



# Математическая статистика

## и основы анализа данных

### Основы машинного обучения



# План занятия

- ▶ Введение в машинное обучение: задача регрессии



# План занятия

- ▶ Введение в машинное обучение: задача регрессии
  - ▶ Линейная регрессия



# План занятия

- ▶ Введение в машинное обучение: задача регрессии
  - ▶ Линейная регрессия
  - ▶ Метрики качества для задачи регрессии



# План занятия

- ▶ Введение в машинное обучение: задача регрессии
  - ▶ Линейная регрессия
  - ▶ Метрики качества для задачи регрессии
- ▶ Свойства оценок



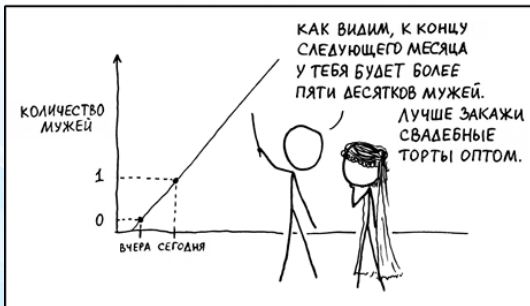
# План занятия

- ▶ Введение в машинное обучение: задача регрессии
  - ▶ Линейная регрессия
  - ▶ Метрики качества для задачи регрессии
- ▶ Свойства оценок

# Линейная регрессия

## Метод наименьших квадратов

МОЁ ХОББИ: ЭКСТРАПОЛИРОВАТЬ





# Первое упоминание регрессии

Впервые регрессия упоминается в работе Гальтона

*"Регрессия к середине в наследственности роста", 1885 г.*





## Первое упоминание регрессии

Впервые регрессия упоминается в работе Гальтона

*"Регрессия к середине в наследственности роста"*, 1885 г.

$x$  — рост родителей,  $y$  — рост детей



## Первое упоминание регрессии

Впервые регрессия упоминается в работе Гальтона

*"Регрессия к середине в наследственности роста"*, 1885 г.

$x$  — рост родителей,  $y$  — рост детей

Установлена зависимость  $y - \bar{y} \approx \frac{2}{3}(x - \bar{x})$ , т.е. регрессия к середине.



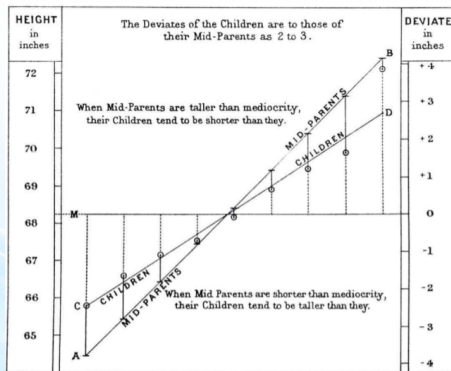
# Первое упоминание регрессии

Впервые регрессия упоминается в работе Гальтона

*"Регрессия к середине в наследственности роста", 1885 г.*

$x$  — рост родителей,  $y$  — рост детей

Установлена зависимость  $y - \bar{y} \approx \frac{2}{3}(x - \bar{x})$ , т.е. регрессия к середине.





# Задача регрессии: интуиция

Есть объект, обладающий признаками  $x$ .



## Задача регрессии: интуиция

Есть объект, обладающий признаками  $x$ .

*Примеры признаков: рост песика, экспрессия белка, энергия частицы.*



## Задача регрессии: интуиция

Есть объект, обладающий признаками  $x$ .

*Примеры признаков: рост песика, экспрессия белка, энергия частицы.*

Мы предполагаем, что есть зависимость какой-то численной характеристики объекта  $y$  от его признаков:

$$y \approx f(x)$$



## Задача регрессии: интуиция

Есть объект, обладающий признаками  $x$ .

*Примеры признаков: рост песика, экспрессия белка, энергия частицы.*

Мы предполагаем, что есть зависимость какой-то численной характеристики объекта  $y$  от его признаков:

$$y \approx f(x)$$

*Пример: зависимость уровня когнитивных способностей от параметров поражения мозга при рассеянном склерозе.*



## Задача регрессии: интуиция

Есть объект, обладающий признаками  $x$ .

*Примеры признаков: рост песика, экспрессия белка, энергия частицы.*

Мы предполагаем, что есть зависимость какой-то численной характеристики объекта  $y$  от его признаков:

$$y \approx f(x)$$

*Пример: зависимость уровня когнитивных способностей от параметров поражения мозга при рассеянном склерозе.*

Однако мы не знаем, какова эта зависимость на самом деле.





## Задача регрессии: интуиция

Есть объект, обладающий признаками  $x$ .

*Примеры признаков: рост песика, экспрессия белка, энергия частицы.*

Мы предполагаем, что есть зависимость какой-то численной характеристики объекта  $y$  от его признаков:

$$y \approx f(x)$$

*Пример: зависимость уровня когнитивных способностей от параметров поражения мозга при рассеянном склерозе.*

Однако мы не знаем, какова эта зависимость на самом деле.

На основании **данных** – набора объектов, для которых известны  $x$  и  $y$ , мы пытаемся "восстановить" зависимость:



# Пример

Пусть  $x$  — рост песика, а  $y$  — его вес.

Что мы знаем?



# Пример

Пусть  $x$  — рост песика, а  $y$  — его вес.

Что мы знаем?

- ▶ чем крупнее песик, тем больший вес он имеет;
- ▶ песики одинакового роста могут иметь разный вес.



# Пример

Пусть  $x$  — рост песика, а  $y$  — его вес.

Что мы знаем?

- ▶ чем крупнее песик, тем больший вес он имеет;
- ▶ песики одинакового роста могут иметь разный вес.

Выводы:

- ▶ для фиксированного роста песика  $x$  его вес  $y = f(x)$  является случайной величиной;
- ▶ в среднем вес  $f(x)$  возрастает при увеличении роста песика  $x$ .



# Пример

Пусть  $x$  — рост песика, а  $y$  — его вес.

Что мы знаем?

- ▶ чем крупнее песик, тем больший вес он имеет;
- ▶ песики одинакового роста могут иметь разный вес.

Выводы:

- ▶ для фиксированного роста песика  $x$  его вес  $y = f(x)$  является случайной величиной;
- ▶ в среднем вес  $f(x)$  возрастает при увеличении роста песика  $x$ .



# Пример

Простая зависимость:

$$y = \theta_0 + \theta_1 x + \varepsilon,$$



# Пример

Простая зависимость:

$$y = \theta_0 + \theta_1 x + \varepsilon,$$

$x$  — рост песика,



# Пример

Простая зависимость:

$$y = \theta_0 + \theta_1 x + \varepsilon,$$

$x$  — рост песика,

$y$  — вес песика,





# Пример

Простая зависимость:

$$y = \theta_0 + \theta_1 x + \varepsilon,$$

$x$  — рост песика,

$y$  — вес песика,

$\theta_0, \theta_1$  — неизвестные параметры,



# Пример

Простая зависимость:

$$y = \theta_0 + \theta_1 x + \varepsilon,$$

$x$  — рост песика,

$y$  — вес песика,

$\theta_0, \theta_1$  — неизвестные параметры,

$\varepsilon$  — случайная составляющая с нулевым средним (погрешность).



# Пример

Простая зависимость:

$$y = \theta_0 + \theta_1 x + \varepsilon,$$

$x$  — рост песика,

$y$  — вес песика,

$\theta_0, \theta_1$  — неизвестные параметры,

$\varepsilon$  — случайная составляющая с нулевым средним (погрешность).

Зависимость **линейна по параметрам**, линейна по аргументу.



# Пример

Простая зависимость:

$$y = \theta_0 + \theta_1 x + \varepsilon,$$

$x$  — рост песика,

$y$  — вес песика,

$\theta_0, \theta_1$  — неизвестные параметры,

$\varepsilon$  — случайная составляющая с нулевым средним (погрешность).

Зависимость **линейна по параметрам**, линейна по аргументу.



## Пример

Более сложная зависимость:

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_2^2 + \varepsilon,$$



## Пример

Более сложная зависимость:

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_2^2 + \varepsilon,$$

$x_1$  — рост песика,

$x_2$  — обхват туловища песика,

$y$  — вес песика,



## Пример

Более сложная зависимость:

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_2^2 + \varepsilon,$$

$x_1$  — рост песика,

$x_2$  — обхват туловища песика,

$y$  — вес песика,

$\theta_0, \theta_1, \theta_2, \theta_3$  — неизвестные параметры,

$\varepsilon$  — случайная составляющая с нулевым средним.

Зависимость **линейна по параметрам**, квадратична по аргументам.



## Пример

Более сложная зависимость:

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_2^2 + \varepsilon,$$

$x_1$  — рост песика,

$x_2$  — обхват туловища песика,

$y$  — вес песика,

$\theta_0, \theta_1, \theta_2, \theta_3$  — неизвестные параметры,

$\varepsilon$  — случайная составляющая с нулевым средним.

Зависимость **линейна по параметрам**, квадратична по аргументам.





# Модель линейной регрессии

Рассматриваем функциональную зависимость вида

$$y = y(x) = \theta_1 x_1 + \dots + \theta_d x_d$$



# Модель линейной регрессии

Рассматриваем функциональную зависимость вида

$$y = y(x) = \theta_1 x_1 + \dots + \theta_d x_d$$

$x_1, \dots, x_d$  — признаки ,

$\theta = (\theta_1, \dots, \theta_d)^T$  — вектор параметров.



# Модель линейной регрессии

Рассматриваем функциональную зависимость вида

$$y = y(x) = \theta_1 x_1 + \dots + \theta_d x_d$$

$x_1, \dots, x_d$  — признаки ,

$\theta = (\theta_1, \dots, \theta_d)^T$  — вектор параметров.

Для оценки  $\theta$  производится  $n$  испытаний вида

$$Y_i = \theta_1 x_{i1} + \dots + \theta_d x_{id} + \varepsilon_i, \quad i = 1, \dots, n,$$



# Модель линейной регрессии

Рассматриваем функциональную зависимость вида

$$y = y(x) = \theta_1 x_1 + \dots + \theta_d x_d$$

$x_1, \dots, x_d$  — признаки ,

$\theta = (\theta_1, \dots, \theta_d)^T$  — вектор параметров.

Для оценки  $\theta$  производится  $n$  испытаний вида

$$Y_i = \theta_1 x_{i1} + \dots + \theta_d x_{id} + \varepsilon_i, \quad i = 1, \dots, n,$$

$x_i = (x_{i1}, \dots, x_{id})$  — признаковые описания объекта  $i$   
(обычно неслучайные),



# Модель линейной регрессии

Рассматриваем функциональную зависимость вида

$$y = y(x) = \theta_1 x_1 + \dots + \theta_d x_d$$

$x_1, \dots, x_d$  — признаки ,

$\theta = (\theta_1, \dots, \theta_d)^T$  — вектор параметров.

Для оценки  $\theta$  производится  $n$  испытаний вида

$$Y_i = \theta_1 x_{i1} + \dots + \theta_d x_{id} + \varepsilon_i, \quad i = 1, \dots, n,$$

$x_i = (x_{i1}, \dots, x_{id})$  — признаковые описания объекта  $i$   
(обычно неслучайные),

$\varepsilon_i$  — случайная ошибка измерений.



# Модель линейной регрессии

Рассматриваем функциональную зависимость вида

$$y = y(x) = \theta_1 x_1 + \dots + \theta_d x_d$$

$x_1, \dots, x_d$  — признаки ,

$\theta = (\theta_1, \dots, \theta_d)^T$  — вектор параметров.

Для оценки  $\theta$  производится  $n$  испытаний вида

$$Y_i = \theta_1 x_{i1} + \dots + \theta_d x_{id} + \varepsilon_i, \quad i = 1, \dots, n,$$

$x_i = (x_{i1}, \dots, x_{id})$  — признаковые описания объекта  $i$   
(обычно неслучайные),

$\varepsilon_i$  — случайная ошибка измерений.



# Модель линейной регрессии

Введем обозначения

$$Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & \dots & x_{1d} \\ \dots & & \\ x_{n1} & \dots & x_{nd} \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$



# Модель линейной регрессии

Введем обозначения

$$Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & \dots & x_{1d} \\ \dots & & \\ x_{n1} & \dots & x_{nd} \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

Матричная форма записи проведенных испытаний





# Модель линейной регрессии

Введем обозначения

$$Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & \dots & x_{1d} \\ \dots & & \\ x_{n1} & \dots & x_{nd} \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

Матричная форма записи проведенных испытаний

$$Y = X\theta + \varepsilon.$$



# Модель линейной регрессии

Введем обозначения

$$Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & \dots & x_{1d} \\ \dots & & \\ x_{n1} & \dots & x_{nd} \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

Матричная форма записи проведенных испытаний

$$Y = X\theta + \varepsilon.$$

$X \in \mathbb{R}^{n \times d}$  — регрессоры (или матрица плана эксперимента),



# Модель линейной регрессии

Введем обозначения

$$Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & \dots & x_{1d} \\ \dots & & \\ x_{n1} & \dots & x_{nd} \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

Матричная форма записи проведенных испытаний

$$Y = X\theta + \varepsilon.$$

$X \in \mathbb{R}^{n \times d}$  — регрессоры (или матрица плана эксперимента),

$Y \in \mathbb{R}^n$  — отклик.



# Модель линейной регрессии

Введем обозначения

$$Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & \dots & x_{1d} \\ \dots & & \\ x_{n1} & \dots & x_{nd} \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

Матричная форма записи проведенных испытаний

$$Y = X\theta + \varepsilon.$$

$X \in \mathbb{R}^{n \times d}$  — регрессоры (или матрица плана эксперимента),

$Y \in \mathbb{R}^n$  — отклик.

Матричный вид зависимости:  $y(x) = x^T \theta$ .



# Модель линейной регрессии

Введем обозначения

$$Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & \dots & x_{1d} \\ \dots & & \\ x_{n1} & \dots & x_{nd} \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

Матричная форма записи проведенных испытаний

$$Y = X\theta + \varepsilon.$$

$X \in \mathbb{R}^{n \times d}$  — регрессоры (или матрица плана эксперимента),

$Y \in \mathbb{R}^n$  — отклик.

Матричный вид зависимости:  $y(x) = x^T \theta$ .



## Замечание

Зависимость  $y = y(x)$  должна быть **линейна по параметрам**,  
но не обязана быть линейной по признакам.



## Замечание

Зависимость  $y = y(x)$  должна быть **линейна по параметрам**,  
но не обязана быть линейной по признакам.

Пусть  $z_1, \dots, z_k$  — набор "независимых" переменных.



## Замечание

Зависимость  $y = y(x)$  должна быть **линейна по параметрам**,  
но не обязана быть линейной по признакам.

Пусть  $z_1, \dots, z_k$  — набор "независимых" переменных.

Можно рассматривать модель

$$y(x) = \theta_1 x_1(z_1, \dots, z_k) + \dots + \theta_d x_d(z_1, \dots, z_k),$$

где  $x_j(z_1, \dots, z_k)$  — некоторые функции (м.б. нелинейные).





## Замечание

Зависимость  $y = y(x)$  должна быть **линейна по параметрам**, но не обязана быть линейной по признакам.

Пусть  $z_1, \dots, z_k$  — набор "независимых" переменных.

Можно рассматривать модель

$$y(x) = \theta_1 x_1(z_1, \dots, z_k) + \dots + \theta_d x_d(z_1, \dots, z_k),$$

где  $x_j(z_1, \dots, z_k)$  — некоторые функции (м.б. нелинейные).

Примеры:

►  $x(z_1, \dots, z_k) = 1;$



## Замечание

Зависимость  $y = y(x)$  должна быть **линейна по параметрам**, но не обязана быть линейной по признакам.

Пусть  $z_1, \dots, z_k$  — набор "независимых" переменных.

Можно рассматривать модель

$$y(x) = \theta_1 x_1(z_1, \dots, z_k) + \dots + \theta_d x_d(z_1, \dots, z_k),$$

где  $x_j(z_1, \dots, z_k)$  — некоторые функции (м.б. нелинейные).

Примеры:

- ▶  $x(z_1, \dots, z_k) = 1$ ;
- ▶  $x(z_1, \dots, z_k) = \ln z_1$ ;



## Замечание

Зависимость  $y = y(x)$  должна быть **линейна по параметрам**, но не обязана быть линейной по признакам.

Пусть  $z_1, \dots, z_k$  — набор "независимых" переменных.

Можно рассматривать модель

$$y(x) = \theta_1 x_1(z_1, \dots, z_k) + \dots + \theta_d x_d(z_1, \dots, z_k),$$

где  $x_j(z_1, \dots, z_k)$  — некоторые функции (м.б. нелинейные).

Примеры:

▶  $x(z_1, \dots, z_k) = 1;$

▶  $x(z_1, \dots, z_k) = z_1;$

▶  $x(z_1, \dots, z_k) = \ln z_1;$



## Замечание

Зависимость  $y = y(x)$  должна быть **линейна по параметрам**, но не обязана быть линейной по признакам.

Пусть  $z_1, \dots, z_k$  — набор "независимых" переменных.

Можно рассматривать модель

$$y(x) = \theta_1 x_1(z_1, \dots, z_k) + \dots + \theta_d x_d(z_1, \dots, z_k),$$

где  $x_j(z_1, \dots, z_k)$  — некоторые функции (м.б. нелинейные).

Примеры:

▶  $x(z_1, \dots, z_k) = 1;$

▶  $x(z_1, \dots, z_k) = z_1;$

▶  $x(z_1, \dots, z_k) = \ln z_1;$

▶  $x(z_1, \dots, z_k) = z_1^2 z_2.$



## Замечание

Зависимость  $y = y(x)$  должна быть **линейна по параметрам**, но не обязана быть линейной по признакам.

Пусть  $z_1, \dots, z_k$  — набор "независимых" переменных.

Можно рассматривать модель

$$y(x) = \theta_1 x_1(z_1, \dots, z_k) + \dots + \theta_d x_d(z_1, \dots, z_k),$$

где  $x_j(z_1, \dots, z_k)$  — некоторые функции (м.б. нелинейные).

Примеры:

▶  $x(z_1, \dots, z_k) = 1;$

▶  $x(z_1, \dots, z_k) = z_1;$

▶  $x(z_1, \dots, z_k) = \ln z_1;$

▶  $x(z_1, \dots, z_k) = z_1^2 z_2.$



# Определение моментов инерции твёрдых тел с помощью трифилярного подвеса



## Определение моментов инерции твёрдых тел с помощью трифилярного подвеса

- ▶ На платформу помещается тело — диск, разрезанный по диаметру;



## Определение моментов инерции твёрдых тел с помощью трифилярного подвеса

- ▶ На платформу помещается тело — диск, разрезанный по диаметру;
  - ▶  $I$  — момент инерции тела;





## Определение моментов инерции твёрдых тел с помощью трифилярного подвеса

- ▶ На платформу помещается тело — диск, разрезанный по диаметру;
  - ▶  $I$  — момент инерции тела;
  - ▶  $m$  — масса тела;



## Определение моментов инерции твёрдых тел с помощью трифилярного подвеса

- ▶ На платформу помещается тело — диск, разрезанный по диаметру;
  - ▶  $I$  — момент инерции тела;
  - ▶  $m$  — масса тела;
  - ▶  $h$  — расстояние от половинок до оси вращения;



## Определение моментов инерции твёрдых тел с помощью трифилярного подвеса

- ▶ На платформу помещается тело — диск, разрезанный по диаметру;
  - ▶  $I$  — момент инерции тела;
  - ▶  $m$  — масса тела;
  - ▶  $h$  — расстояние от половинок до оси вращения;
  - ▶  $I_0$  — момент инерции нераздвинутого диска.



## Определение моментов инерции твёрдых тел с помощью трифилярного подвеса

- ▶ На платформу помещается тело — диск, разрезанный по диаметру;
  - ▶  $I$  — момент инерции тела;
  - ▶  $m$  — масса тела;
  - ▶  $h$  — расстояние от половинок до оси вращения;
  - ▶  $I_0$  — момент инерции нераздвинутого диска.
- ▶ Половинки диска постепенно раздвигаются;



## Определение моментов инерции твёрдых тел с помощью трифилярного подвеса

- ▶ На платформу помещается тело — диск, разрезанный по диаметру;
  - ▶  $I$  — момент инерции тела;
  - ▶  $m$  — масса тела;
  - ▶  $h$  — расстояние от половинок до оси вращения;
  - ▶  $I_0$  — момент инерции нераздвинутого диска.
- ▶ Половинки диска постепенно раздвигаются;
- ▶ Снимается зависимость момента инерции системы  $I$  от  $h$ .



## Определение моментов инерции твёрдых тел с помощью трифилярного подвеса

- ▶ На платформу помещается тело — диск, разрезанный по диаметру;
  - ▶  $I$  — момент инерции тела;
  - ▶  $m$  — масса тела;
  - ▶  $h$  — расстояние от половинок до оси вращения;
  - ▶  $I_0$  — момент инерции нераздвинутого диска.
- ▶ Половинки диска постепенно раздвигаются;
- ▶ Снимается зависимость момента инерции системы  $I$  от  $h$ .

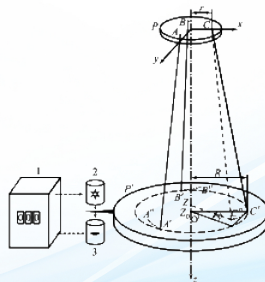


Рис. 2. Трифилярный подвес

По материалам "Модели и концепции физики: механика. Лабораторный практикум"



## Пример: Момент инерции

Согласно теореме Гюйгенса-Штейнера должно выполняться:

$$I = I_0 + mh^2$$



## Пример: Момент инерции

Согласно теореме Гюйгенса-Штейнера должно выполняться:

$$I = I_0 + mh^2$$

Итого, предполагается линейная зависимость момента инерции  $I$  от квадрата расстояния  $h^2$ . Мы хотим найти неизвестные  $m$  и  $I_0$ .





## Пример: Момент инерции

Согласно теореме Гюйгенса-Штейнера должно выполняться:

$$I = I_0 + mh^2$$

Итого, предполагается линейная зависимость момента инерции  $I$  от квадрата расстояния  $h^2$ . Мы хотим найти неизвестные  $m$  и  $I_0$ .

Наблюдения:  $I_i = I_0 + mh_i^2 + \varepsilon_i$ , где  $\varepsilon_i$  — погрешность.



## Пример: Момент инерции

Согласно теореме Гюйгенса-Штейнера должно выполняться:

$$I = I_0 + mh^2$$

Итого, предполагается линейная зависимость момента инерции  $I$  от квадрата расстояния  $h^2$ . Мы хотим найти неизвестные  $m$  и  $I_0$ .

Наблюдения:  $I_i = I_0 + mh_i^2 + \varepsilon_i$ , где  $\varepsilon_i$  — погрешность.

В данном примере  $x_1(h) = 1$ ,  $x_2(h) = h^2$ ,

$$X = \begin{pmatrix} 1 & h_1^2 \\ \dots & \dots \\ 1 & h_n^2 \end{pmatrix}, Y = \begin{pmatrix} I_1 \\ \dots \\ I_n \end{pmatrix}, \theta = \begin{pmatrix} I_0 \\ m \end{pmatrix}.$$



## Пример: Момент инерции

Согласно теореме Гюйгенса-Штейнера должно выполняться:

$$I = I_0 + mh^2$$

Итого, предполагается линейная зависимость момента инерции  $I$  от квадрата расстояния  $h^2$ . Мы хотим найти неизвестные  $m$  и  $I_0$ .

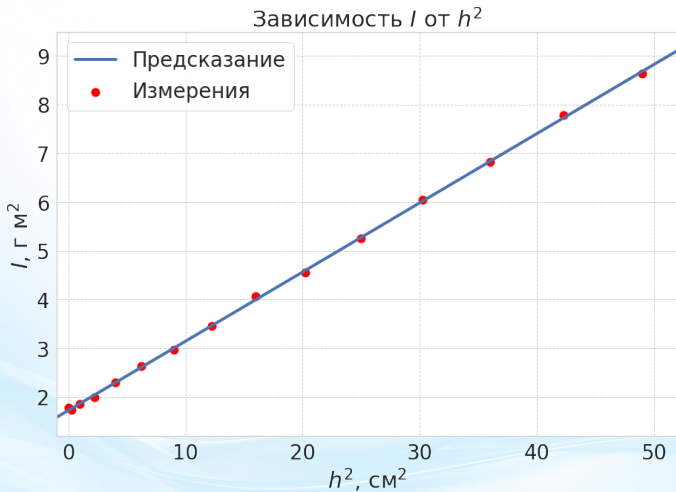
Наблюдения:  $I_i = I_0 + mh_i^2 + \varepsilon_i$ , где  $\varepsilon_i$  — погрешность.

В данном примере  $x_1(h) = 1$ ,  $x_2(h) = h^2$ ,

$$X = \begin{pmatrix} 1 & h_1^2 \\ \dots & \dots \\ 1 & h_n^2 \end{pmatrix}, Y = \begin{pmatrix} I_1 \\ \dots \\ I_n \end{pmatrix}, \theta = \begin{pmatrix} I_0 \\ m \end{pmatrix}.$$

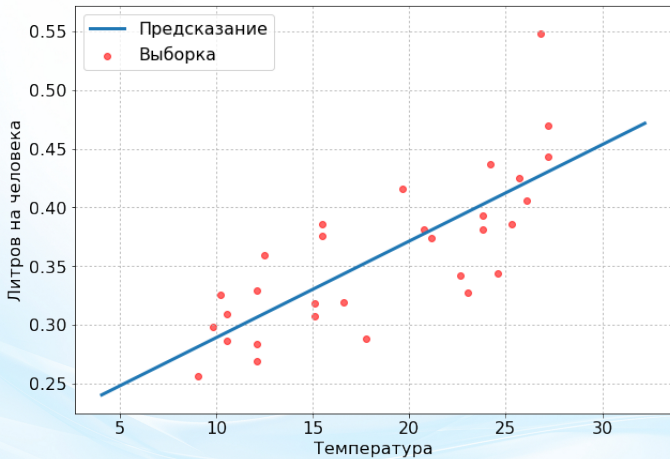


## Пример: Момент инерции





## Пример: Потребление мороженого





# Метод наименьших квадратов

Зависимость:  $y(x) = x^T \theta$ ,  $\theta \in \mathbb{R}^d$ .



# Метод наименьших квадратов

Зависимость:  $y(x) = x^T \theta$ ,  $\theta \in \mathbb{R}^d$ .

Испытания:  $Y = X\theta + \varepsilon$ ,  $X \in \mathbb{R}^{n \times d}$ ,  $Y \in \mathbb{R}^n$ .



# Метод наименьших квадратов

Зависимость:  $y(x) = x^T \theta$ ,  $\theta \in \mathbb{R}^d$ .

Испытания:  $Y = X\theta + \varepsilon$ ,  $X \in \mathbb{R}^{n \times d}$ ,  $Y \in \mathbb{R}^n$ .

Хотим как-то **оценить** параметр  $\theta$  на основании полученных данных.





# Метод наименьших квадратов

Зависимость:  $y(x) = x^T \theta$ ,  $\theta \in \mathbb{R}^d$ .

Испытания:  $Y = X\theta + \varepsilon$ ,  $X \in \mathbb{R}^{n \times d}$ ,  $Y \in \mathbb{R}^n$ .

Хотим как-то **оценить** параметр  $\theta$  на основании полученных данных.

Пусть  $\hat{\theta} = \hat{\theta}(X, Y)$  — наша оценка  $\theta$ .



# Метод наименьших квадратов

Зависимость:  $y(x) = x^T \theta$ ,  $\theta \in \mathbb{R}^d$ .

Испытания:  $Y = X\theta + \varepsilon$ ,  $X \in \mathbb{R}^{n \times d}$ ,  $Y \in \mathbb{R}^n$ .

Хотим как-то **оценить** параметр  $\theta$  на основании полученных данных.

Пусть  $\hat{\theta} = \hat{\theta}(X, Y)$  — наша оценка  $\theta$ .

Как понять, что она хорошая?



# Метод наименьших квадратов

Зависимость:  $y(x) = x^T \theta$ ,  $\theta \in \mathbb{R}^d$ .

Испытания:  $Y = X\theta + \varepsilon$ ,  $X \in \mathbb{R}^{n \times d}$ ,  $Y \in \mathbb{R}^n$ .

Хотим как-то **оценить** параметр  $\theta$  на основании полученных данных.

Пусть  $\hat{\theta} = \hat{\theta}(X, Y)$  — наша оценка  $\theta$ .

Как понять, что она хорошая?

Метрика MSE:

$$MSE(\hat{\theta}) = \left\| Y - X\hat{\theta} \right\|^2$$



# Метод наименьших квадратов

Зависимость:  $y(x) = x^T \theta$ ,  $\theta \in \mathbb{R}^d$ .

Испытания:  $Y = X\theta + \varepsilon$ ,  $X \in \mathbb{R}^{n \times d}$ ,  $Y \in \mathbb{R}^n$ .

Хотим как-то **оценить** параметр  $\theta$  на основании полученных данных.

Пусть  $\hat{\theta} = \hat{\theta}(X, Y)$  — наша оценка  $\theta$ .

Как понять, что она хорошая?

Метрика MSE:

$$MSE(\hat{\theta}) = \left\| Y - X\hat{\theta} \right\|^2$$

Оценка  $\hat{\theta} = \arg \min_{\theta} MSE(\hat{\theta})$  называется **оценкой по методу наименьших квадратов** параметра  $\theta$ .



# Метод наименьших квадратов

**Теорема.** Если матрица  $X^T X$  невырождена, то  $\hat{\theta} = (X^T X)^{-1} X^T Y$ .



# Метод наименьших квадратов

**Теорема.** Если матрица  $X^T X$  невырождена, то  $\hat{\theta} = (X^T X)^{-1} X^T Y$ .

$$MSE(\theta) = \|Y - X\theta\|^2 = (Y - X\theta)^T (Y - X\theta) = Y^T Y - 2Y^T X\theta + \theta^T X^T X\theta$$



# Метод наименьших квадратов

**Теорема.** Если матрица  $X^T X$  невырождена, то  $\hat{\theta} = (X^T X)^{-1} X^T Y$ .

$$MSE(\theta) = \|Y - X\theta\|^2 = (Y - X\theta)^T (Y - X\theta) = Y^T Y - 2Y^T X\theta + \theta^T X^T X\theta$$

Берем производную по  $\theta$  и приравниваем ее к нулю.



# Метод наименьших квадратов

**Теорема.** Если матрица  $X^T X$  невырождена, то  $\hat{\theta} = (X^T X)^{-1} X^T Y$ .

$$MSE(\theta) = \|Y - X\theta\|^2 = (Y - X\theta)^T (Y - X\theta) = Y^T Y - 2Y^T X\theta + \theta^T X^T X\theta$$

Берем производную по  $\theta$  и приравниваем ее к нулю.

$$\frac{\partial MSE(\theta)}{\partial \theta} = -2Y^T X + 2\theta^T X^T X = 0$$

Отсюда получается утверждение теоремы. □





# Метод наименьших квадратов

**Теорема.** Если матрица  $X^T X$  невырождена, то  $\hat{\theta} = (X^T X)^{-1} X^T Y$ .

$$MSE(\theta) = \|Y - X\theta\|^2 = (Y - X\theta)^T (Y - X\theta) = Y^T Y - 2Y^T X\theta + \theta^T X^T X\theta$$

Берем производную по  $\theta$  и приравниваем ее к нулю.

$$\frac{\partial MSE(\theta)}{\partial \theta} = -2Y^T X + 2\theta^T X^T X = 0$$

Отсюда получается утверждение теоремы. □

Предсказанием отклика на новом объекте  $x$  будет величина  $\hat{y}(x) = x^T \hat{\theta}$ .



# Реализация в sklearn

```
m = sklearn.linear_model.LinearRegression(fit_intercept=True)
```

Обучение модели:

```
m.fit(X, Y)
```

Вектор коэффициентов:

```
m.coef_
```

Свободный коэффициент:

```
m.intercept_
```

Предсказания:

```
m.predict(X)
```



# Метрики качества в задаче регрессии



# Обозначения

Пусть  $x_1, \dots, x_n$  — признаковые описания объектов;

$Y = (Y_1, \dots, Y_n)^T$  — наблюдения.



# Обозначения

Пусть  $x_1, \dots, x_n$  — признаковые описания объектов;

$Y = (Y_1, \dots, Y_n)^T$  — наблюдения.

Пусть  $\hat{f}(x)$  — оцененная нами зависимость.

*В случае линейной регрессии  $\hat{f}(x) = x^T \hat{\theta}$ .*



# Обозначения

Пусть  $x_1, \dots, x_n$  — признаковые описания объектов;

$Y = (Y_1, \dots, Y_n)^T$  — наблюдения.

Пусть  $\hat{f}(x)$  — оцененная нами зависимость.

*В случае линейной регрессии  $\hat{f}(x) = x^T \hat{\theta}$ .*

Пусть  $\hat{Y}_i = \hat{f}(x_i)$  — предсказание нашей модели на  $i$ -м объекте;

$\hat{Y} = (\hat{Y}_1, \dots, \hat{Y}_n)^T$ .



# Метрики качества в задаче регрессии

$Y$  — реальные наблюдения,  $\hat{Y}$  — предсказания.



# Метрики качества в задаче регрессии

$Y$  — реальные наблюдения,  $\hat{Y}$  — предсказания.

► MSE (Mean Squared Error):

$$MSE(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$





# Метрики качества в задаче регрессии

$Y$  — реальные наблюдения,  $\hat{Y}$  — предсказания.

- ▶ MSE (Mean Squared Error):

$$MSE(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

- ▶ MAE (Mean Absolute Error):

$$MAE(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$



# Метрики качества в задаче регрессии

$Y$  — реальные наблюдения,  $\hat{Y}$  — предсказания.

- ▶ MSE (Mean Squared Error):

$$MSE(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

- ▶ MAE (Mean Absolute Error):

$$MAE(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

- ▶ MAPE (Mean Absolute Percentage Error):

$$MAPE(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| * 100\%$$



# Недообучение vs Переобучение

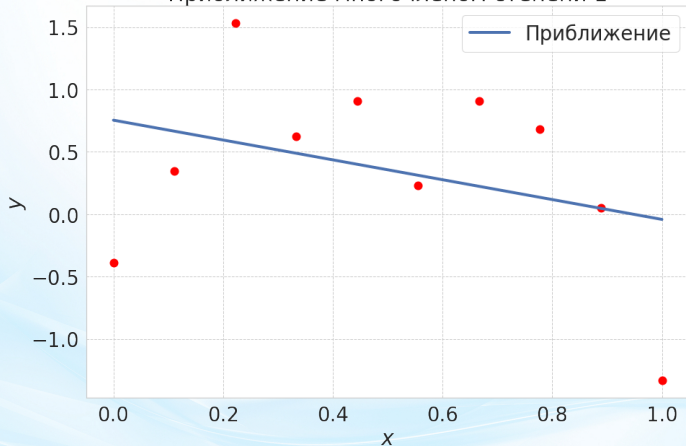
Зависимость:  $y = 5x - 6x^2$ , имеется погрешность



# Недообучение vs Переобучение

Зависимость:  $y = 5x - 6x^2$ , имеется погрешность

Приближение многочленом степени 1

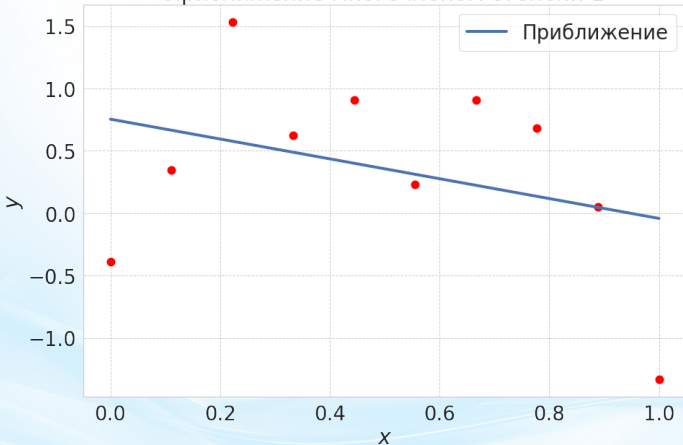




# Недообучение vs Переобучение

Зависимость:  $y = 5x - 6x^2$ , имеется погрешность

Приближение многочленом степени 1



Недообучение



# Недообучение vs Переобучение

Зависимость:  $y = 5x - 6x^2$ , имеется погрешность

Приближение многочленом степени 10





# Недообучение vs Переобучение

Зависимость:  $y = 5x - 6x^2$ , имеется погрешность

Приближение многочленом степени 10



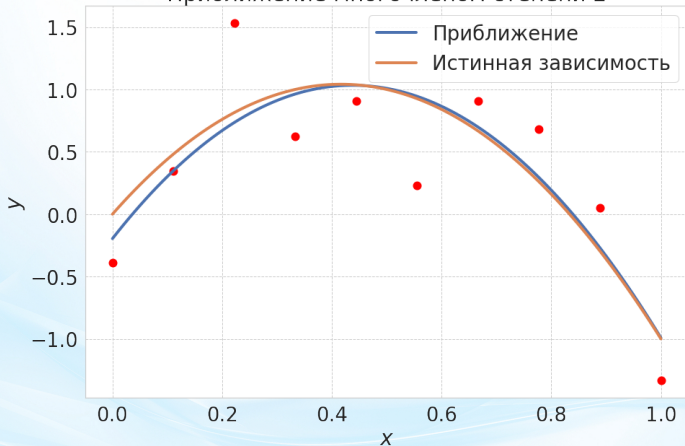
Переобучение



# Недообучение vs Переобучение

Зависимость:  $y = 5x - 6x^2$ , имеется погрешность

Приближение многочленом степени 2



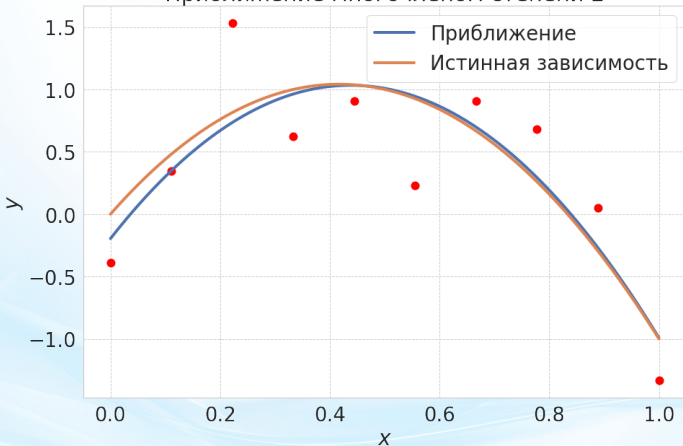




# Недообучение vs Переобучение

Зависимость:  $y = 5x - 6x^2$ , имеется погрешность

Приближение многочленом степени 2



Нормально!



# Тренировочная и тестовая выборки

Если все время работать с одной и той же выборкой  
(это жаргон, корректно понимать "реализацией выборки")  
и все больше улучшать модель, "подгонять" ее под выборку,  
может возникнуть переобучение.



# Тренировочная и тестовая выборки

Если все время работать с одной и той же выборкой (это жаргон, корректно понимать "реализацией выборки") и все больше улучшать модель, "подгонять" ее под выборку, может возникнуть переобучение.

Предсказание на **новом** объекте может быть неадекватным.



# Тренировочная и тестовая выборки

Если все время работать с одной и той же выборкой (это жаргон, корректно понимать "реализацией выборки") и все больше улучшать модель, "подгонять" ее под выборку, может возникнуть переобучение.

Предсказание на **новом** объекте может быть неадекватным.

Поэтому перед началом работы имеющиеся данные делят на две части:

**тренировочную (обучающую)** и **тестовую** выборки.



# Тренировочная и тестовая выборки

Если все время работать с одной и той же выборкой (это жаргон, корректно понимать "реализацией выборки") и все больше улучшать модель, "подгонять" ее под выборку, может возникнуть переобучение.

Предсказание на **новом** объекте может быть неадекватным.

Поэтому перед началом работы имеющиеся данные делят на две части:

**тренировочную (обучающую)** и **тестовую** выборки.





# Тренировочная и тестовая выборки

Если все время работать с одной и той же выборкой (это жаргон, корректно понимать "реализацией выборки") и все больше улучшать модель, "подгонять" ее под выборку, может возникнуть переобучение.

Предсказание на **новом** объекте может быть неадекватным.

Поэтому перед началом работы имеющиеся данные делят на две части:

**тренировочную (обучающую)** и **тестовую** выборки.



На тренировочной выборке происходит **обучение** моделей (например, оценка коэффициентов в линейной регрессии).



# Тренировочная и тестовая выборки

Если все время работать с одной и той же выборкой (это жаргон, корректно понимать "реализацией выборки") и все больше улучшать модель, "подгонять" ее под выборку, может возникнуть переобучение.

Предсказание на **новом** объекте может быть неадекватным.

Поэтому перед началом работы имеющиеся данные делят на две части:

**тренировочную (обучающую)** и **тестовую** выборки.



На тренировочной выборке происходит **обучение** моделей (например, оценка коэффициентов в линейной регрессии).

На тестовой выборке происходит **оценка качества** итоговой модели с использованием метрик качества.

