

Основы машинного обучения

Лекция 5

Линейная регрессия

Евгений Соколов

esokolov@hse.ru

НИУ ВШЭ, 2022

Обучение линейной регрессии

$$\frac{1}{\ell} \|Xw - y\|^2 \rightarrow \min_w$$

- Вычисление MSE в NumPy:

```
np.square(X.dot(w) - y).mean()
```

Обучение линейной регрессии

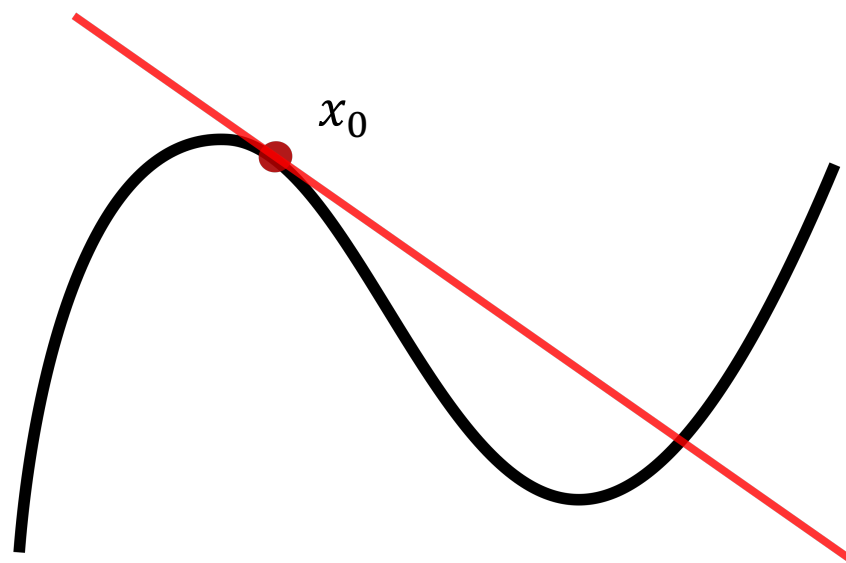
Среднеквадратичная ошибка

- MSE для линейной регрессии:

$$Q(w_1, \dots, w_d) = \frac{1}{\ell} \sum_{i=1}^{\ell} (\mathbf{w}_1 x_1 + \dots + \mathbf{w}_d x_d - y_i)^2$$

Производная

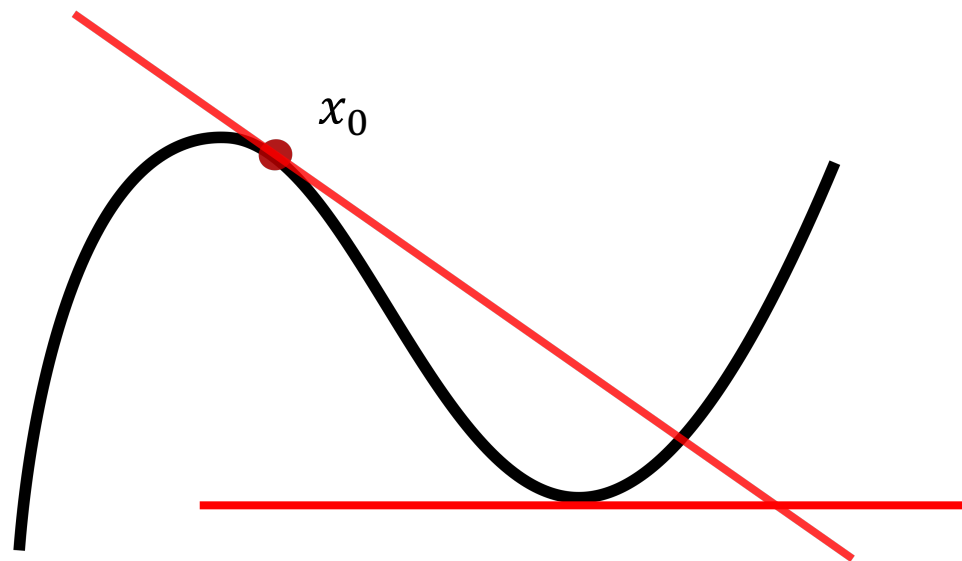
$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$$



Производная

- Если точка x_0 — экстремум и в ней существует производная, то

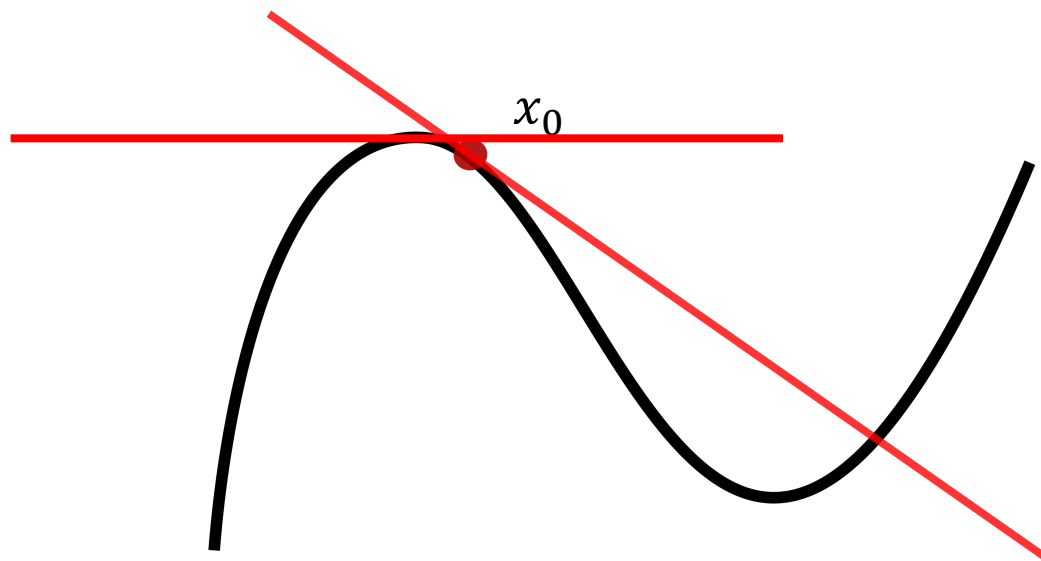
$$f'(x_0) = 0$$



Производная

- Если точка x_0 — экстремум и в ней существует производная, то

$$f'(x_0) = 0$$



Градиент

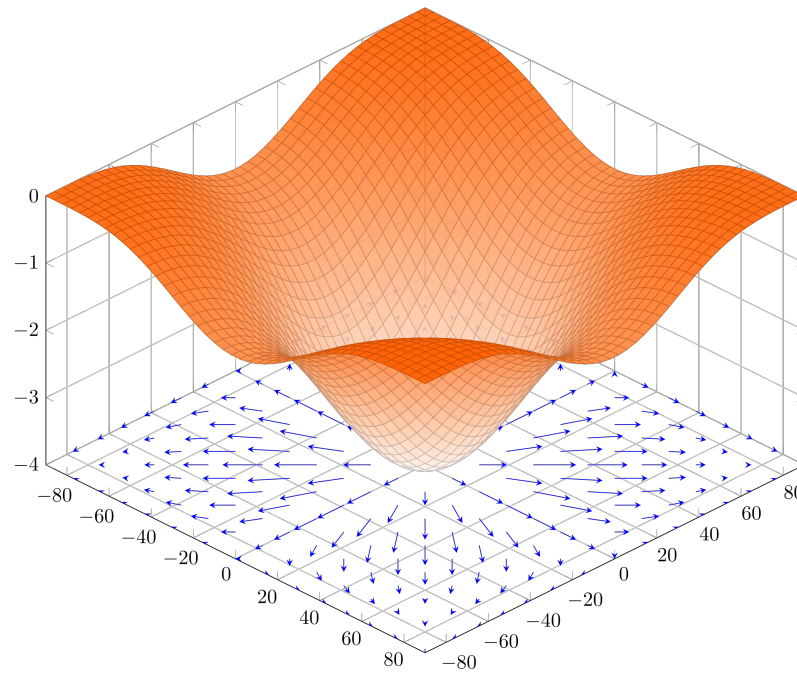
- Градиент — вектор частных производных

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_d} \right)$$

- У градиента есть важное свойство!

Важное свойство

- Зафиксируем точку x_0
- В какую сторону функция быстрее всего растёт?



Важное свойство

- Зафиксируем точку x_0
- В какую сторону функция быстрее всего растёт?
- В направлении градиента!
- Если градиент равен нулю, то это экстремум

Условие экстремума

- Если точка x_0 — экстремум и в ней существует производная, то

$$\nabla f(x_0) = 0$$

Условие экстремума

- Если точка x_0 — экстремум и в ней существует производная, то

$$\nabla f(x_0) = 0$$

- Если функция выпуклая, то экстремум один
- MSE для линейной регрессии — выпуклая!
 - (при некоторых условиях)

Обучение линейной регрессии

- Можно посчитать градиент MSE:

$$\nabla \frac{1}{\ell} \|Xw - y\|^2 = \frac{2}{\ell} X^T (Xw - y)$$

- Приравниваем нулю и решаем систему линейных уравнений:

$$w = (X^T X)^{-1} X^T y$$

Аналитическое решение

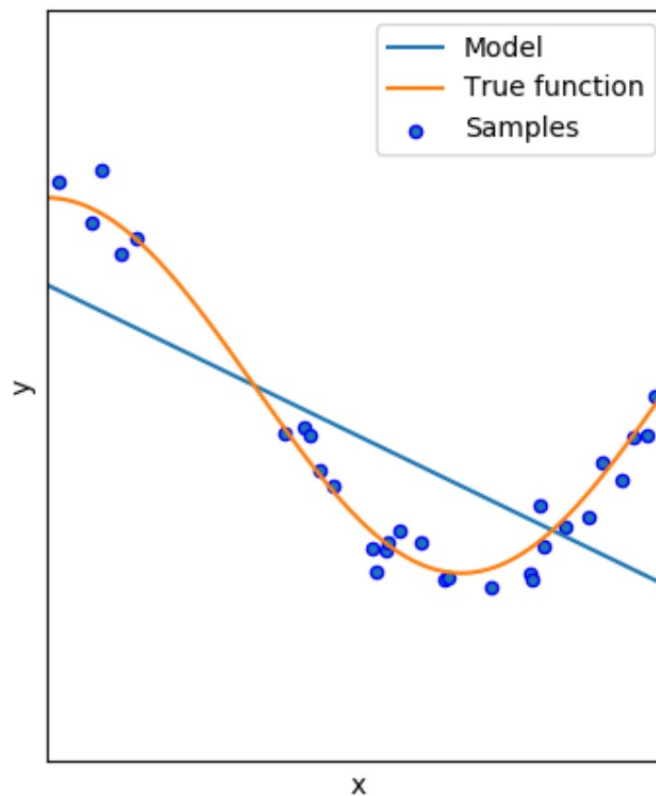
$$w = (X^T X)^{-1} X^T y$$

- Если матрица $X^T X$ вырожденная, то будут проблемы
- Даже если она почти вырожденная, всё равно будут проблемы
- Если признаков много, то придётся долго ждать

Переобучение и регуляризация линейных моделей

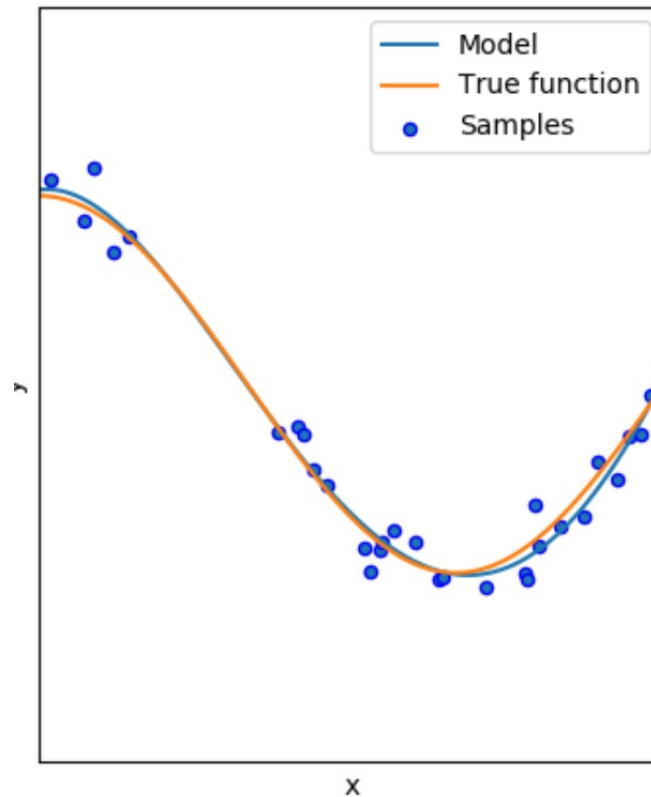
Нелинейная задача

$$a(x) = w_0 + w_1 x$$



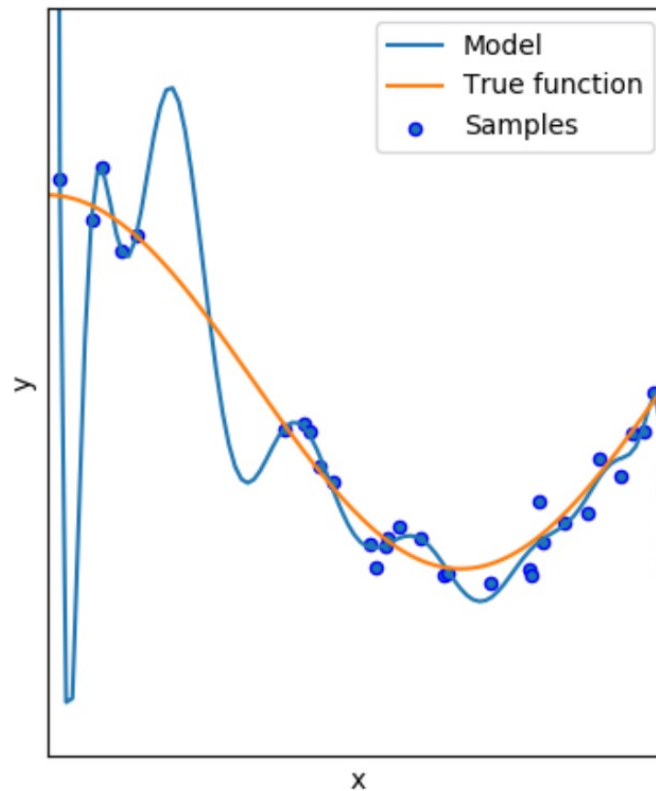
Нелинейная задача

$$a(x) = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4$$



Нелинейная задача

$$a(x) = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4 + \dots + w_{15}x^{15}$$



Симптом переобучения

$$a(x) = 0.5 + 13458922x - 43983740x^2 + \dots$$

- Большие коэффициенты — симптом переобучения
- Эмпирическое наблюдение

Симптом переобучения

- Большие коэффициенты в линейной модели — это плохо
- Пример: предсказание роста по весу

$$a(x) = 698x - 41714$$

- Изменение веса на 0.01 кг приведет к изменению роста на 7 см
- Не похоже на правильную зависимость

Регуляризация

- Будем штрафовать за большие веса!
- Пример функционала:

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

- Регуляризатор:

$$\|w\|^2 = \sum_{j=1}^d w_j^2$$

Регуляризация

- Регуляризованный функционал

$$\frac{1}{\ell} \sum_{i=1}^{\ell} (\langle w, x_i \rangle - y_i)^2 + \lambda \|w\|^2 \rightarrow \min_w$$

- λ — коэффициент регуляризации

Регуляризация

- Регуляризованный функционал

$$\frac{1}{\ell} \sum_{i=1}^{\ell} (\langle w, x_i \rangle - y_i)^2 + \lambda \|w\|^2 \rightarrow \min_w$$

- Аналитическое решение:

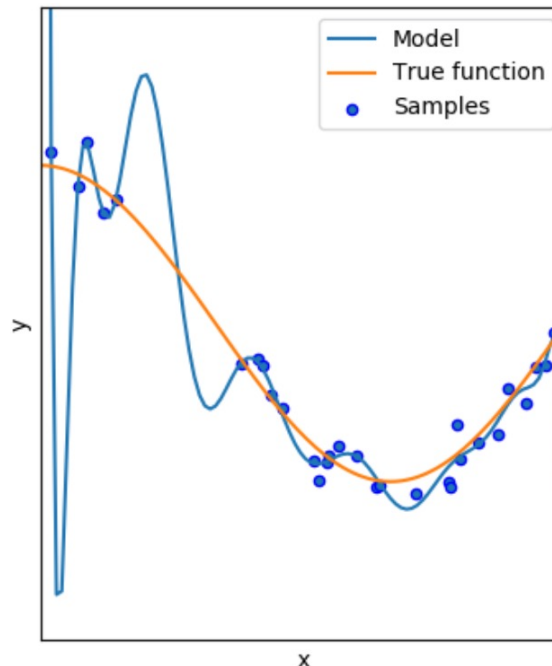
$$w = (X^T X + \lambda I)^{-1} X^T y$$

- Гребневая регрессия (Ridge regression)

Эффект регуляризации

$$a(x) = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4 + \dots + w_{15}x^{15}$$

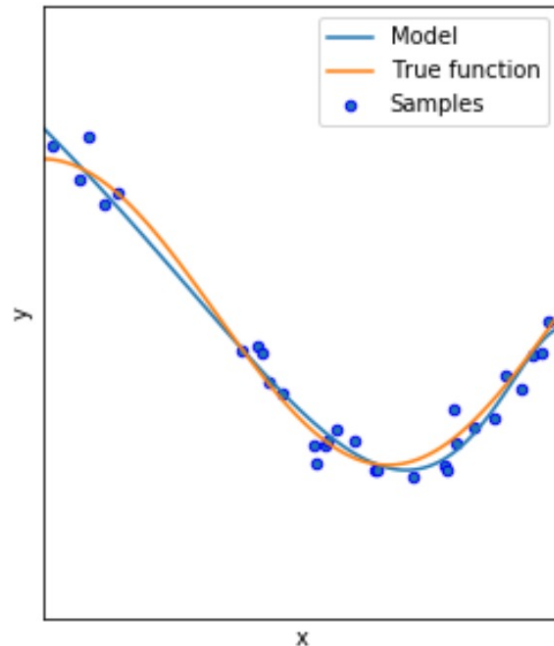
$$\frac{1}{\ell} \sum_{i=1}^{\ell} (a(x_i) - y_i)^2 \rightarrow \min_w$$



Эффект регуляризации

$$a(x) = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4 + \dots + w_{15}x^{15}$$

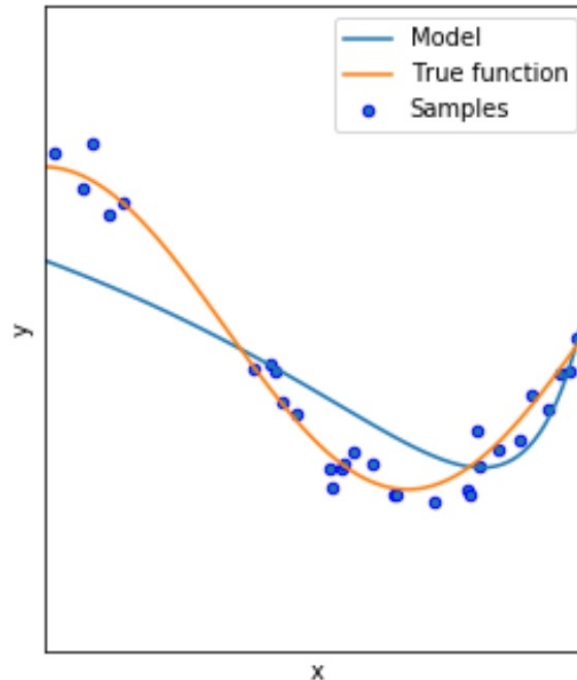
$$\frac{1}{\ell} \sum_{i=1}^{\ell} (a(x_i) - y_i)^2 + \mathbf{0.01} \|w\|^2 \rightarrow \min_w$$



Эффект регуляризации

$$a(x) = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4 + \dots + w_{15}x^{15}$$

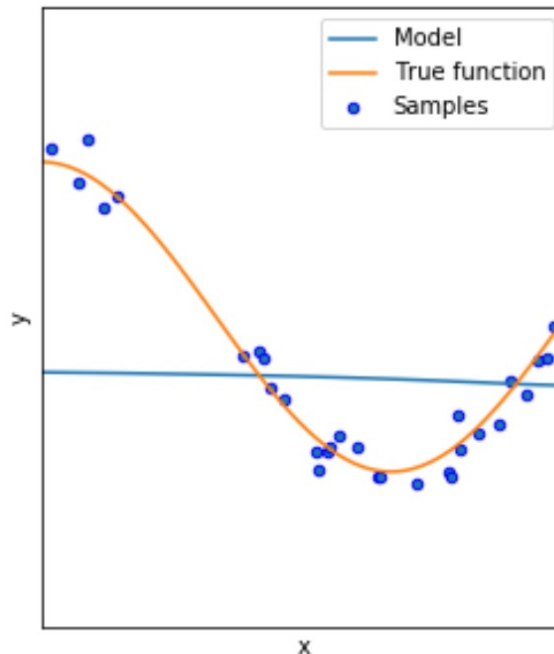
$$\frac{1}{\ell} \sum_{i=1}^{\ell} (a(x_i) - y_i)^2 + \textcolor{red}{1} \|w\|^2 \rightarrow \min_w$$



Эффект регуляризации

$$a(x) = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4 + \dots + w_{15}x^{15}$$

$$\frac{1}{\ell} \sum_{i=1}^{\ell} (a(x_i) - y_i)^2 + \mathbf{100} \|w\|^2 \rightarrow \min_w$$



Лассо

- Регуляризованный функционал

$$\frac{1}{\ell} \sum_{i=1}^{\ell} (\langle w, x_i \rangle - y_i)^2 + \lambda \sum_{j=1}^d |w_j| \rightarrow \min_w$$

- LASSO (Least Absolute Shrinkage and Selection Operator)
- Некоторые веса зануляются
- Приводит к отбору признаков

Регуляризаторы

- $\|z\|_2 = \sqrt{\sum_{j=1}^d z_j^2}$ — L_2 -норма
- $\|z\|_1 = \sum_{j=1}^d |z_j|$ — L_1 -норма