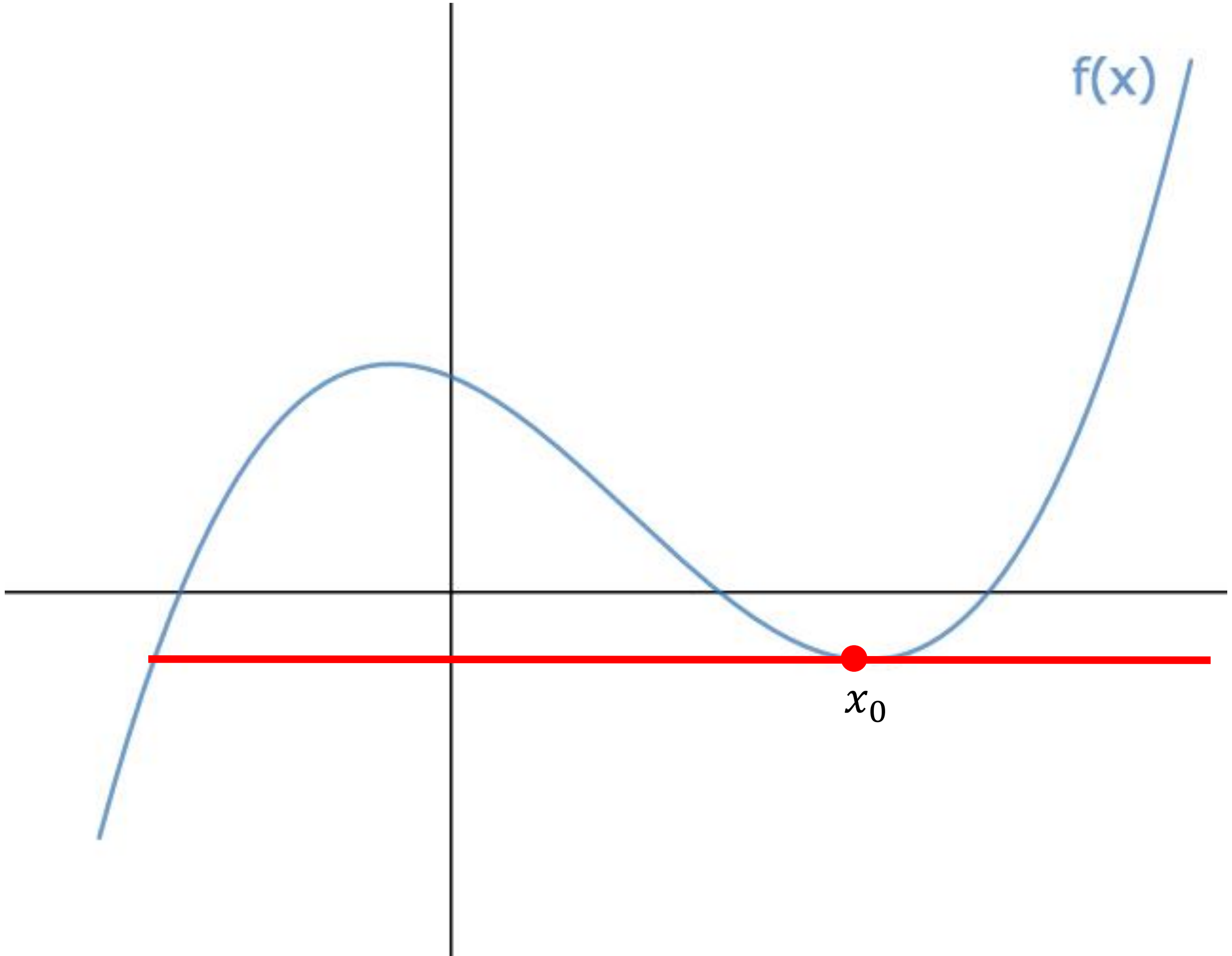
Производная

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$$



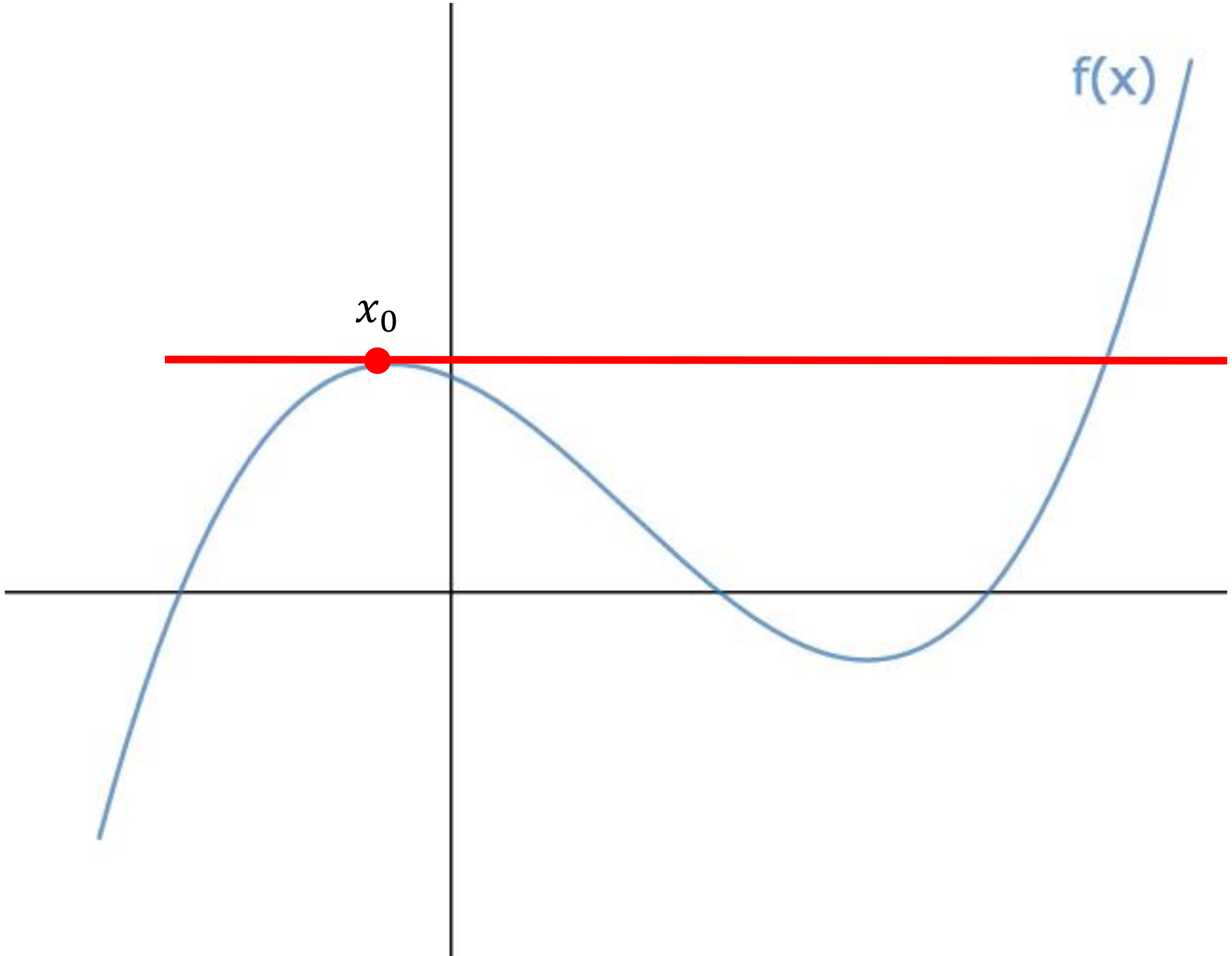
Условие экстремума

Если точка x_0 экстремум и в ней существует производная, то $f^{'}(x_0) = 0$



Условие экстремума

Если точка x_0 экстремум и в ней существует производная, то $f^{'}(x_0) = 0$



Условие экстремума

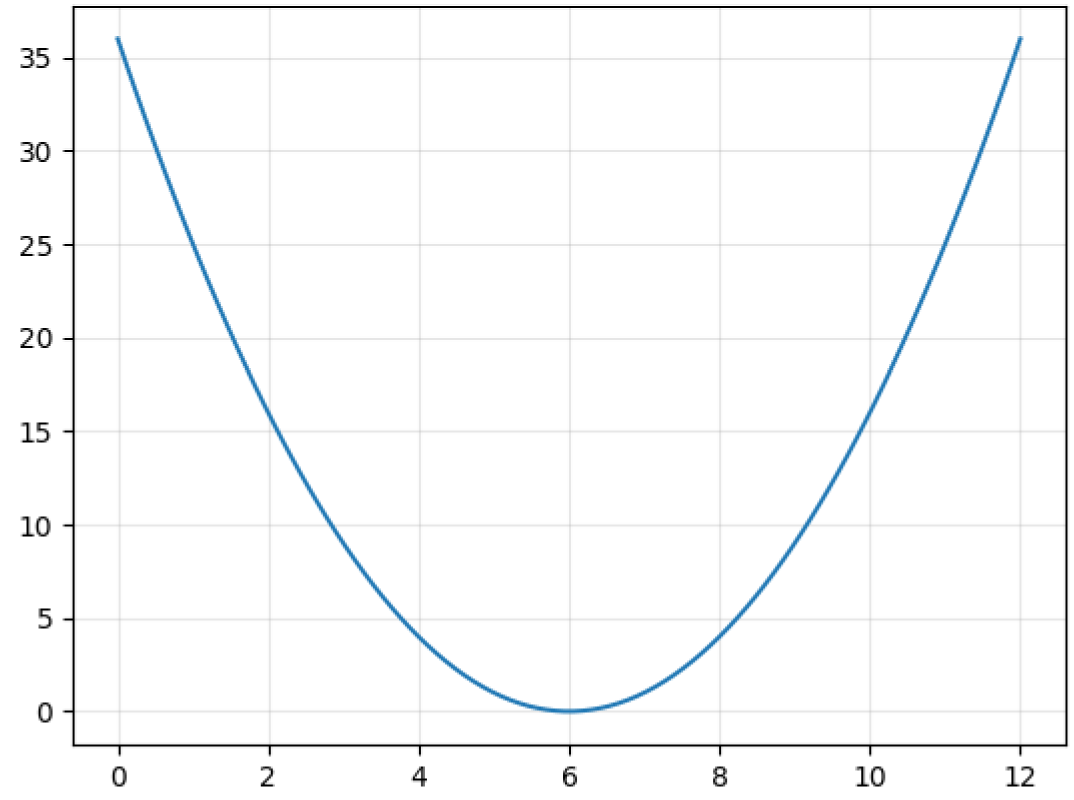
1- Вычисляем производную функции

2- Приравниваем производную к нулю

3- Находим корни этого уравнения

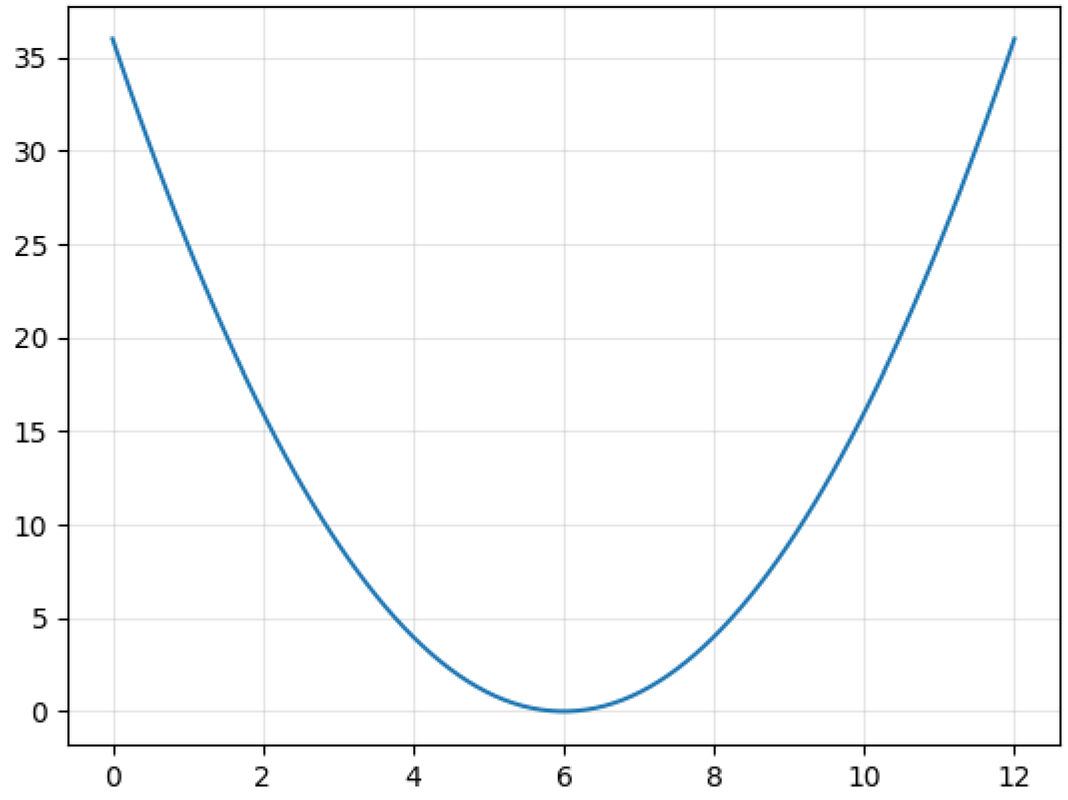
Производная

$$f(x) = (x - 6)^2$$



Производная

$$f(x) = (x-6)^2 \qquad \frac{\partial f}{\partial x} = 2(x-6)$$



Производная

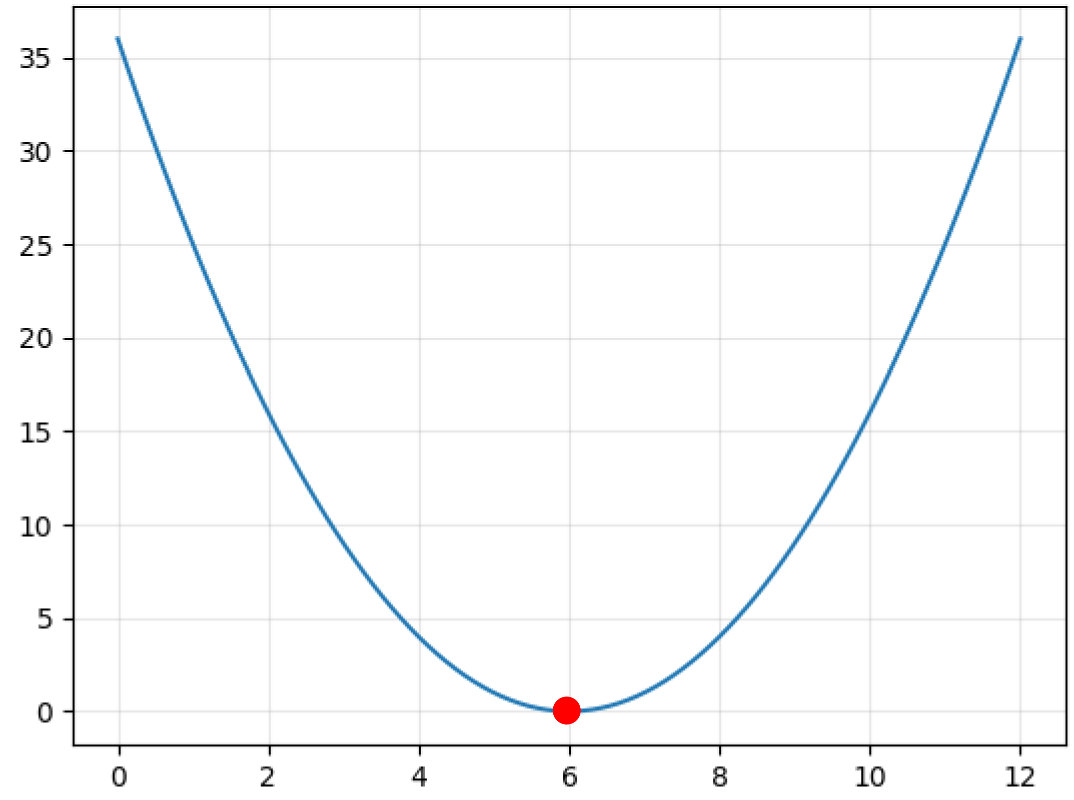
$$f(x) = (x-6)^2$$

$$\frac{\partial f}{\partial x} = 0$$

$$\frac{\partial f}{\partial x} = 2(x-6)$$

$$2(x-6) = 0$$

$$x = 6$$



Градиент

Вектор частных производных

$$Q(a, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(\boldsymbol{\theta_1} \boldsymbol{x_{i1}} + \ldots + \boldsymbol{\theta_d}\, \boldsymbol{x_{id}})^2$$

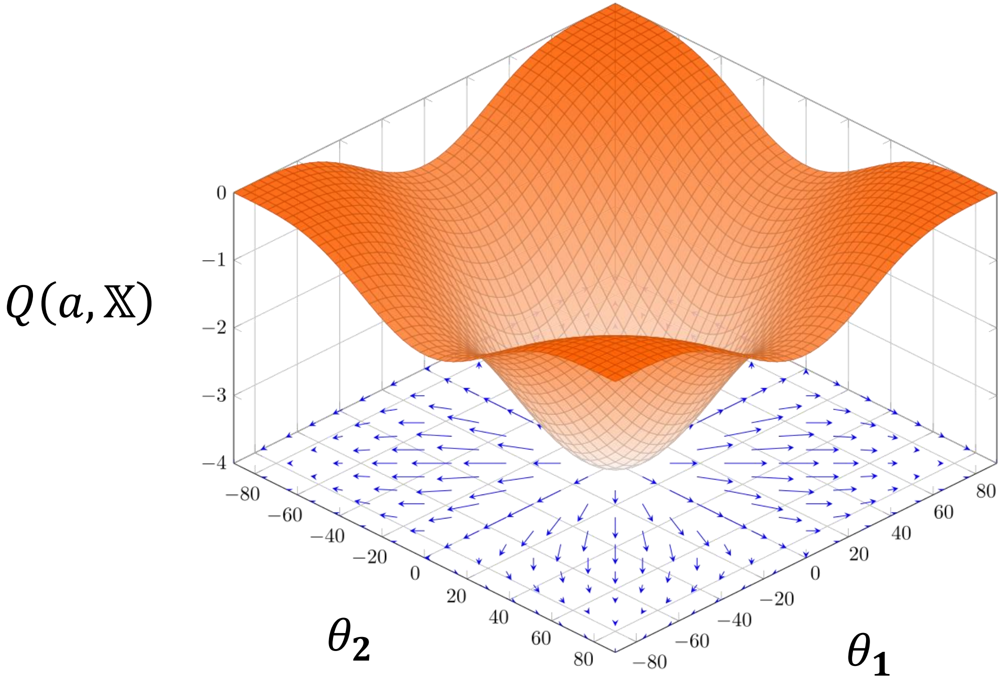
$$\boldsymbol{\nabla f}(\boldsymbol{X}) = \left(\frac{\boldsymbol{\partial f}}{\boldsymbol{\partial x_1}}, \ldots, \frac{\boldsymbol{\partial f}}{\boldsymbol{\partial x_d}}\right)$$

Важное свойство

Вектор частных производных

$$\nabla f(X) = \left(\frac{\partial f}{\partial \theta_1}, \dots, \frac{\partial f}{\partial \theta_d}\right)$$

В какую сторону функция быстрее всего растёт?



Важное свойство

Если точка $\boldsymbol{x_0}$ - экстремум, в ней существует производная, то

$$\nabla \boldsymbol{f}(\boldsymbol{x_0}) = \boldsymbol{0}$$

Если функция выпуклая, то экстемум один

MSE для линейной регрессии – выпуклая! (при некоторых условиях)

Метод наименьших квадратов

Модель линейной регрессии

Модель

$$\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{\theta_0} \boldsymbol{x_0} + \boldsymbol{\theta_1} \boldsymbol{x_1} + \ldots + \boldsymbol{\theta_d}\, \boldsymbol{x_d}$$

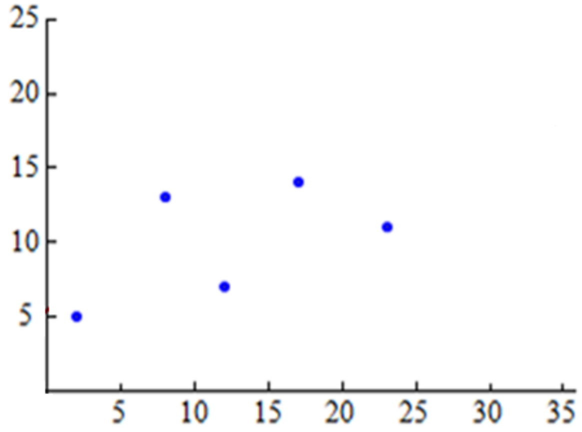
Функционал ошибки

$$\mathrm{Q}(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell} \sum_{i=1}^{\ell} (a(x_i) - y_i)^2$$

Среднеквадратичная ошибка

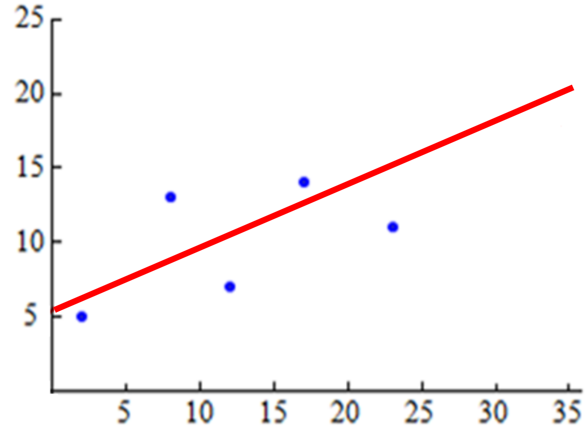
Цель: $\mathrm{Q}(\mathrm{a}, \mathbb{X}) \to \underset{\boldsymbol{\theta}}{min}$

Метод наименьших квадратов (Ordinary least squares)

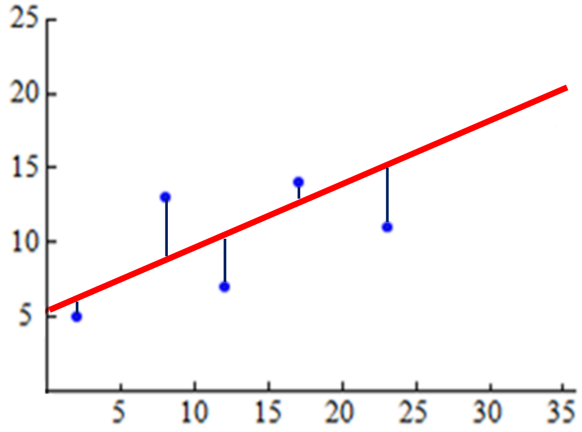


Метод наименьших квадратов (Ordinary least squares)

$$\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{\theta_0} + \boldsymbol{\theta_1 x_1}$$

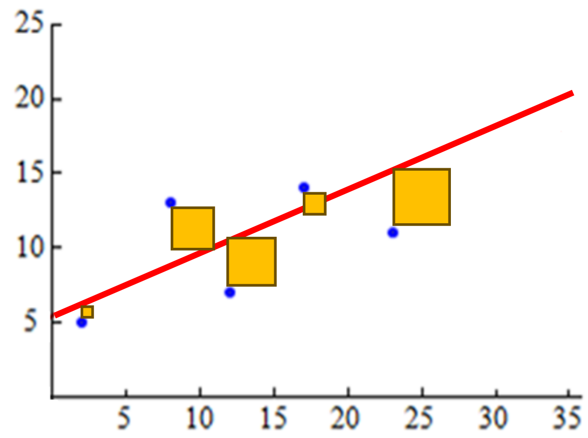


Метод наименьших квадратов (Ordinary least squares)



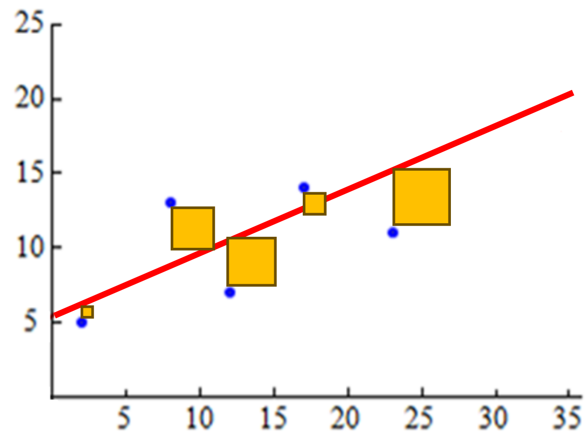
$$Q(a, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 \to \min_{\boldsymbol{\theta}}$$

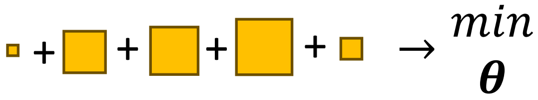
Метод наименьших квадратов (Ordinary least squares)



$$Q(a, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 \to \min_{\boldsymbol{\theta}}$$

Метод наименьших квадратов (Ordinary least squares)





Суть МНК: в качестве оценок взять такие параметры $\boldsymbol{\theta}$, при которых сумма квадратов ошибок прогноза Q минимальна

Метод наименьших квадратов (Ordinary least squares)

Модель

$$\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{\theta_0} + \boldsymbol{\theta_1} \boldsymbol{x_1}$$

Функционал ошибки

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell} \sum_{i=1}^{\ell} (a(x_i) - y_i)^2 \rightarrow \min_{\boldsymbol{\theta}}$$

Метод наименьших квадратов (Ordinary least squares)

Функционал ошибки

$$Q = \frac{1}{\ell}\sum_{i=1}^{\ell}(\boldsymbol{y_i} - \boldsymbol{\theta_0} + \boldsymbol{\theta_1 x_1})^2 \to \min_{\boldsymbol{\theta}}$$

Нахождение производных

$$\frac{\partial Q}{\partial \theta_0} \qquad \frac{\partial Q}{\partial \theta_1}$$

Метод наименьших квадратов (Ordinary least squares)

Функционал ошибки

$$Q = \frac{1}{\ell}\sum_{i=1}^{\ell}(\boldsymbol{y_i} - \boldsymbol{\theta_0} + \boldsymbol{\theta_1 x_1})^2 \to \min_{\boldsymbol{\theta}}$$

Нахождение производных

$$\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{\theta_0}} = \sum_{\boldsymbol{i=1}}^{\ell} \boldsymbol{2(y_i - \theta_0 + \theta_1 x_i).(1)}$$

Метод наименьших квадратов (Ordinary least squares)

Функционал ошибки

$$Q = \frac{1}{\ell}\sum_{i=1}^{\ell}(\boldsymbol{y_i} - \boldsymbol{\theta_0} + \boldsymbol{\theta_1 x_1})^2 \to \min_{\boldsymbol{\theta}}$$

Нахождение производных

$$\frac{\partial Q}{\partial \theta_0} = \sum_{i=1}^{\ell} 2(y_i - \theta_0 + \theta_1 x_i).(1)$$

$$\frac{\partial Q}{\partial \theta_1} = \sum_{i=1}^{\ell} 2(y_i - \theta_0 + \theta_1 x_i).(x_i)$$

Метод наименьших квадратов (Ordinary least squares)

Нахождение производных

$$\frac{\partial Q}{\partial \theta_0} = 0 \qquad \frac{\partial Q}{\partial \theta_1}=0$$

$$\begin{cases} \sum_{i=1}^{\ell} 2(y_i - \theta_0 + \theta_1 x_i).(1) = 0 \\ \sum_{i=1}^{\ell} 2(y_i - \theta_0 + \theta_1 x_i).(x_i) = 0 \end{cases}$$

Метод наименьших квадратов (Ordinary least squares)

$$\begin{cases} 2\left(\theta_1 \sum_{i=1}^{\ell} x_i + \theta_0 \ell - \sum_{i=1}^{\ell} y_i\right) = 0 \\ 2\left(\theta_1 \sum_{i=1}^{\ell} {x_i}^2 + \theta_0 \sum_{i=1}^{\ell} x_i - \sum_{i=1}^{\ell} x_i y_i\right) = 0 \end{cases}$$

Метод наименьших квадратов (Ordinary least squares)

$$\boldsymbol{\theta_0} = \frac{\sum_{i=1}^{\ell} y_i - \theta_1 \sum_{i=1}^{\ell} x_i}{\boldsymbol{n}}$$

$$\boldsymbol{\theta_1} = \frac{\ell \sum_{i=1}^{\ell} x_i y_i - \sum_{i=1}^{\ell} x_i \cdot \sum_{i=1}^{\ell} y_i}{\ell \sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{\ell} x_i\right)^{\mathbf{2}}}$$

Метод наименьших квадратов (Ordinary least squares)

Функционал ошибки

$$Q = \frac{1}{\ell}\sum_{i=1}^{\ell}(\boldsymbol{y_i} - \boldsymbol{\theta_0} + \boldsymbol{\theta_1 x_1} + \dots + \boldsymbol{\theta_d x_d})^2 \to \min_{\boldsymbol{\theta}}$$

$$\begin{cases} \dfrac{\partial \boldsymbol{Q}}{\partial \boldsymbol{\theta_0}} = \mathbf{0} \\ \vdots \\ \dfrac{\partial \boldsymbol{Q}}{\partial \boldsymbol{\theta_d}} = \mathbf{0} \end{cases} \Longrightarrow \begin{cases} \boldsymbol{\theta_0} \\ \vdots \\ \boldsymbol{\theta_d} \end{cases}$$

Линейная регрессия в векторном виде

Функция потерь (Lost function)

Модель

$$\boldsymbol{a}(\boldsymbol{x}) = \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle$$

Среднеквадратичная ошибка

$$Q(a, \mathbb{X}) = \frac{1}{\ell} \sum_{i=1}^{\ell} (\langle \theta, x \rangle - y_i)^2$$

Цель

$$\underset{\theta}{minimize}\, Q(\mathrm{a}, \mathbb{X})$$

Матрица

Матрица объектов-признаков

$$\mathbb{X} = \begin{pmatrix} x_{11} & x_{11} & \cdots & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{\ell 1} & x_{\ell 1} & \cdots & \cdots & x_{\ell d} \end{pmatrix} \in \mathbb{R}^{\ell \times d}$$

Матрица

Матрица объектов-признаков

$$\mathbb{X} = \begin{pmatrix} x_{11} & x_{11} & \cdots & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{\ell 1} & x_{\ell 1} & \cdots & \cdots & x_{\ell d} \end{pmatrix} \in \mathbb{R}^{\ell \times d}$$

Объект и его признаки

Матрица

Матрица объектов-признаков

$$\mathbb{X} = \begin{pmatrix} x_{11} & x_{11} & \cdots & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{\ell 1} & x_{\ell 1} & \cdots & \cdots & x_{\ell d} \end{pmatrix} \in \mathbb{R}^{\ell \times d}$$

Значения признака на всех объектах

Вектор

Вектор размера d тоже матрица

Вектор-строка:

$$\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \dots, \boldsymbol{\theta_d}) \in \mathbb{R}^{\mathbf{1}\times \boldsymbol{d}}$$

Вектор-столбец:

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta_1} \\ \vdots \\ \boldsymbol{\theta_d} \end{pmatrix} \in \mathbb{R}^{\boldsymbol{d}\times \mathbf{1}}$$

Матричное умножение

$$A \in \mathbb{R}^{m\times \boldsymbol{k}} \qquad \mathrm{B} \in \mathbb{R}^{\boldsymbol{k}\times n}$$

$$A\mathrm{B} = \mathrm{C} \in \mathbb{R}^{m\times n}$$

$$C_{ij} = \sum_{p=1}^{k} a_{ij} b_{pj}$$

Матричное умножение

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 10 & 0 \end{pmatrix} * \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix}$$

Матричное умножение

$$\begin{pmatrix} \boxed{1 \quad 2} \\ 0 \quad 1 \\ 10 \quad 0 \end{pmatrix} * \begin{pmatrix} \boxed{\begin{matrix}1\\0\end{matrix}} & \begin{matrix}0\\0\end{matrix} & \begin{matrix}0\\2\end{matrix} \end{pmatrix} = \begin{pmatrix} 1 & & \\ & & \\ & & \end{pmatrix}$$

1*1 + 2*0=1

Матричное умножение

$$\begin{pmatrix} \boxed{1 \quad 2} \\ 0 \quad 1 \\ 10 \quad 0 \end{pmatrix} * \begin{pmatrix} 1 & \boxed{\begin{matrix} 0 \\ 0 \end{matrix}} & 0 \\ 0 & & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \\ & & \\ & & \end{pmatrix}$$

1*0 + 2*0=0

Матричное умножение

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 10 & 0 \end{pmatrix} * \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 4 \\ & & \\ & & \end{pmatrix}$$

1*0 + 2*2=4

Матричное умножение

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 10 & 0 \end{pmatrix} * \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 4 \\ 0 & & \\ & & \end{pmatrix}$$

0*1 + 1*0=0

Применение линейной модели

$$\boldsymbol{a}(\boldsymbol{x}) \ = \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle = \boldsymbol{\theta_0} \boldsymbol{x_0} + \boldsymbol{\theta_1} \boldsymbol{x_1} + \ldots + \boldsymbol{\theta_d}\, \boldsymbol{x_d}$$

$$\mathrm{X} = \begin{pmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{\ell 1} & \cdots & x_{\ell d} \end{pmatrix} \in \mathbb{R}^{\ell \times d}$$

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{pmatrix} \in \mathbb{R}^{\boldsymbol{d} \times \boldsymbol{1}}$$

Результат применения линейной модели к выборке $\mathbb{X}$

$$\mathrm{X}\theta = \begin{pmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{\ell 1} & \cdots & x_{\ell d} \end{pmatrix} * \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{d} \theta_i x_{1i} \\ \sum_{i=2}^{d} \theta_i x_{2i} \\ \vdots \\ \sum_{i=1}^{d} \theta_i x_{\ell i} \end{pmatrix} = \begin{pmatrix} \langle \theta, x_1 \rangle \\ \langle \theta, x_2 \rangle \\ \vdots \\ \langle \theta, x_\ell \rangle \end{pmatrix}$$

Вычисление ошибки

Функционал ошибки

$$Q(a, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(\langle \boldsymbol{\theta}, \boldsymbol{x}\rangle - y_i)^2 \to \min_{\boldsymbol{\theta}}$$

Отклонения прогнозов от ответа

$$X\theta - y = \begin{pmatrix} \langle \theta, x_1\rangle - y_1 \\ \langle \theta, x_2\rangle - y_2 \\ \vdots \\ \langle \theta, x_\ell\rangle - y_\ell \end{pmatrix}$$

Вычисление ошибки

Евклидова норма:

$$\|z\| = \sqrt{\sum_{j=1}^{n} z_j^2} \qquad \|z\|^2 = \sum_{j=1}^{n} z_j^2$$

Отклонения прогнозов от ответа

$$X\theta - y = \begin{pmatrix} \langle \theta, x_1 \rangle - y_1 \\ \langle \theta, x_2 \rangle - y_2 \\ \vdots \\ \langle \theta, x_\ell \rangle - y_\ell \end{pmatrix}$$

Вычисление ошибки

Отклонения прогнозов от ответа

$$\mathsf{X}\theta - y = \begin{pmatrix} \langle\theta, x_1\rangle - y_1 \\ \langle\theta, x_2\rangle - y_2 \\ \vdots \\ \langle\theta, x_\ell\rangle - y_\ell \end{pmatrix}$$

Среднеквадратичная ошибка

$$\frac{1}{\ell}\sum_{i=1}^{\ell}(\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle - y_i)^2 = \frac{\mathbf{1}}{\ell}\|\mathbf{X}\boldsymbol{\theta} - \boldsymbol{y}\|^{\mathbf{2}}$$

Цель:

$$\frac{\mathbf{1}}{\ell}\|\mathbf{X}\boldsymbol{\theta} - \boldsymbol{y}\|^{\mathbf{2}} \to \underset{\boldsymbol{\theta}}{min}$$

Среднеквадратичная ошибка

MSE для линейной регрессии:

$$Q(a, \mathbb{X}) = \frac{1}{\ell} \sum_{i=1}^{\ell} (\boldsymbol{\theta_1} \boldsymbol{x_{i1}} + \ldots + \boldsymbol{\theta_d} \boldsymbol{x_{id}})^2$$

Нормальные уравнения (Normal equation)

$$\frac{\mathbf{1}}{\ell}\|\mathsf{X}\boldsymbol{\theta}-\boldsymbol{y}\|^{2} \to \min_{\boldsymbol{\theta}}$$

Вычисление градиента

$$\nabla\frac{\mathbf{1}}{\ell}\|\mathsf{X}\boldsymbol{\theta}-\boldsymbol{y}\|^{2} = \frac{2}{\ell}\boldsymbol{X}^{\boldsymbol{T}}(\mathsf{X}\boldsymbol{\theta}-\boldsymbol{y})$$

Приравнивание градиента к нулю

$$\frac{2}{\ell}\boldsymbol{X}^{\boldsymbol{T}}(\mathsf{X}\boldsymbol{\theta}-\boldsymbol{y}) = \mathbf{0}$$

Нормальные уравнения (Normal equation)

$$\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

$\boldsymbol{\theta}$ – вектор коэффициентов (весов) модели

$\boldsymbol{X}^T$ – транспонированная матрица признаков

$\boldsymbol{y}$ – вектор меток

Нормальные уравнения (Normal equation)

$$\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

$\boldsymbol{\theta}$ – вектор коэффициентов (весов) модели

$\boldsymbol{X}^T$ – транспонированная матрица признаков

$\boldsymbol{y}$ – вектор меток

```
np.linalg.inv(X.T @ X) @ X.T @ Y
```

Нормальные уравнения (Normal equation)

$$\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

у метода нормальных уравнений есть несколько проблем

Высокая вычислительная сложность

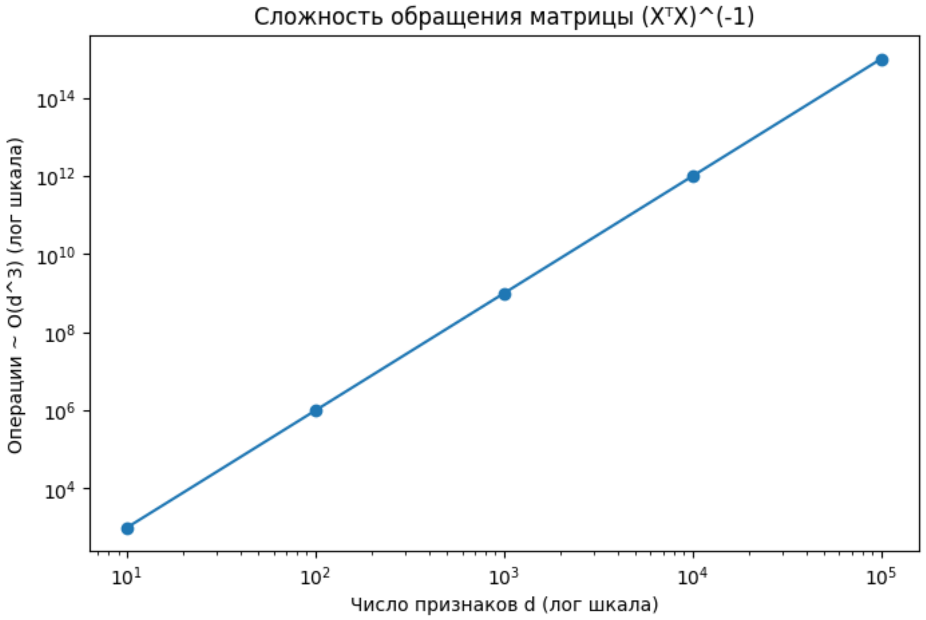
Проблемы с обратимостью матрицы

Когда количество объектов меньше количества признаков

Высокая вычислительная сложность

$$(X^T X)^{-1}$$

Вычисляется за $O(d^3)$



Если процессор выполняет $\sim 10^9$ операций / сек:

$d = 10 \rightarrow \sim 10^{-6}$ с микросекунды

$d = 100 \rightarrow \sim 10^{-3}$ с микросекунды

$d = 1000 \rightarrow 1$ с

$d = 10000 \rightarrow 1000$ с ~ 17 минут

$d = 100000 \rightarrow \sim 10^6$с ~ 11.6 дней

Проблемы с обратимостью матрицы

$$\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

Линейная зависимость признаков (мультиколлинеарность)

- Какой-то столбец X, линейная комбинация других, тогда rank(x) < число признаков

Признак с нулевой дисперсией: столбец константа

Проблемы с обратимостью матрицы

$$A^{-1}, \qquad AA^{-1} = I$$

`np.linalg.inv`

Работает только для квадратных и невырожденных (полного ранга) матриц.

Если матрица вырожденная (det = 0), то np.linalg.inv выдаст ошибку LinAlgError: Singular matrix.

`np.linalg.pinv` (псевдообратную Мура–Пенроуза)

Определена для любых матриц: прямоугольных, сингулярных и плохо обусловленных.

Реализована через SVD (сингулярное разложение)

Нормальные уравнения (Normal equation)

Постановка

- Даны данные $\{(x_i, y_i)\}_{i=1}^{\ell}$
- Соберём матрицу признаков $X \in \mathbb{R}^{\ell \times d}$
- Вектор ответов $y \in \mathbb{R}^{\ell}$
- Ищем всеа $\theta \in \mathbb{R}^{\boldsymbol{d}}$ для линейной модели

$$\boldsymbol{a}(\boldsymbol{x}) = \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle$$

Функция качества (средняя квадратичная ошибка)

$$\boldsymbol{Q}(\boldsymbol{\theta}) = \frac{\mathbf{1}}{\ell}\left(\boldsymbol{y} - \boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{T}}\boldsymbol{\theta}\right)^{\mathbf{2}} = \frac{\mathbf{1}}{\ell}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_{\mathbf{2}}^{\mathbf{2}} = \frac{\mathbf{1}}{\ell}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})^{\boldsymbol{T}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) \to \min_{\boldsymbol{\theta}}$$

Нормальные уравнения (Normal equation)

1) Раскрываем скобки

- Используем тождества для скаляров : $a^T x = x^T a$

$$(\boldsymbol{y} - \boldsymbol{X\theta})^{\boldsymbol{T}}(\boldsymbol{y} - \boldsymbol{X\theta}) = \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y} - \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{X\theta} - (\boldsymbol{X\theta})^T y + (\boldsymbol{X\theta})^T(\boldsymbol{X\theta})$$

$$= \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y} - \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{X\theta} - \boldsymbol{\theta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} + \boldsymbol{\theta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X\theta}$$

$$= \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y} - 2\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{X\theta} + \boldsymbol{\theta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X\theta}$$

$\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{X\theta}$ равно своему транспонированию $\left(\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{X\theta}\right)^T = \boldsymbol{\theta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}$

$$\boldsymbol{Q}(\boldsymbol{\theta}) = \frac{\boldsymbol{1}}{\ell}\left(\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y} - 2\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{X\theta} + \boldsymbol{\theta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X\theta}\right)$$

Нормальные уравнения (Normal equation)

2) Считаем градиент по $\boldsymbol{\theta}$

- Правила матричного дифференцирования:

$$\nabla_{\boldsymbol{\theta}}(b^T\boldsymbol{\theta}) = b$$

$$\nabla_{\boldsymbol{\theta}}(\boldsymbol{\theta}^T b) = b$$

$\nabla_{\boldsymbol{\theta}}(\boldsymbol{\theta}^T A\boldsymbol{\theta}) = (A + A^T)\boldsymbol{\theta}$ (в частности, если A симметрична, то $= 2A\boldsymbol{\theta}$)

Здесь $A = X^T X$ $-$ симметрична.

$$\nabla_{\boldsymbol{\theta}}\boldsymbol{Q}(\boldsymbol{\theta}) = \frac{\mathbf{1}}{\ell}\left(\mathbf{0} - \mathbf{2}\boldsymbol{X}^T\boldsymbol{y} + \mathbf{2}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta}\right) = \frac{\mathbf{2}}{\ell}\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{X}^T\boldsymbol{y}\right)$$

Эквивалентно можно запомнить короткое правило: $\nabla_{\boldsymbol{\theta}}(\|y - X\boldsymbol{\theta}\|^2) = 2\boldsymbol{X}^T(\boldsymbol{X\theta} - \boldsymbol{y})$

Нормальные уравнения (Normal equation)

2) Условие минимума (Нормальные уравнения)

- Ставим градиент равным нулю:

$$\frac{\mathbf{2}}{\ell}\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}\right) = \mathbf{0}$$

$$\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{\theta} = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}$$

Это линейная система вида A$\boldsymbol{\theta} = \mathrm{b}$ c A: $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}$, $\boldsymbol{b} \coloneqq \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}$

Если матрица А обранима, то умножаем на A^{-1}

$$A^{-1}A\boldsymbol{\theta} = A^{-1}\mathrm{b} \Rightarrow \boldsymbol{\theta} = A^{-1}\mathrm{b}$$

$$\boldsymbol{\theta} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}$$