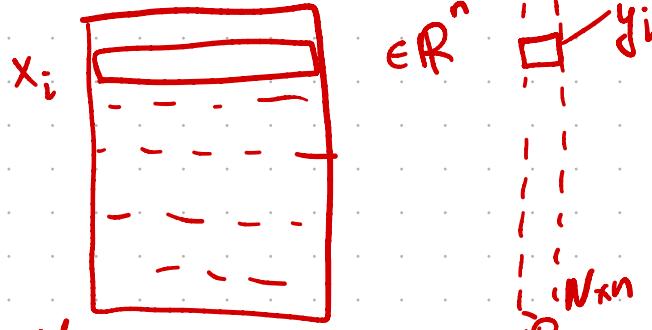
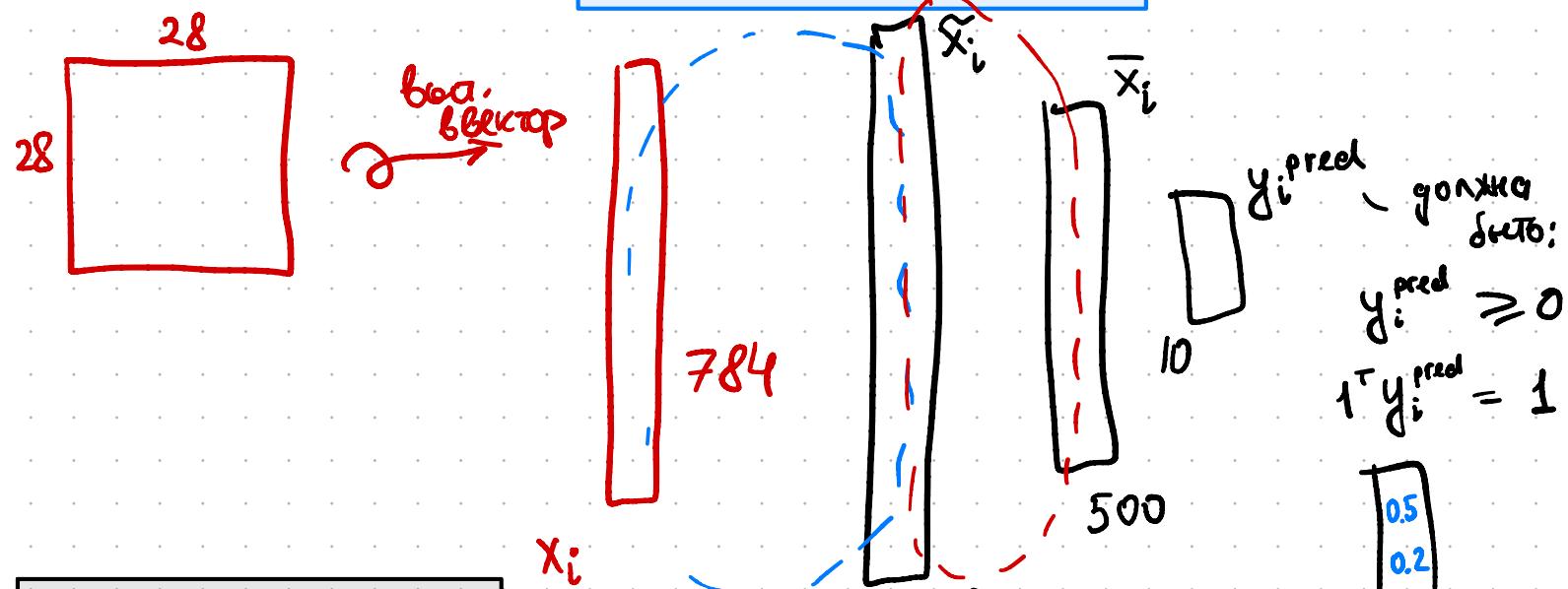
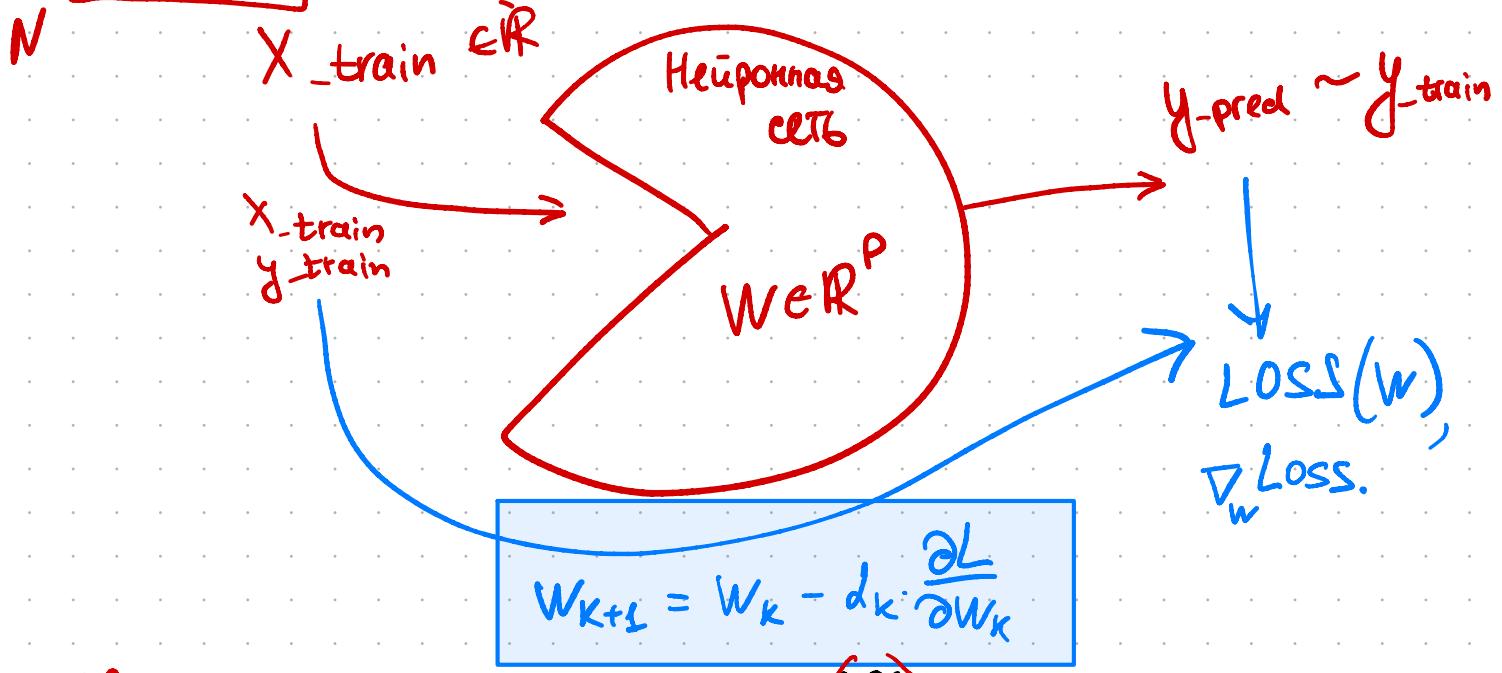


Classification



MNIST

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



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

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

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

$$\tilde{x}_i = g_2(W_2 \cdot \tilde{x}_i)$$

$$y_i^{\text{pred}} = g_3(W_3 \cdot \tilde{x}_i)$$

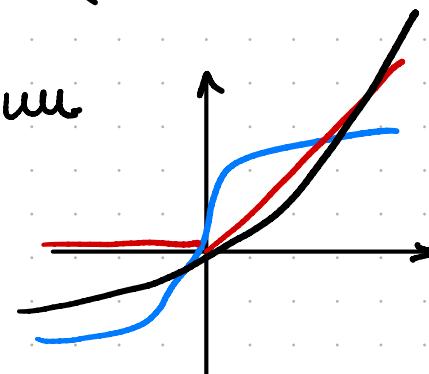
$$\begin{bmatrix} 0.5 \\ 0.2 \\ 0.3 \end{bmatrix}$$

на самом деле

$$y_{pred} = \text{softmax}(\bar{x}_i) = \sigma(\bar{x}_i) = \begin{pmatrix} e^{\bar{x}_1} \\ \vdots \\ e^{\bar{x}_n} \end{pmatrix} \cdot \frac{1}{\sum_i e^{\bar{x}_i}}$$

$$y^{pred} = \sigma_N \left(W_N \cdot \tilde{G}_{N-1} \left(W_{N-1} \cdot \tilde{G}_{N-2} (\dots \cdot W_1 \cdot x) \right) \right)$$

σ - функция активации



① Задача подбора оптимальных W_i НЕ ВЫПУКЛА

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

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

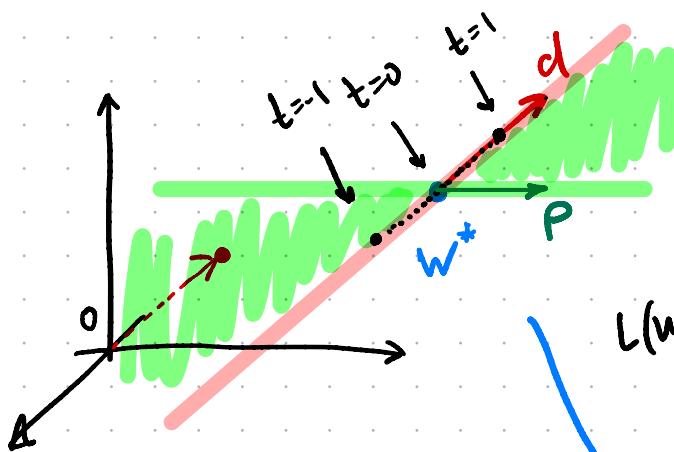
③ Обучаем модель на X_{train} , а проверяем на X_{test}
generalization error

Пусть есть $L(w)$

$w \in \mathbb{R}^P$, $P=10^6$

Зададим вектор w^* , который получился в результате обучения.

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



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

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

$$t = np.linspace(-1, 1)$$

$$L(w(t))$$

$$L(w)$$

$$w(t, s) =$$

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

за сколько t и s

$$t = np.linspace(-1, 1)$$

$$s = \dots$$

