

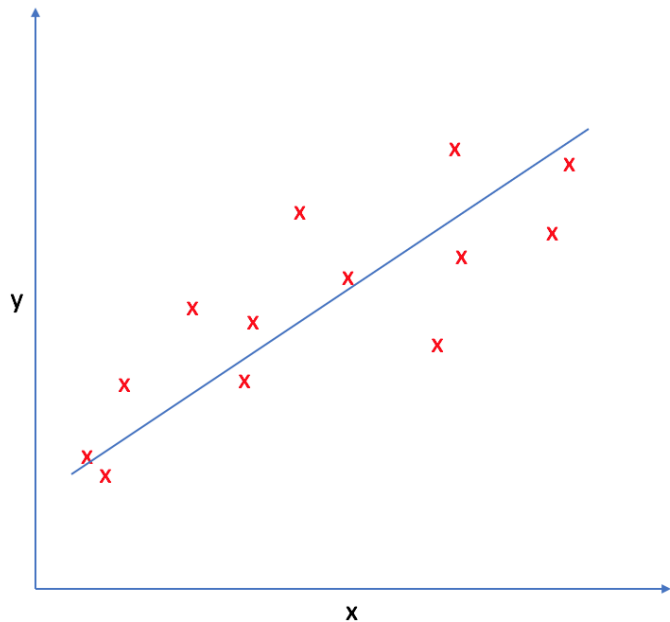
Линейная регрессия

Зюзин Юрий

План занятия

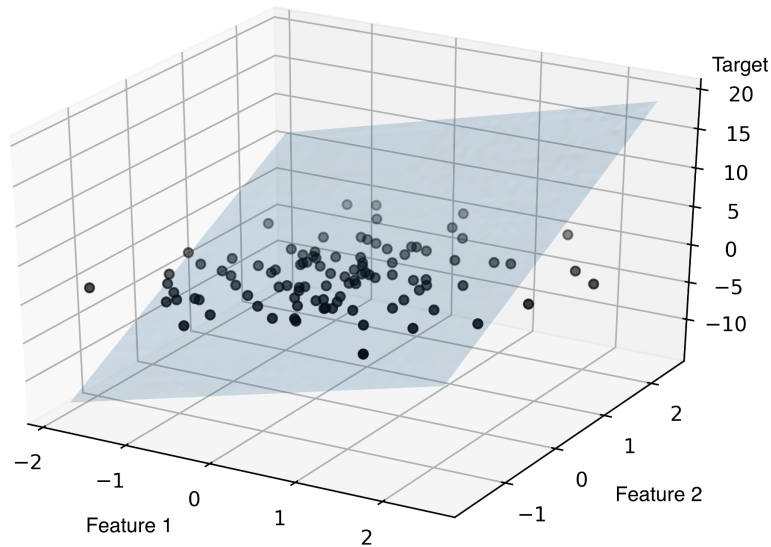
- Задачи линейной регрессии
- Критерии качества
- Подготовка признаков
- Как обучается
- Проблема переобучения и недообучения
- Кроссвалидация
- Настройка сложности
- Регуляризация
- Интерпретация линейной модели
- Практика

Задачи линейной регрессии



$$y = w_1x + w_0$$

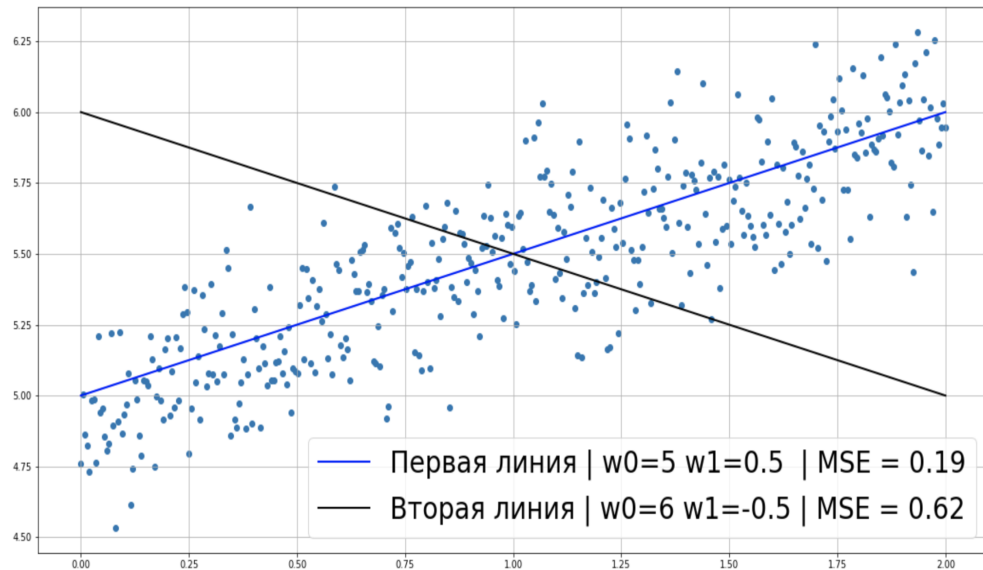
Задачи линейной регрессии



$$y = w_0 + w_1 x_1 + w_2 x_2$$

$$y = \begin{pmatrix} 1 & 0.3 & 1.3 \\ 1 & 2.1 & 0.7 \\ \vdots & \vdots & \vdots \\ 1 & 3.8 & -1.6 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} = Xw$$

Критерии качества

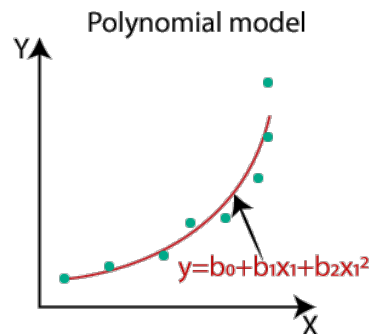
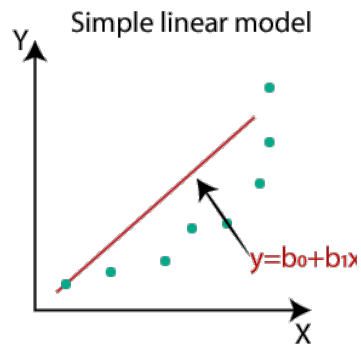


$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{true} - Xw)^2$$

Q - функция потерь

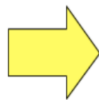
Подготовка признаков

Полиномиальные



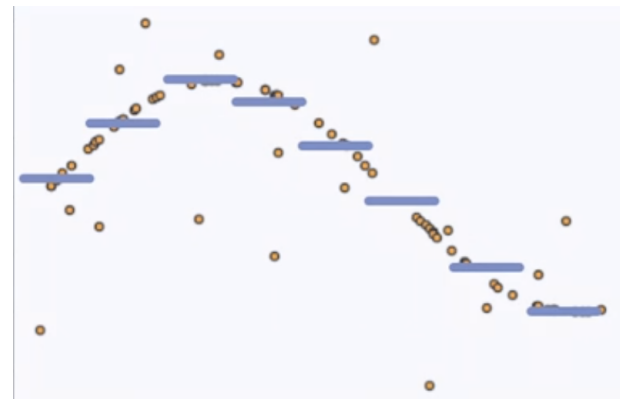
Категориальные

Color
Red
Red
Yellow
Green
Yellow



Red	Yellow	Green
1	0	0
1	0	0
0	1	0
0	0	1

Разбиение



Как обучается

1. Аналитическое

$$\begin{aligned} Q &= \|y - Xw\|_2^2 = (y - Xw)^T (y - Xw) = \\ &= y^T y - w^T X^T y - y^T Xw + w^T X^T Xw \end{aligned}$$

$$\nabla Q = 2X^T Xw - 2X^T y = 0$$

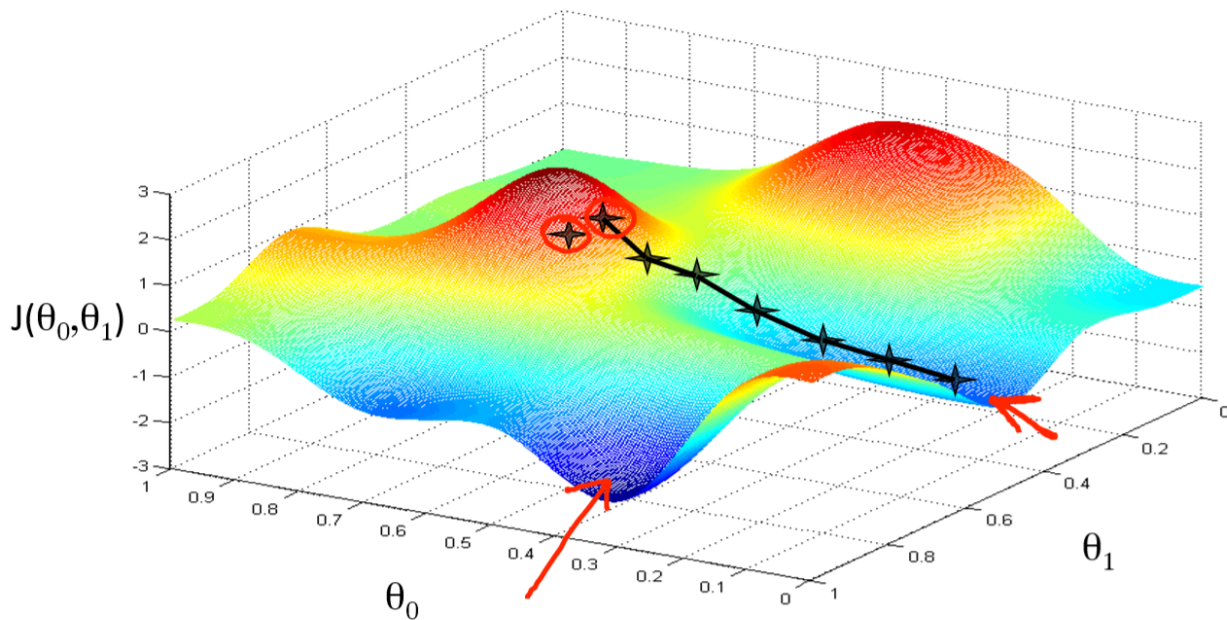
$$X^T Xw = X^T y$$

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

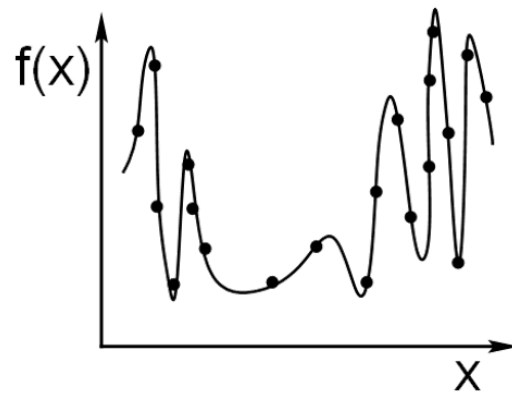
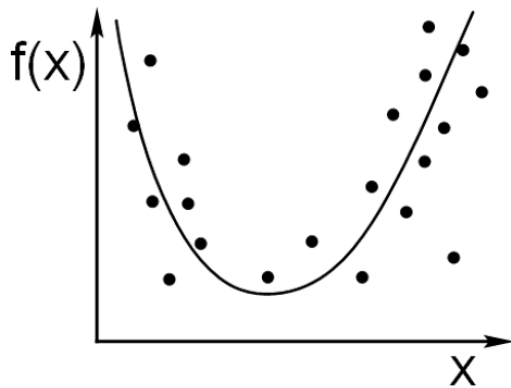
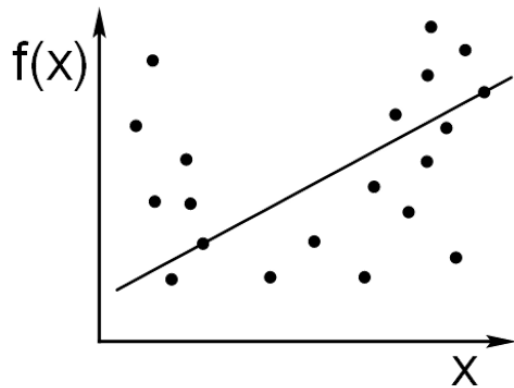
Как обучается

2. Градиентный спуск

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

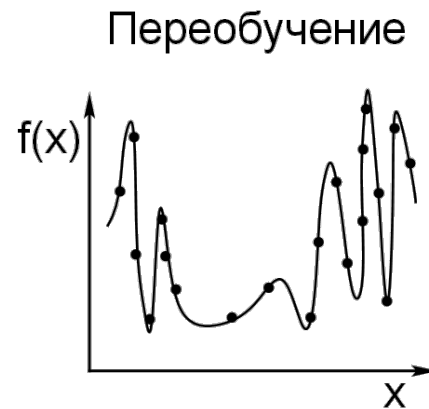
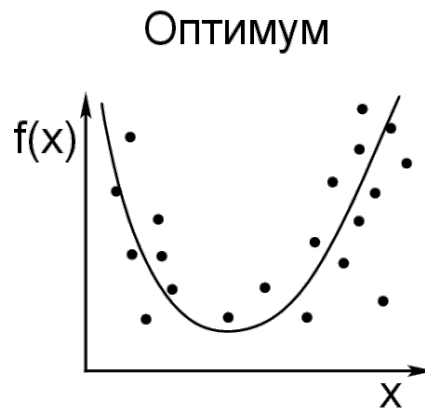
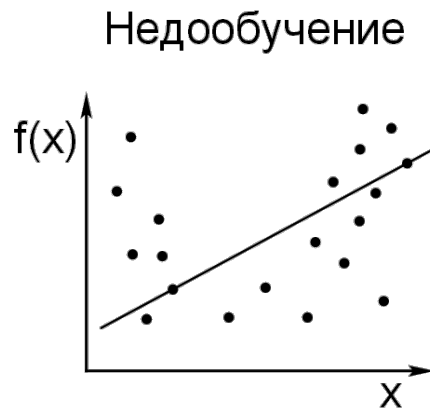


Переобучение

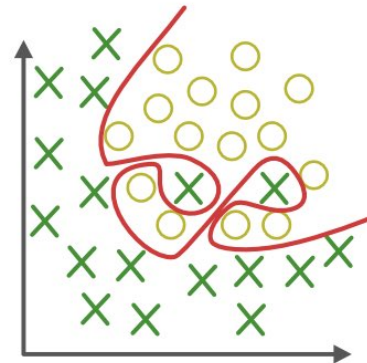
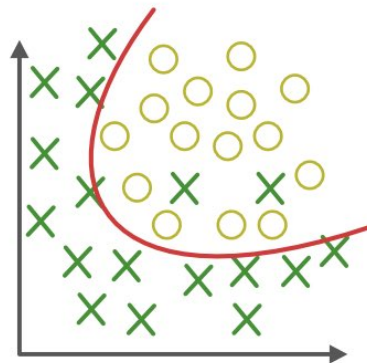
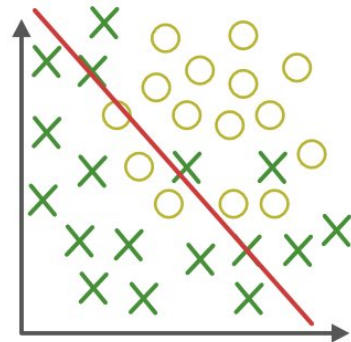


Проблема переобучения и недообучения

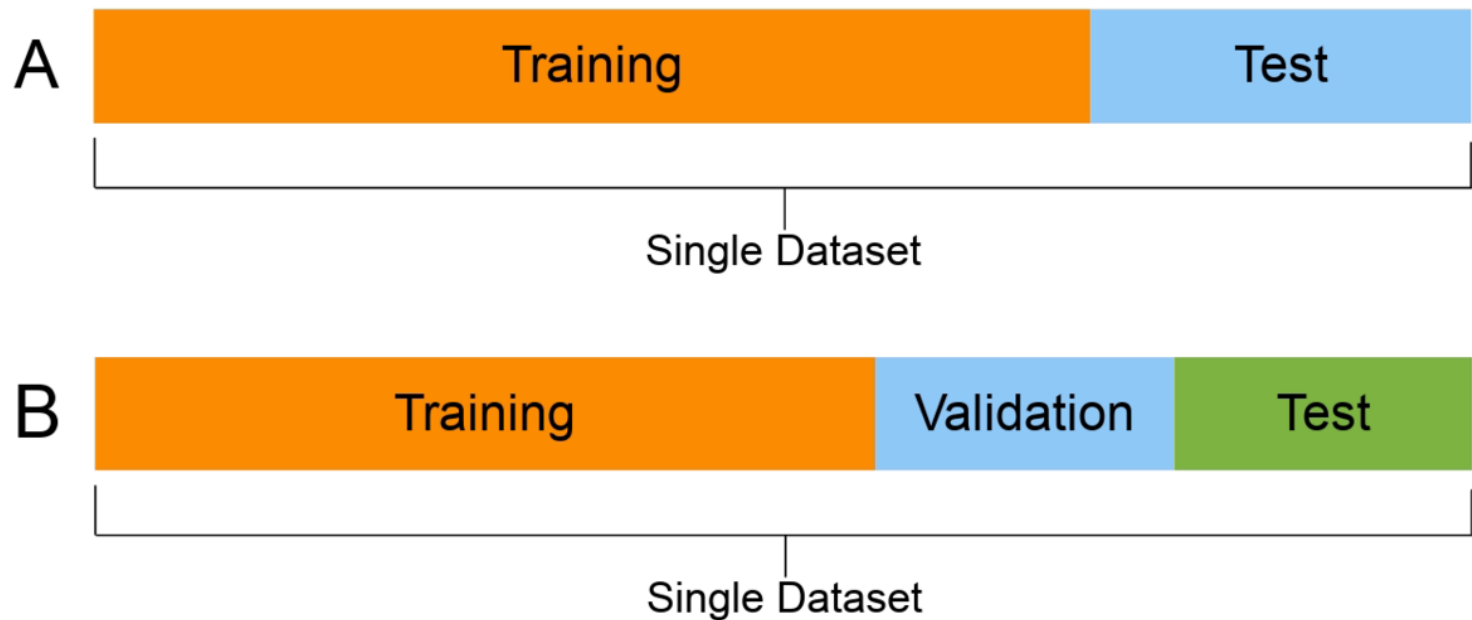
Регрессия



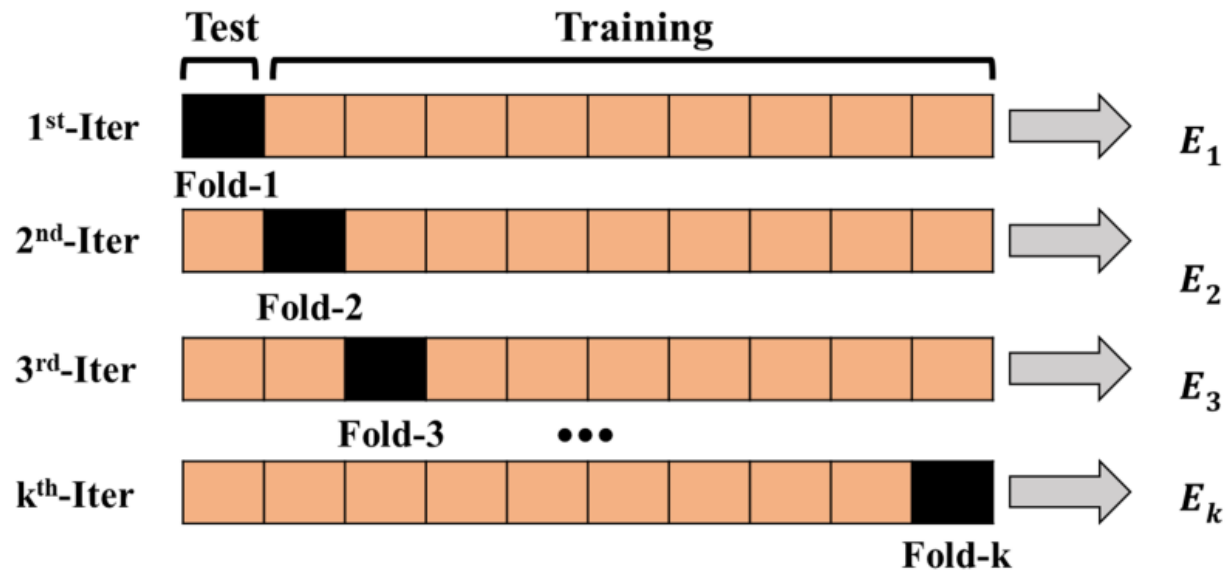
Классификация



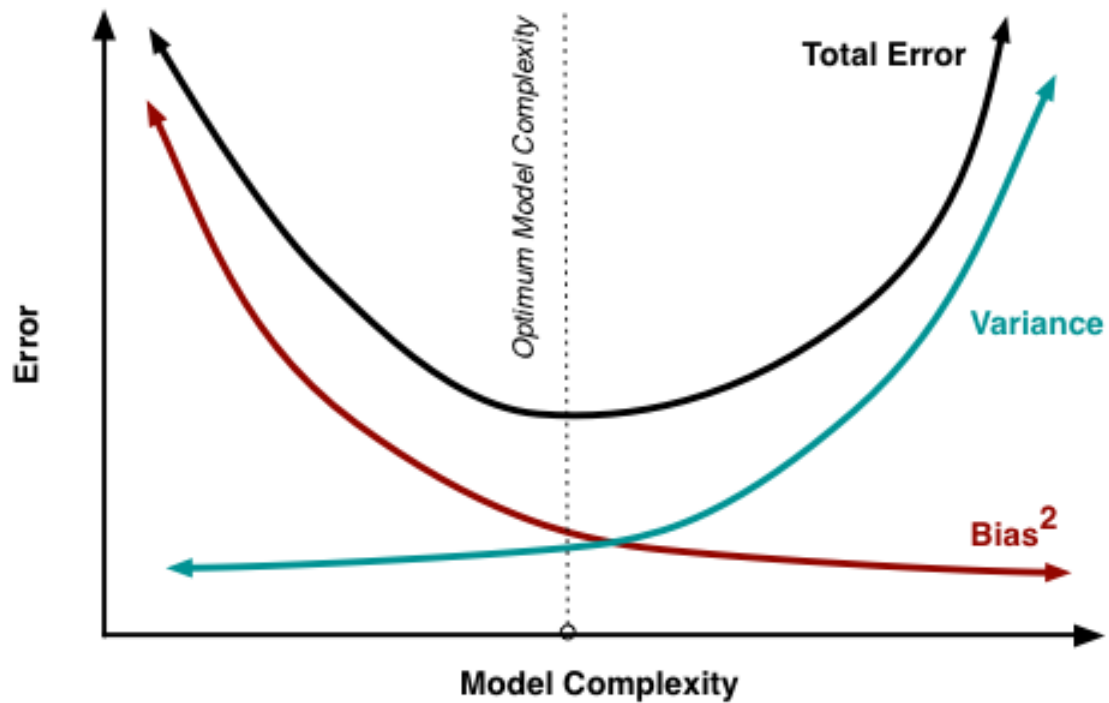
Кроссвалидация



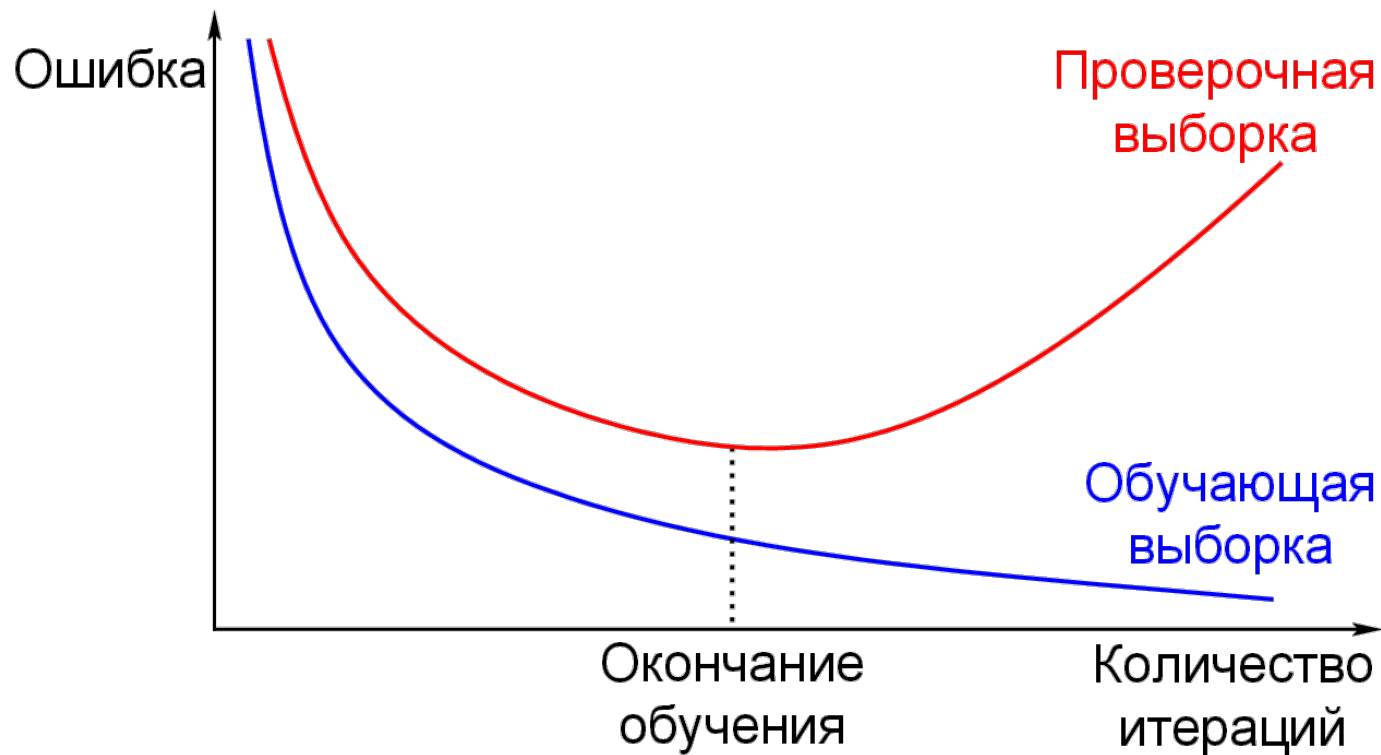
Кроссвалидация



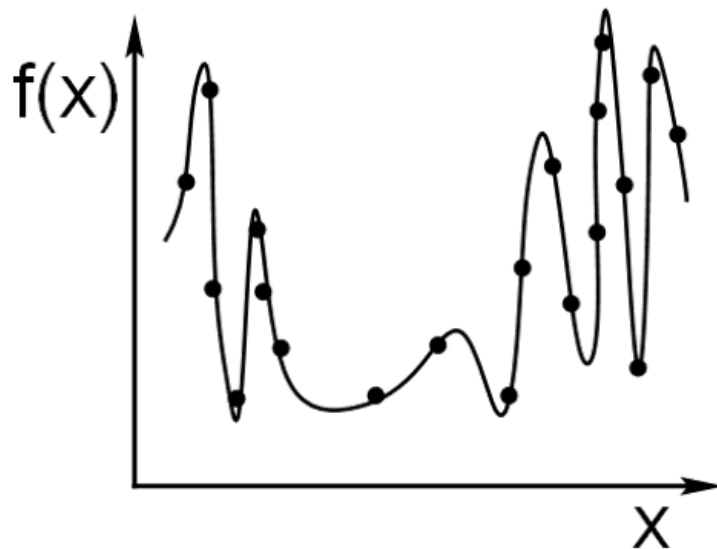
Настройка сложности



Настройка сложности



Регуляризация



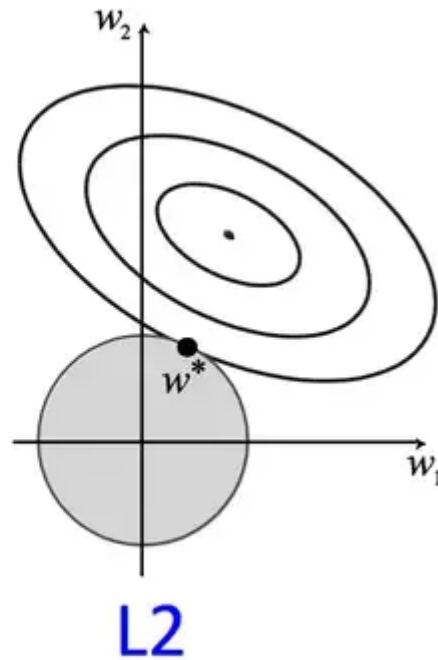
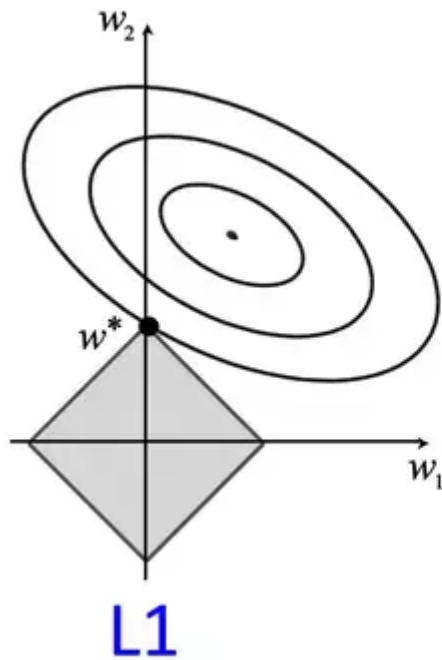
$$y = 93324528x + 23132497x^2 - 23196733x^3 + \dots$$

Регуляризация

$$L1 \quad Q = \frac{1}{n} \sum_{i=1}^n (y_{true} - Xw)^2 + \lambda |w|$$

$$L2 \quad Q = \frac{1}{n} \sum_{i=1}^n (y_{true} - Xw)^2 + \lambda |w|^2$$

Регуляризация



Интерпретация линейной модели

$$y = 100x_1 + 0.3x_2 - 200x_3$$

Чем важнее признак тем больше вес?

Интерпретация линейной модели

$$y = 100x_1 + 0.3x_2 - 200x_3$$

Чем важнее признак тем больше вес?

Только если признаки отмасштабированы и отрегуляризованы

Интерпретация линейной модели

Обучаем модель

Считаем MSE – mse_all

for признак in list(пространство_признаков):

 удаляем признак

 Обучаем модель

 Считаем MSE – mse_без_признака

 Считаем дельту $\text{delta} = \text{mse_all} - \text{mse_без_признака}$

Ваши вопросы

