Логистическая регрессия

$\{x_i, y_i\}_{i=1}^n$, $x_i \in \mathbb{R}^d$, $y_i \in \{-1, +1\}$

$y_i \in \{1 \dots k\}$

$y(x) = \operatorname{sign} w^T x$ $\quad p(y=1|x,w) = \frac{1}{1+\exp(-w^T x)} = \sigma(w^T x)$

$y(x) = \underset{k \in \{1 \dots k\}}{\arg\max}\, w_k^T x$ $\quad p(y=k|x,W) = \frac{\exp(w_k^T x)}{\sum_{j=1}^k \exp(w_j^T x)}$

$$p(y|x,w) = \prod_{i=1}^n p(y_i|x_i,w) \to \max_w$$

$$-\log p(y|x,w) = -\sum_{i=1}^n \log p(y_i|x_i,w) =$$

$$= \sum_{i=1}^n L_1(y_i, Wx_i) = \sum_{i=1}^n L_2(y_i, \{p(y=k|x_i,w)\}_{i=1}^n) \to \min_w$$

$L_1(y,z) = -z_y + \log\left(\sum_{i=1}^k \exp(z_j)\right)$ мультиномиальные потери

$L_2(y,p) = -\sum_{j=1}^k [y=j] \log p_j$ кросс-энтропия

$r(x)$ и $q(x)$ распределения

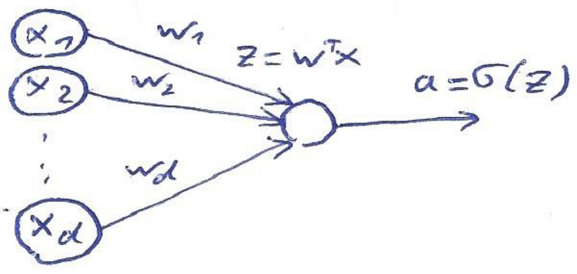
$KL(r\|q) = \int r(x) \log \frac{r(x)}{q(x)}\, dx$

$r(x)$: $\begin{array}{cccc} 1 & 2 & \dots & k \\ [y=1] & [y=2] & \dots & [y=k] \end{array}$

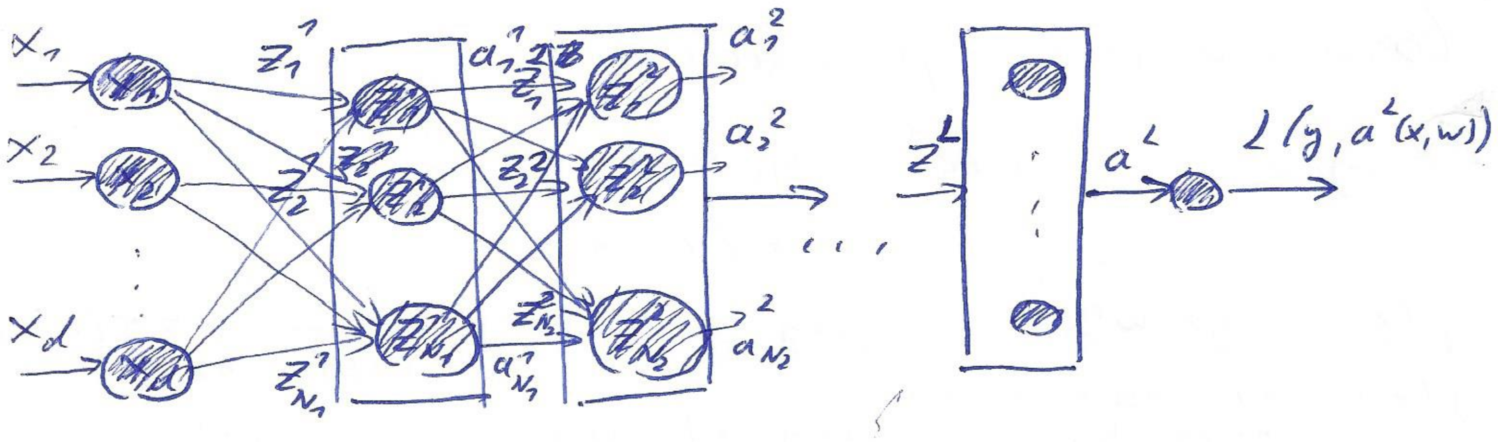
$q(x)$: $\begin{array}{ccc} 1 & 2 \dots & k \\ p(y=1|x,w) & \dots & p(y=k|x,w) \end{array}$

$$KL(r(x)\|q(x)) = \underbrace{\int r(x)\log r(x)\,dx}_{0} - \int r(x) \log q(x)\, dx =$$

$$= -\sum_{j=1}^k [y=j] \log p(y=j|x,w)$$

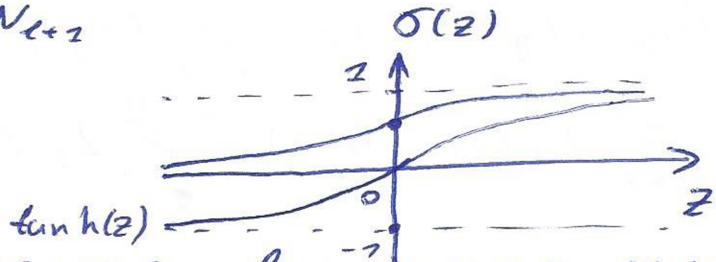


Многослойный перцептрон, MLP



$$\begin{cases} a_i^0 = x_i, \quad i = 1 \dots d \\ z^{l+1} = W^{l+1} a^l + b^{l+1}, \quad l = 0 \dots L-1 \\ a_i^{l+1} = g(z_i^{l+1}), \quad i = 1 \dots N_{l+1} \end{cases}$$

Функции активации

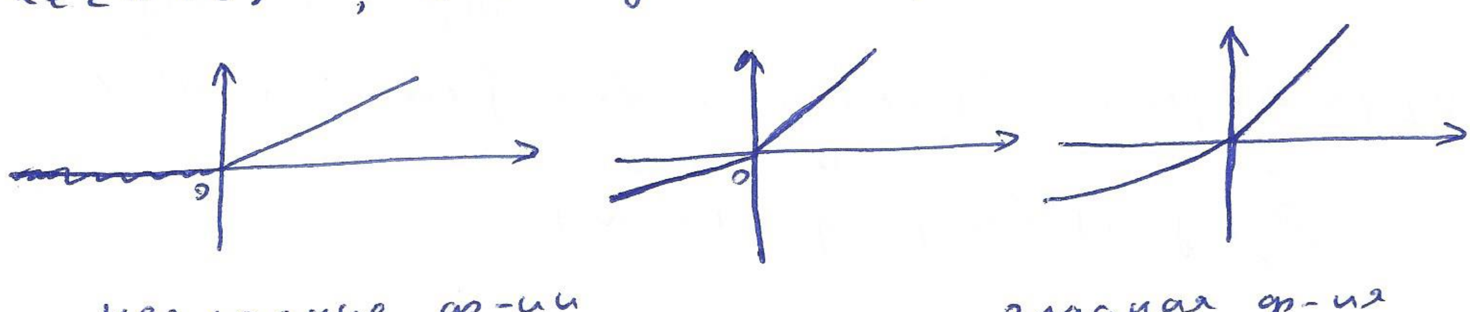


Сигмоида – плохо шкалируемая вытянутая ф-ия.

1) $\sigma(z) \in [0, 1]$, 2) $\tanh(z) = 2(\sigma(2z) - \frac{1}{2}) \in [-1, 1]$

гиперболический тангенс не смещает выборку в область положительных значений, но не решает проблему затухания градиентов

3) $ReLu(z)$, 4) $Leaky\,ReLu(z)$, 5) $ELU(z)$

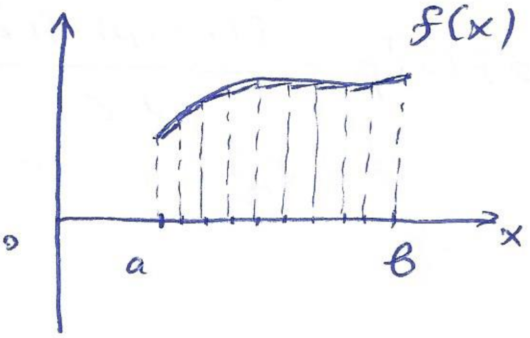


негладкие ф-ии   гладкая ф-ия

легко считать

решена проблема затухания градиентов

Универсальный аппроксиматор



$[x \in [a_i, b_i]]$

$[x \geq a_i] \approx \sigma(\alpha(x - a_i))$

$\alpha >> 1$

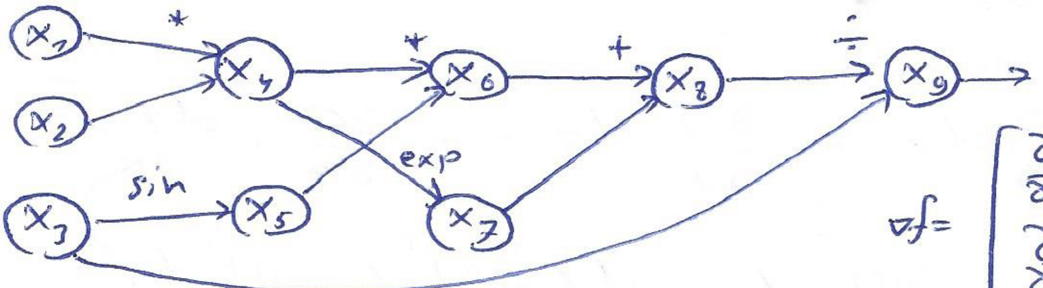
$[x \leq b_i]$

$$\left[[x \geq a_i] + [x \leq b_i] > \frac{3}{2}\right], \quad f(x) = \sum_i f(a_i)[x \in [a_i, b_i]]$$

Автоматическое дифференцирование

$$F(w) = \sum_{i=1}^{n} L(y_i, a^2(x_i, w)) \longrightarrow \min_w$$

$$f(x) = \frac{x_1 x_2 \sin(x_3) + \exp(x_1 x_2)}{x_3}$$



$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix}$$

сложность $O(n)$ по числу входных переменных

Проход вперёд $\frac{\partial x_4}{\partial x_1} = x_2$,

$$\frac{\partial x_6}{\partial x_1} = \{x_6(x_4(x_1))\} = \underbrace{\frac{\partial x_6}{\partial x_4}}_{x_5}\frac{\partial x_4}{\partial x_1} = x_5 \cdot x_2$$

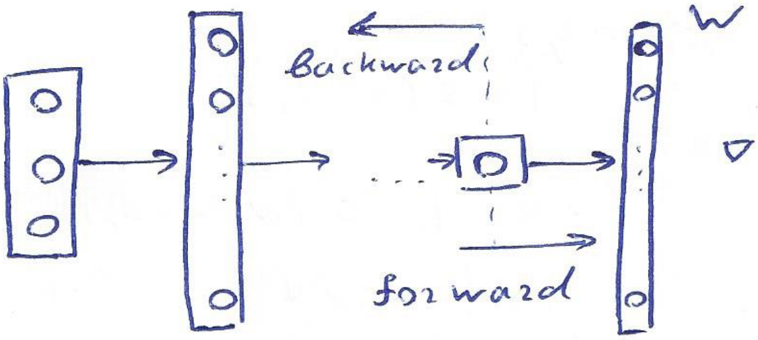
$$\frac{\partial x_8}{\partial x_1} = \{x_8(x_6(x_1), x_7(x_1))\} = \frac{\partial x_8}{\partial x_6}\frac{\partial x_6}{\partial x_1} + \frac{\partial x_8}{\partial x_7}\frac{\partial x_7}{\partial x_1}$$

Проход назад, $\frac{\partial f}{\partial x_9} = 1$

$$\frac{\partial f}{\partial x_8} = \{f(x_9(x_8))\} = \underbrace{\frac{\partial f}{\partial x_9}}_{=1}\underbrace{\frac{\partial x_9}{\partial x_8}}_{=\frac{1}{x_3}} = \frac{1}{x_3}$$

$$\frac{\partial f}{\partial x_4} = \{f(x_6(x_4), x_7(x_4))\} = \frac{\partial f}{\partial x_6}\frac{\partial x_6}{\partial x_4} + \frac{\partial f}{\partial x_7}\frac{\partial x_7}{\partial x_4}$$

Для градиентов нужен только один проход, необходимо хранить в памяти все переменные.

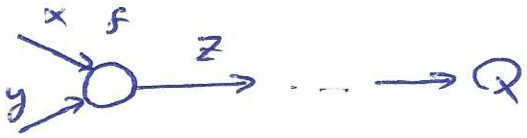


разностное диф-е

$$\nabla f(x)^T p \approx \frac{f(x+\varepsilon p)-f(x-\varepsilon p)}{2\varepsilon}$$

$$\varepsilon = \varepsilon_M^{\frac{1}{3}}$$

семинар



$$\nabla_x Q = \left(\frac{\partial Q}{\partial x_i}\right)_{i=1}^{n}$$

$$z = f(x,y) \in \mathbb{R}^k \qquad Q(z) \qquad J(x) = \frac{\partial f}{\partial x} = \left(\frac{\partial f_i}{\partial x_j}\right)_{i,j=1}^{k,n}$$

$x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $z \in \mathbb{R}^k$

$$\frac{\partial Q}{\partial z}, \quad \underset{n}{\frac{\partial Q}{\partial x}} = \underset{k}{\frac{\partial Q}{\partial z}} \underset{k\times n}{\frac{\partial f}{\partial x}} \leftarrow \text{докажем}$$

$$dQ = \sum_{i=1}^{k} \frac{\partial Q}{\partial z_i} dz_i = \frac{\partial Q}{\partial z}^T dz = \left\langle \frac{\partial Q}{\partial z}, dz \right\rangle$$

$$dQ = \frac{\partial Q}{\partial x}^T dx + \frac{\partial Q}{\partial y}^T dy, \quad dz = \frac{\partial f}{\partial x}^T dx + \frac{\partial f}{\partial y}^T dy$$

$$dQ = \frac{\partial Q}{\partial z}^T \left(\frac{\partial f}{\partial x}^T dx + \frac{\partial f}{\partial y}^T dy\right) =$$

$$= \left(\frac{\partial Q}{\partial z}\frac{\partial f}{\partial x}\right)^T dx + \left(\frac{\partial Q}{\partial z}\frac{\partial f}{\partial y}\right)^T dy \quad \blacksquare$$

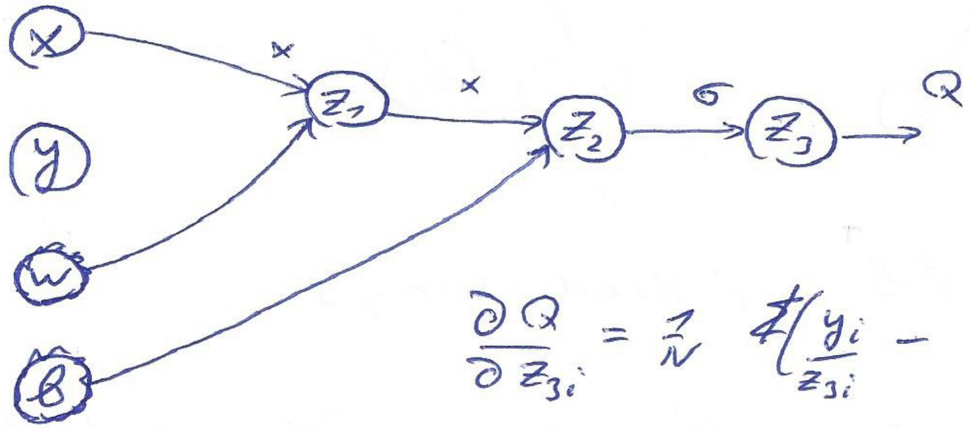
$$dQ = \sum_{i,j} \frac{\partial Q}{\partial z_{ij}} dz_{ij} = tr\left(\frac{\partial Q}{\partial Z}^T dZ\right) \leftarrow \text{в случае, } Z \text{ - матрица}$$

Логистическая регрессия

$X \in \mathbb{R}^{N\times D}$, $y \in \{0, 1\}^N$ $\qquad w \in \mathbb{R}^D$

$$\hat{y} = \sigma(\underset{N}{Xw} + \underset{N}{b\bar{1}}) \qquad b \in \mathbb{R}$$

$$Q(w,b) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right) \to \max_{w,b}$$



$$\frac{\partial Q}{\partial z_{3i}} = \frac{1}{N}\cancel{\sum}\left(\frac{y_i}{z_{3i}} - \frac{1-y_i}{1-z_{3i}}\right)$$

$$\frac{\partial Q}{\partial z_3} = \frac{1}{N}\left(\frac{y}{z_3} - \frac{1-y}{1-z_3}\right)$$

$$\frac{\partial Q}{\partial z_2} = \underbrace{\frac{\partial Q}{\partial z_3}}_{N}\underbrace{\frac{\partial z_3}{\partial z_2}}_{N\times N}, \quad z_3 = \sigma(z_2)$$

$$\frac{\partial z_{3i}}{\partial z_{2j}} = \begin{cases} 0, & i \neq j \\ \dots, & i = j \end{cases}$$ т.к. покомпонентная ф-ия $\sigma(z_2)$, то диагональная матрица

$$\sigma'(t) = \sigma(t)(1-\sigma(t))$$

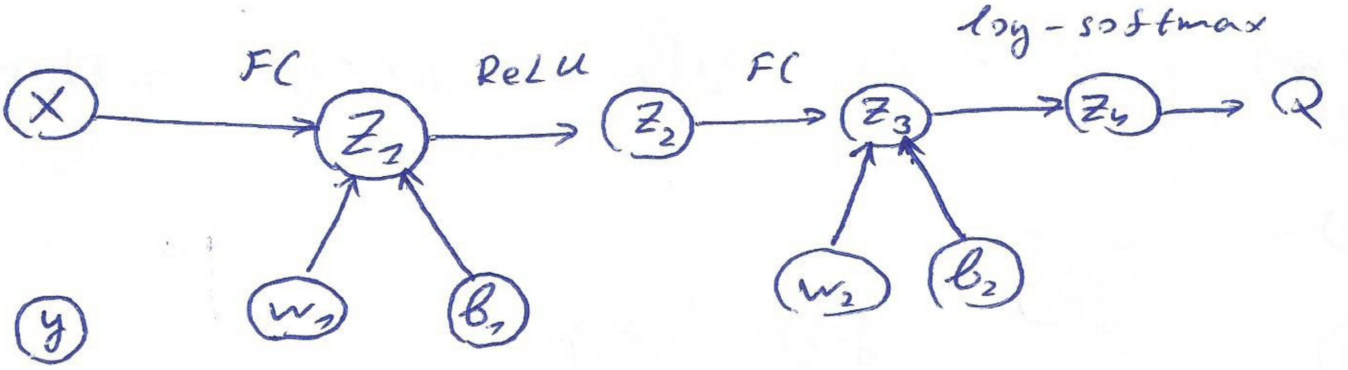
$$\frac{\partial z_3}{\partial z_2} = \mathrm{diag}(z_3(1-z_3))$$

$$\frac{\partial Q}{\partial z_2} = \frac{\partial Q}{\partial z_3} \odot z_3 \odot (1-z_3)$$

$$\frac{\partial Q}{\partial z_1} = \frac{\partial Q}{\partial z_2}\frac{\partial z_2}{\partial z_1} = \frac{\partial Q}{\partial z_2}, \quad \{z_2 = z_1 + b\bar{1}\}$$

$$\frac{\partial Q}{\partial b} = \frac{\partial Q}{\partial z_2}\frac{\partial z_2}{\partial b} = \frac{\partial Q}{\partial z_2}\cdot \bar{1}$$

$$\frac{\partial Q}{\partial w} = \frac{\partial Q}{\partial z_1}\frac{\partial z_1}{\partial w} = \frac{\partial Q}{\partial z_1}\cdot X, \quad \{z_1 = Xw\}$$



$$\underset{N\times M}{Z} = \underset{N\times D}{X}\,\underset{D\times M}{W} + \underset{N\times 1}{1}\cdot \underset{1\times M}{b^T} \qquad \{\text{broadcasting}\}$$

$$\frac{\partial Q}{\partial Z_i} - ? \qquad dQ = tr\left(\frac{\partial Q}{\partial W}^T dW\right) + tr\left(\frac{\partial Q}{\partial X}^T dX\right) + tr\left(\frac{\partial Q}{\partial b}^T db\right)$$

$$dQ = tr\left(\frac{\partial Q}{\partial Z}^T dZ\right) \qquad \left\{ \begin{array}{l} tr\,AB = tr\,BA \\ tr\,A^T = tr\,A \end{array} \right\}$$

$$dZ = dX \cdot W + X\,dW + 1\,db^T$$

$$dQ = tr\left(\frac{\partial Q}{\partial Z}^T dX \cdot W\right) + tr\left(\frac{\partial Q}{\partial Z}^T X\,dW\right) +$$

$$+ tr\left(\underset{M\times N}{\frac{\partial Q}{\partial Z}^T}\,\underset{N}{1}\,\underset{M}{db^T}\right) = tr\left(\left(\frac{\partial Q}{\partial Z} W^T\right)^T dX\right) + tr\left(\left(X^T \frac{\partial Q}{\partial Z}\right)^T dW\right) + tr\left(\left(\frac{\partial Q}{\partial Z}^T 1\right)^T db\right)$$

$$\frac{\partial Q}{\partial X} = \underset{N\times M}{\frac{\partial Q}{\partial Z}}\,\underset{M\times D}{W^T}, \qquad \underset{D\times M}{\frac{\partial Q}{\partial W}} = \underset{D\times N}{X^T}\,\underset{N\times M}{\frac{\partial Q}{\partial Z}}, \qquad \underset{M}{\frac{\partial Q}{\partial b}} = \underset{M\times N}{\left(\frac{\partial Q}{\partial Z}\right)^T}\,\underset{N}{1}$$

K K

$$Z_4 = \log softmax(Z_3), \qquad Y \in \{0, 1\}^K$$

N

$$Q = \sum_{i,k=1}^{N,K} y_{ik} Z_{4ik} \to \max, \qquad \frac{\partial Q}{\partial Z_4} = Y$$

$$Z_{4i} = Z_{3i} - \log \sum_{k=1}^{K} \exp(Z_{3,k})$$

$$Z_4 = Z_3 - \log \sum_{k=1}^{K} \exp(Z_{3k})$$

$$dQ = \frac{\partial Q}{\partial Z_4}^T dZ_4, \qquad dZ_4 = dZ_3 - \frac{\exp(Z_3)\odot dZ_3}{\sum_{k=1}^{K} \exp(Z_{3k})}$$

$$dQ = \frac{\partial Q}{\partial Z_4}^T dZ_3 - \frac{\sum_{k=1}^{K} \frac{\partial Q}{\partial Z_{4k}} \exp(Z_{3k}) \cdot dZ_{3k}}{\sum_{k=1}^{K} \exp(Z_{3k})}$$

$$\left(\frac{\partial Q}{\partial z_4} - \frac{\frac{\partial Q}{\partial z_4} \odot \exp(z_3)}{\sum_k \exp(z_{3k})} \right)^T dz_3 = dQ$$

$$\frac{\partial Q}{\partial z_3} = \frac{\partial Q}{\partial z_4} - \frac{\frac{\partial Q}{\partial z_4} \odot \exp(z_3)}{1^T \exp(z_3)}$$