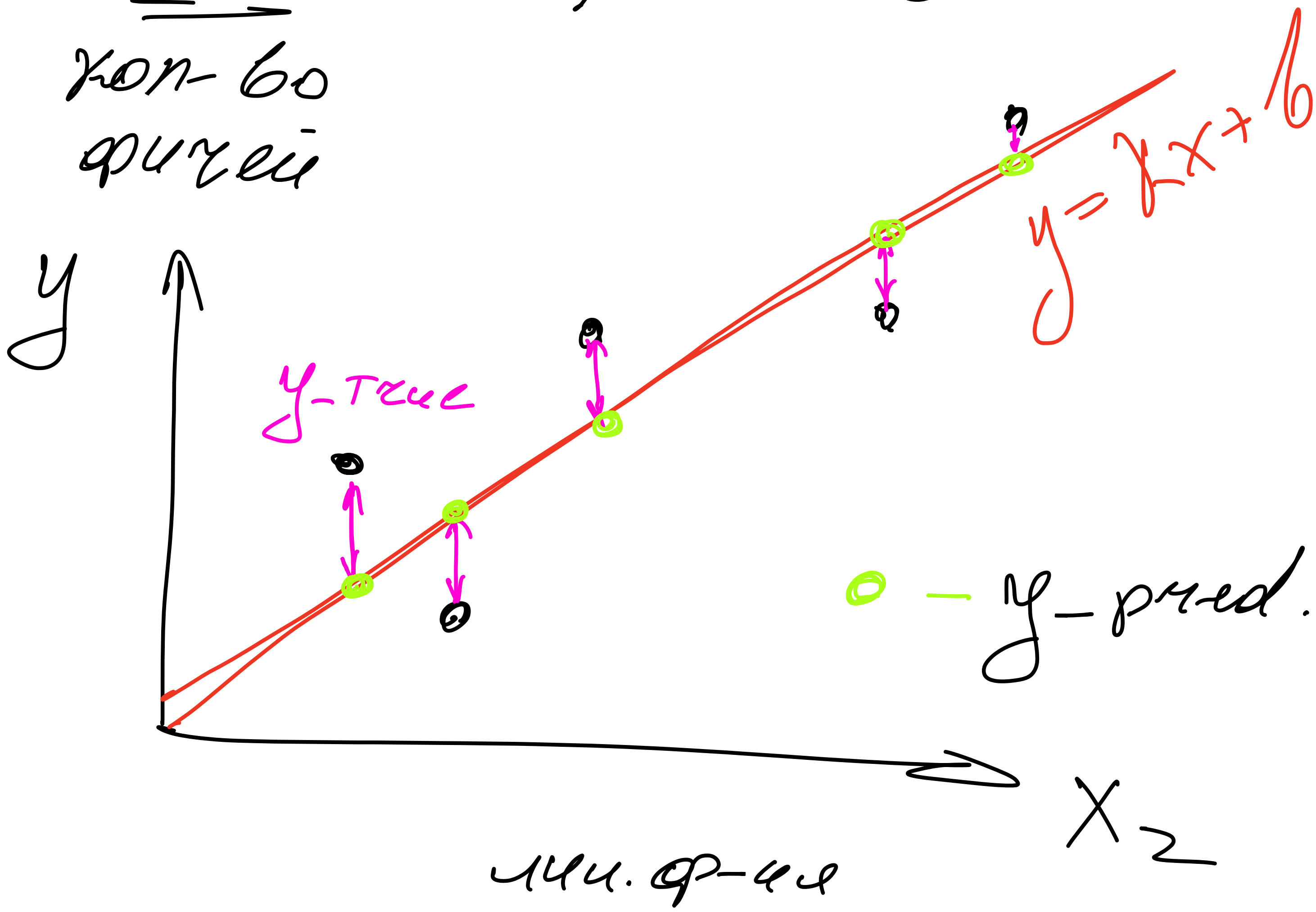
$$W = \underset{W}{\operatorname{argmin}}\ MSE(X \cdot W; y)$$

Посмотрим на плоскость предсказание — 1 признак и спроецируем на m — пространство

m — кол-во фичей



лин. ф-ия

$$MSE = \sum_{m} (y - y\_pred)^2 \cdot \frac{1}{n}$$

$$y\_pred = kx + b$$

$$MSE = \sum_{m} \left(y - (kx+b)\right)^2 \cdot \frac{1}{n}$$

$$grad_{MSE} = \left(\frac{\partial MSE}{\partial k} ; \frac{\partial MSE}{\partial b}\right)$$

$$(*) \qquad (**)$$

$$(*) \quad \frac{\partial MSE}{\partial k} = 2(y - kx - b)\cdot(-x)$$

$$(**) \quad \frac{\partial MSE}{\partial b} = 2(y - kx - b)\cdot(-1)$$

$$k = k + \underset{lr}{\underline{\underline{0{,}01}}} \cdot (-b - kx + y)\cdot(-2x)$$

$\cdot(-1)$

шаг в пр. grad

$$b = b + \underset{\text{в.ч}}{\underline{\underline{0{,}01}}} \cdot \left(-2\left(y - kx - b\right)\right) \cdot (-1)$$

обратную сторону

$$y_{pred} = \underline{\underline{k}}\,\underline{x} + \underline{\underline{b}}$$

матрица весов

матрица признаков

матрица св. членов

При умножении матрицы признаков на число →

→ изменяется положение точки в n-пространстве,

но масштаб сохранится.
градиент так или иначе
будет стремиться к min
(теоретич. вывод).

Практический вывод:

$$(X\cdot w - y)^T (X\cdot w - y) \to \min_w$$

$$w = (X^T\cdot X)^{-1}\cdot X^T\cdot y$$

Проверка умножения на число

$$w' = (b\cdot X^T\cdot X\cdot b)^{-1}\cdot b\cdot X^T\cdot y =$$

$$= (b^2 \cdot X^T \cdot X)^{-1} \cdot bX^T \cdot y =$$

$$= (b^2)^{-1} \cdot (X^T \cdot X)^{-1} \cdot b \cdot X^T \cdot y =$$

$$= \frac{1}{b^{\cancel{2}}} \cdot \cancel{b} \, \underbrace{(X^T \cdot X)^{-1} \cdot X^T \cdot y}_{} =$$

$$= \frac{1}{b} w$$

$$w_{new} = \underbrace{X'}_{X \cdot b} \cdot \underbrace{w'}_{\frac{1}{b} w} = X b \cdot \frac{1}{b} w = X \cdot w$$

ч.т.д. (ничто не поменялось, обновление весов умножением на число — возможно)