Мультиномиальная логистическая регрессия

$$p(\theta_k, k, x_i) \approx \exp(\theta_k x_i)$$

$$p(\theta_k, x_i, k) = \frac{e^{\theta_k x_i}}{\sum_{j=1}^{K} e^{\theta_j x_i}}$$

$$\mathcal{L}(x_i, y_i) = -\sum_{k=1}^{K} \boxed{y_{ik}} \log p_{ik}$$

y_{oh}

0	0	0	1	0	0

one-hot encoding

$y = 3$

y_i

* * *

$\log$

p_{i0}	p_{i1}				p_{i5}

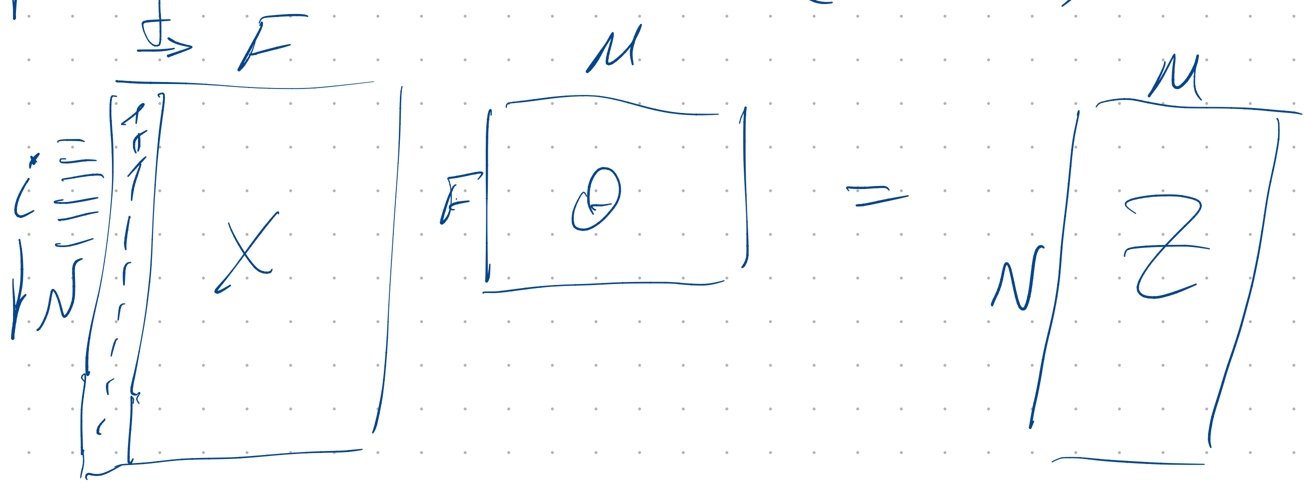
$$\mathcal{L}_{bin} = -\left(y_i \log p_i + (1-y_i)\log(1-p_i)\right)$$

$$\mathcal{L}_{bin} = -\left(y_{i1}\log p_1 + y_{i0}\log p_0\right) = -\sum_{k=0}^{1} y_{ik}\log p_k$$

$$\mathcal{L}(x_i, y_i) = -\sum_{k=1}^{K} \boxed{y_{ik}} \log p_{ik}$$

$$p(x_i, \theta_k, k) = \frac{e^{\theta_k x_i}}{\sum_{j=1}^{K} e^{\theta_j x_i}}$$

$$p(x_i, \theta) = \mathrm{Softmax}(\theta^T x_i)$$



$$N\times F \cdot F\times M \to N\times M$$

$$P = \mathrm{Softmax}(Z)$$

$$\mathcal{L} = -\sum_k \log P \odot Y$$

$$\underset{N\times M}{Y} \odot \underset{N\times M}{\log P} \to \sum_{i,k} Y \log P = \mathcal{L}$$

$\mathcal{L} \to \nabla_\theta \mathcal{L} \to$ градиентная оптимизация ф-ии потерь $\mathcal{L}(X, Y, \theta)$