

Линейная регрессия

$$y = kx + b$$

$$Q(x) = w_0 + w_1 x_1 + w_2 x_2 + \dots$$

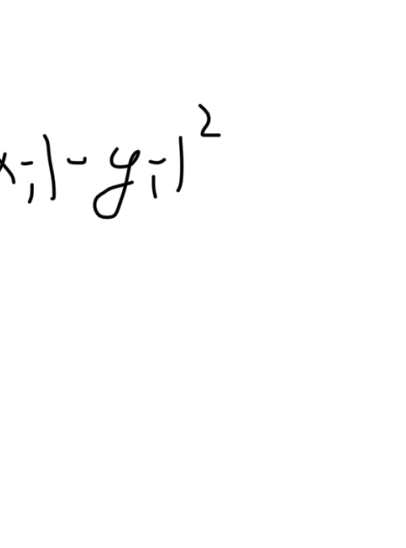
x_1 - площадь

x_2 - количество комнат

$$Q(x) = w_0 + \sum_{i=1}^n w_i x_i$$

$$x_0 = 1$$

$$Q(x) = \sum_{i=0}^n w_i x_i = (w, x)$$



MSE

$$Q(w, x) = \frac{1}{L} \sum_{i=1}^L (w_i x_i - y_i)^2$$

min
w

$$Q(x) = w_0 x_0 + w_1 x_1 + w_2 x_2 + \dots$$

w_0 - район k, b, c, d

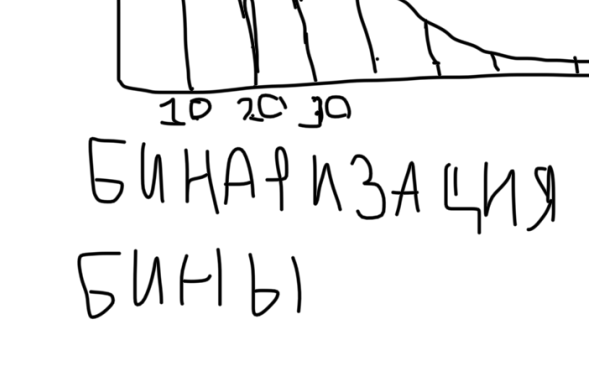
One hot encoding (OHE)

x_p	$x_{p1}, x_{p1}, x_{p3}, x_{p4}$
A	1, 0, 0, 0
B	0, 1, 0, 0
B	0, 1, 0, 0
A	1, 0, 0, 0
C	0, 0, 1, 0
D	0, 0, 0, 1

x_{p1} - район = A?

w_{p1} - добавка к цене за район = A

РАССТОЯНИЕ ДО МКАД



Бинаризация

Биты

$$Q(0-10), Q(10-20), Q(20-30), Q(>30)$$

1	0	0	0
0	1	0	0
-	-	-	-

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

Аналитика: $w = (X^T X)^{-1} X^T y$

X - матрица объектов

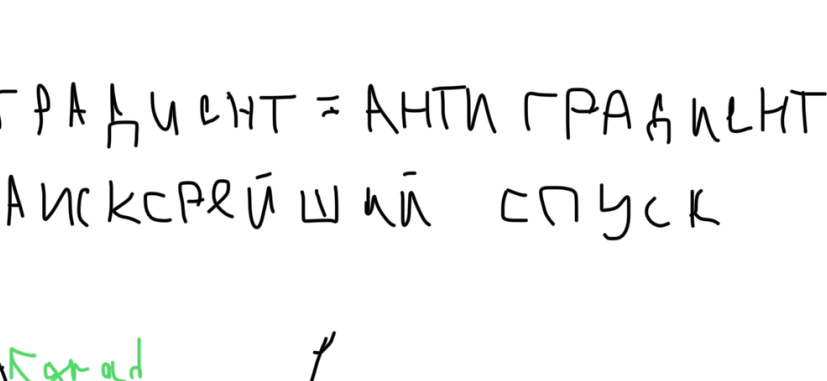
$O(N^3)$ - обращение матрицы

от числа признаков

$X^T X$ - может необратима?

X :

	S	d	h	-	-	-
кв1	20	1	1			
кв2	30	5	12			
кв3						
кв4						

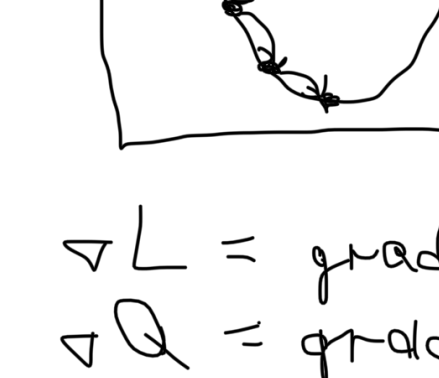


ГРАДИЕНТ = ВЕКТОР

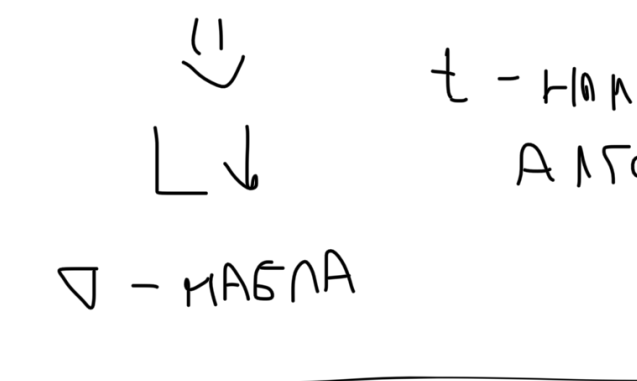
НАИСКОРЕЙШЕГО РОСТА

-ГРАДИЕНТ = АНТИ ГРАДИЕНТ =

НАИСКРЕЙШИЙ СПУСК



ГРАД. СПУСК:



$\nabla L = \text{grad } L$ - градиент

$\nabla Q = \text{grad } Q$ - функции потерь

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

t - номер итерации

L - алгоритм

η - шаг

$$\frac{\partial f}{\partial x} > 0 \quad f \uparrow$$

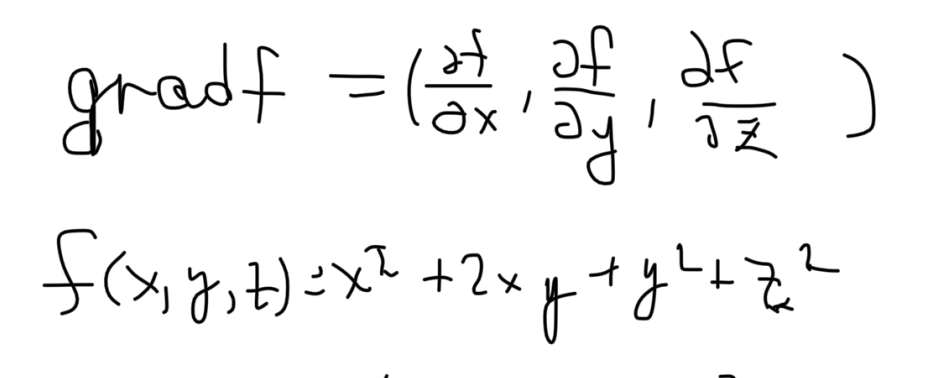
$$\frac{\partial f}{\partial x} < 0 \quad f \downarrow$$

$$\frac{\partial f}{\partial x} = 0 \quad \text{экстремум}$$

$$f(x, y, z)$$

$$\text{grad } f(x, y, z):$$

$$\frac{\partial f}{\partial x} \vec{i} + \frac{\partial f}{\partial y} \vec{j} + \frac{\partial f}{\partial z} \vec{k}$$



$$\text{grad } f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)$$

$$f(x, y, z) = x^2 + 2xy + y^2 + z^2$$

$$\text{grad } f = (2x + 2y, 2x + 2y, 2z)$$

$$\text{grad } f =$$

$$(2x + 2y + 0 + 0,$$

$$0 + 2x + 2y + 0,$$

$$0 + 0 + 0 + 2z)$$

$$Q(w, x) = w_0 x_0 + w_1 x_1 + w_2 x_2$$

$$x_0 = 1$$

$$L(w, x, y) =$$

$$= (w_0 x_0 + w_1 x_1 + w_2 x_2 - y)^2 =$$

$$= w_0^2 x_0^2 + w_1^2 x_1^2 + w_2^2 x_2^2 +$$

$$- 2w_0 x_0 y - 2w_1 x_1 y - 2w_2 x_2 y +$$

$$2w_0 w_1 x_0 x_1 + 2w_0 w_2 x_0 x_2 +$$

$$2w_1 w_2 x_1 x_2 + w_1^2 x_1^2 + w_2^2 x_2^2 +$$

$$y^2 - 2w_0 x_0 y - 2w_1 x_1 y - 2w_2 x_2 y$$

$$= w_0^2 x_0^2 + 2w_0 w_1 x_0 x_1 + 2w_0 w_2 x_0 x_2 +$$

$$2w_1 w_2 x_1 x_2 + w_1^2 x_1^2 + w_2^2 x_2^2 +$$

$$y^2 - 2w_0 x_0 y - 2w_1 x_1 y - 2w_2 x_2 y$$

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

$$\frac{1}{L} \sum_{i=1}^L (w_i x_i - y_i)^2$$

$$\text{grad } L$$

$$(2w_0 x_0^2 + 2w_1 x_0 x_1 + 2w_2 x_0 x_2 - 2x_0 y,$$

$$2w_1 x_1^2 + 2w_0 x_0 x_1 + 2w_2 x_1 x_2 - 2x_1 y,$$

$$2w_2 x_2^2 + 2w_0 x_0 x_2 + 2w_1 x_1 x_2 - 2x_2 y)$$

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

$$w_0^{(t)} = w_0^{(t-1)} -$$

$$- (2w_0 x_0^2 + 2w_1 x_0 x_1 + 2w_2 x_0 x_2 - 2x_0 y)$$

$$w_j^{(0)} = U[-a, a]$$

uniform

a - noise

Learning rate

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

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

$$\eta = \text{const}$$

$$\eta_t = \frac{\text{const}}{t+1}$$

$$\eta_t = g(t)$$

SGD

Stochastic GD

1 объект на шаг

mini batch SGD

N объектов

$$N \sim 32, 64$$

Adam, AdamW

