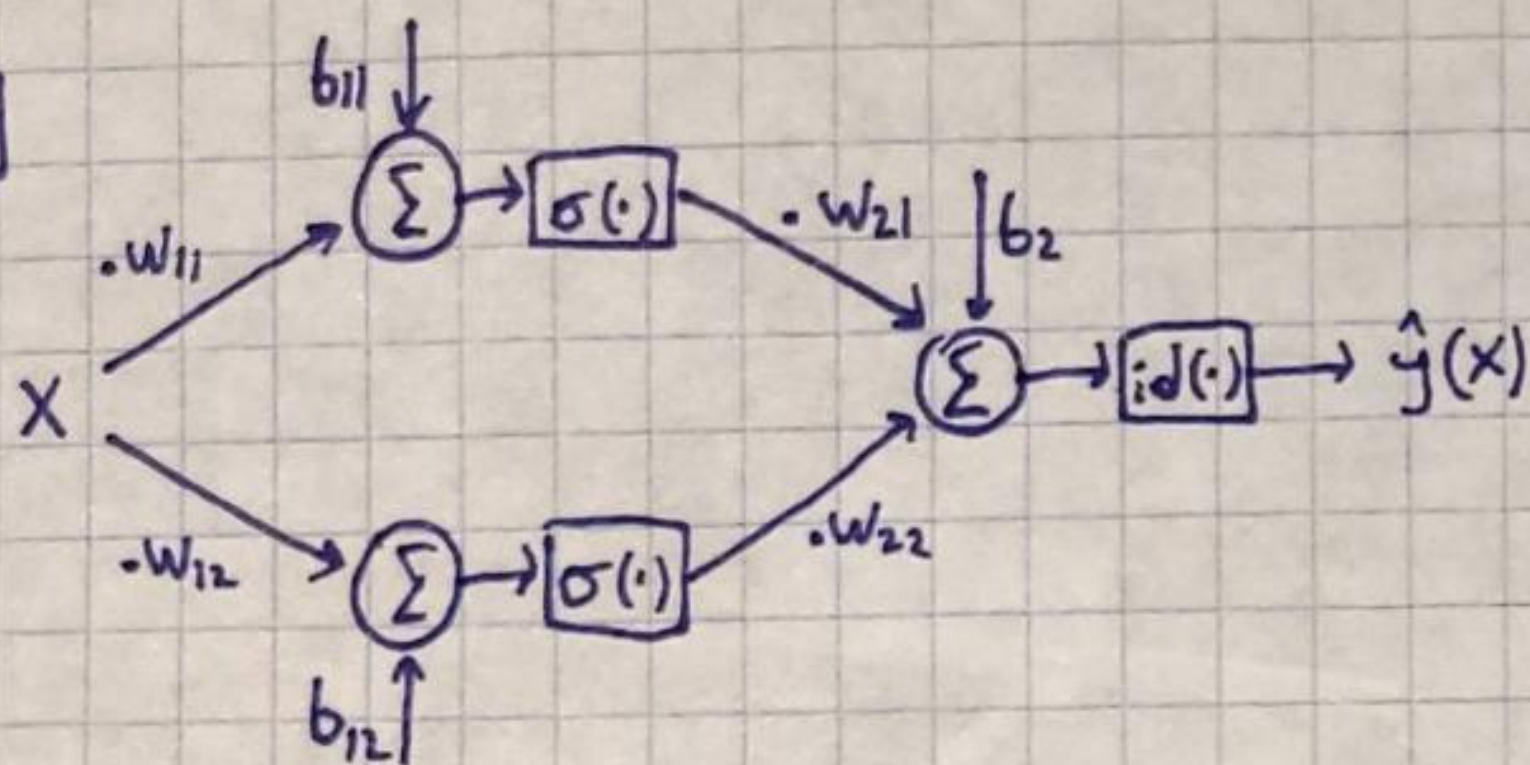


ЗАДАЧА 1

1



ОБУЧАЕМЫЕ ПАРАМЕТРЫ:

$w_{11}, w_{12}, w_{21}, w_{22}$
 b_{11}, b_{12}, b_2 } СУММАРНО 7 ПАРАМЕТРОВ

2.1

$$\frac{\partial \text{MSE}}{\partial \hat{y}(x_i)} = \frac{2}{n} (\hat{y} - y)^T, \text{ где } n - \text{РАЗМЕР ВЫБОРКИ (кол-во точек)}$$

2.2

$$\frac{\partial \hat{y}(x_i)}{\partial w_{2h}} = u_h \quad \frac{\partial \text{MSE}}{\partial w_{2h}} = \sum_{i=1}^n \frac{\partial \text{MSE}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_{2h}}$$

$$\frac{\partial \hat{y}(x_i)}{\partial b_2} = 1 \quad \frac{\partial \text{MSE}}{\partial b_2} = \sum_{i=1}^n \frac{\partial \text{MSE}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b_2}$$

$$\frac{\partial \hat{y}(x_i)}{\partial u_h} = w_{2h} \quad \frac{\partial \text{MSE}}{\partial u_h} = \sum_{i=1}^n \frac{\partial \text{MSE}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial u_h}$$

2.3 пусть для удобства $z_h = w_{1h}x + b_{1h}$

$$\frac{\partial \sigma(x)}{\partial x} = \left(\frac{1}{1+e^{-x}} \right)' = \frac{e^{-x}}{(1+e^{-x})^2} = \sigma(x) \cdot (1 - \sigma(x))$$

$$\frac{\partial u_h(x_i)}{\partial w_{1h}} = \frac{\partial \sigma(z_h)}{\partial z_h} \cdot \frac{\partial z_h}{\partial w_{1h}} = \sigma(z_h) \cdot (1 - \sigma(z_h)) \cdot x$$

$$\frac{\partial \text{MSE}}{\partial w_{1h}} = \sum_{i=1}^n \frac{\partial \text{MSE}}{\partial u_h} \cdot \frac{\partial u_h}{\partial w_{1h}}$$

$$\frac{\partial u_h(x_i)}{\partial b_{1h}} = \frac{\partial \sigma(z_h)}{\partial z_h} \cdot \frac{\partial z_h}{\partial b_{1h}} = \sigma(z_h) \cdot (1 - \sigma(z_h)) \cdot 1$$

$$\frac{\partial \text{MSE}}{\partial b_{1h}} = \sum_{i=1}^n \frac{\partial \text{MSE}}{\partial u_h} \cdot \frac{\partial u_h}{\partial b_{1h}}$$

3 Например, можно использовать mini-batch GD (разобьем большую обучающую выборку

на выборки поменьше, размера B). Правило обновления параметров:

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla \left(\frac{1}{B} \sum_{b=1}^B \text{MSE}(\hat{y}_t(x_{ib}), y_{ib}) \right), \text{ где сумма считается по батчу } \{x_{i1}, \dots, x_{iB}\}$$

Вывод: вычисление градиентов - достаточно затратная операция, однако на ГПУ и

с использованием SGD / mini-batch GD ее можно проводить более эффективно,

что позволяет применять алгоритм ко многим практическим задачам.