



«Градиентный спуск»

ML Professional

Цели вебинара | После занятия вы узнаете

1

Ресар: линейная регрессия

2

Градиентный спуск

3

Регуляризация

4

Применение на практике

Задачи машинного обучения

С учителем
Supervised learning



Классификация



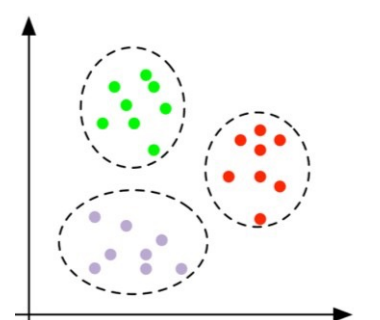
Регрессия



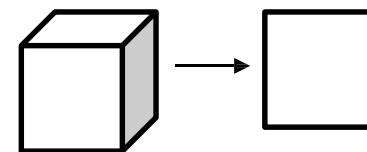
Без учителя
Unsupervised learning



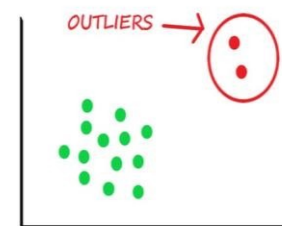
Кластеризация



Снижение
размерности



Поиск
аномалий

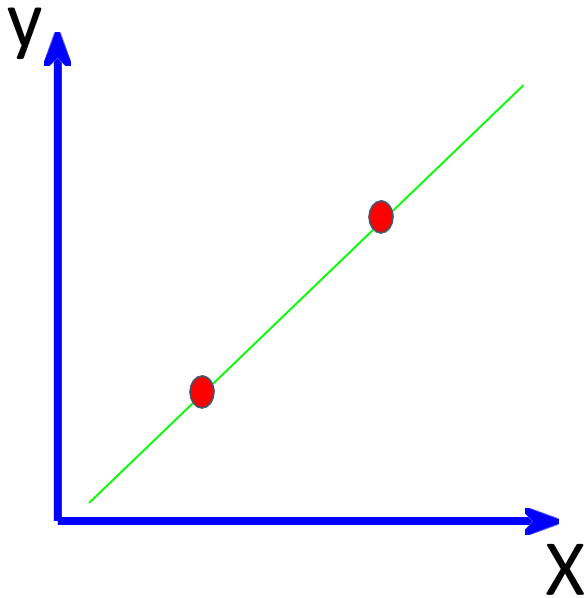


The background of the slide is an aerial photograph of a city skyline, likely New York City, with numerous skyscrapers. The entire image is filtered with a blue color scheme. A horizontal band across the middle of the image features a network of white lines and dots, resembling a data or neural network diagram, which serves as a backdrop for the title text.

Ресар: линейная регрессия

Линейная регрессия

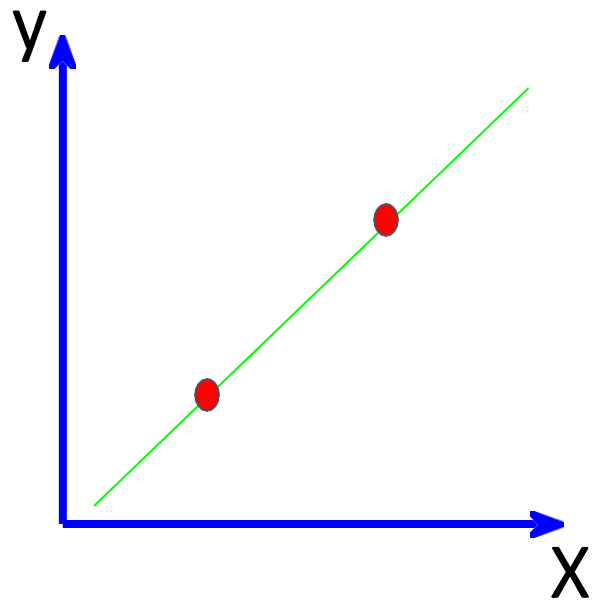
Через две точки на плоскости
можно легко провести прямую и
только одну



$$y = \omega_0 + \omega_1 x$$

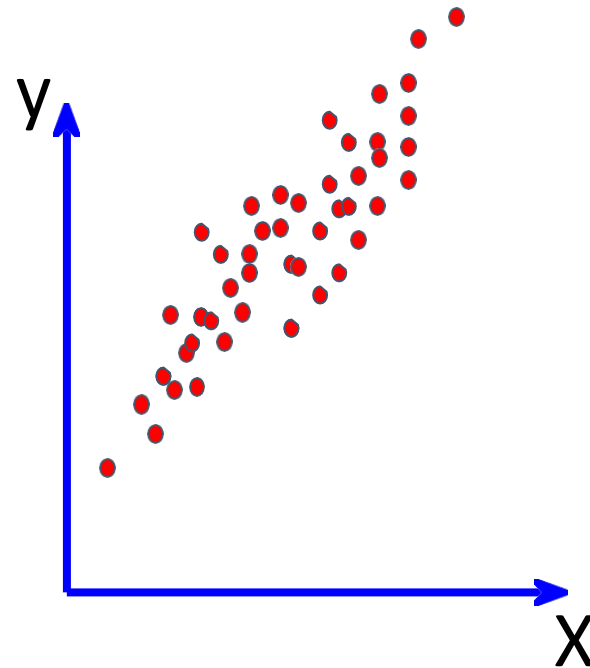
Линейная регрессия

Через две точки на плоскости можно легко провести прямую и только одну



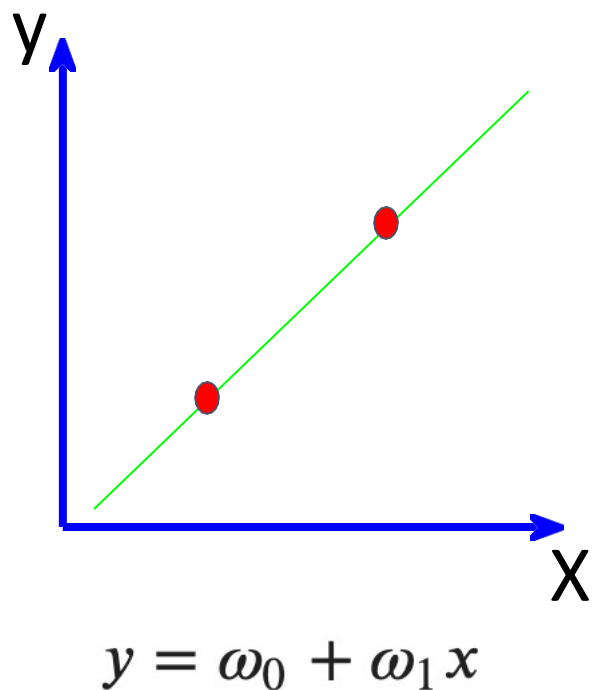
$$y = \omega_0 + \omega_1 x$$

А если точек три или более?



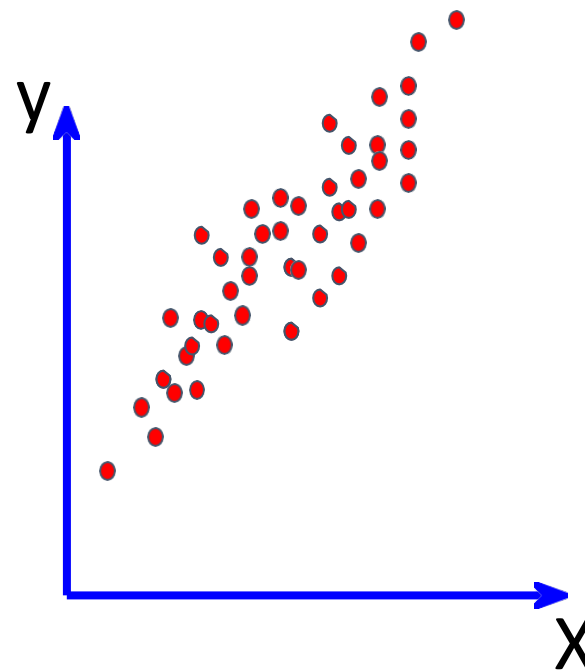
Линейная регрессия

Через две точки на плоскости можно легко провести прямую и только одну



А если точек три или более?

$$y = \omega_0 + \omega_1 x + \epsilon$$

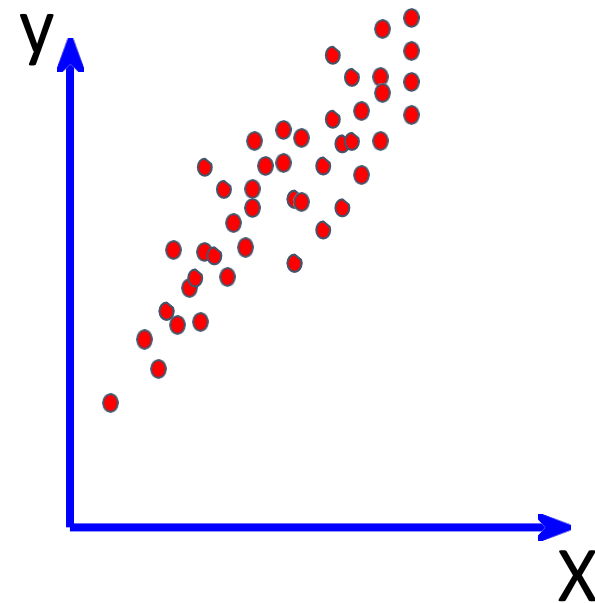


Линейная регрессия

Картина мира:

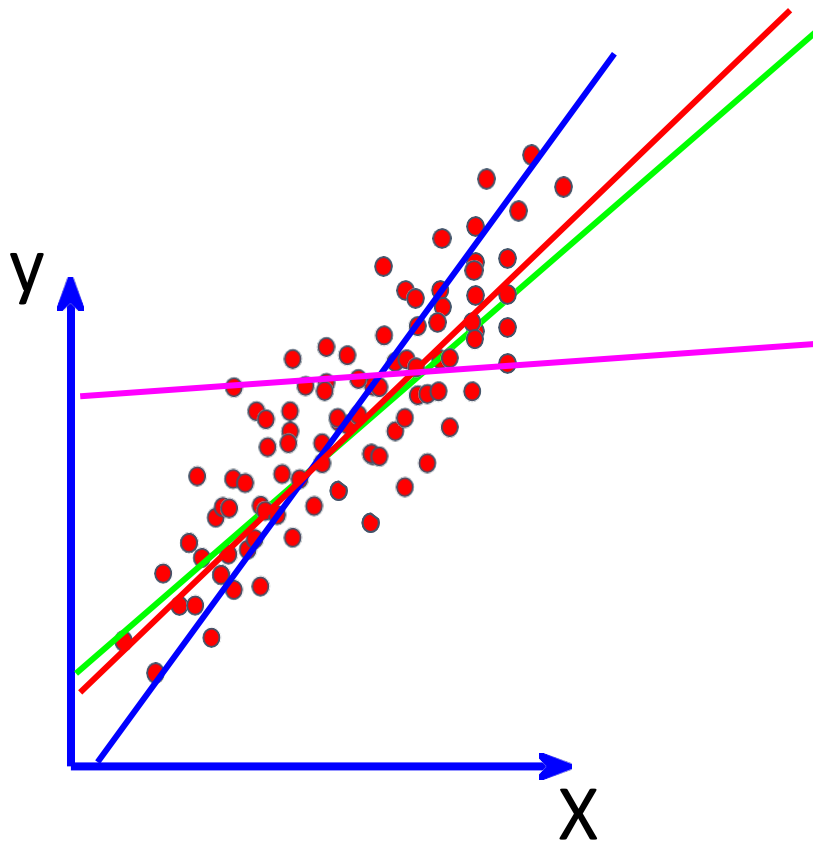
$$y = \omega_0 + \omega_1 x + \epsilon$$

- матожидание случайных ошибок равно нулю: $\forall i : \mathbb{E} [\epsilon_i] = 0$;
- дисперсия случайных ошибок одинакова и конечна, это свойство называется **гомоскедастичностью**: $\forall i : \text{Var} (\epsilon_i) = \sigma^2 < \infty$;
- случайные ошибки не скоррелированы: $\forall i \neq j : \text{Cov} (\epsilon_i, \epsilon_j) = 0$.



Линейная регрессия

А если точек три или более?

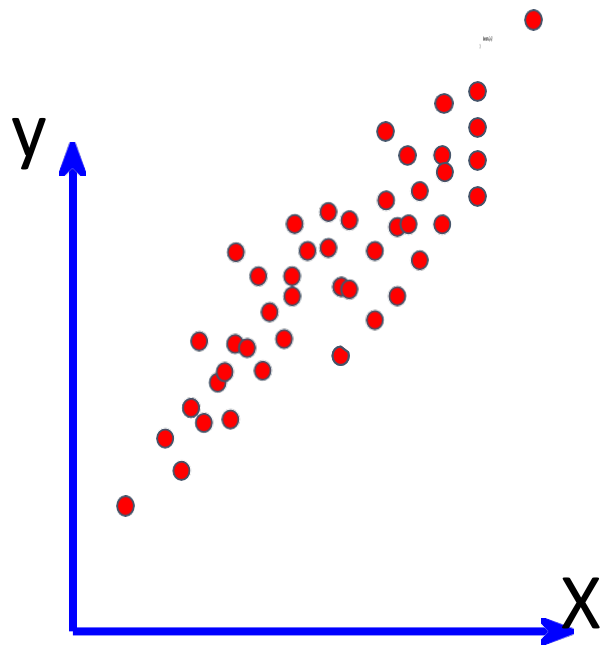


Можно конечно попробовать провести ее руками, опираясь зрительно.

Но как понять, что вы проверили ее хорошо?

Линейная регрессия

А если точек три или более?



Можно сделать умнее.
Пусть зависимость линейная,
тогда прямая будет иметь вид:

$$\hat{y} = \hat{\omega}_0 + \hat{\omega}_1 x$$

Какие взять $\hat{\omega}_0$ и $\hat{\omega}_1$?

Линейная регрессия

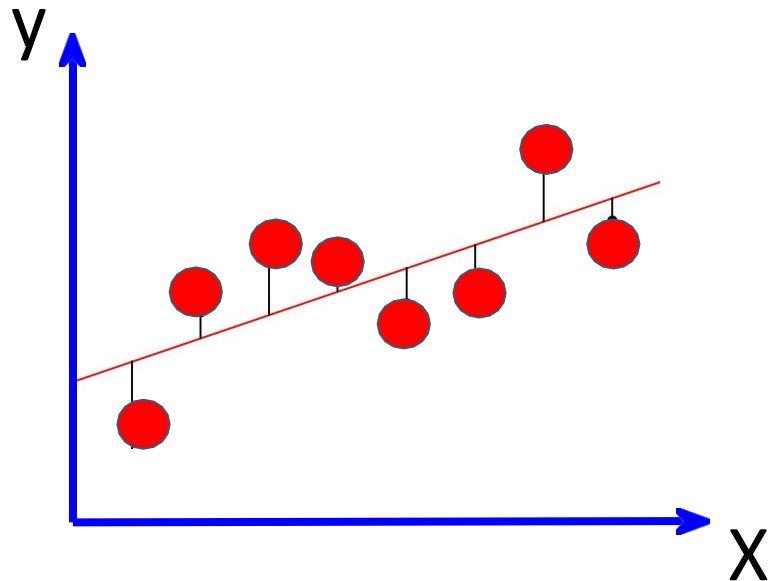
**Нам нужен какой-то показатель качества
проведенной прямой**



Метод наименьших квадратов

$$\mathcal{L}(y, \hat{y}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min$$

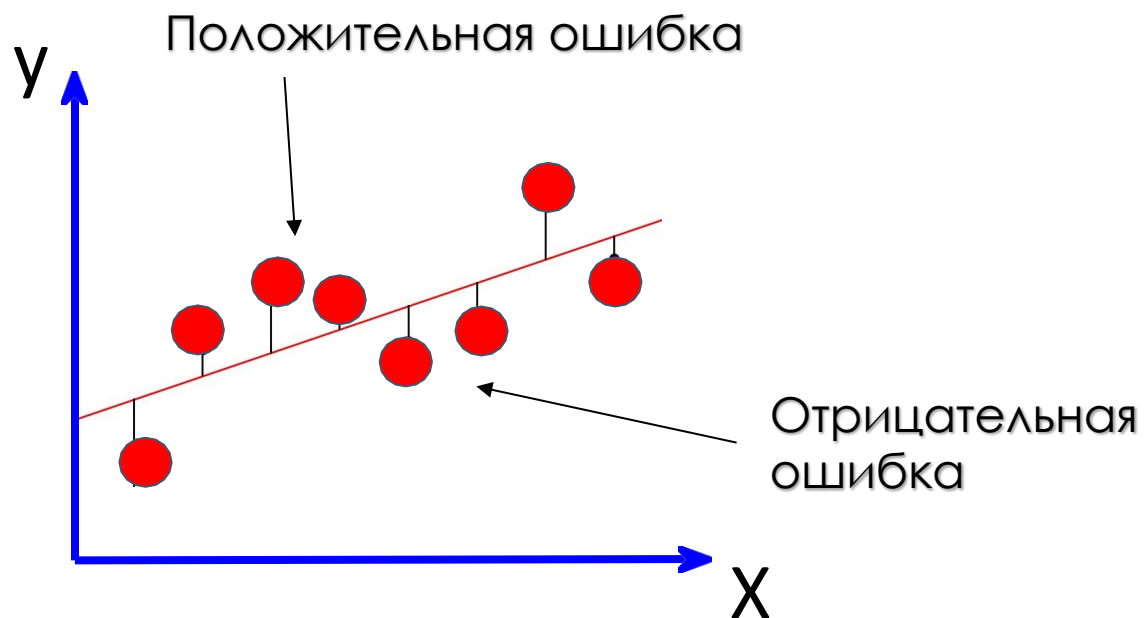
В чем проблема с такой функцией ошибок?



Метод наименьших квадратов

$$\mathcal{L}(y, \hat{y}) = \sum_{i=1}^n (y_i - \hat{y}_i) \rightarrow \min$$

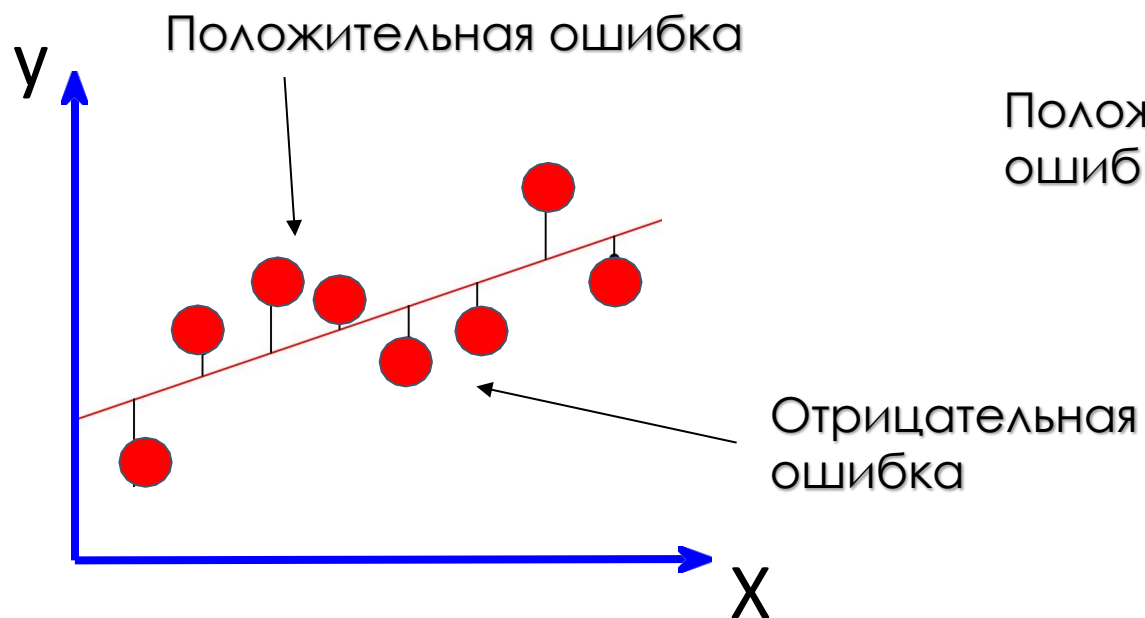
В чем проблема с такой функцией ошибок?



Метод наименьших квадратов

$$\mathcal{L}(y, \hat{y}) = \sum_{i=1}^n (y_i - \hat{y}_i) \rightarrow \min$$

В чем проблема с такой функцией ошибок?



Положительная
ошибка + Отрицательная
ошибка ≈ 0

Метод наименьших квадратов

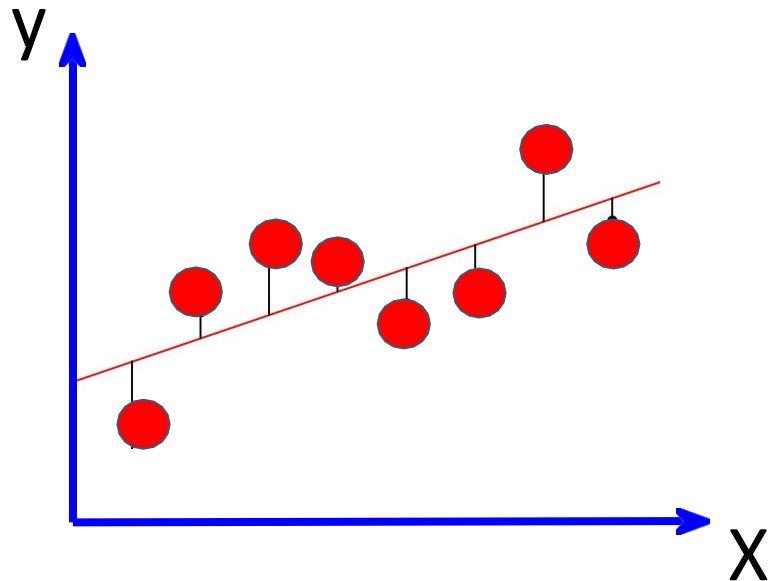
Идея: брать квадраты!



Метод наименьших квадратов

$$RSS = \mathcal{L}(y, \hat{y}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min$$

RSS - Residual Sum of Squares (сумма квадратов ошибок)



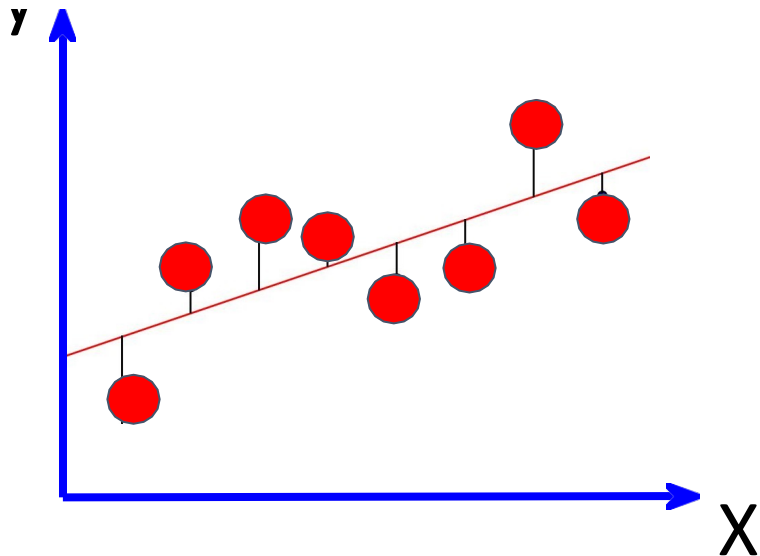
Простыми словами - это сумма квадратов разности **между точкой и значением проведенной прямой**

Метод наименьших квадратов

$$RSS = \mathcal{L}(y, \hat{y}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min$$

$$RSS = \mathcal{L}(y, \hat{y}) = \sum_{i=1}^n (y_i - (\hat{\omega}_0 + \hat{\omega}_1 x))^2 \rightarrow \min$$

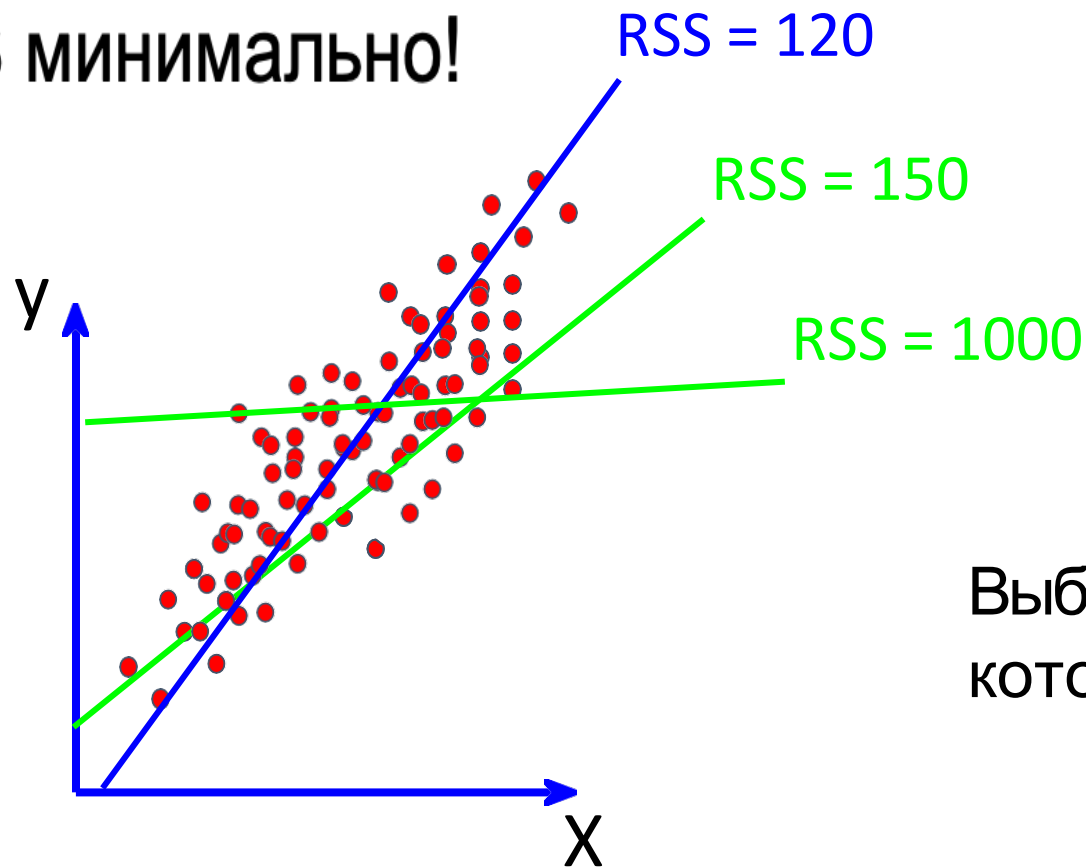
RSS - Residual Sum of Squares (сумма квадратов ошибок)



Простыми словами - это сумма квадратов разности **между точкой и значением проведенной прямой**

Метод наименьших квадратов

Выбираем $\hat{\omega}_0$ и $\hat{\omega}_1$, для которых RSS минимально!



Выбираем $\hat{\omega}_0$ и $\hat{\omega}_1$, для которых RSS минимально!

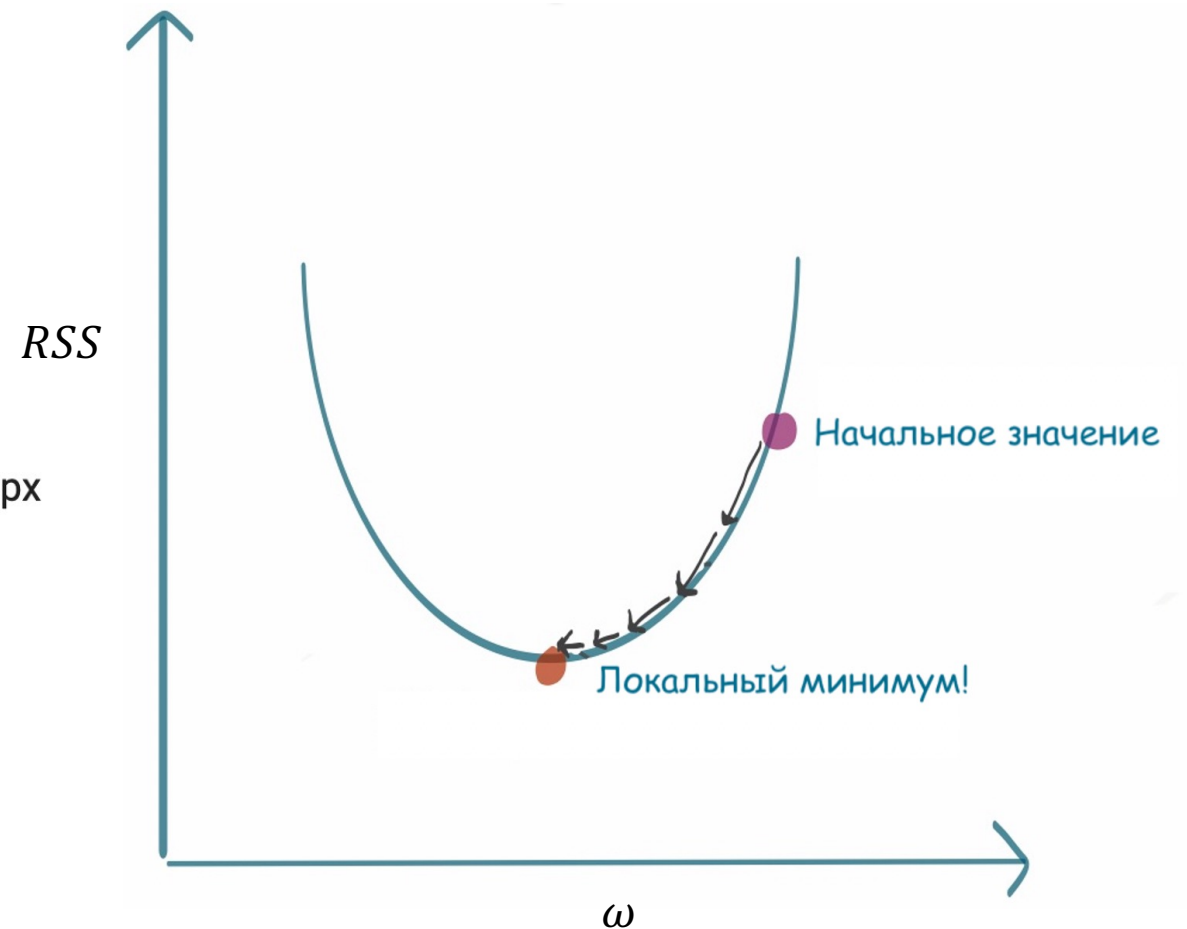
Градиентный спуск

Оптимизируем RSS градиентным спуском

$$RSS = \mathcal{L}(y, \hat{y}) = \sum_{i=1}^n (y_i - (\hat{\omega}_0 + \hat{\omega}_1 x))^2 \rightarrow \min$$

$$RSS = \mathcal{L}(y, \hat{y}) = \sum_{i=1}^n ((\hat{\omega}_0 + \hat{\omega}_1 x) - y_i)^2 \rightarrow \min$$

Квадратичная функция от ω_0 и ω_1 - парабола ветвями вверх



Градиентный спуск

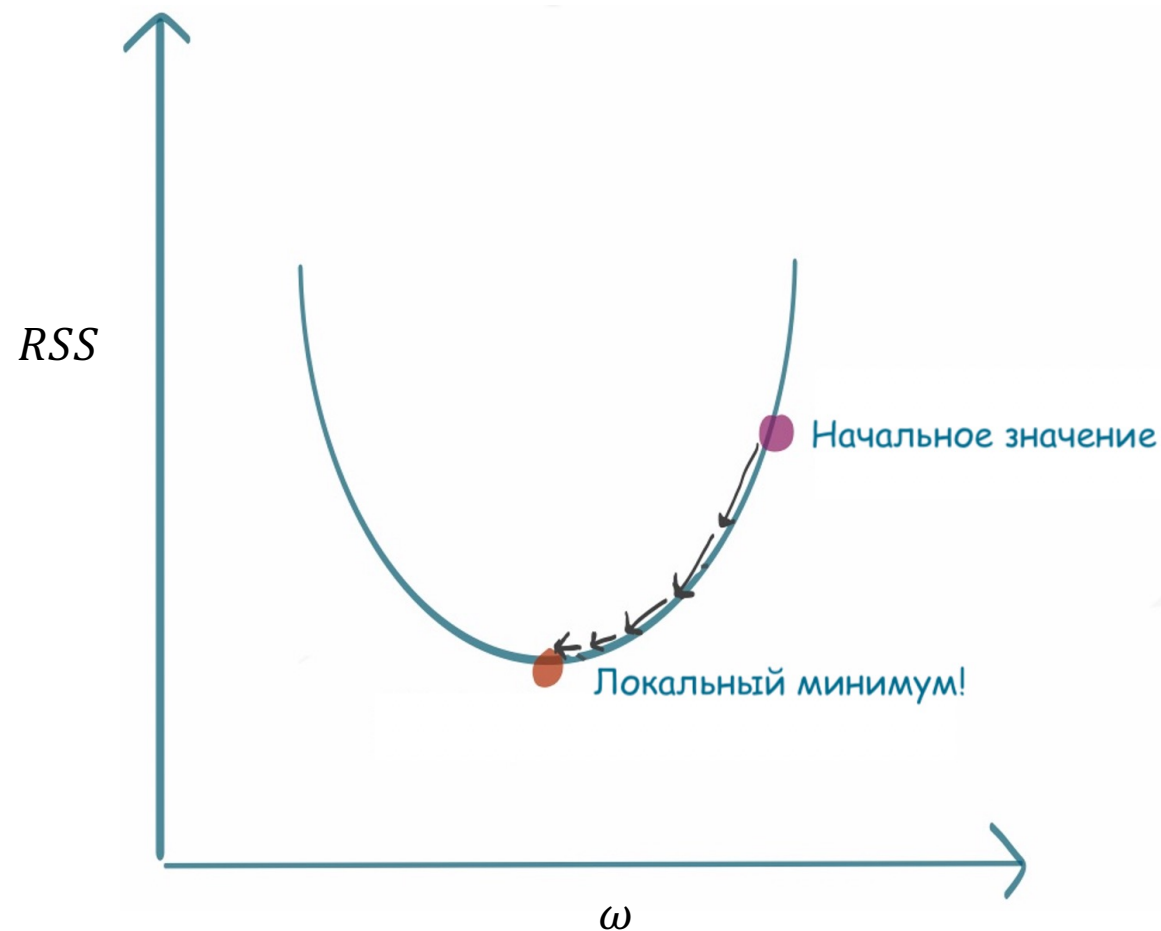
Оптимизируем RSS градиентным спуском

$$RSS = \mathcal{L}(y, \hat{y}) = \sum_{i=1}^n (y_i - (\hat{\omega}_0 + \hat{\omega}_1 x))^2 \rightarrow \min$$

$$RSS = \mathcal{L}(y, \hat{y}) = \sum_{i=1}^n ((\hat{\omega}_0 + \hat{\omega}_1 x) - y_i)^2 \rightarrow \min$$

Квадратичная функция от ω_0 и ω_1 - парабола ветвями вверх

$$w^{(t)} = w^{(t-1)} - \eta L'(w^{(t-1)})$$



Градиентный спуск

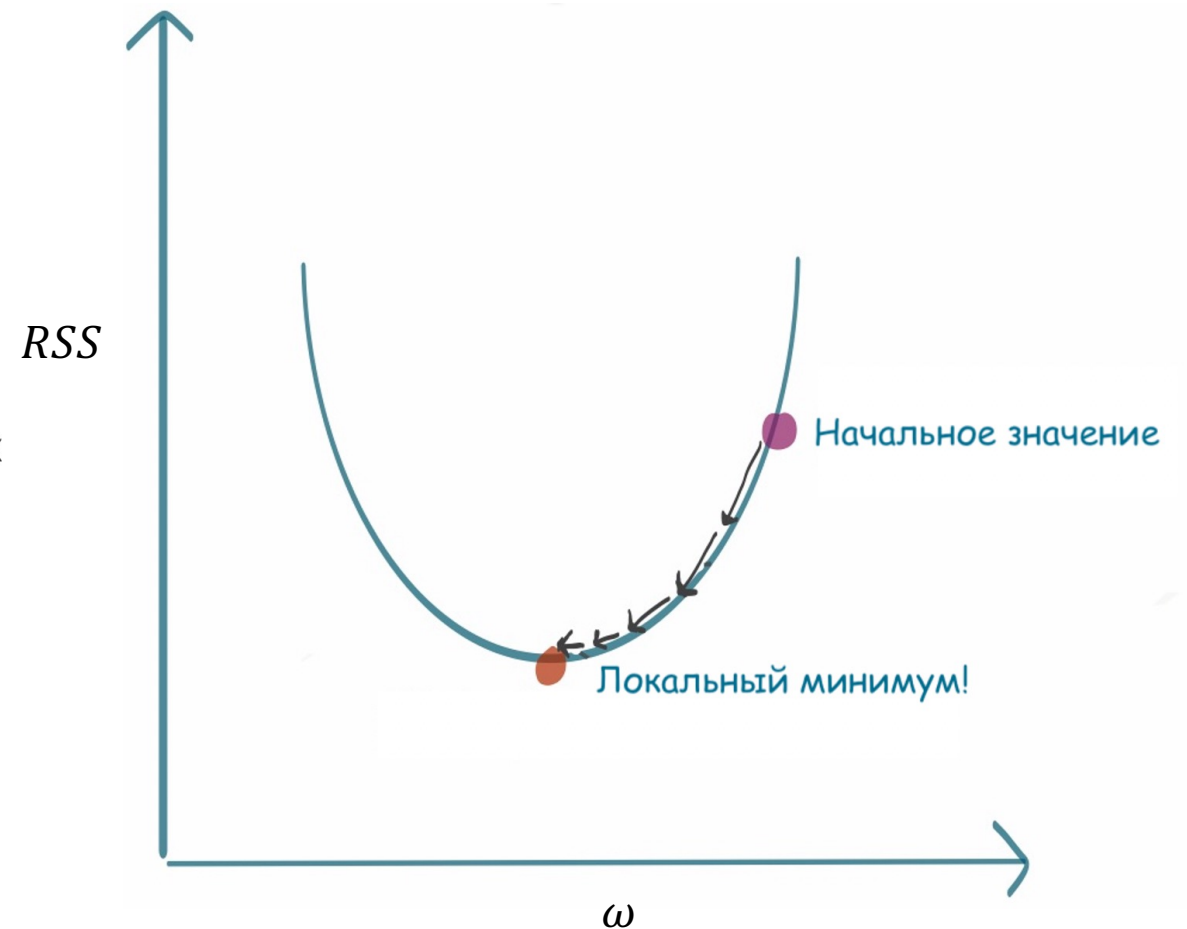
Оптимизируем RSS градиентным спуском

$$RSS = \mathcal{L}(y, \hat{y}) = \sum_{i=1}^n (y_i - (\hat{\omega}_0 + \hat{\omega}_1 x))^2 \rightarrow \min$$

$$RSS = \mathcal{L}(y, \hat{y}) = \sum_{i=1}^n ((\hat{\omega}_0 + \hat{\omega}_1 x) - y_i)^2 \rightarrow \min$$

Квадратичная функция от ω_0 и ω_1 - парабола ветвями вверх

$$w^{(t)} = w^{(t-1)} - \eta \nabla L(w^{(t-1)})$$

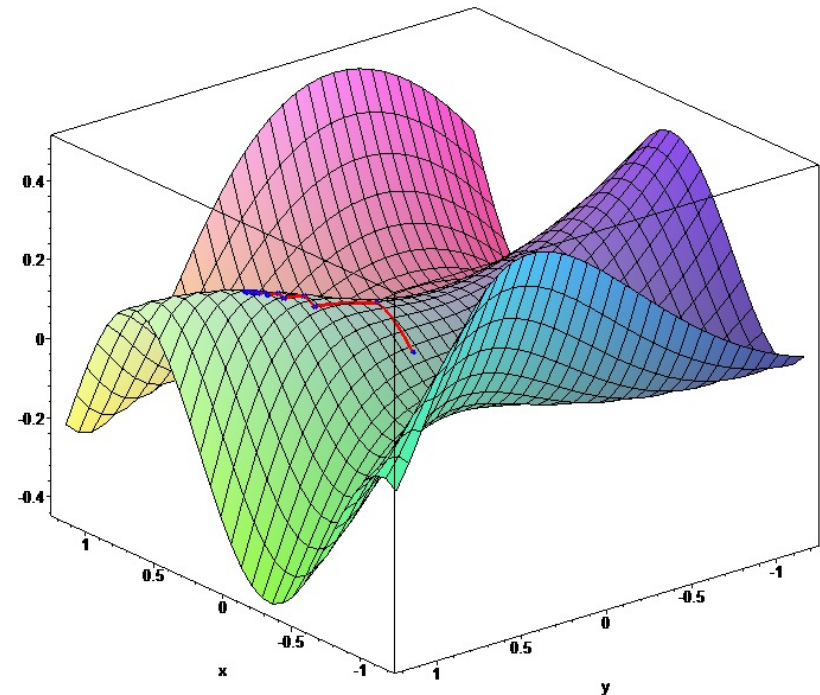
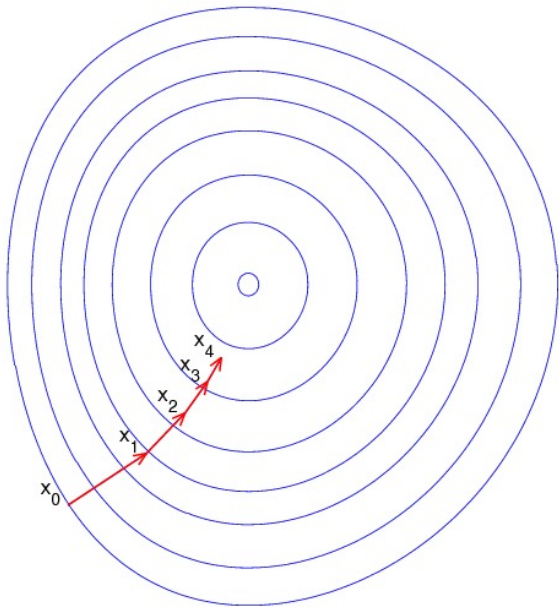


Градиентный спуск

Градиентный спуск работает и когда признаков много!

$$\hat{y} = \omega_0 + \omega_1 x_1 + \omega_2 x_2 + \dots + \omega_p x_p = x^T \omega$$

$$w^{(t)} = w^{(t-1)} - \eta \nabla L(w^{(t-1)})$$



The background of the slide is a high-angle, aerial photograph of a city skyline, likely New York City, featuring numerous skyscrapers. The image is tinted with a blue and green color palette. A semi-transparent network of white lines and dots is overlaid on the image, particularly concentrated in the central area where the text is located. The text "Метрики качества" is written in a large, white, sans-serif font, centered horizontally and partially overlaid by the network pattern.

Метрики качества

Метрики качества

Средняя абсолютная ошибка

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

Средняя квадратичная ошибка

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$

Корень из средней квадратичной ошибки

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

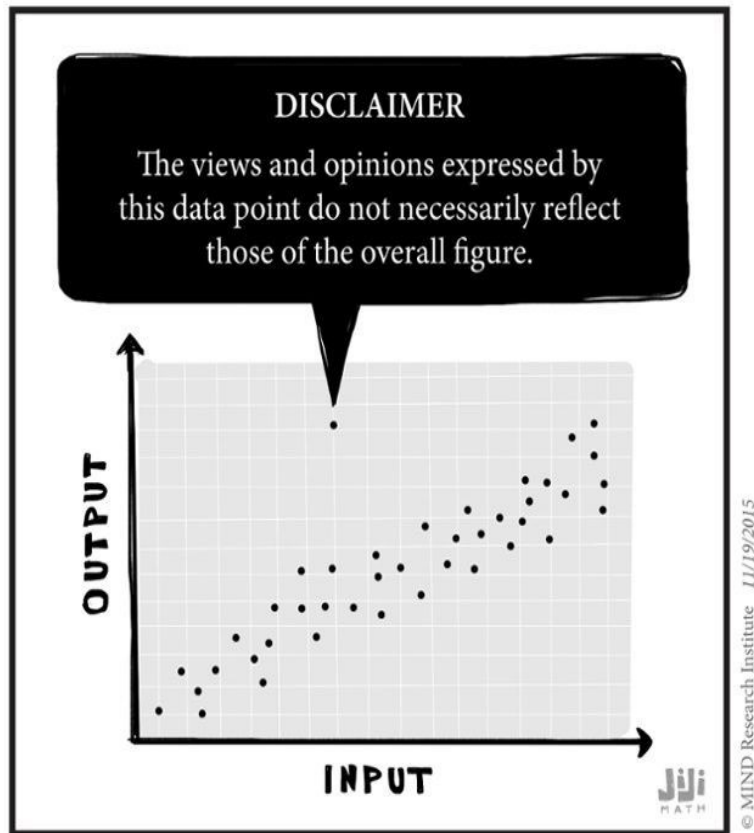




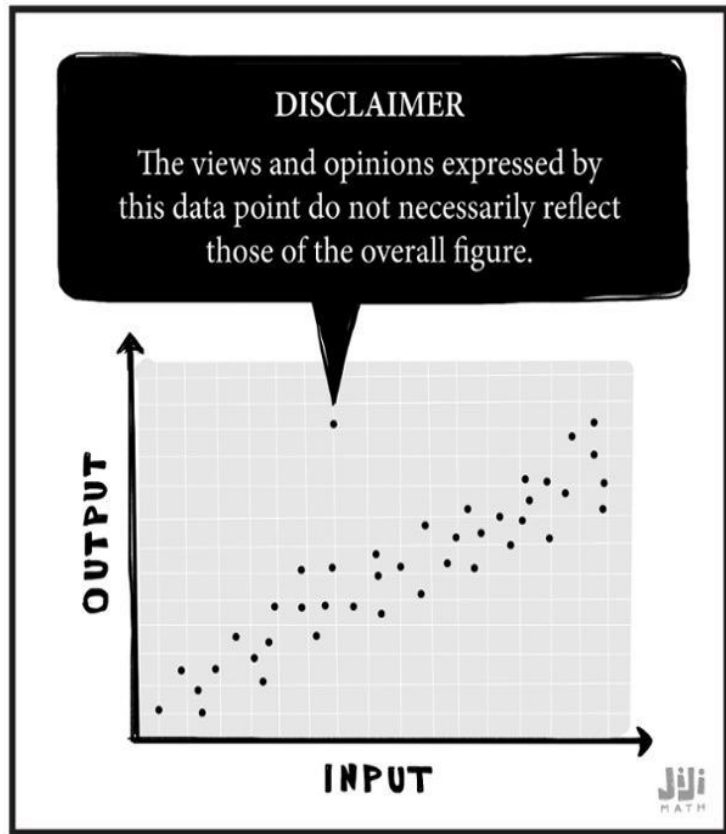
Регуляризация

Регуляризация

Проблема: выбросы и переобучение



Регуляризация



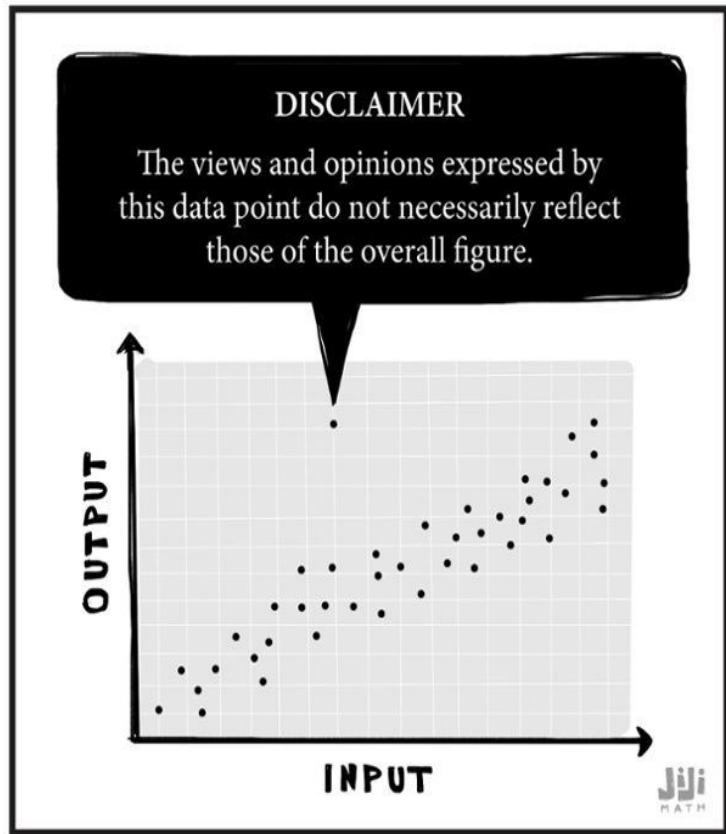
Проблема: выбросы и переобучение

Идея: ввести штраф за излишнюю сложность

Регуляризация – ограничение на норму вектора весов

$$\sum_{n=1}^N (x_n^T \omega - y_n)^2 + \lambda R(\omega) \rightarrow \min_{\beta}$$

Регуляризация



Регуляризация – ограничение на норму вектора весов

$$\sum_{n=1}^N (x_n^T \omega - y_n)^2 + \lambda R(\omega) \rightarrow \min_{\beta}$$

$$R(\omega) = ||\omega||_1 \quad (\text{L1}) \text{ Lasso regression}$$

$$R(\omega) = ||\omega||_2^2 \quad (\text{L2}) \text{ Ridge regression}$$

