

Линейная классификация

алгоритм: $a(x) = \langle w, x \rangle + w_0$
 $x \in \mathbb{R}^d, y$ $a(x) \in (-\infty, +\infty)$

предсказание вероятности:

Б: $p(1|x) = 1 - p(0|x) = \sigma(a(x))$

М: $p(k|x) = \text{softmax}_{k \in K} a(x)$ $\leftarrow a(x) = Wx + w_0$

Функция потерь: $\prod_{i=1}^n p(y_i | x_i) \rightarrow \max_{w, w_0}$
 $-\sum_{i=1}^n \log p(y_i | x_i) \rightarrow \min_{w, w_0}$

SGD

o nymu.

$i \sim \text{Unif} \{1, \dots, n\}$

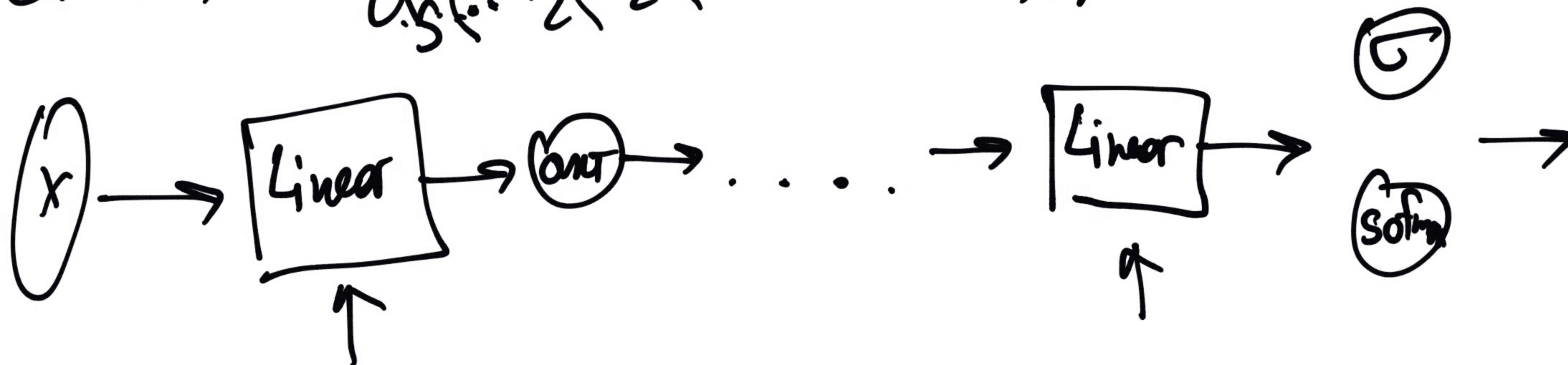
$w_{\text{new}} := w_{\text{old}} - \alpha \nabla_w \log P(y_i | x_i)$

$\frac{1}{c} \sum_j \log P(y_j | x_j)$

$$\begin{aligned}
 a_1(x) &= W_1 x + w_1^0 \\
 a_2(x) &= W_2 x + w_2^0 \\
 &\vdots \\
 a_L(x) &= W_L x + w_L^0
 \end{aligned}$$

$W_i \in \mathbb{R}^{L_i \times d}$
 $w_i^0 \in \mathbb{R}^{L_i}$
 $\dots L_e$

$$a(x) = a_L(\dots f_2(a_2(f_1(a_1(x))))))$$



SGD и аналог

Adam

$$\nabla_{\theta} \log p(y_i | x_i)$$

$$f: X \times W \rightarrow Z_1$$

$$g: Z_1 \times \Theta \rightarrow Z_2$$

$$S: Z_2 \rightarrow Z_3$$

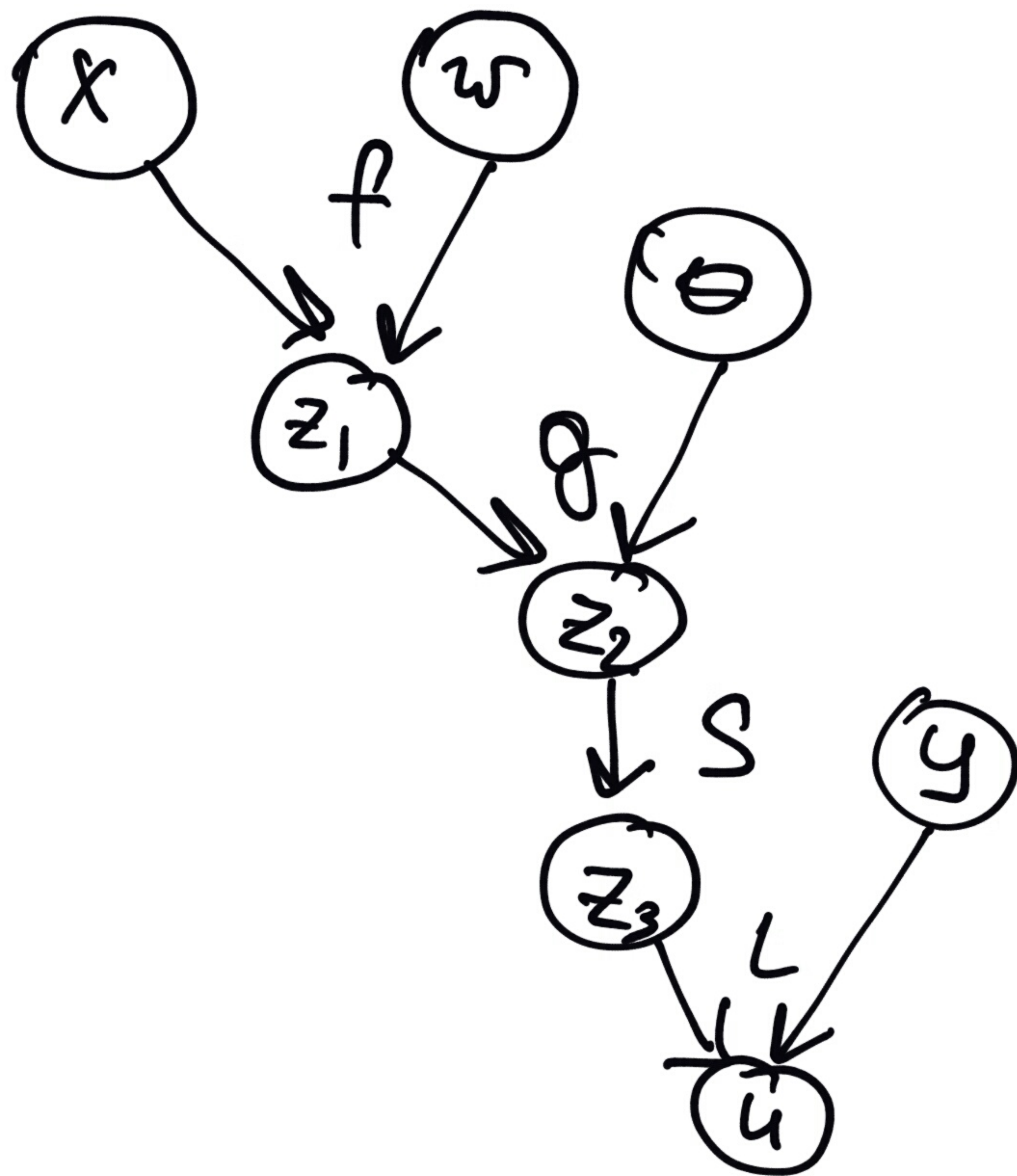
$$L: Z_3 \times Y \rightarrow \mathbb{R}$$

$$Z_1 = f(x, w)$$

$$Z_2 = g(Z_1, \theta)$$

$$Z_3 = S(Z_2)$$

$$u = L(Z_3, y)$$



$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial \mathcal{L}}{\partial z_3} (z_3).$$

$$\cdot \frac{\partial z_3}{\partial z_2} (z_2) \frac{\partial z_2}{\partial \theta} (\theta)$$

Backpropagation

