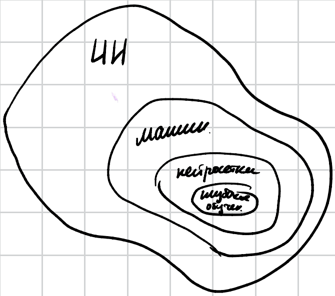


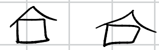
1.  – 5

№ недели	кол-во посещ.
1	5
2	6
3	4
4	5
5	?

равновероятные

4–6 – случ.
5 – среднее арифмет.

5. потому что хочет есть в течении недели одинаково, 5 посещение

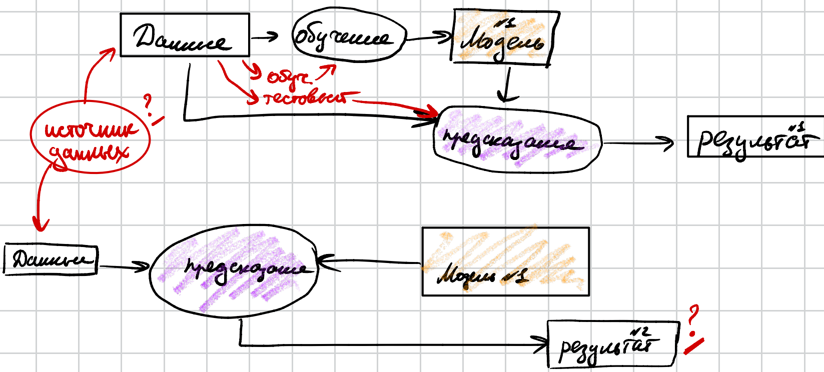
2. 

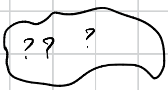
№ нед.	счет	
1	3:2	5
2	3:3	6
3	1:3	4
4	2:3	5
5		? 4–6 5

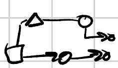
набор данных тот же

неравновероятные

ответ 5 неверен
не равномерное распределение
а пуассона



1.  Постановка задачи

2. $\int_a^b f(x)\,dx$, , 

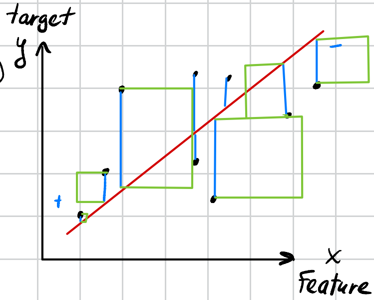
3. Программные конструкции – атомарное решение

4. Собственные силы

5. $\boxed{\text{Scikit}}$ — O

библиотеки

2 лек.

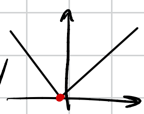


Определение: остатки (отклонения, ошибки) – это расс-ние между точками данных и вычисленными (по вертикали) точками на прямой

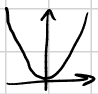
$$y = w_0 + w_1 x$$

$R_1, R_2, \dots, R_{10}$

$\cancel{\sum_i R_i}$ $\quad \sum_i |R_i|$



★ $\sum_i R_i^2$



Задача состоит в минимизации остатков $\Rightarrow$ разрыв между прямой и точками будет миним.

Обучение модели

$\downarrow$

минимизация функции потерь

$$y = w_0 + w_1 x$$

1) Аналитическое решение

Решение $\to$ числен. – проще, доступнее (вычисл), не точное решение

$\hookrightarrow$ аналит. – точное (мат. теория)

(x_i, y_i)

n - число точек

$$w_1 = \frac{n \cdot \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n \cdot \sum_i x_i^2 - \left(\sum x_i\right)^2}$$

$$w_0 = \frac{\sum_i y_i}{n} - w_1 \frac{\sum x_i}{n}$$

2) Метод обратных матриц

$$y = w_1 \cdot x + w_0 \qquad \vec{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \qquad \vec{w} = (X^T \cdot X)^{-1} X^T \cdot y$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

3) Разложение матриц

$$X = Q \cdot R \rightarrow \vec{w} = R^{-1} \cdot Q^{T} \cdot y$$

QR-разложение — приближ. вычисл. → миним. ошибку вычисл.

4) Градиентный спуск

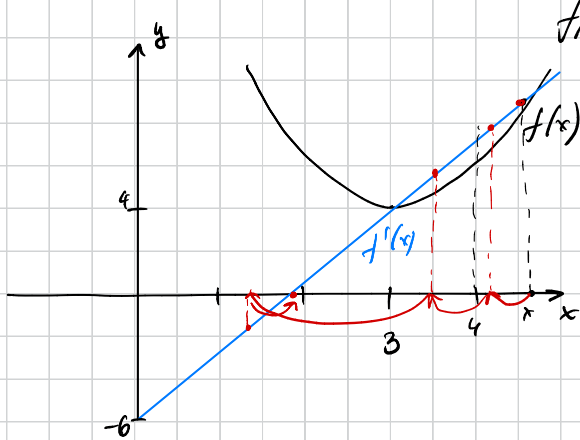
Метод оптимизации, где использ. производ. и итерации

Частная производная (по одному из параметров) позволяет определить угловой коэффициент и изменению параметра выполняется в ту сторону, где он max/min

Для больших угловых коэфф. делается более широкий шаг, для маленьких меньше. Ширина шага вычисляется как доля от углов. коэфф. — скорость обучения. Чем выше скорость, тем быстрее будет работать система, за счёт снижения точности, чем ниже скорость, тем больше времени потратим на обучение, но точность будет выше. 41:44

$f(x) = (x-3)^2 + 4$

Найти x, где $f(x)$ — min



$f(x) = x^2 - 6x + 9 - 4$

$f'(x) = 2x - 6 = 2(x-3)$

Градиентный спуск работает с линейной регрессией

$y = w_0 + w_1 x$ миним. остатки $\sum_i R_i^2$,

(x_1, y_1)
(x_2, y_2)
(x_3, y_3)
$\vdots$
(x_n, y_n)

$$E(w_1, w_0) = \sum R_i^2 = \sum_{i=1}^{n} \left(y_i - (w_0 + w_1 x)\right)^2$$

функция потерь.

$$\frac{d}{dw_1} E(w_1, w_0) \quad ; \quad \frac{d}{dw_0} E(w_1, w_0)$$

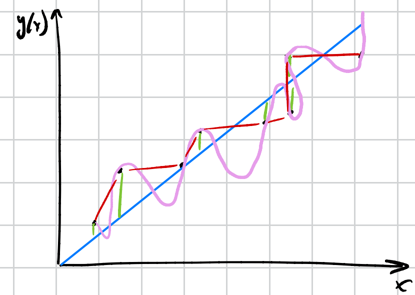
$$E(w_1 w_0) = \sum_{i=1}^{n} \left(y_i^2 - 2y_i(w_0 + w_1 x_i) + (w_0 + w_1 x_i)^2\right) = \sum_{i=1}^{n} y_i^2 - 2y_i w_0 - 2y_i w_1 x_i + w_0^2 + 2w_0 w_1 x_i + w_1^2 x_i^2)$$

$$\frac{d}{dw_0} E(w_1 w_0) = \sum_{i=1}^{n} -2y_i + 2w_0 + 2w_1 x_i) = -2 \sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)$$

$$\frac{d}{dw_1} E(w_1 w_0) = \sum_{i=1}^{n} (-2y_i x_i + 2w_0 x_i + 2w_1 x_i^2) = 2 \sum_{i=1}^{n} -y_i x_i + w_0 x_i + w_1 x_i^2 = 2 \sum_{i=1}^{n} x_i (y_i + w_0 + w_1 x_i)$$

1.04 lec_3.

Переобучение и дисперсия



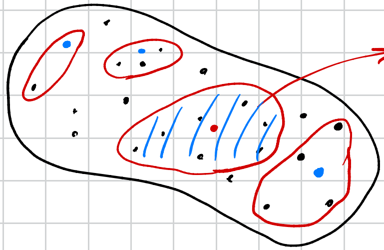
Линейная регрессия $\min \sum R^2$

сильно переобученная, но точно подстроил регрессию под обучающие данные

Цель состоит не в минимизац., а в том, чтобы делать правильн. предсказ. на новых данных.

Переобуч. модели очень чувств. к выбросам, которые находятся далеко от остальн. точек, в результате будет высокая дисперсия.

Поэтому к моделям специально добавляется смещение.



Смещение модели означ., что при попытке построить модель отдаётся определ. схеме (например, прямая линия, точка (0,0))

не график со сложной структурой миним. $\sum R^2$

Если в модель добавить смещение, есть риск недообучения.

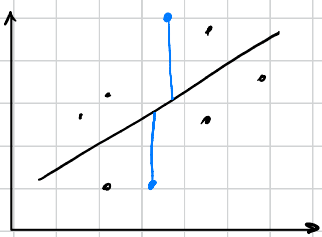
Балансировка: миним. функции потерь → переобучение, смещение → недообучение

2 варианта регрессии:

1) Гребневая регрессия (Ridge) добавляется смещение в виде штрафа, из-за этого хуже идёт подгонка под имеющ. данные

2) Лассо-регрессия. – удаление некот. переменных.

Механизм: применить линейную регрес. к данным, сделать на основе получ. модели прогноз., и думать что все в порядке.



Коэффициент корреляции. (от -1 до 1)

$$r = \frac{n\sum_i^n x_i y_i - \left(\sum_i^n x_i\right)\left(\sum_i^n y_i\right)}{\sqrt{n\cdot\sum_i^n (x_i)^2 - \left(\sum_i^n x_i\right)^2}\cdot\sqrt{n\cdot\sum_i^n y_i^2 - \left(\sum_i^n y_i\right)^2}}$$

Коэфф. детерминации

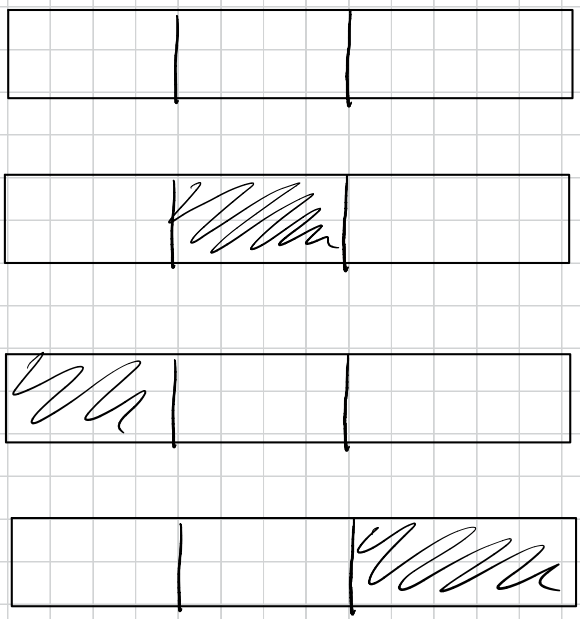
$$r^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}$$

y_i — факт. знач. (тест. выбор.)

$\hat{y}_i$ — предсказан. знач.

$\bar{y}_i$ — средн. знач. для всех точек y.

$0 < r^2 < 1$ чем ближе к 1, тем лучше регрессия работает на тестовых данных



3 раза обучают и 3 тестируют на тестир. данных.

$$y = a_0 + a_1 x_1 + a_2 x_2$$