

# Машинное обучение

Лекция 2

Линейная регрессия

НИУ ВШЭ, 2021

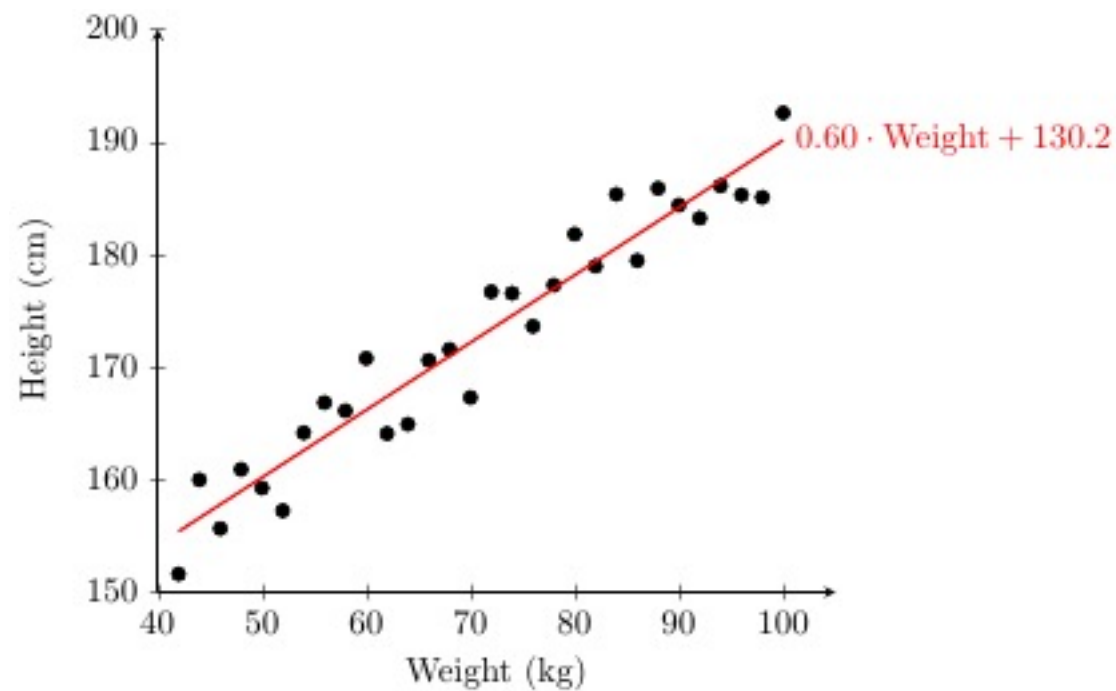
# Напоминание

- $\mathbb{X}$  — пространство объектов,  $\mathbb{Y}$  — пространство ответов
- $x = (x^1, \dots, x^d)$  — признаковое описание
- $X = (x_i, y_i)_{i=1}^{\ell}$  — обучающая выборка
- $a(x)$  — алгоритм, модель
- $L(a, X)$  — функция потерь для алгоритма  $a$  на выборке  $X$
- Обучение:  $a(x) = \arg \min_{a \in \mathcal{A}} L(a, X)$

Типы ответов

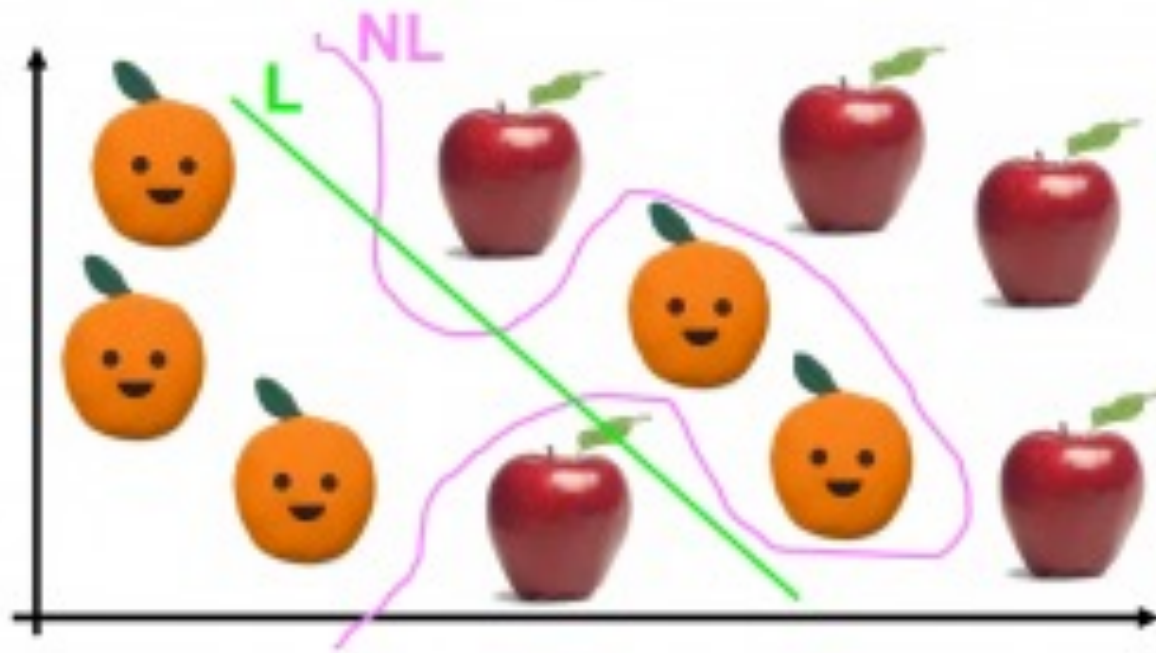
# Регрессия

- Вещественные ответы:  $\mathbb{Y} = \mathbb{R}$
- Пример: предсказание роста по весу



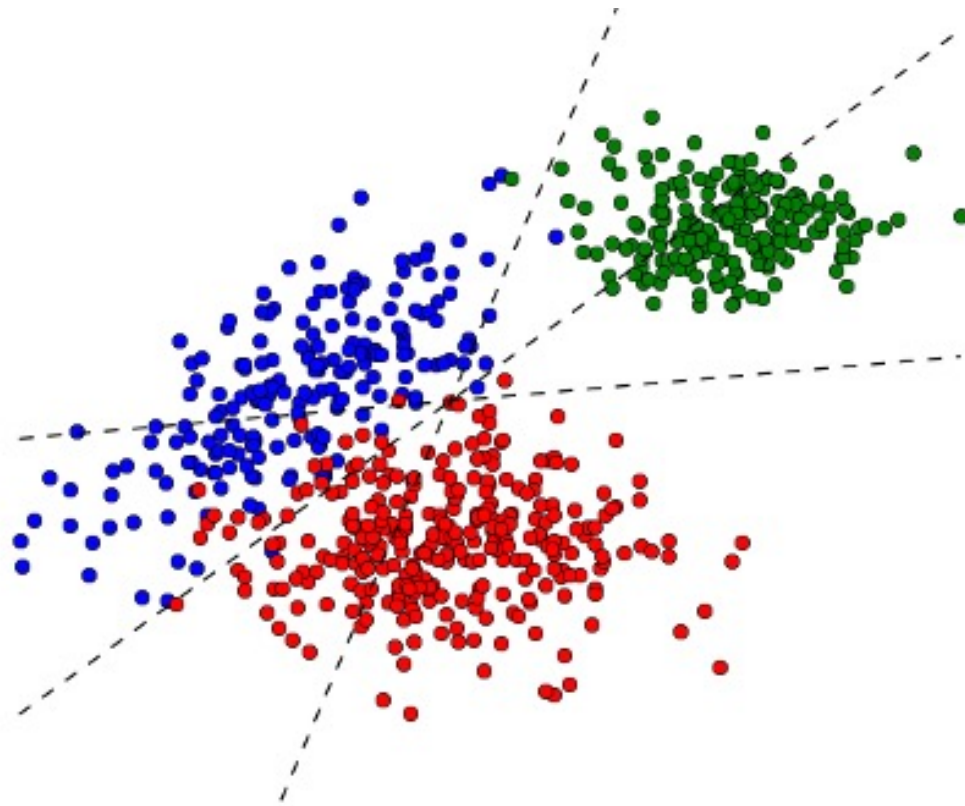
# Классификация

- Конечное число ответов:  $|\mathbb{Y}| < \infty$
- Бинарная классификация:  $\mathbb{Y} = \{-1, +1\}$



# Классификация

- Многоклассовая классификация:  $\mathbb{Y} = \{1, 2, \dots, K\}$



# Классификация

- Классификация с пересекающимися классами:  $\mathbb{Y} = \{0, 1\}^K$ 
  - (multi-label classification)
- Ответ — набор из  $K$  нулей и единиц
- $i$ -й элемент ответа — принадлежит ли объект  $i$ -му классу
  
- Какие темы присутствуют в статье?
- (математика, биология, экономика)

# Ранжирование

- Набор документов  $d_1, \dots, d_n$
- Запрос  $q$
- Задача: отсортировать документы по *релевантности* запросу
- $a(q, d)$  — оценка релевантности



# Ранжирование

Яндекс

картинки с котиками — 5 млн ответов



Найти

Поиск

Картинки

Видео

Карты

Маркет

Ещё



## Картинки с кошками | Fun Cats — Забавные коты

[funcats.by](#) > [pictures/](#) ▼

**Картинки с кошками.** Прикольные коты. 777 **изображений**. ... 32 **изображения**. Кошки Стамбула. 41 **изображение**. Веселые котята.



## Уморные котики (57 фото) » Бяки.нет | Картинки

[byaki.net](#) > **Картинки** > [14026-umornye-kotiki-57...](#) ▼

Бяки нет! . NET. Уморные **котики (57 фото)**. 223. Комментариев:9Автор:4ertonok  
Просмотров:161 395 **Картинки**28-10-2008, 00:03.



## Смешные картинки кошек с надписями | Лолкот.Ру

[lolkot.ru](#) ▼

Смешные **картинки** для новых приколов! Сделать свой прикол очень просто. ... **Котик** верит в чудеса. Он в носке подарок ищет...



## Красивые картинки и фото кошек, котят и котов

[foto-zverey.ru](#) > **Кошки** ▼

**Фото и картинки** кошек и котят потрясающей красоты и нежности. Здесь мы собрали такие **изображения**, которые всегда вызывают море положительных эмоций...



## Обои для рабочего стола Котят | картинки на стол Котят

[7fon.ru](#) > Чёрные обои и **картинки** > Обои котят ▼

**Картинки** Котят с 1 по 15. **Обои** для рабочего стола Котят. ... Скачать **Картинки** Котят на рабочий стол бесплатно.

# Кластеризация

- $Y$  — отсутствует
  - Нужно найти группы похожих объектов
  - Сколько таких групп?
  - Как измерить качество?
- 
- Пример: сегментация пользователей мобильного оператора

# Типы признаков

# Типы признаков

- $D_j$  — множество значений признака

# Бинарные признаки

- $D_j = \{0, 1\}$
- Доход клиента выше среднего по городу?
- Цвет фрукта — зеленый?

# Вещественные признаки

- $D_j = \mathbb{R}$
- Возраст
- Площадь квартиры
- Количество звонков в колл-центр

# Категориальные признаки

- $D_j$  — неупорядоченное множество
  - Цвет глаз
  - Город
  - Образование (может быть упорядоченным)
- 
- Нуждаются в обработке

# Порядковые признаки

- $D_j$  — упорядоченное множество
- Военское звание
- Роль в фильме (первого плана, второго плана, массовка)
- Тип населенного пункта

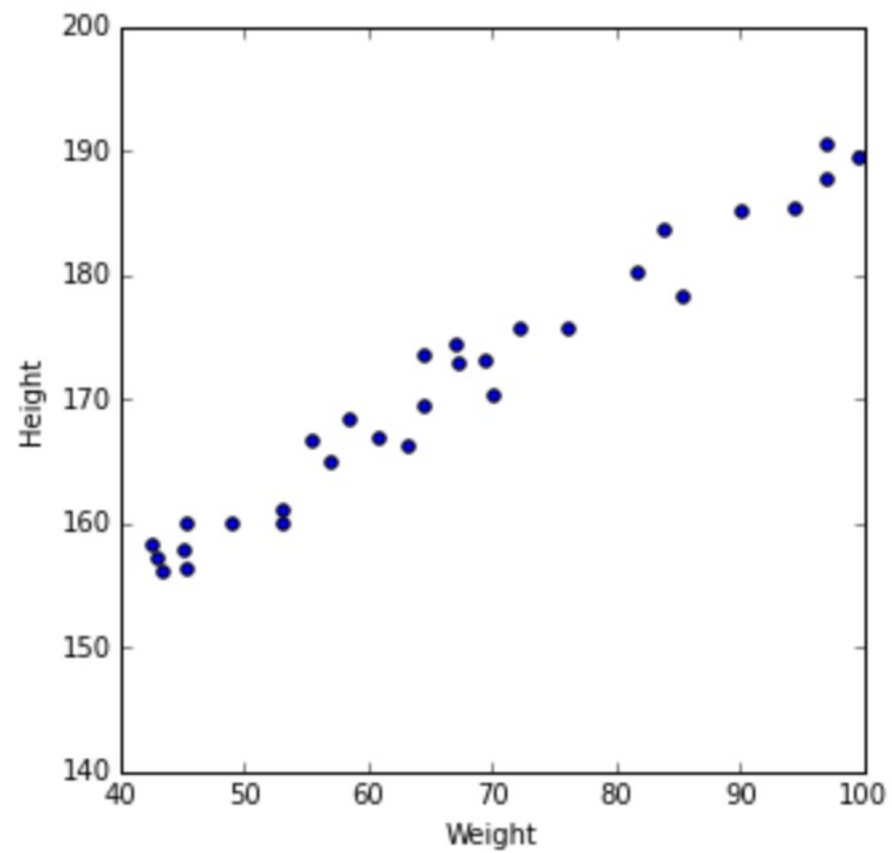


# Другие типы данных

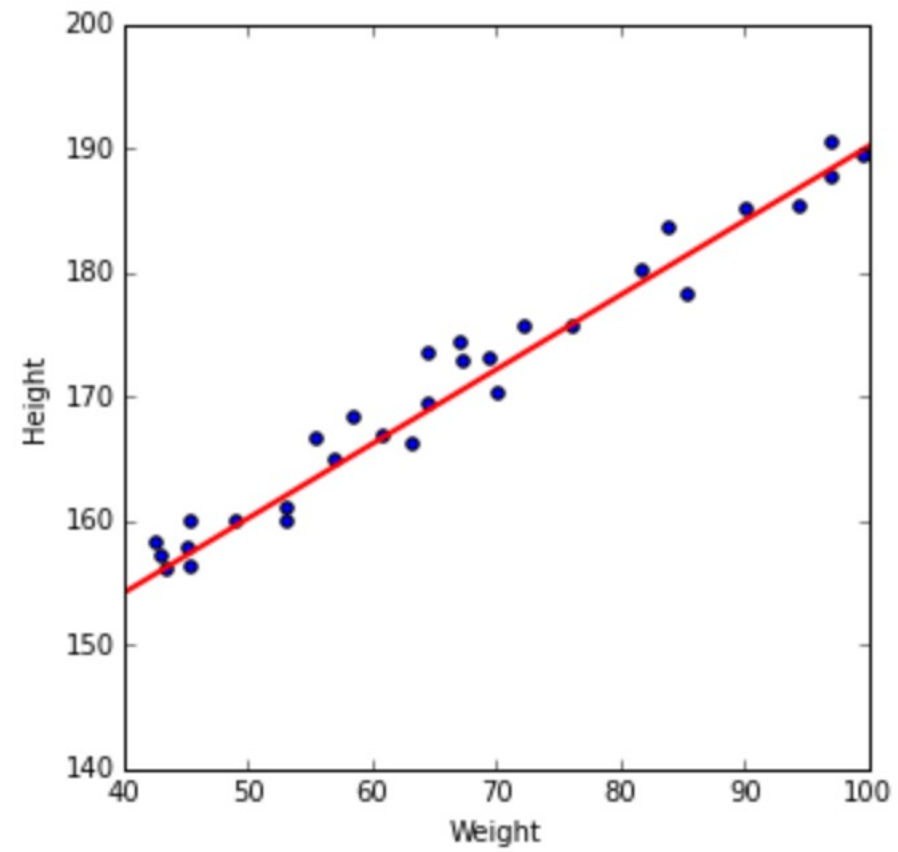
- Тексты
- Изображения
- Видео
- Звук
- Временные ряды
- Графы

# Линейная регрессия

# Парная регрессия



# Парная регрессия



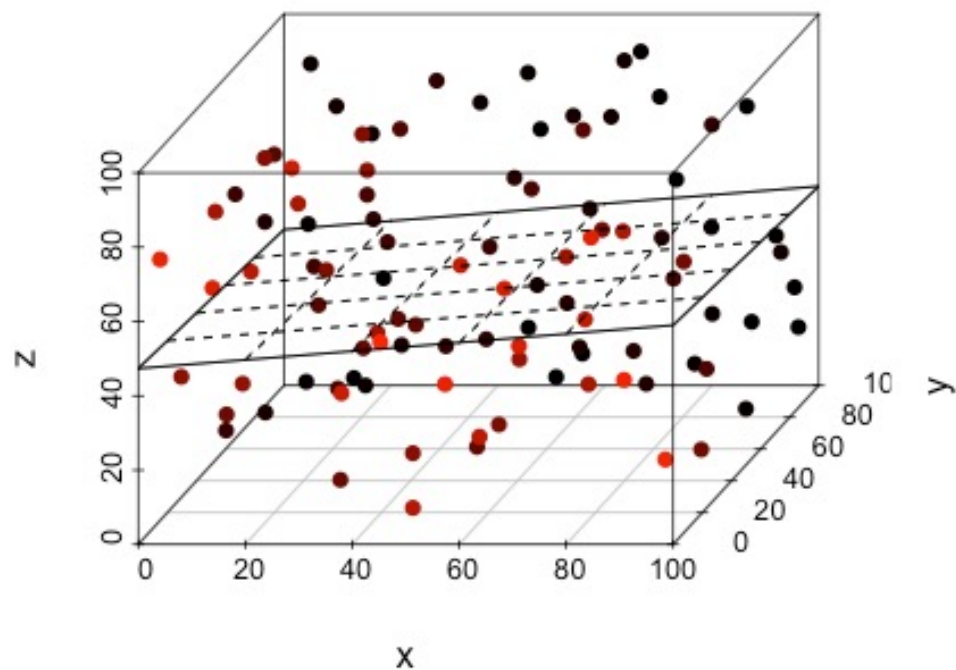
# Парная регрессия

- Простейший случай: один признак
- Модель:  $a(x) = w_1 x + w_0$
- Два параметра:  $w_1$  и  $w_0$
- $w_1$  — тангенс угла наклона
- $w_0$  — где прямая пересекает ось ординат

# Два признака

- Чуть более сложный случай: два признака
- Модель:  $a(x) = w_0 + w_1 x_1 + w_2 x_2$
- Три параметра

# Два признака



# Много признаков

- Общий случай:  $d$  признаков
- Модель

$$a(x) = w_0 + w_1x_1 + \dots + w_dx_d$$

- Количество параметров:  $d + 1$



# Много признаков

- Общий случай:  $d$  признаков
- Модель

$$a(x) = w_0 + w_1x_1 + \dots + w_dx_d$$

Свободный коэффициент/сдвиг/bias

Веса/коэффициенты

- Количество параметров:  $d + 1$

# Много признаков

$$\begin{aligned} a(x) &= w_0 + w_1x_1 + \dots + w_dx_d = \\ &= w_0 + \langle w, x \rangle \end{aligned}$$

- Будем считать, что есть признак, всегда равный единице:

$$\begin{aligned} a(x) &= w_1x_1 + \dots + w_dx_d = \\ &= w_1 * 1 + w_2x_2 + \dots + w_dx_d = \\ &= \langle w, x \rangle \end{aligned}$$

# Применимость линейной регрессии

# Модель линейной регрессии

$$a(x) = w_1x_1 + \dots + w_dx_d = \langle w, x \rangle$$

- Нет гарантий, что целевая переменная именно так зависит от признаков
- Надо формировать признаки так, чтобы модель подходила

# Предсказание стоимости квартиры

- Признаки: площадь, район, расстояние до метро
- Целевая переменная: рыночная стоимость квартиры
- Линейная модель:

$$\begin{aligned} a(x) = & w_0 + w_1 * (\text{площадь}) \\ & + w_2 * (\text{район}) \\ & + w_3 * (\text{расстояние до метро}) \end{aligned}$$

# Предсказание стоимости квартиры

$$\begin{aligned} a(x) = & w_0 + w_1 * (\text{площадь}) \\ & + w_2 * (\text{район}) \\ & + w_3 * (\text{расстояние до метро}) \end{aligned}$$

# Предсказание стоимости квартиры

$$\begin{aligned} a(x) = & w_0 + w_1 * (\text{площадь}) \\ & + w_2 * (\text{район}) \\ & + w_3 * (\text{расстояние до метро}) \end{aligned}$$

- За каждый квадратный метр добавляем  $w_1$  к прогнозу

# Предсказание стоимости квартиры

$$\begin{aligned} a(x) = & w_0 + w_1 * (\text{площадь}) \\ & + w_2 * (\text{район}) \\ & + w_3 * (\text{расстояние до метро}) \end{aligned}$$

- Что-то странное

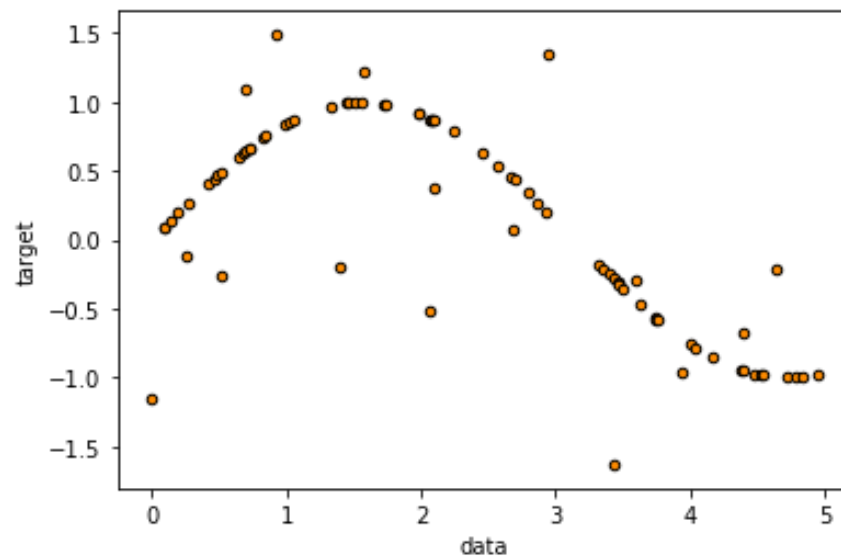


# Предсказание стоимости квартиры

$$a(x) = w_0 + w_1 * (\text{площадь})$$

$$+ w_2 * (\text{район})$$

$$+ w_3 * (\text{расстояние до метро})$$



# Кодирование категориальных признаков

- Значения признака «район»:  $U = \{u_1, \dots, u_m\}$
- Новые признаки вместо  $x_j$ :  $[x_j = u_1], \dots, [x_j = u_m]$
- One-hot кодирование

# Кодирование категориальных признаков

| Район |  | ЦАО | ЮАО | САО |
|-------|--|-----|-----|-----|
| ЦАО   |  | 1   | 0   | 0   |
| ЮАО   |  | 0   | 1   | 0   |
| ЦАО   |  | 1   | 0   | 0   |
| САО   |  | 0   | 0   | 1   |
| ЮАО   |  | 0   | 1   | 0   |

# Кодирование категориальных признаков

| Район |   | ЦАО | ЮАО | САО |
|-------|---|-----|-----|-----|
| ЦАО   |   | 1   | 0   | 0   |
| ЮАО   |   | 0   | 1   | 0   |
| ЦАО   | → | 1   | 0   | 0   |
| САО   |   | 0   | 0   | 1   |
| ЮАО   |   | 0   | 1   | 0   |

$$a(x) = w_0 + w_1 * (\text{площадь})$$

$$+ w_2 * (\text{квартира в ЦАО?})$$

$$+ w_3 * (\text{квартира в ЮАО?})$$

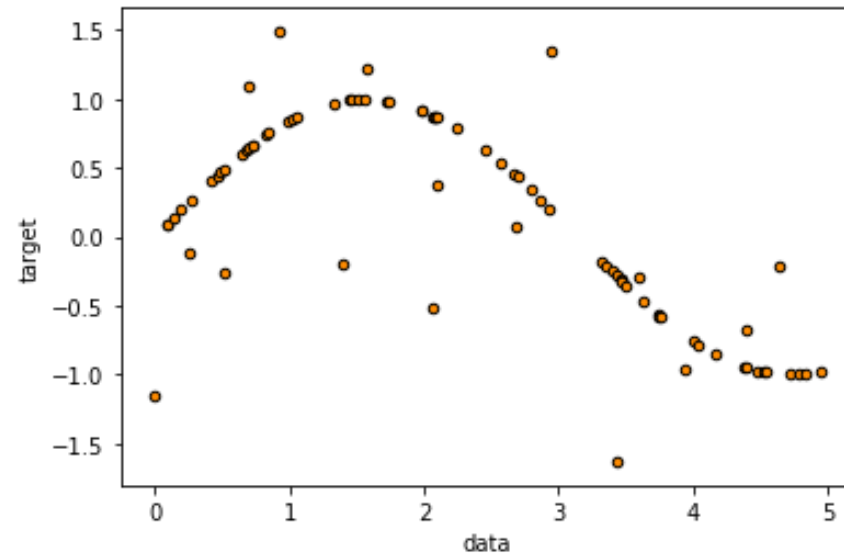
$$+ w_4 * (\text{квартира в САО?})$$

# Предсказание стоимости квартиры

$$a(x) = w_0 + w_1 * (\text{площадь})$$

$$+ w_2 * (\text{район})$$

$$+ w_3 * (\text{расстояние до метро})$$

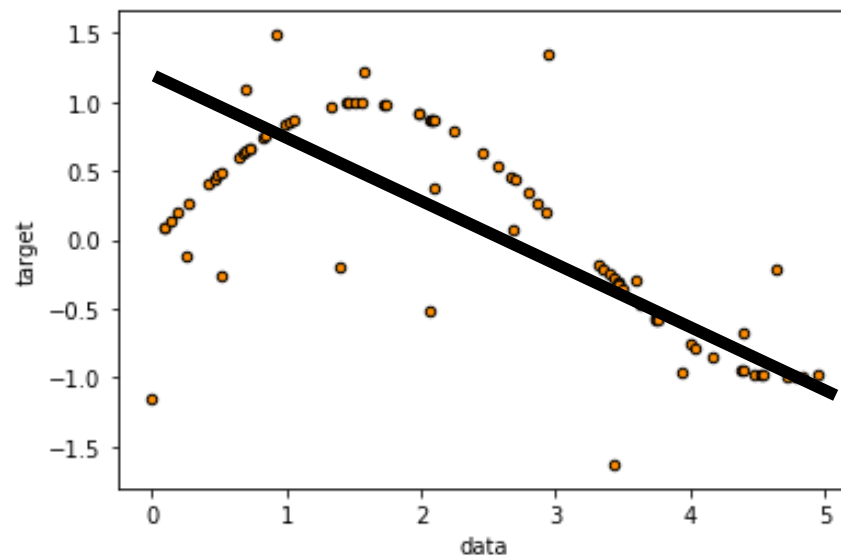


# Предсказание стоимости квартиры

$$a(x) = w_0 + w_1 * (\text{площадь})$$

$$+ w_2 * (\text{район})$$

$$+ w_3 * (\text{расстояние до метро})$$

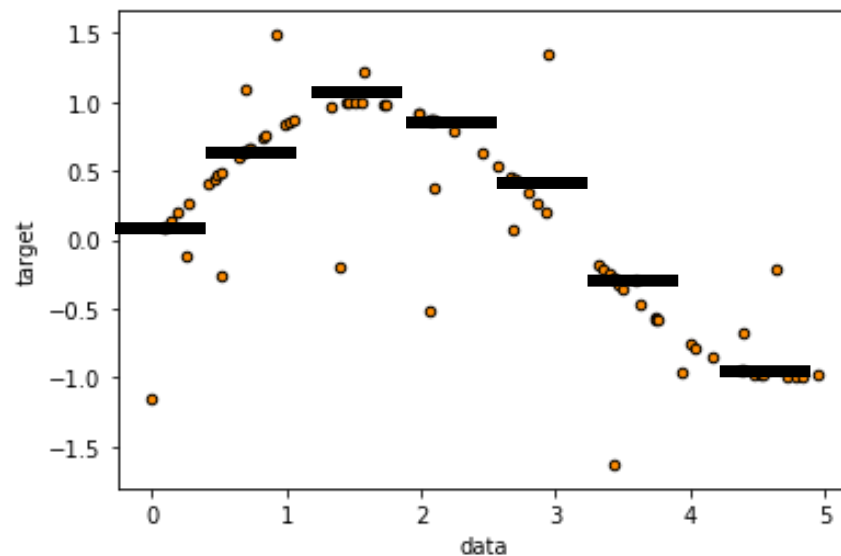


# Предсказание стоимости квартиры

$$a(x) = w_0 + w_1 * (\text{площадь})$$

$$+ w_2 * (\text{район})$$

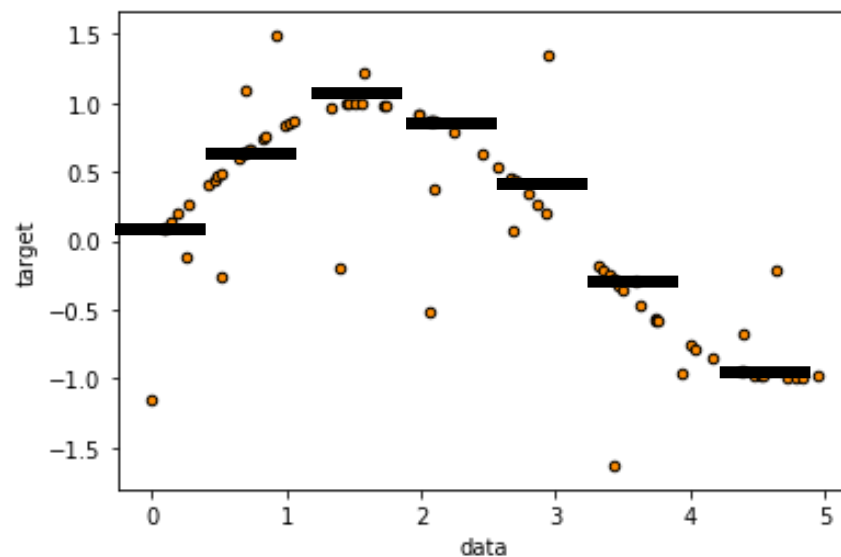
$$+ w_3 * (\text{расстояние до метро})$$



# Предсказание стоимости квартиры

$$a(x) = w_0 + w_1 * (\text{площадь}) \\ + w_2 * (\text{район})$$

$$+ w_3 * [t_0 \leq x_3 < t_1] + \dots + w_{3+n} [t_{n-1} \leq x_3 < t_n]$$





# Линейные модели

- Модель линейной регрессии хороша, если признаки сделаны специально под неё
- Пример: one-hot кодирование категориальных признаков или бинаризация числовых признаков

Измерение ошибки модели

# Функция потерь для регрессии

- Частый выбор — квадратичная функция потерь

$$L(y, a) = (a - y)^2$$

- Функционал ошибки — среднеквадратичная ошибка (mean squared error, MSE)

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

# Функция потерь для регрессии

- Ещё один вариант — средняя абсолютная ошибка (mean absolute error, MAE)

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

- Слабее штрафует за серьёзные отклонения от правильного ответа

# Линейная регрессия в векторном виде

# Модель линейной регрессии

$$a(x) = \langle w, x \rangle$$

- Среднеквадратичная ошибка и задача обучения:

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

# Матрицы

- Матрица — таблица с числами
- Матрица «объекты-признаки»:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\ell 1} & x_{\ell 2} & \cdots & x_{\ell d} \end{pmatrix} \in \mathbb{R}^{\ell \times d}$$

# Матрицы

- Матрица — таблица с числами (для простоты)
- Матрица «объекты-признаки»:

объект и его признаки

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\ell 1} & x_{\ell 2} & \cdots & x_{\ell d} \end{pmatrix}$$



# Матрицы

- Матрица — таблица с числами (для простоты)
- Матрица «объекты-признаки»:

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\ell 1} & x_{\ell 2} & \cdots & x_{\ell d} \end{pmatrix}$$

значения признака на всех объектах

# Векторы

- Вектор размера  $d$  — тоже матрица
- Вектор-строка:  $w = (w_1, \dots, w_d) \in \mathbb{R}^{1 \times d}$
- Вектор-столбец:  $w = \begin{pmatrix} w_1 \\ \dots \\ w_d \end{pmatrix} \in \mathbb{R}^{d \times 1}$

# Применение линейной модели

- $a(x) = \langle w, x \rangle = w_1 x_1 + \dots + w_d x_d$
- Как применить модель к обучающей выборке?

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\ell 1} & x_{\ell 2} & \cdots & x_{\ell d} \end{pmatrix}$$

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix}$$

$$\begin{pmatrix} \sum_{i=1}^d w_i x_{1i} \\ \sum_{i=1}^d w_i x_{2i} \\ \vdots \\ \sum_{i=1}^d w_i x_{\ell i} \end{pmatrix}$$

# Матричное умножение

- Только для матриц  $A \in \mathbb{R}^{m \times k}$  и  $B \in \mathbb{R}^{k \times n}$
- Результат:  $AB = C \in \mathbb{R}^{m \times n}$
- Правило:

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

Пример

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

Пример

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

Пример

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

Пример

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



Пример

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

# Модель линейной регрессии

- Среднеквадратичная ошибка и задача обучения:

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

# Вычисление ошибки

- Отклонения прогнозов от ответов:

$$Xw - y = \begin{pmatrix} \langle w, x_1 \rangle - y_1 \\ \vdots \\ \langle w, x_\ell \rangle - y_\ell \end{pmatrix}$$

# Вычисление ошибки

- Евклидова норма:

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

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

# Вычисление ошибки

- Отклонения прогнозов от ответов:

$$Xw - y = \begin{pmatrix} \langle w, x_1 \rangle - y_1 \\ \vdots \\ \langle w, x_\ell \rangle - y_\ell \end{pmatrix}$$

- Среднеквадратичная ошибка:

$$\frac{1}{\ell} \|Xw - y\|^2 = \frac{1}{\ell} \sum_{i=1}^{\ell} (\langle w, x_i \rangle - y_i)^2$$

# Обучение линейной регрессии

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

Обобщающая способность

# Обобщающая способность

Как готовиться к экзамену?

Заучить все примеры с  
занятий

Разобраться в предмете и  
усвоить алгоритмы решения  
задач



# Обобщающая способность

Как готовиться к экзамену?

Заучить все примеры с  
занятий

Разобраться в предмете и  
усвоить алгоритмы решения  
задач

Переобучение (overfitting)

Обобщение (generalization)

# Обобщающая способность

Как готовиться к экзамену?

Заучить все примеры с  
занятий

Переобучение (overfitting)

Хорошее качество на обучении  
Низкое качество на новых данных

Разобраться в предмете и  
усвоить алгоритмы решения  
задач

Обобщение (generalization)

Хорошее качество на обучении  
Хорошее качество на новых  
данных

# Отложенная выборка



Обучение



Тест

# Отложенная выборка



- Слишком большое обучение — тестовая выборка нерепрезентативна
- Слишком большой тест — модель не сможет обучиться
- Обычно: 70/30, 80/20

# Кросс-валидация



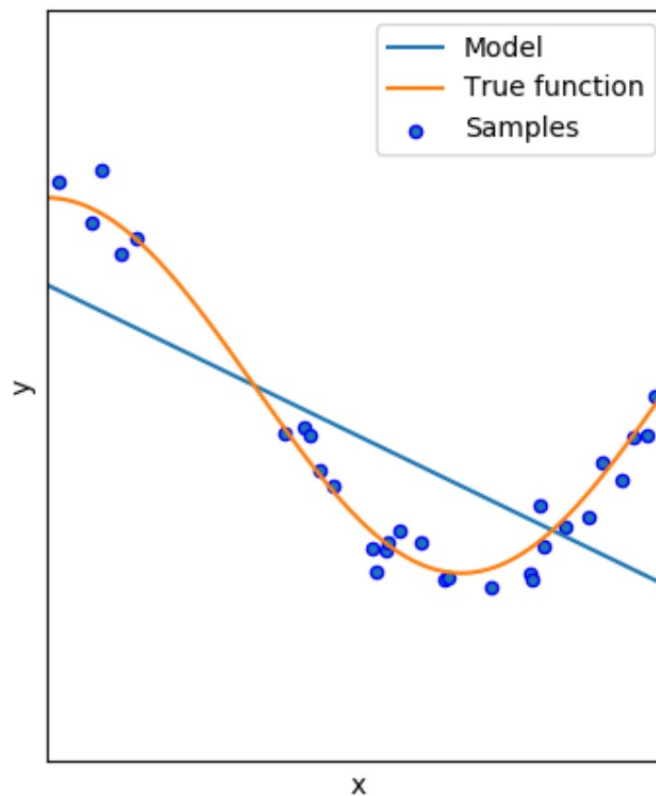
# Кросс-валидация

- Надёжнее отложенной выборки, но медленнее
- Параметр — количество разбиений  $n$  (фолдов, folds)
- Хороший, но медленный вариант —  $n = \ell$  (leave-one-out)
- Обычно:  $n = 3$  или  $n = 5$  или  $n = 10$

# Переобучение и регуляризация линейных моделей

# Нелинейная задача

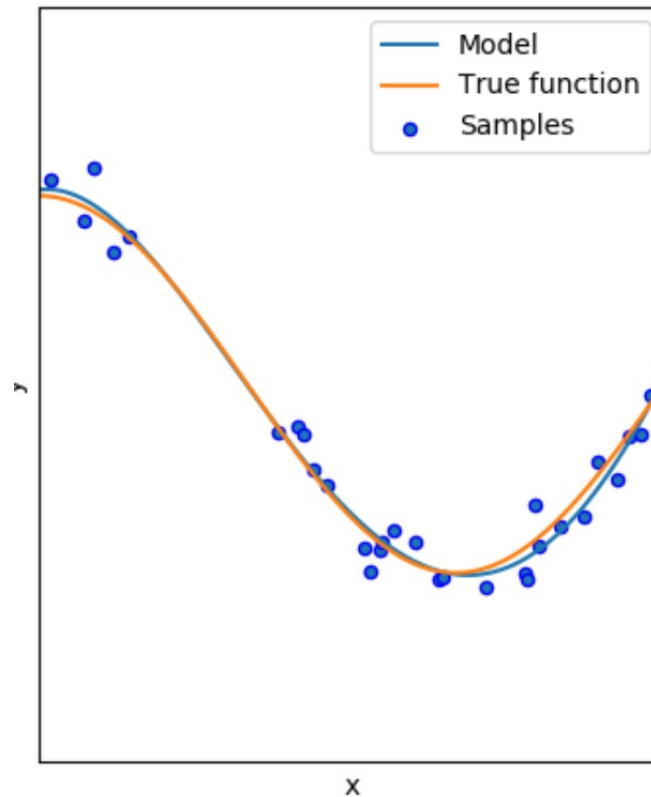
$$a(x) = w_0 + w_1 x$$





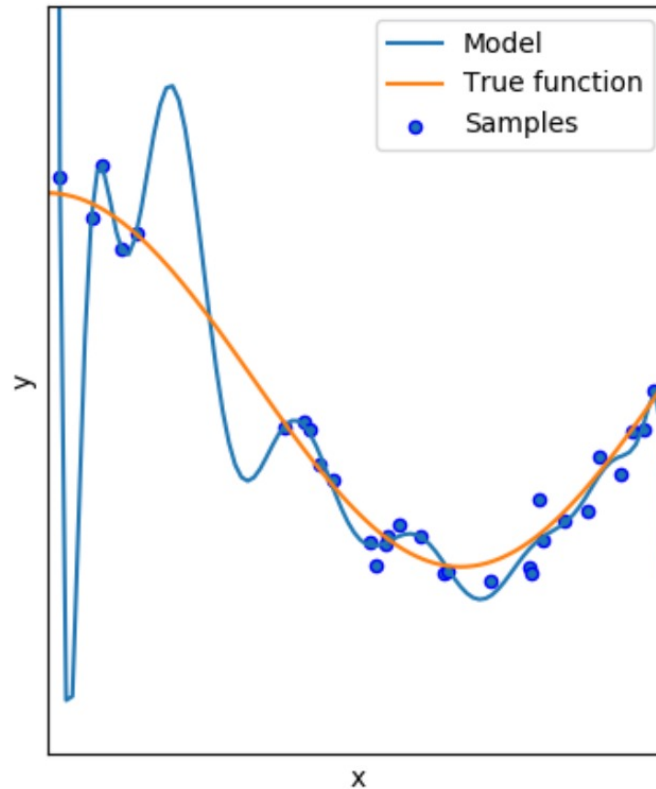
# Нелинейная задача

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



# 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

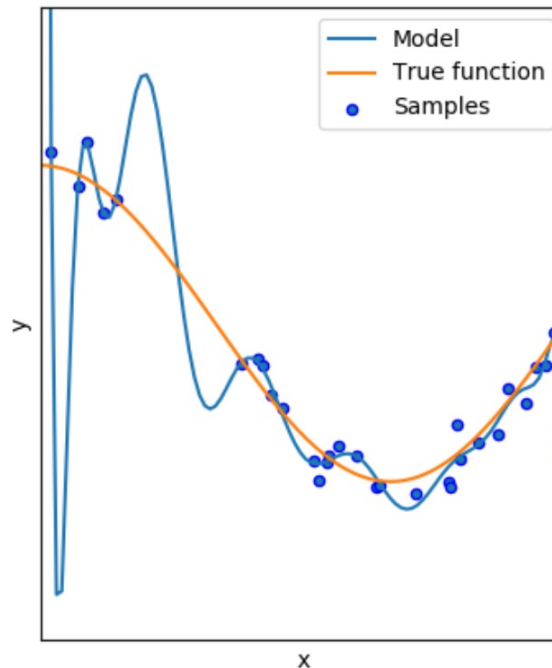
$$\frac{1}{\ell} \sum_{i=1}^{\ell} (\langle w, x_i \rangle - y_i)^2 + \lambda \|w\|^2 \rightarrow \min_w$$

- $\lambda$  — коэффициент регуляризации
- Ridge (гребневая регрессия)

# Эффект регуляризации

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

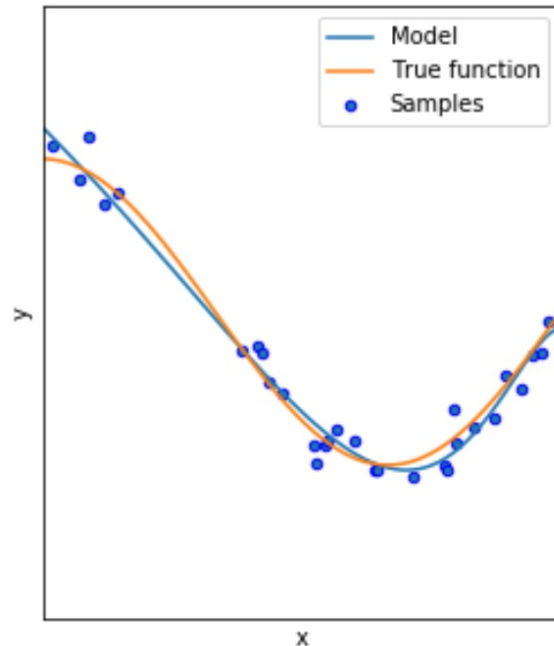
$$\frac{1}{\ell} \sum_{i=1}^{\ell} (a(x_i) - y_i)^2 \rightarrow \min_w$$



# Эффект регуляризации

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

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

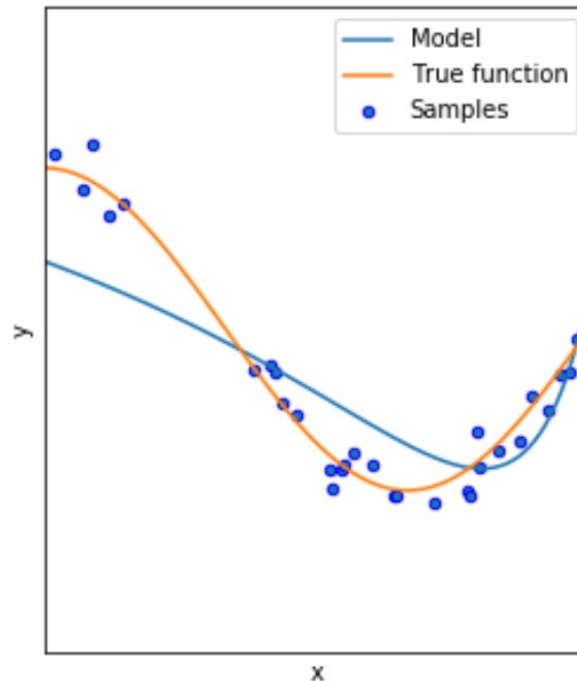




# Эффект регуляризации

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

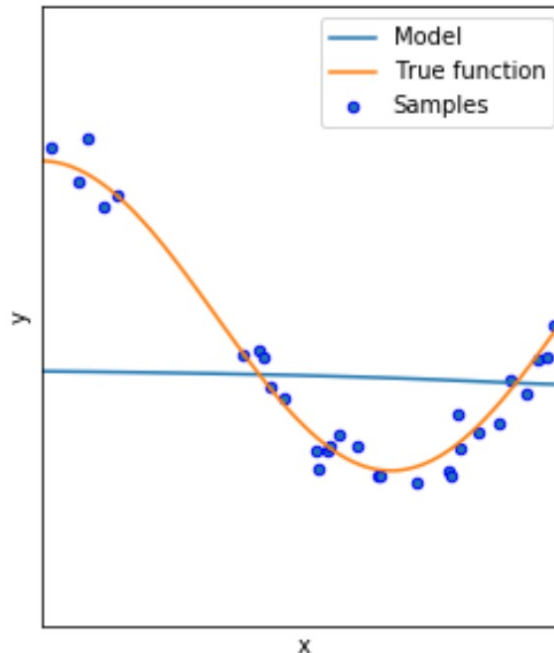
$$\frac{1}{\ell} \sum_{i=1}^{\ell} (a(x_i) - y_i)^2 + \mathbf{1} \|w\|^2 \rightarrow \min_w$$



# Эффект регуляризации

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

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



# 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

# Гиперпараметры

- **Гиперпараметр** алгоритма — то, что задается вручную
  - Пример: коэффициент регуляризации
- **Параметр** алгоритма — то, что определяется моделью
  - Пример: веса регрессии
- Как подбирать гиперпараметры алгоритма?
  - Нельзя подбирать по обучающей выборке — это приведет к переобучению
  - Нужно использовать дополнительные данные (валидация)

# Чуть больше терминов

- После подбора всех гиперпараметров стоит проверить на совсем новых данных, что модель работает
- Обучающая выборка — построение модели
- Валидационная выборка — подбор гиперпараметров модели
- Тестовая выборка — финальная оценка качества модели

# Интерпретация линейных моделей

# Предсказание стоимости квартиры

$$\begin{aligned} a(x) = & 100.000 * (\text{площадь}) \\ & + 500.000 * (\text{число магазинов рядом}) \\ & + 100 * (\text{средний доход жильцов дома}) \end{aligned}$$



# Предсказание стоимости квартиры

$$\begin{aligned} a(x) = & 100.000 * (\text{площадь}) \\ & + 500.000 * (\text{число магазинов рядом}) \\ & + 100 * (\text{средний доход жильцов дома}) \end{aligned}$$

- Чем больше вес, тем важнее признак?

# Предсказание стоимости квартиры

$$\begin{aligned} a(x) = & 100.000 * (\text{площадь в кв. м.}) \\ & + 500.000 * (\text{число магазинов рядом}) \\ & + 100 * (\text{средний доход жильцов дома}) \end{aligned}$$

- Чем больше вес, тем важнее признак?

# Предсказание стоимости квартиры

$$\begin{aligned} a(x) = & 10 * (\text{площадь в кв. см.}) \\ & + 500.000 * (\text{число магазинов рядом}) \\ & + 100 * (\text{средний доход жильцов дома}) \end{aligned}$$

- Чем больше вес, тем важнее признак?

# Предсказание стоимости квартиры

$$\begin{aligned} a(x) = & 100.000 * (\text{площадь в кв. м.}) \\ & + 500.000 * (\text{число магазинов рядом}) \\ & + 100 * (\text{средний доход жильцов дома}) \end{aligned}$$

- Чем больше вес, тем важнее признак?

# Предсказание стоимости квартиры

$$\begin{aligned} a(x) = & 100.000 * (\text{площадь в кв. м.}) \\ & + 500.000 * (\text{число магазинов рядом}) \\ & + 100 * (\text{средний доход жильцов дома}) \end{aligned}$$

- Чем больше вес, тем важнее признак?
- Только если признаки масштабированы!

# Масштабирование признаков

- Отмасштабируем  $j$ -й признак
- Вычисляем среднее и стандартное отклонение признака на обучающей выборке:

$$\mu_j = \frac{1}{\ell} \sum_{i=1}^{\ell} x_i^j$$

$$\sigma_j = \sqrt{\frac{1}{\ell} \sum_{i=1}^{\ell} (x_i^j - \mu_j)^2}$$

# Масштабирование признаков

- Вычтем из каждого значения признака среднее и поделим на стандартное отклонение:

$$x_i^j := \frac{x_i^j - \mu_j}{\sigma_j}$$

# Регуляризация

- Если модель переобучается, то веса используются для запоминания обучающей выборки
- Правильнее масштабировать признаки и регуляризовать модель перед изучением весов



# Пример

- 1000 объектов
- Два признака
- Первый принимает значения от 0 до 1
- Второй равен единице на 10 объектах и нулю на 990 объектах
- $y = x_1 + 2x_2$

# Пример

```
[0.3175037 , 1.      ],
[0.59558502, 1.      ],
[0.48660609, 1.      ],
[0.69255463, 1.      ],
[0.81968981, 1.      ],
[0.48844247, 1.      ],
[0.13426702, 1.      ],
[0.850628   , 1.      ],
[0.57499033, 1.      ],
[0.73993748, 1.      ],
[0.70466465, 0.      ],
[0.96821177, 0.      ],
[0.29530732, 0.      ],
[0.70530677, 0.      ],
[0.36567633, 0.      ],
[0.39541072, 0.      ],
[0.23059464, 0.      ],
[0.34401018, 0.      ],
[0.94829675, 0.      ],
[0.29257085, 0.      ],
[0.24599061, 0.      ],
[0.58313798, 0.      ],
```

# Пример

$$a(x) = x_1 + 2x_2$$

- Удаляем первый признак, получаем  $MSE = 0.08$
- Удаляем второй признак, получаем  $MSE = 0.04$
- Правильнее удалить признак и посмотреть, как сильно растёт ошибка без него