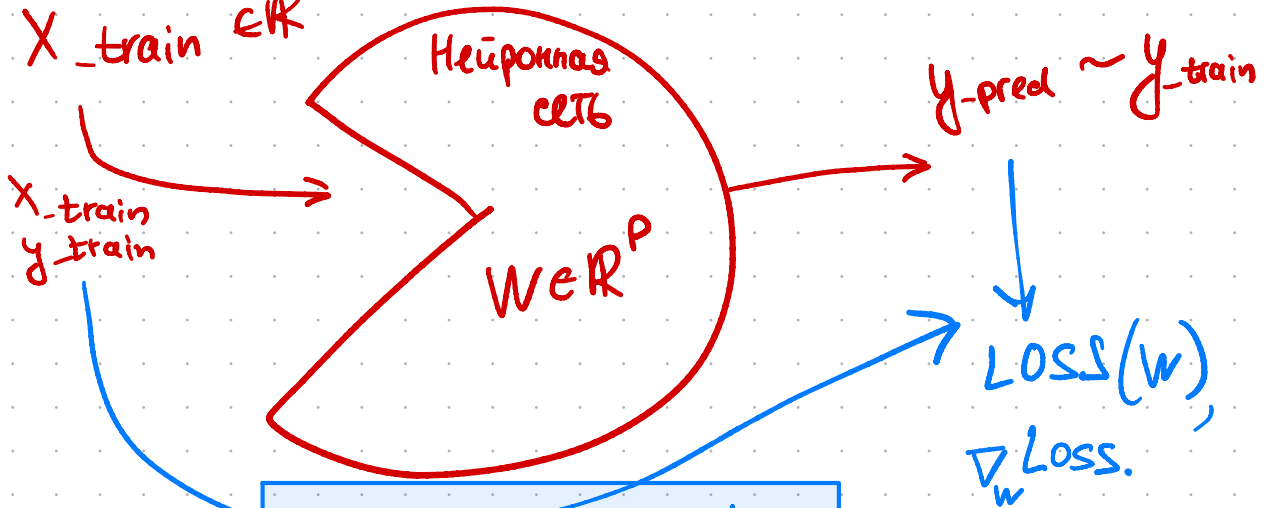
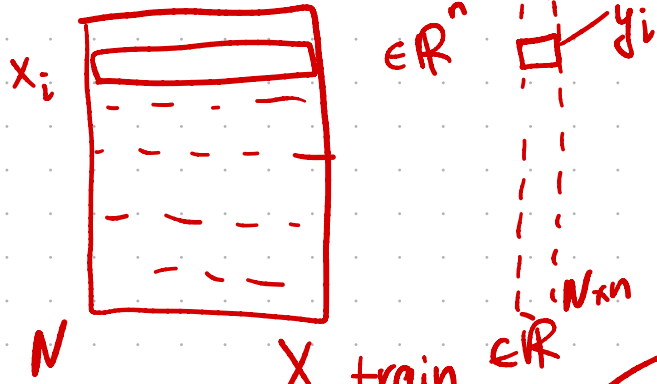


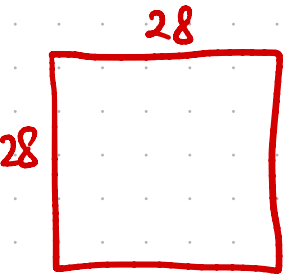
Classification

MNIST

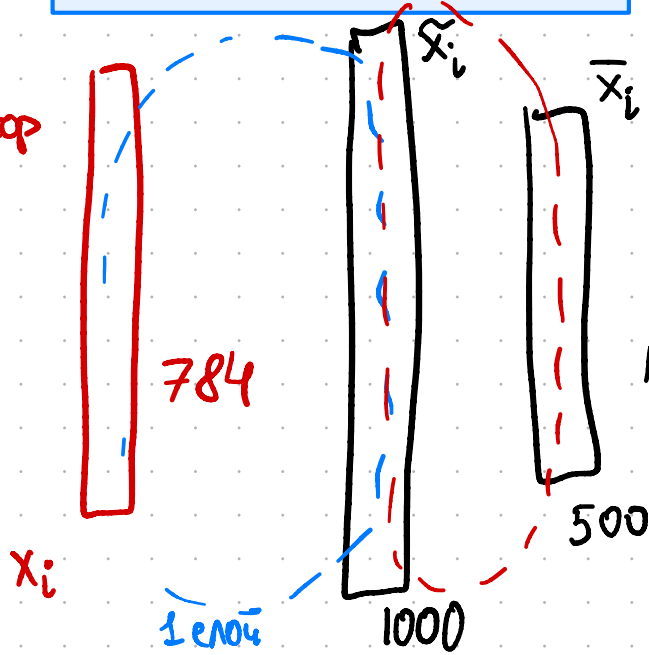
$y_i \in \{0, \dots, 9\}$
 $y_i \in$ ONE-HOT ENCODING



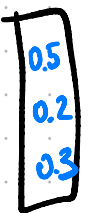
$$W_{k+1} = W_k - d_k \cdot \frac{\partial L}{\partial W_k}$$



баз. вектор



голова сети:
 $y_i^{pred} \geq 0$
 $1^T y_i^{pred} = 1$



$$y_i^{pred} = W_3 \cdot W_2 \cdot W_1 \cdot x_i$$

$$y_i^{pred} = \tilde{W} x_i$$

$$\tilde{x}_i = \sigma_1(W_1 \cdot x_i)$$

$$\bar{x}_i = \sigma_2(W_2 \cdot \tilde{x}_i)$$

$$y_i^{pred} = \sigma_3(W_3 \cdot \bar{x}_i)$$

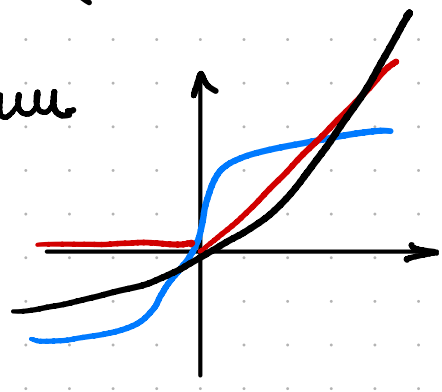
на самом деле

$$y_i^{pred} = \text{softmax}(\bar{x}_i) = \frac{1}{e^{\sum_j \bar{x}_{ij}}}$$

$$\sigma(\bar{x}_i)$$

$$y^{pred} = \sigma_N(W_N \cdot \sigma_{N-1}(W_{N-1} \cdot \sigma_{N-2}(\dots W_1 \cdot X)))$$

σ - функция активации



① Задача подбора оптимальных W_i НЕ ВЫПУКЛА

$$L(W) \rightarrow \min_{W \in \mathbb{R}^D}$$

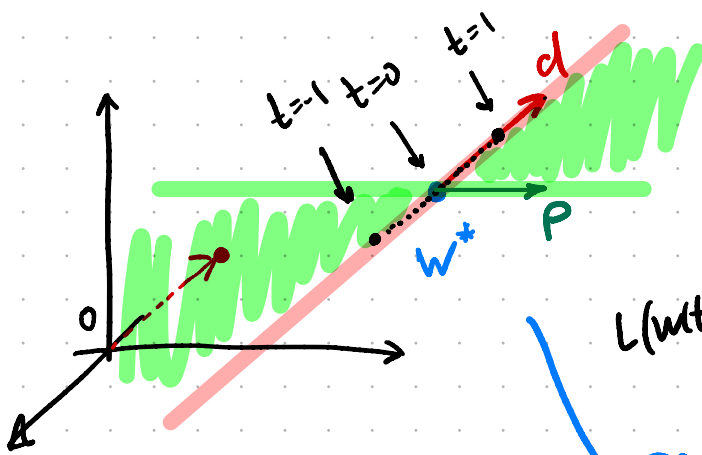
② N, p - могут быть "огромными".
данные - модель

③ Обучаем модель на X_{train} , и проверяем на X_{test}
generalization error

Пусть есть $L(w)$ $w \in \mathbb{R}^p$, $p=10^6$

Зафиксируем вес w^* , который получили в результате обучения.

Возьмём случайное направление $d \in \mathbb{R}^p$

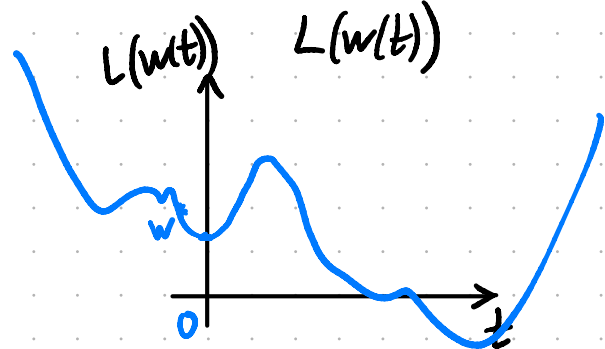


$$w(t) = w^* + t \cdot d$$

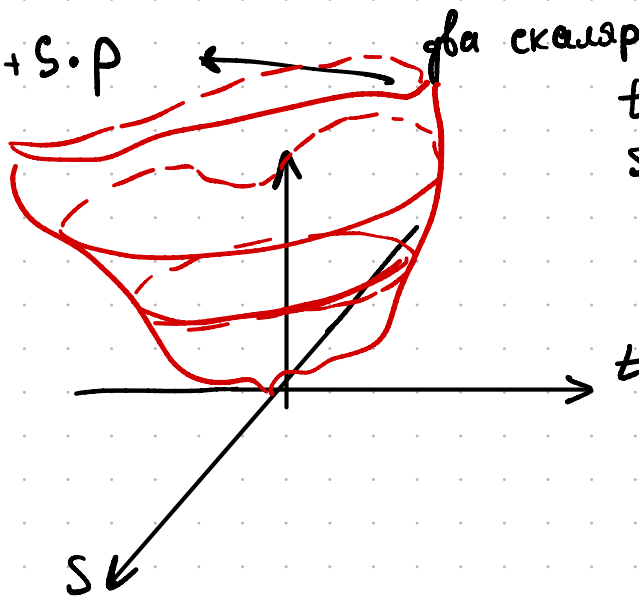
$$t=0 \quad w(t) = w^*$$

$$t = \text{np.linspace}(-1, 1)$$

$$L(w(t))$$



$$w(t, s) = w^* + t \cdot d + s \cdot p$$



оба скачка t и s
 $t = \text{np.linspace}(-1, 1)$
 $s = \dots$

