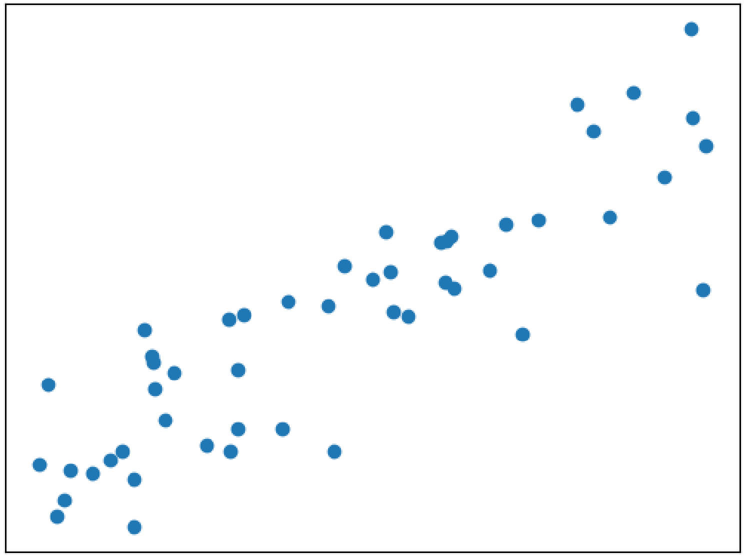
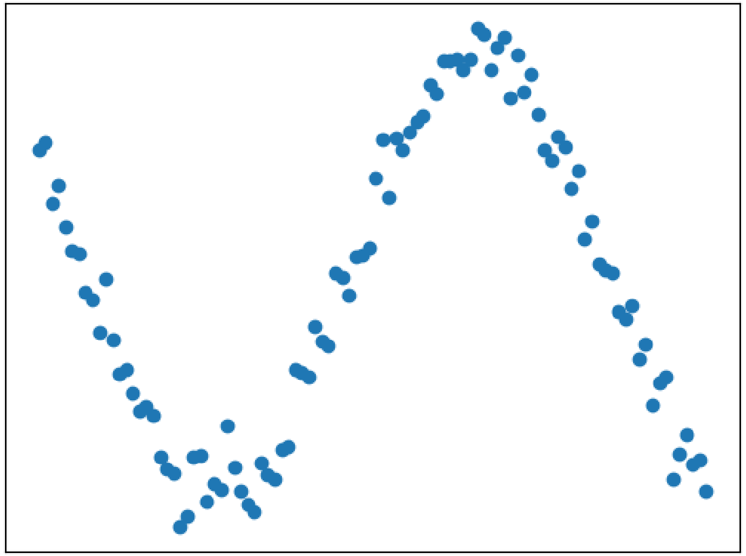
Машинное обучение

Полиномиальная регрессия и переобучение

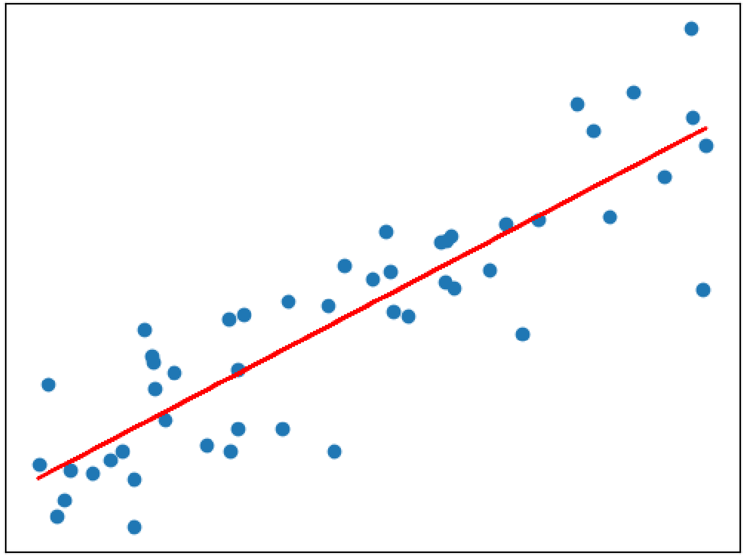
Нелинейная задача

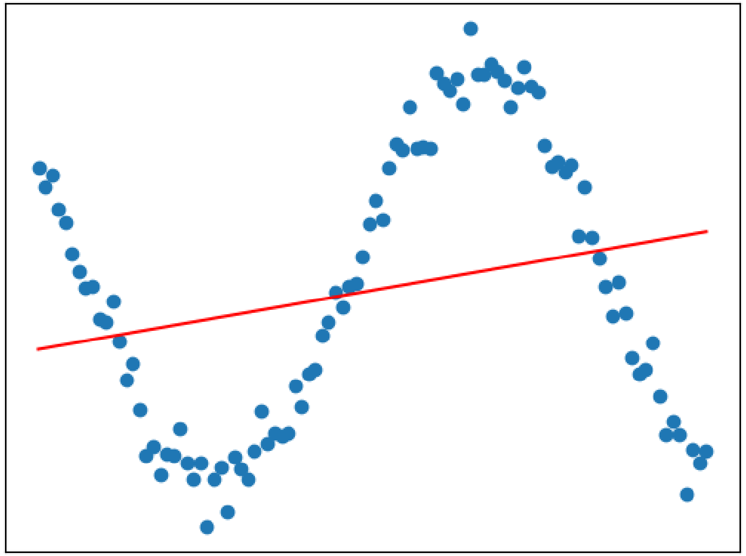




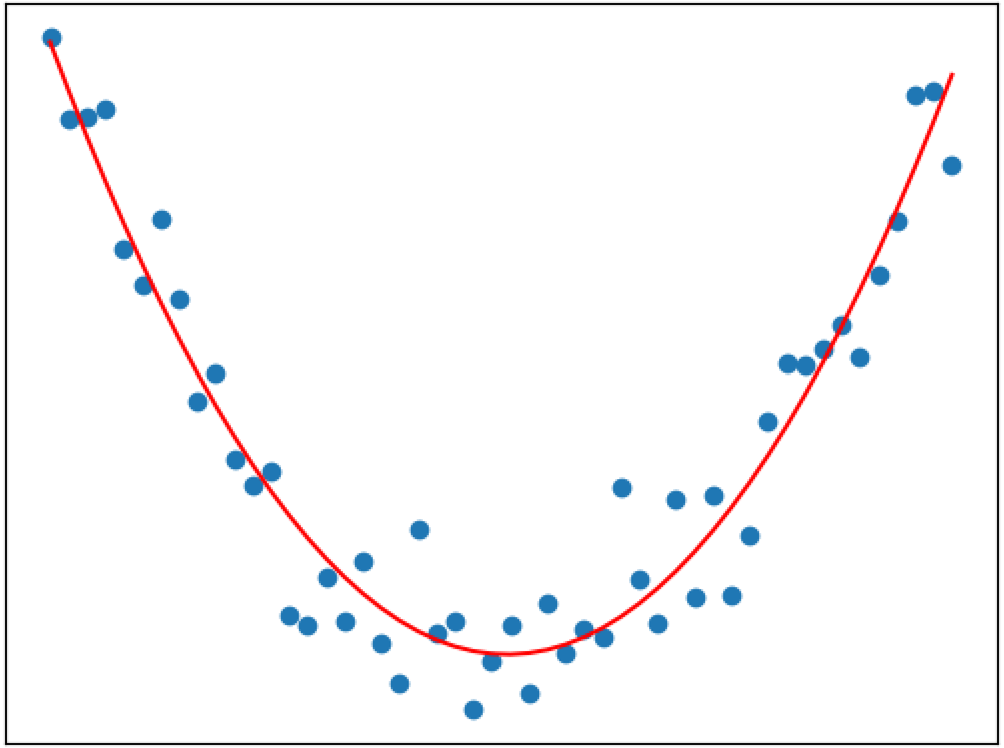
Нелинейная задача

$$a(x) = \theta_0 + \theta_1 x_1$$

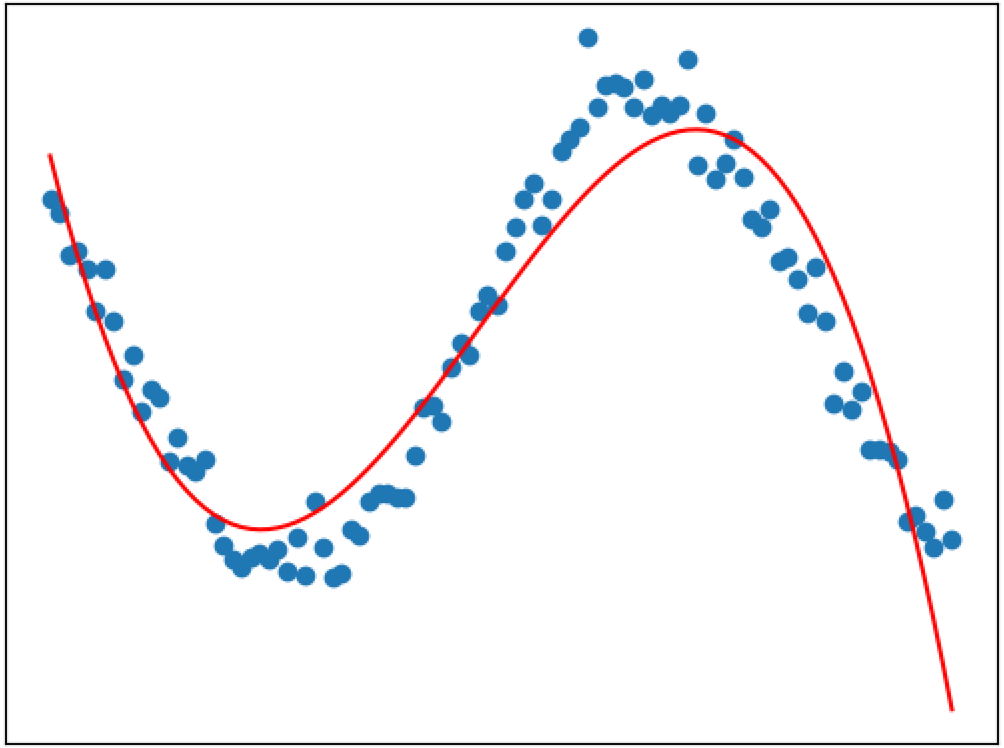




Полиномиальная регрессия



$$a(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2$$



$$a(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + + \theta_3 x_1^3$$

Полиномиальная регрессия

Для регрессии с двумя признаками

Линейная модель (полином степени 1)

$$\boxed{\boldsymbol{a(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2}}$$

Квадратичная модель (полином степени 2)

$$\boldsymbol{a(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2}$$

Кубическая модель (полином степени 3)

$$\boldsymbol{a_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2 + \theta_6 x_1^3 + \theta_7 x_2^3 + \theta_8 x_1^2 x_2 + \theta_8 x_1 x_2^2}$$

Полиномиальная регрессия

Для регрессии с двумя признаками

Линейная модель (полином степени 1)

$$a(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

Квадратичная модель (полином степени 2)

$$\boxed{a(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2}$$

Кубическая модель (полином степени 3)

$$a_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2 + \theta_6 x_1^3 + \theta_7 x_2^3 + \theta_8 x_1^2 x_2 + \theta_8 x_1 x_2^2$$

Полиномиальная регрессия

Для регрессии с двумя признаками

Линейная модель (полином степени 1)

$$a(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

Квадратичная модель (полином степени 2)

$$a(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2$$

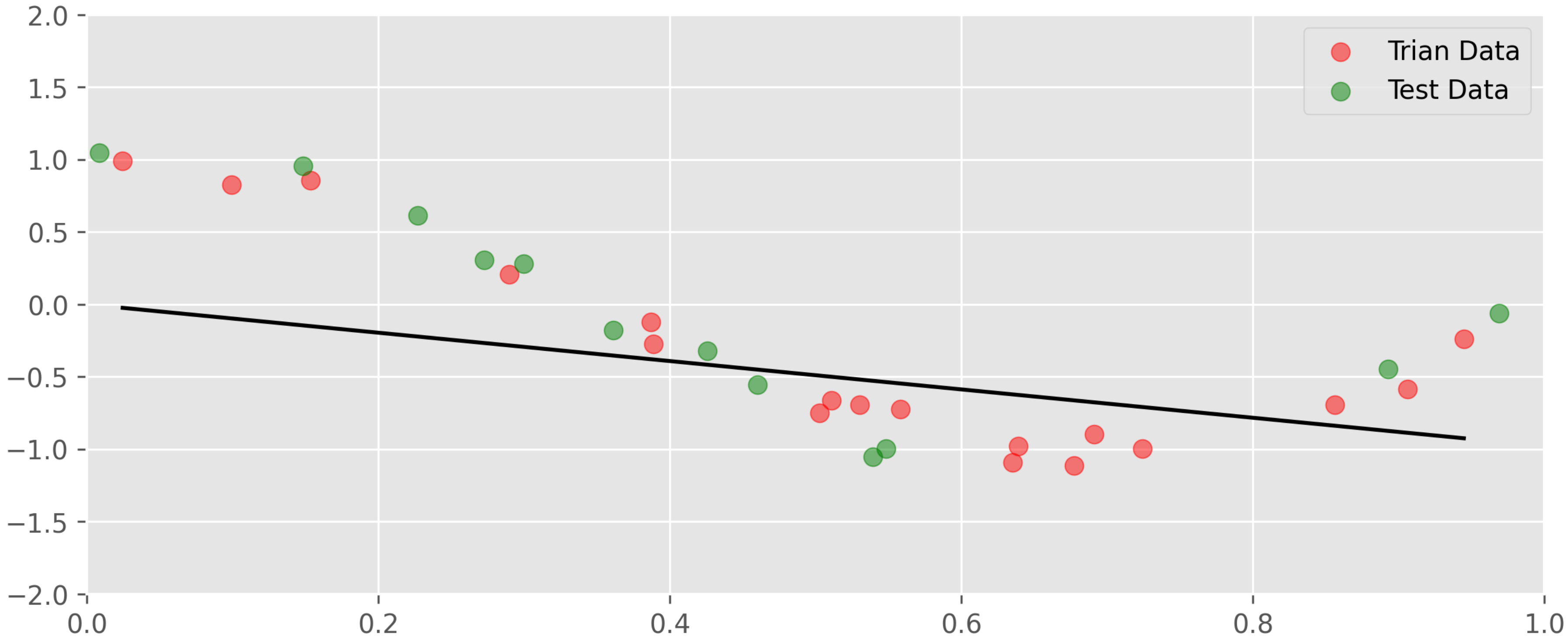
Кубическая модель (полином степени 3)

$$\boxed{a_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2 + \theta_6 x_1^3 + \theta_7 x_2^3 + \theta_8 x_1^2 x_2 + \theta_8 x_1 x_2^2}$$

Переобучение

полином степени 1

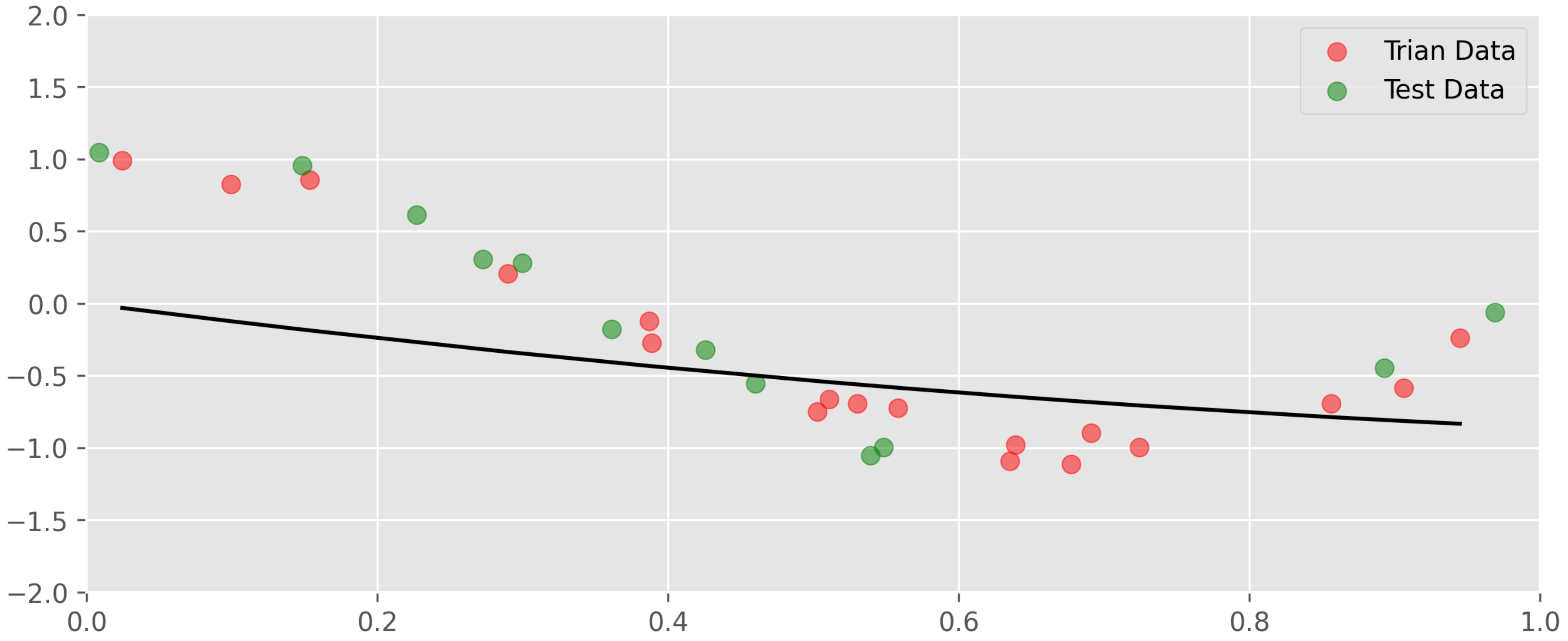
$$\boldsymbol{\theta} = [\mathbf{0.98}]$$



Переобучение

полином степени 2

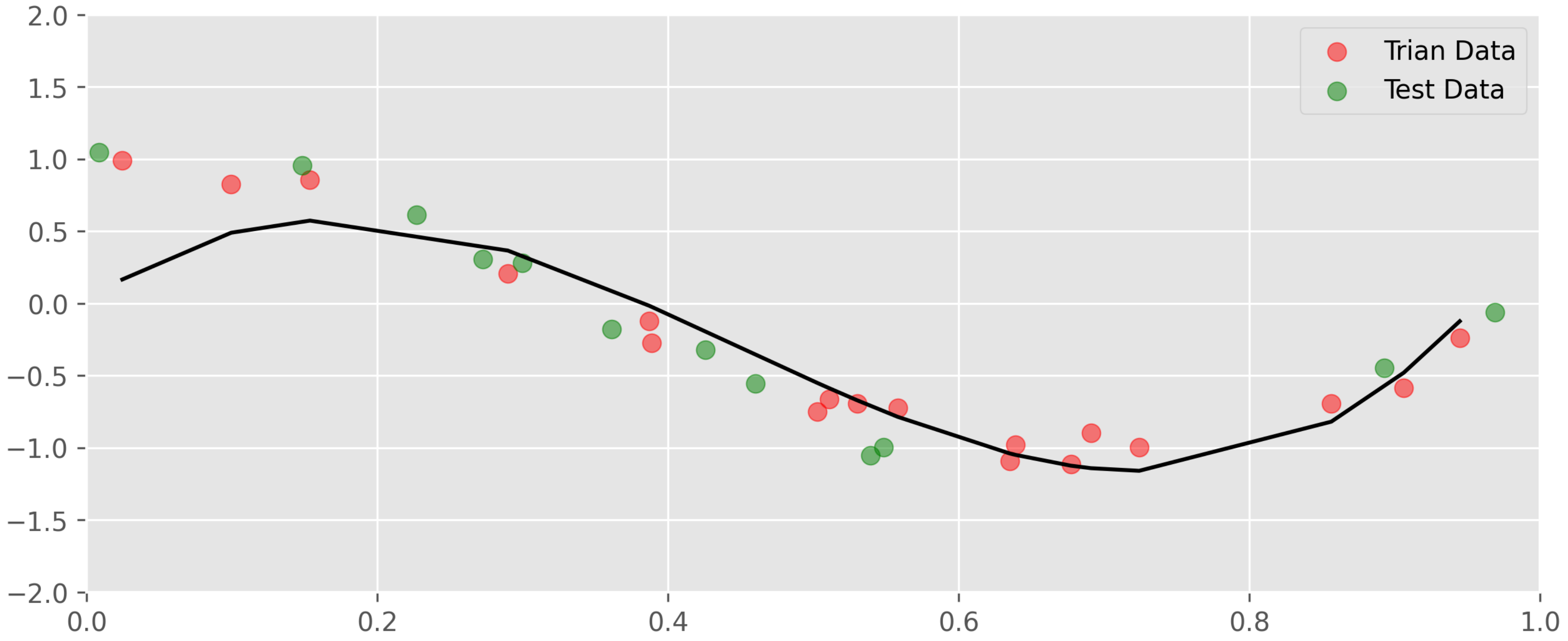
$$\boldsymbol{\theta} = [\mathbf{-1.28, 0.42}]$$



Переобучение

полином степени 3

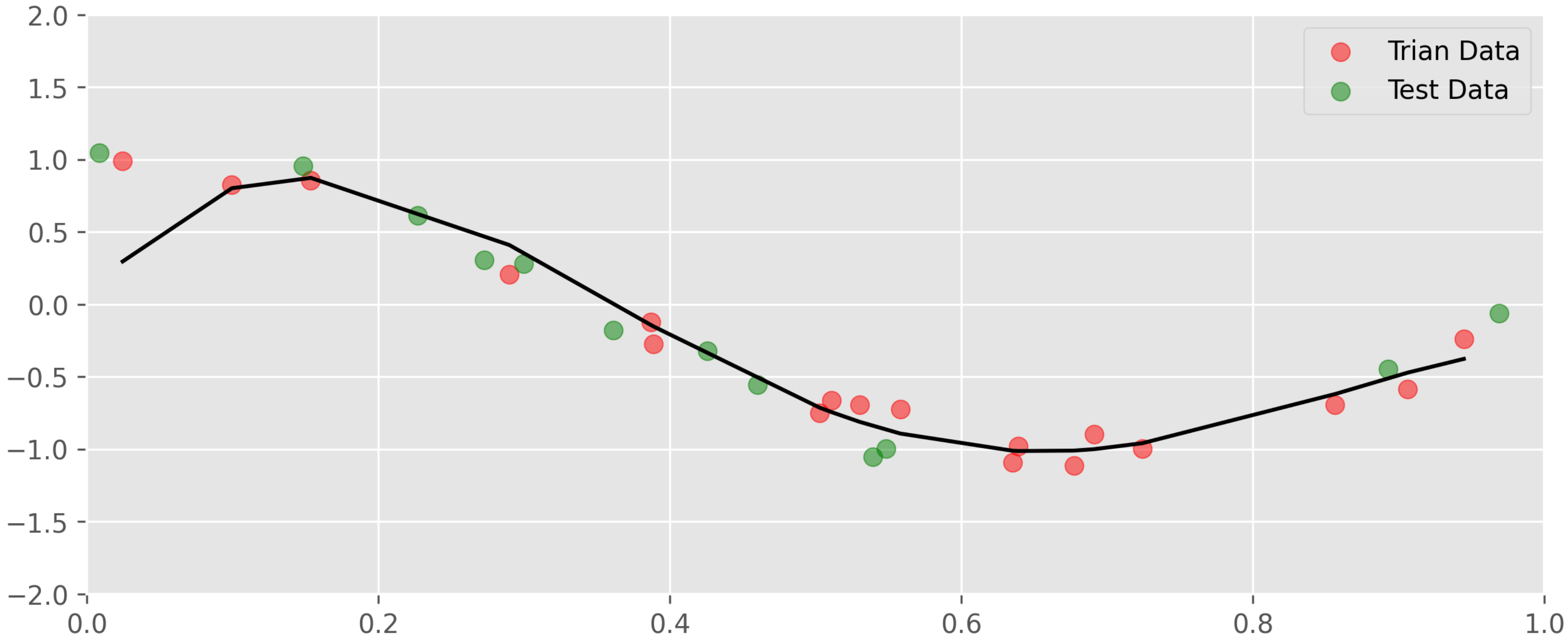
$$\boldsymbol{\theta} = [\mathbf{7.42}, \mathbf{-27.13}, \mathbf{20.25}]$$



Переобучение

полином степени 4

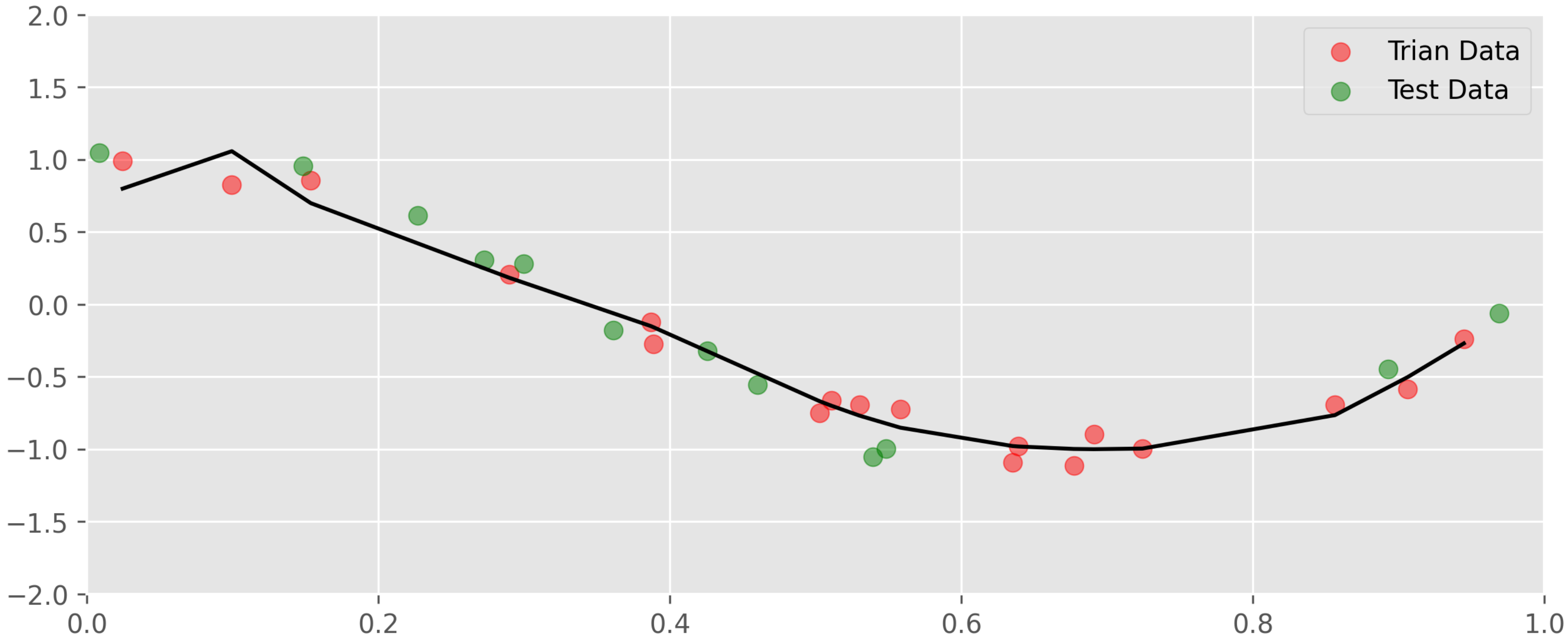
$$\boldsymbol{\theta = [13.6, -63.74, 85.1, -35.26]}$$



Переобучение

полином степени 8

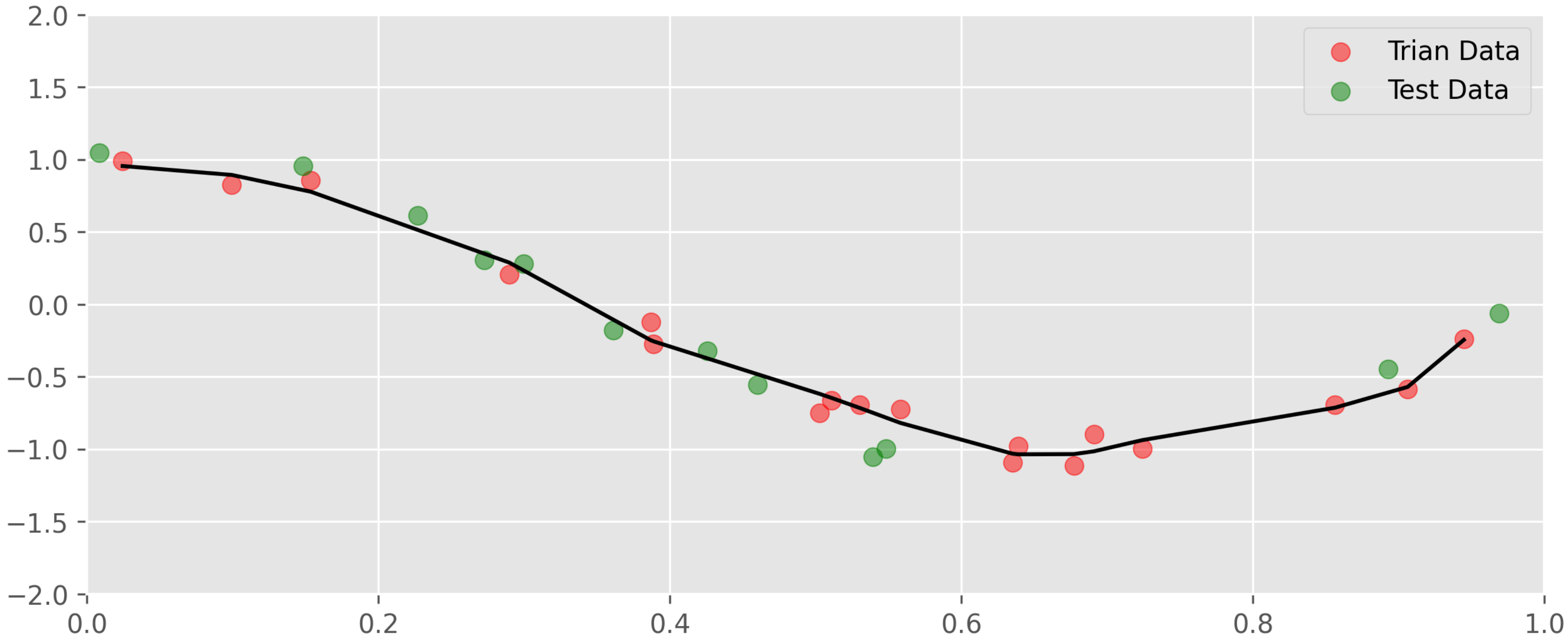
$$\boldsymbol{\theta} = [\mathbf{45.78}, -\mathbf{621.25}, \mathbf{3675.67}, -\mathbf{11844.29}, \mathbf{21978.15}, -\mathbf{23449.19}, \mathbf{13375.74}, -\mathbf{3160.86}]$$



Переобучение

полином степени 10

$$\boldsymbol{\theta} = [\mathbf{6.64}\boldsymbol{e} + \mathbf{01}, -\mathbf{1.43}\boldsymbol{e} + \mathbf{03}, \mathbf{1.44}\boldsymbol{e} + \mathbf{04}, -\mathbf{8.18}\boldsymbol{e} + \mathbf{04}\ \ \mathbf{2.80}\boldsymbol{e} + \mathbf{05}, -\mathbf{5.99}\boldsymbol{e} + \mathbf{05}, \mathbf{8.09}\boldsymbol{e} + \mathbf{05}, -\mathbf{6.69}\boldsymbol{e} + \mathbf{05}, \mathbf{3.09}\boldsymbol{e} + \mathbf{05}, -\mathbf{6.10}\boldsymbol{e} + \mathbf{04}]$$



Переобучение

$$a(x) = 0.5x_0 + 2397432x_1 - 83292139x_2 + \cdots$$

Эмпирическое наблюдение

Большие коэффициенты

Переобучение

$$\boldsymbol{\theta} = [\mathbf{0.98}]$$

$$\boldsymbol{\theta} = [\mathbf{-1.28, 0.42}]$$

$$\boldsymbol{\theta} = [\mathbf{7.42, -27.13, 20.25}]$$

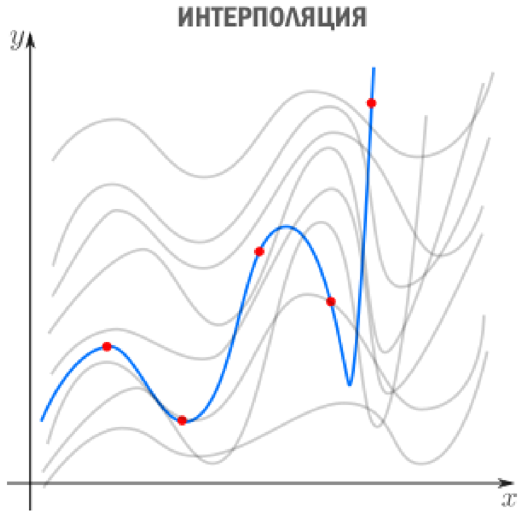
$$\boldsymbol{\theta} = [\mathbf{13.6, -63.74, 85.1, -35.26}]$$

$$\boldsymbol{\theta} = [\mathbf{45.78, -621.25, 3675.67, -11844.29, 21978.15, -23449.19, 13375.74, -3160.86}]$$

$$\boldsymbol{\theta} = [\mathbf{6.64}\boldsymbol{e} + \mathbf{01, -1.43}\boldsymbol{e} + \mathbf{03, 1.44}\boldsymbol{e} + \mathbf{04, -8.18}\boldsymbol{e} + \mathbf{04\ \ 2.80}\boldsymbol{e} + \mathbf{05, -5.99}\boldsymbol{e} + \mathbf{05, 8.09}\boldsymbol{e} + \mathbf{05, -6.69}\boldsymbol{e} + \mathbf{05,}$$
$$\mathbf{3.09}\boldsymbol{e} + \mathbf{05, -6.10}\boldsymbol{e} + \mathbf{04}]$$

Полиномиальная регрессия

Интерполяция — Способ выбрать из семейства функций ту, которая проходит через заданные точки. предсказание поведения функции вне интервала



Регрессия — Способ выбрать из семейства функций ту, которая минимизирует функцию потерь. Последняя характеризует насколько сильно пробная функция отклоняется от значений в заданных точках.

https://habr.com/ru/articles/514818/

Регуляризация

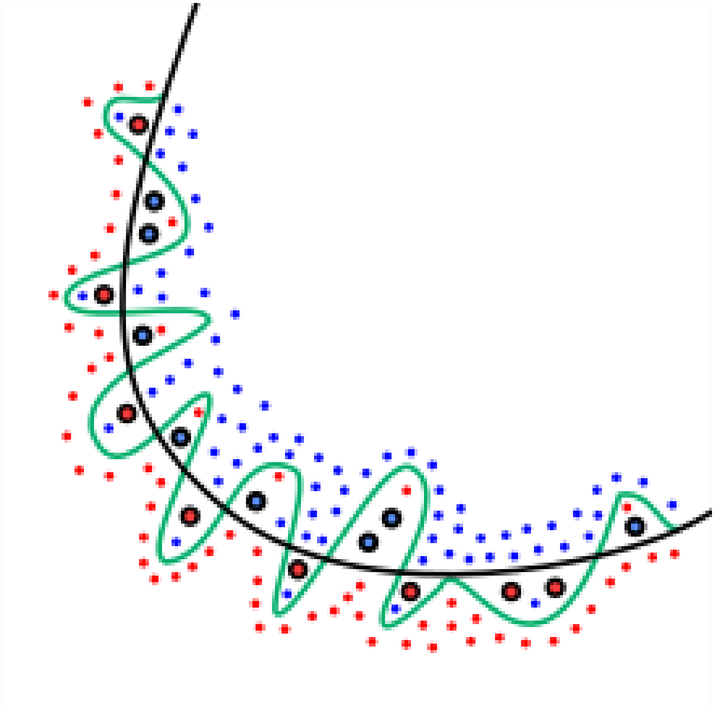
Регуляризация

Симптом переобучения

Большие коэффициенты

Как бороться с переобучением

Будем штрафовать за большие коэффиценты!



Регуляризация L2 (Ridge)

$$\mathrm{L} : (\mathrm{a}(x_i) - y_i)^2$$

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

+

$$\|\theta\|_2^2 = \theta_1^2 + \theta_2^2 + \cdots + \theta_n^2$$

$$L2 = \lambda\|\theta\|^2$$

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

Регуляризация L2 (Ridge)

$$\mathrm{L}: (\mathrm{a}(x_i) - y_i)^2$$

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

+

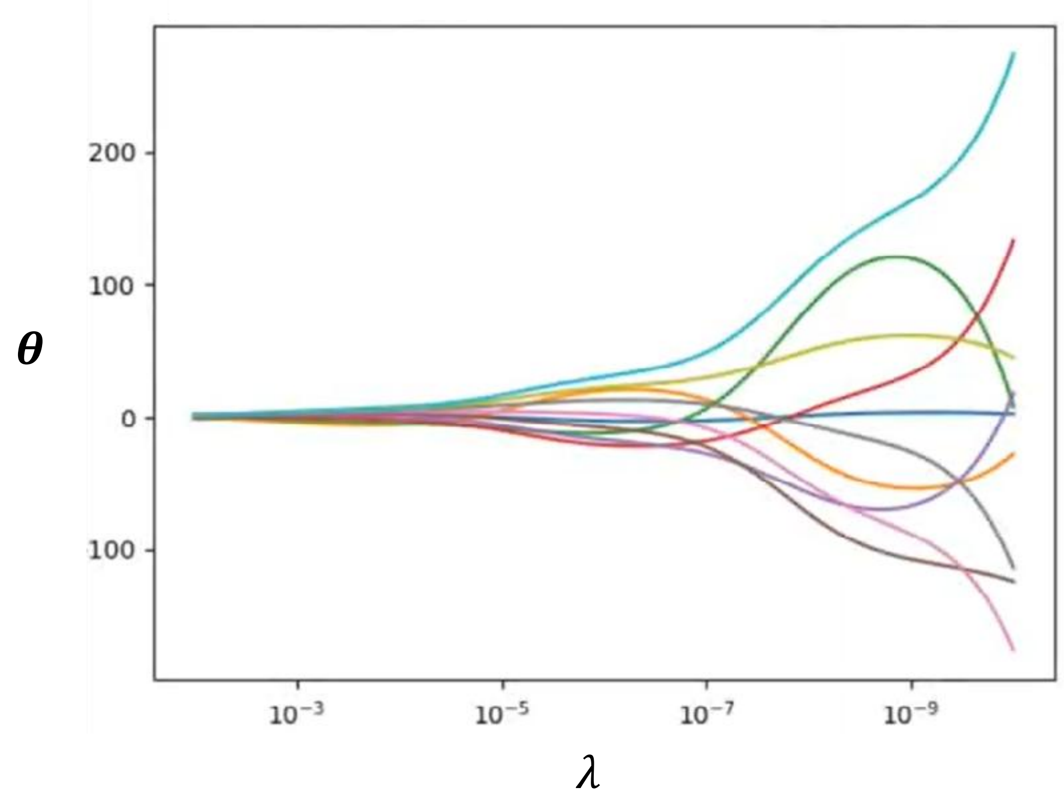
$$\|\theta\|_2^2 = \theta_1^2 + \theta_2^2 + \cdots + \theta_n^2$$

$$L2 = \lambda \|\theta\|^2$$

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell} \sum_{i=1}^{\ell} (a(x_i) - y_i)^2 + \boxed{\lambda \|\theta\|_2} \to \min_{\boldsymbol{\theta}}$$

Регуляризация L2 (Ridge)

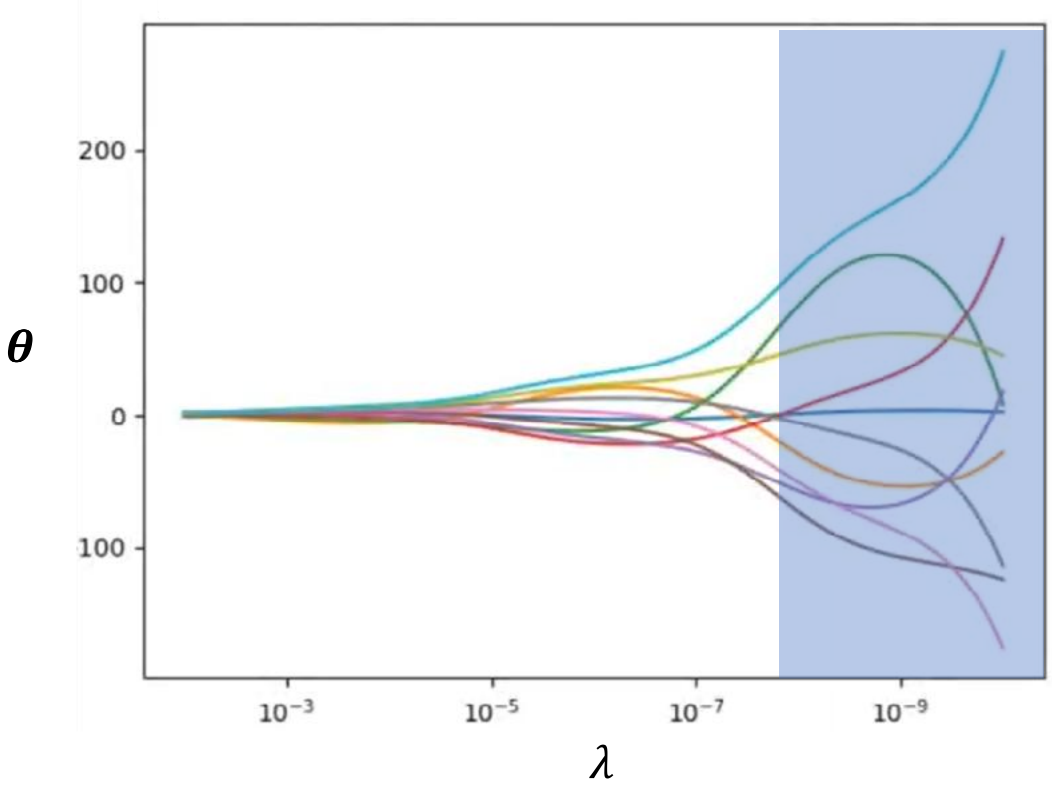
$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell} \sum_{i=1}^{\ell} (a(x_i) - y_i)^2 + \lambda \|\theta\|_2 \to \min_{\boldsymbol{\theta}}$$



Регуляризация L2 (Ridge)

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 + \lambda\|\theta\|_2 \to \min_{\boldsymbol{\theta}}$$

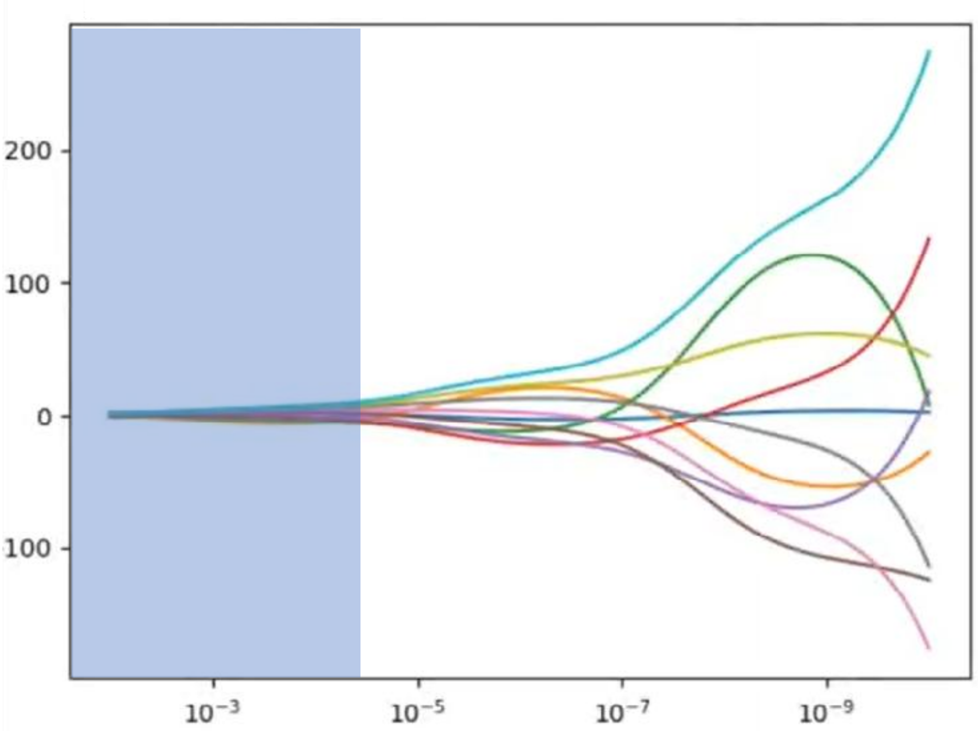
Если λ очень маленькая,
то веса будут большими



Регуляризация L2 (Ridge)

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 + \lambda\|\theta\|_2 \to \min_{\boldsymbol{\theta}}$$

Если λ очень большая, то веса будут маленькими

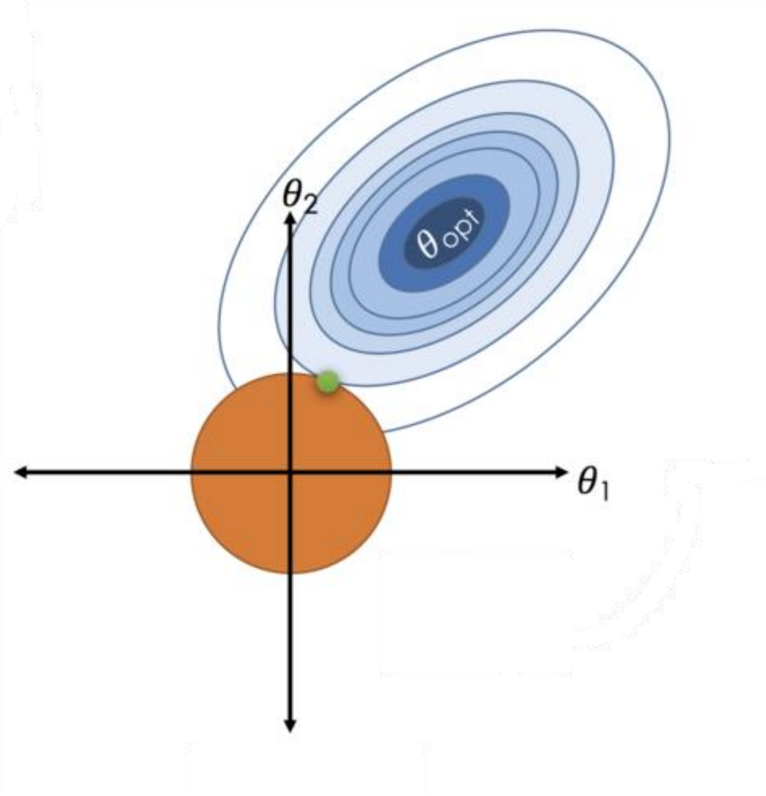


$\boldsymbol{\theta}$

λ

Регуляризация L2 (Ridge)

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 + \lambda\|\theta\|_2 \to \min_{\boldsymbol{\theta}}$$



Регуляризация L1 (Lasso)

$$\mathrm{L} : (a(x^{(i)}) - y^{(i)})^2$$

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

+

$$\|\theta\|_1 = |\theta_1| + |\theta_2| + \ldots + |\theta_n|$$

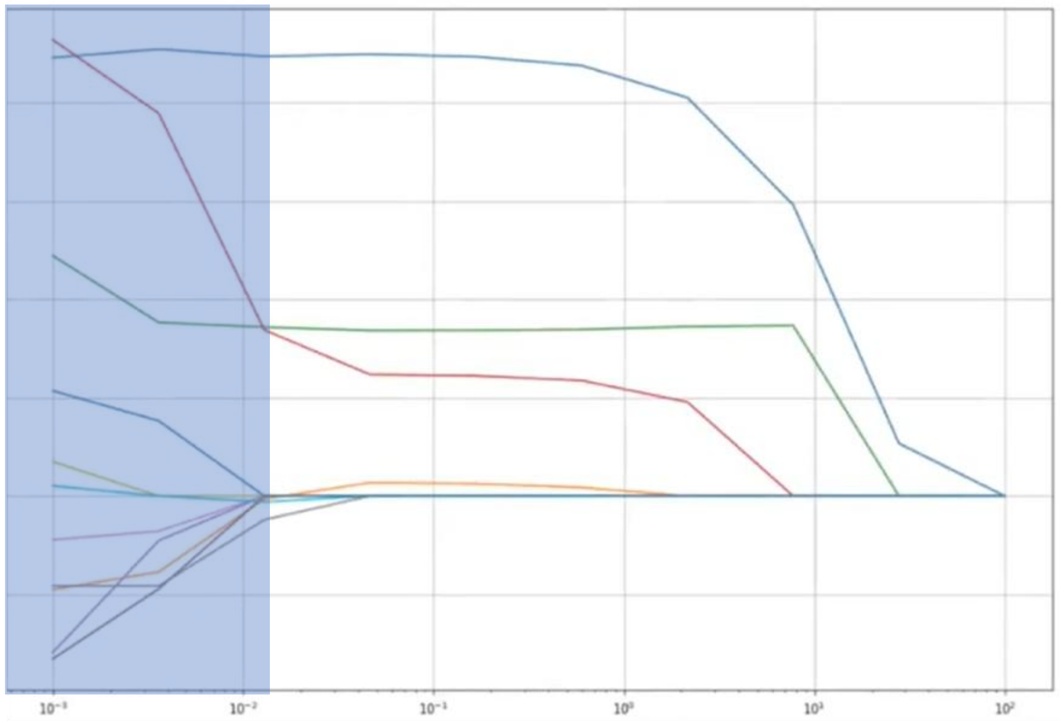
$$L1 = \lambda\|\theta\|_1$$

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 + \lambda\|\theta\|_1 \rightarrow \min_{\boldsymbol{\theta}}$$

Регуляризация L1 (Lasso)

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 + \lambda\|\theta\|_1 \to \min_{\boldsymbol{\theta}}$$

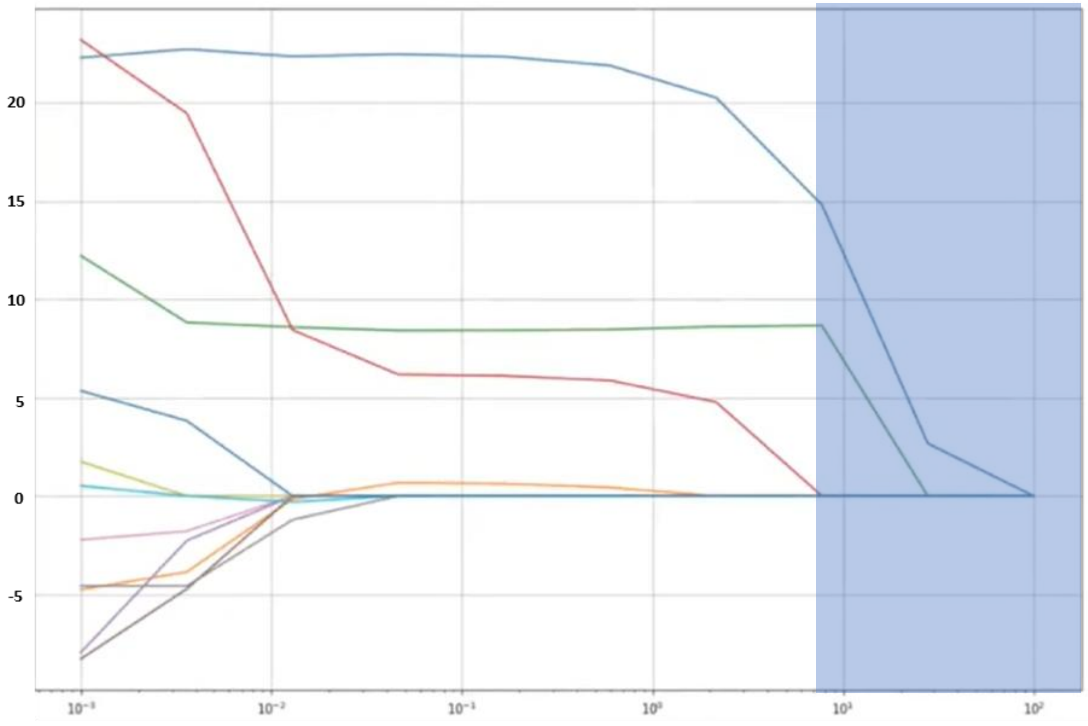
Если λ маленькая, то веса будут большими



Регуляризация L1 (Lasso)

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell} \sum_{i=1}^{\ell} (a(x_i) - y_i)^2 + \lambda \|\theta\|_1 \to \min_{\boldsymbol{\theta}}$$

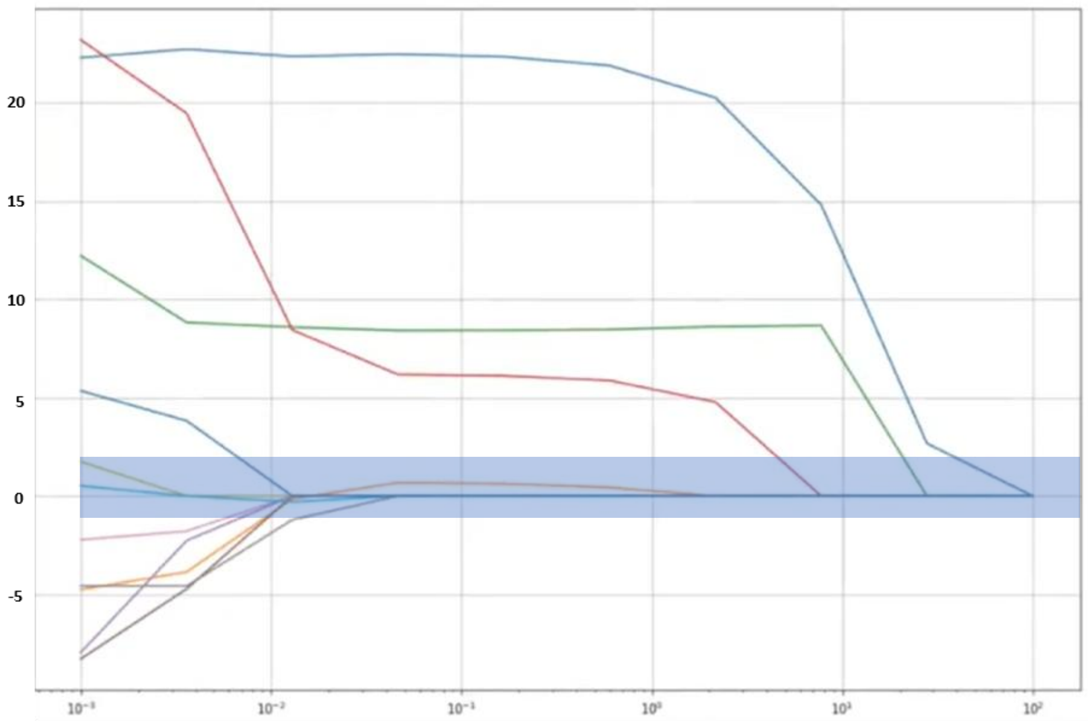
Если λ большая, то веса будут маленькие



Регуляризация L1 (Lasso)

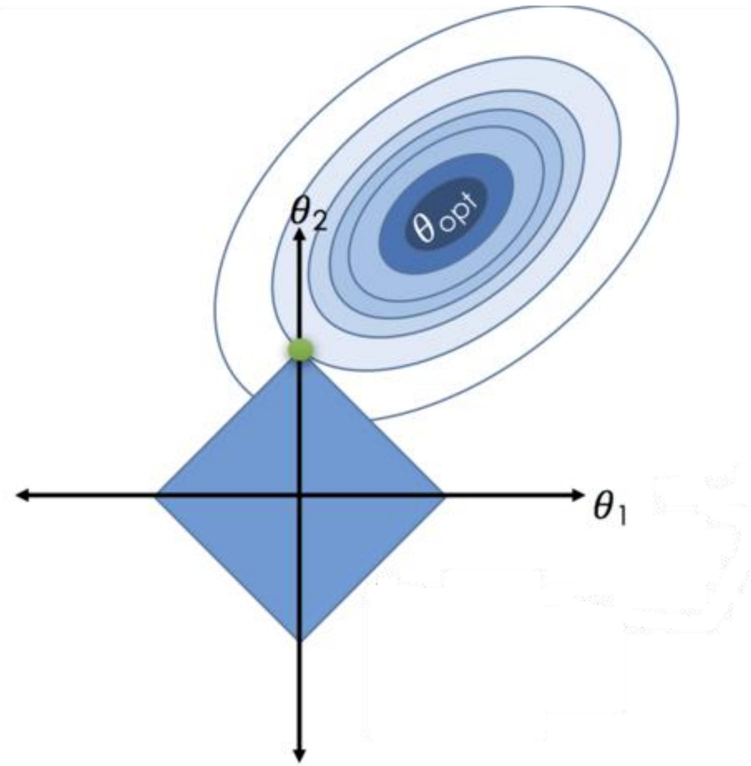
$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 + \lambda\|\theta\|_1 \to \min_{\boldsymbol{\theta}}$$

Если λ большая, то веса будут маленькими, а могут быть нулевыми



Регуляризация L1 (Lasso)

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 + \lambda\|\theta\|_1 \to \min_{\boldsymbol{\theta}}$$



Регуляризация L1+L2 (Elastic Net)

$$\mathrm{L} : (a_\theta(x^{(i)}) - y^{(i)})^2$$

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

+

$$\|\theta\|_2^2 = \theta_1^2 + \theta_2^2 + \cdots + \theta_n^2$$

$$\|\theta\|_1 = |\theta_1| + |\theta_2| + \ldots + |\theta_n|$$

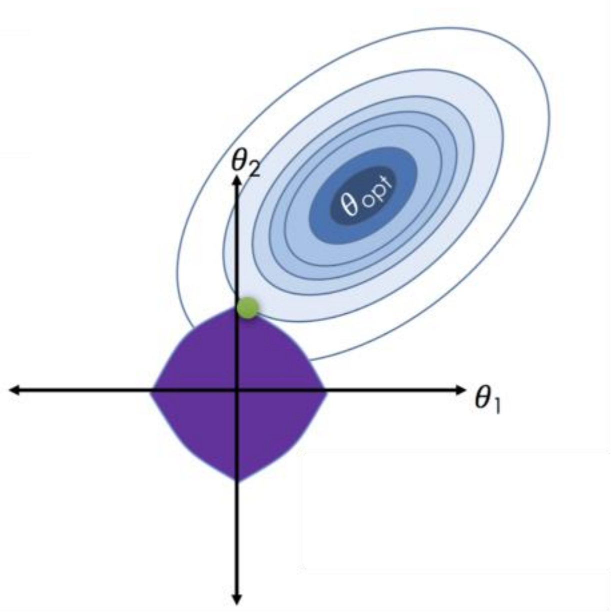
$$L1 = \lambda_1\|\theta\|_1 \qquad L2 = \lambda_2\|\theta\|^2$$

Elastic Net= L1+L2

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 + \lambda_1\|\theta\|_1 + \lambda_2\|\theta\|_2 \rightarrow \min_{\boldsymbol{\theta}}$$

Регуляризация L1+L2 (Elastic Net)

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 + \lambda_1\|\theta\|_1 + \lambda_2\|\theta\|_2 \to \min_{\boldsymbol{\theta}}$$



$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 + 0\,\|\theta\|_1 + \lambda_2\|\theta\|_2 \to \min_{\boldsymbol{\theta}}$$

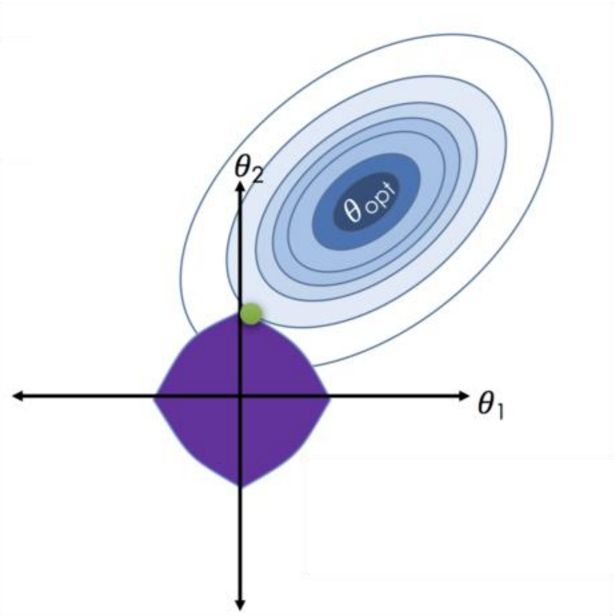
Регуляризация L2 (Ridge)

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 + \lambda_1\|\theta\|_1 + 0\,\|\theta\|_2 \to \min_{\boldsymbol{\theta}}$$

Регуляризация L1 (Lasso)

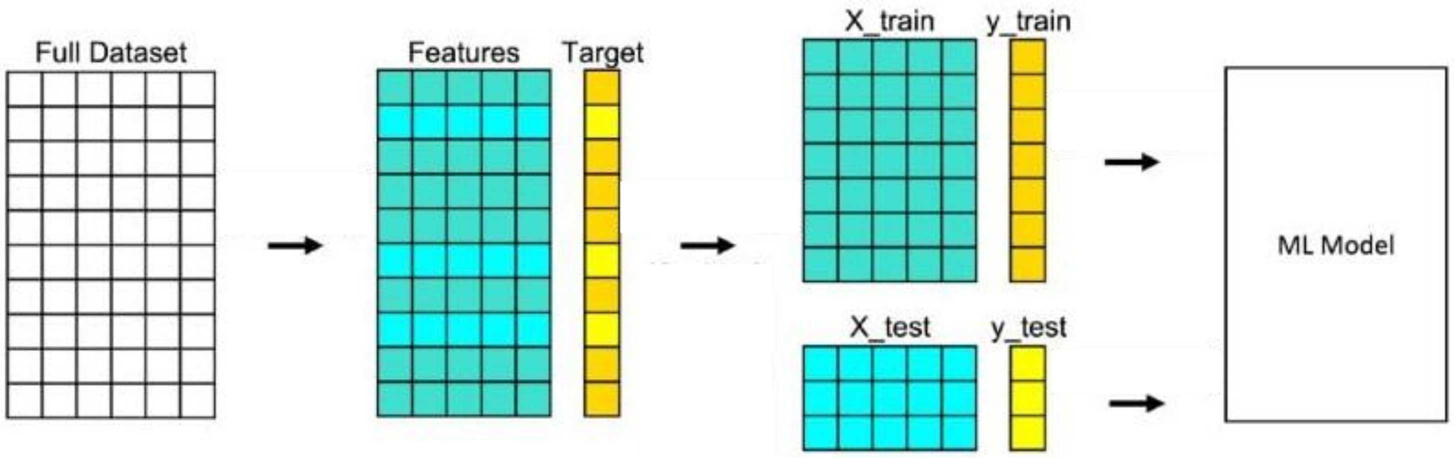
Регуляризация L1+L2 (Elastic Net)

$$Q(\mathrm{a}, \mathbb{X}) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 + \lambda_1\|\theta\|_1 + \lambda_2\|\theta\|_2 \to \min_{\boldsymbol{\theta}}$$



Метрики качества в задачах регрессии

Разделение данных



	bedrooms	bathrooms	sqft_living	sqft_lot	floors	price
0	3	1.000000	1180	5650	1.000000	221900.000000
1	3	2.250000	2570	7242	2.000000	538000.000000
2	2	1.000000	770	10000	1.000000	180000.000000
3	4	3.000000	1960	5000	1.000000	604000.000000
4	3	2.000000	1680	8080	1.000000	510000.000000
5	4	4.500000	5420	101930	1.000000	1225000.000000
6	3	2.250000	1715	6819	2.000000	257500.000000
7	3	1.500000	1060	9711	1.000000	291850.000000
8	3	1.000000	1780	7470	1.000000	229500.000000
9	3	2.500000	1890	6560	2.000000	323000.000000

X_train
X_test
y_train
y_test

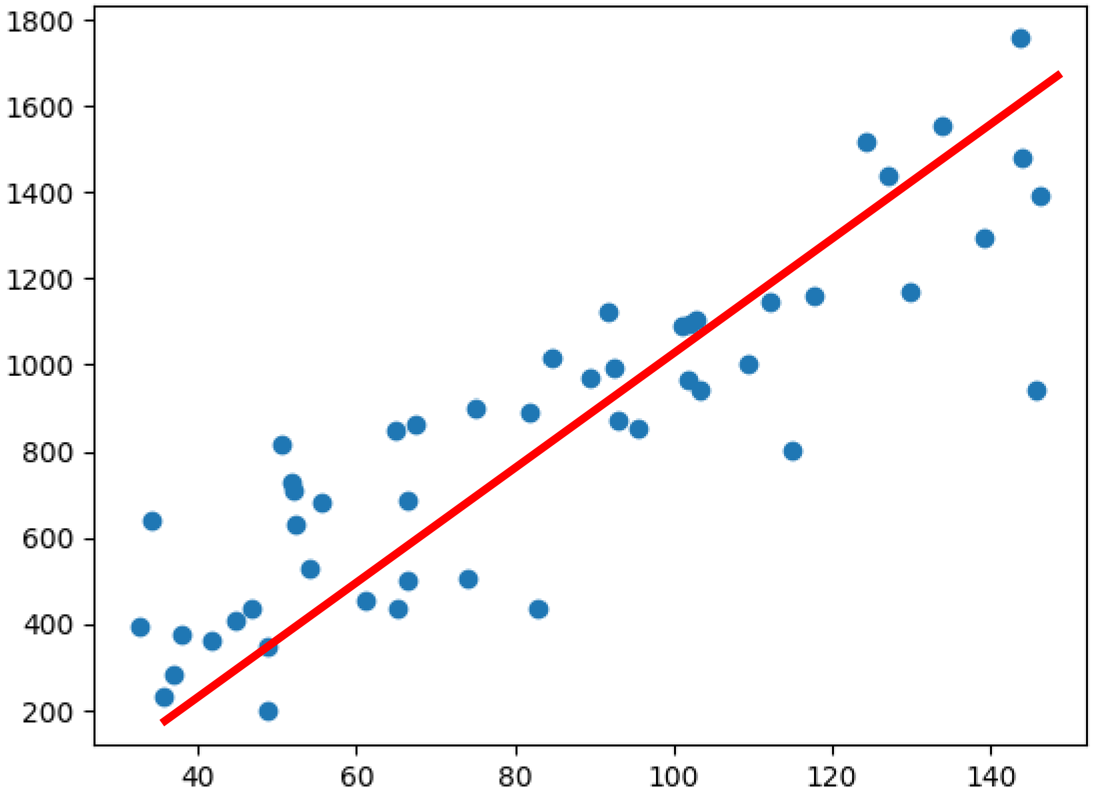
Метрики качества в задачах регрессии

Фунционал ошибки – функция, которую минимизируют в процессе обучения модели для нахождения незвестных параметров

Метрика качества – функция, которую используют для оценки качества построенной модели.

Метрики качества в задачах регрессии

Что такое метрики качества?



Среднеквадратичная ошибка (Mean Squared Error)

MSE рассчитывается по формуле:

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

ℓ - количество наблюдений, по которым строится модель

y_i– фактическое значение

$a(x_i)$ значение зависимой переменной, предсказанное моделью

Среднеквадратичная ошибка (Mean Squared Error)

MSE рассчитывается по формуле:

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

Минусы:

По величине метрики не понятно она хорошая или плохая

Единицы измерения не сохраняются

MSE очень чувствительна к выбросам

Корень из среднеквадратичной ошибки (Root Mean Squared Error)

RMSE рассчитывается по формуле:

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

ℓ - количество наблюдений, по которым строится модель

y_i– фактическое значение

$a(x_i)$ значение зависимой переменной, предсказанное моделью

Корень из среднеквадратичной ошибки (Root Mean Squared Error)

RMSE рассчитывается по формуле:

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

Минусы:

По величине метрики не понятно она хорошая или плохая

Коэффициент детерминации R^2

R^2 рассчитывается по формуле:

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

ℓ - количество наблюдений

y_i– фактическое значение

$a(x_i)$ предсказания алгоритма

$\overline{y} = \frac{1}{\ell}\sum_{i=1}^{\ell} y_i$

Коэффициент детерминации R^2

R^2 рассчитывается по формуле:

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

Коэффициент детерминации это доля дисперсии целевой переменной, объясняемая моделью.

Чем ближе R^2 к 1, тем лучше модель объясняет данные

Чем ближе R^2 к 0, тем ближе модель к константному предсказанию

Отрицательный R^2 говорит о том, что модель плохо решает задачу

Средняя абсолютная ошибка (Mean Absolute Error)

MAE рассчитывается по формуле:

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

ℓ - количество наблюдений, по которым строится модель

y_i– фактическое значение

$a(x_i)$ значение зависимой переменной, предсказанное моделью

Средняя абсолютная ошибка (Mean Absolute Error)

MAE рассчитывается по формуле:

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

Минусы:

По величине метрики не понятно она хорошая или плохая

Средняя абсолютная процентная ошибка (Mean Absolute Percentage Error)

$MAEP$ рассчитывается по формуле:

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

ℓ - количество наблюдений, по которым строится модель

y_i– фактическое значение

$a(x_i)$ значение зависимой переменной, предсказанное моделью

Средняя абсолютная процентная ошибка (Mean Absolute Percentage Error)

$MAEP$ рассчитывается по формуле:

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

Минусы:

Метрика не симметрична относительно ошибок

Симметричная средняя абсолютная процентная ошибка (Symmetric Mean Absolute Percentage Error)

$SMAEP$ рассчитывается по формуле:

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

ℓ - количество наблюдений, по которым строится модель

y_i – фактическое значение

$a(x_i)$ значение зависимой переменной, предсказанное моделью

Симметричная средняя абсолютная процентная ошибка (Symmetric Mean Absolute Percentage Error)

$SMAEP$ рассчитывается по формуле:

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

Минусы:

не полностью симметрична