1) $y = 10x$

x	y
1 стр.	30 р.
2 стр.	60 р.
3 стр.	90 р.

$C = ?$

$y = C \cdot x$ — цена за стр. (C), цена (y), кол-во стр. (x)

2) 5 страниц $\Rightarrow x = 5$ $\Rightarrow y = 150$ — вывод inference

$\hat{y} = f(x, C)$

$\{(1, 30), (2, 60), (3, 90)\} \Rightarrow C = 30$ — обучение learning — поиск параметра, при котором лучше всех описывают данные

3) Функция ошибки — "что такое хорошо / плохо"

$D = \{(1, 30), (2, 60), (3, 90)\}$

$C = 10$

x	$\hat{y} = C \cdot x$	y	Loss
1	10	30	20
2	20	60	40
3	30	90	60

Avg. 40

$$L = \frac{1}{N} \sum_{n=1}^{N} |y - \hat{y}|$$

$f(x)$

MAE — mean absolute error

x	y	$\hat{y}$	Loss
1	30	20	10
2	60	70	10
3	90	90	0

Avg 20/3

$$L(C) = \frac{1}{N} \sum_{n} |y - C \cdot x|$$

x	y	$\hat{y}$ для $C=10$	$C=20$	$C=30$	$\hat{y}$ для $C=40$
1	30	10	20	30	40
2	60	20	40	60	80
3	90	30	60	90	120
		40	$\frac{1}{3}(10+20+30)$	0	20
			20		

4) Подбор с переходом к соседям

5) База данных и разные модели

x	y
1	30
2	60
3	90

$\Rightarrow$ $y = c \cdot x$, $c = 30$

модель $y = c \cdot x$ $\Rightarrow$ $c = 30$

предположение

$y = 30 \cdot x + 10 \cdot (x-1)(x-2)(x-3)$

6) Два параметра

X	y
1	50
2	80
3	110

$y = c \cdot x$

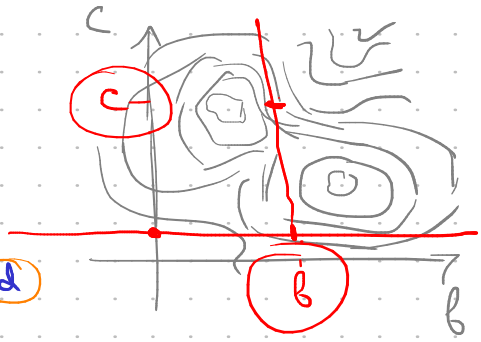
$c = 40 \Rightarrow Avg. = \frac{1}{3}(10+0+10) = 20/3$

$c = 50 \Rightarrow Avg = 1/3(0+20+40) = 20$

$c^* = 36\frac{2}{3} \Rightarrow Avg.MAE > 0$

$y = c \cdot x + b \Rightarrow b = 20,\ c = 30,\ Avg.MAE = 0$

$L(b,c) = \frac{1}{N} \sum_n l(y_n, x_n, b, c) = \frac{1}{N} \sum_n |y_n - (cx_n + b)|$

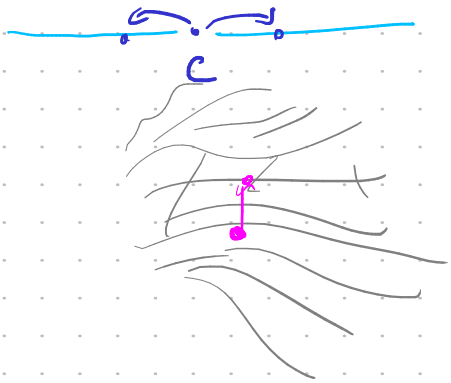


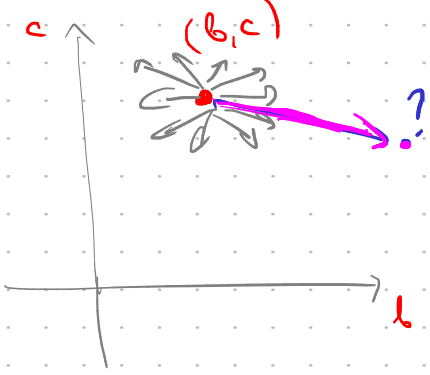
7) Откуда берутся ошибки

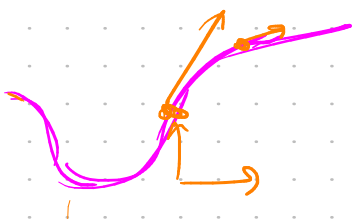
8) Почему бы не перебрать?

$y \sim w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_d x_d$

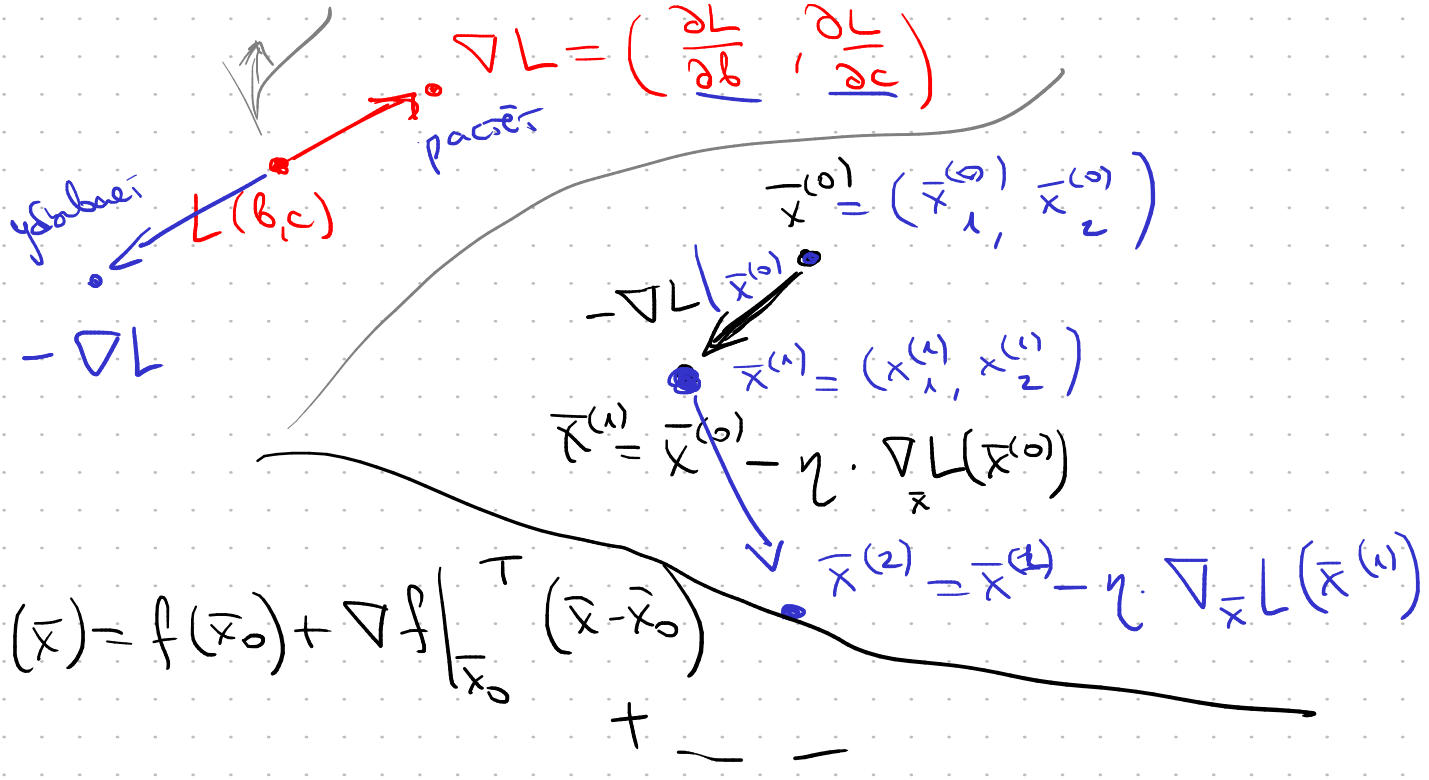
9) Локальный поиск







$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + \ldots$



$$f(\bar{x}) = f(\bar{x}_0) + \nabla f\Big|_{\bar{x}_0}^{T} (\bar{x} - \bar{x}_0) + \dots$$