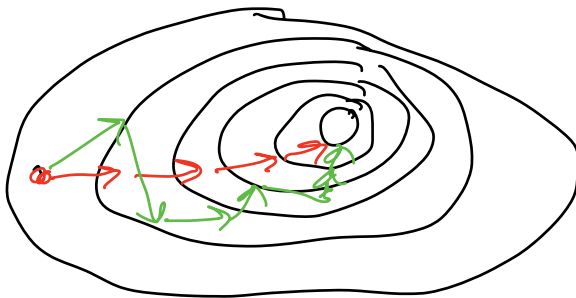


$$\nabla_{\omega} L(f, x, \omega) = - \frac{2}{n} x^T (\underbrace{x\omega}_{K \times 1} - \underbrace{y}_{K \times 1})$$

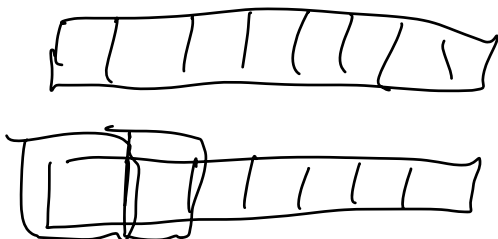


$$K \times P \quad P \times 1 \quad = K \times 1$$

§ stochastic: Вычисляем зр. по 1 наблюдению  
Batch gradient descent: по подвыборкам  
некоторого размера

обычно:

перемешиваем выборку



После каждого фолда  
мат случай на каждом  
блоке

Средне значение по блокам — это то

$$\frac{f(x+dx) - f(x)}{dx}$$

LARS, stepwise regression

# Регуляризаторы

$$L(f, x, y) \Rightarrow \underbrace{\|xw - y\|_2^2}_{L_2} + \underbrace{\lambda \|w\|_k^k}_{L_1} \rightarrow \min$$

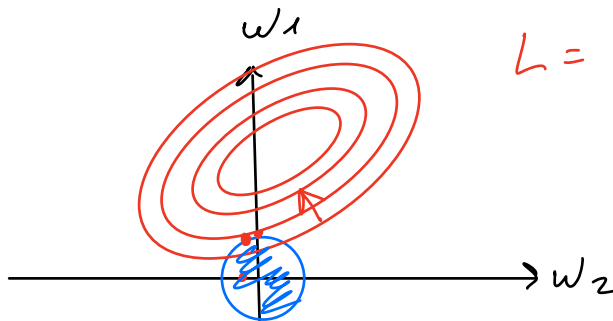
$$L^2: \|w\|_2^2 = w_1^2 + \dots + w_k^2$$

$$L^1: \|w\|_1 = |w_1| + \dots + |w_k|$$



$$L^2: (y - xw)^T (y - xw) + \lambda w^T w$$

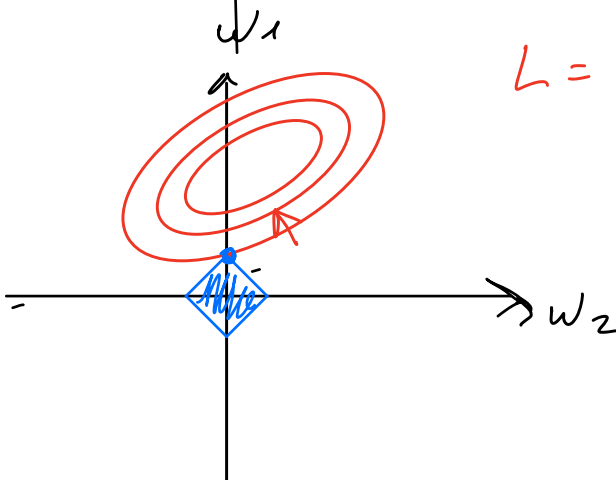
$$w^* = (x^T x + \lambda I)^{-1} x^T y$$



$$L = C$$

1) Регуляризатор с регуляризующим свойством

2) Точная регуляризация



$$L = C \quad L_1$$

Elastic Net

Другие Loss

$$MSE \rightarrow E(Y|X)$$

$$MAE \rightarrow Med(Y|X)$$

$$Q_d \rightarrow Kлеммa_d | X$$

1) Loss

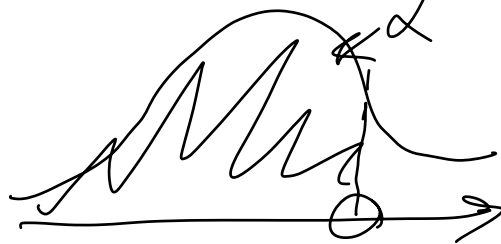
2) Метрики

MAE, MAPE, Huber, SMAPE

$$MAE: \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}| = \frac{1}{n} \sum_{i=1}^n |y_i - \langle x_i, \hat{w} \rangle|$$

нормализ

$$\frac{1}{n} \sum |y_i - \langle x_i, w \rangle| \rightarrow \hat{w}$$



$$L(f, X, y) = \sum_{i=1}^n h_{\delta}(y_i - \langle x_i, w \rangle),$$

$$h_{\delta}(z) = \begin{cases} \frac{1}{2} z^2, & |z| \leq \delta \\ \delta(|z| - \frac{1}{2}\delta), & |z| > \delta \end{cases}$$

- x - v

$\begin{bmatrix} a \\ b \\ c \\ d \\ c \end{bmatrix}$

$\rightarrow$

$\begin{bmatrix} d \\ 1 \\ 0 \\ c \\ 1 \\ c \end{bmatrix}$

$\begin{bmatrix} b \\ 0 \\ 1 \\ c \\ 0 \\ 0 \end{bmatrix}$

$\begin{bmatrix} c \\ 0 \\ 0 \\ 1 \\ c \\ 1 \end{bmatrix}$

$\uparrow$