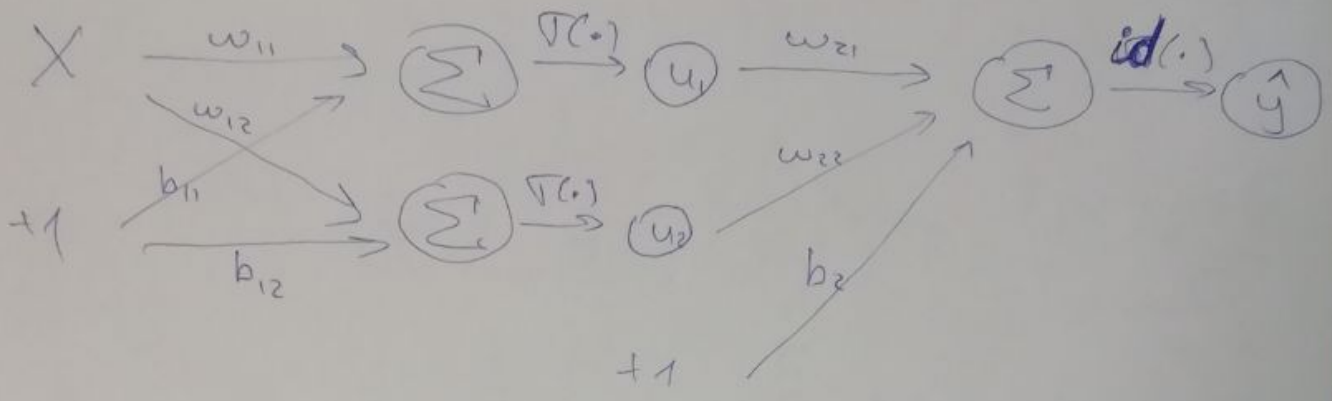
Задача 1

$$\sigma = \frac{1}{1+e^{-x}}$$

1)



7 параметров (обучаемых): $w_{11}, w_{12}, w_{21}, w_{22}, b_{11}, b_{12}, b_2$

2.1) $$L(X, Y) = MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{y}(X_i)\right)^2$$

$$\frac{\partial MSE}{\partial \hat{y}(X_i)} = -\frac{2}{n}\left(Y_i - \hat{y}(X_i)\right) = \frac{2}{n}\left(\hat{y}(X_i) - Y_i\right)$$

2.2) $$\frac{\partial \hat{y}(X_i)}{\partial w_{2h}} = u_h(X_i) \qquad \frac{\partial \hat{y}(X_i)}{\partial b_2} = 1 \qquad \frac{\partial \hat{y}(X_i)}{\partial u_h(X_i)} = w_{2h}$$

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

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

$$\frac{\partial MSE}{\partial u_h} = \sum_{i=1}^{n} \frac{\partial MSE}{\partial \hat{y}(X_i)} \cdot \frac{\partial \hat{y}(X_i)}{\partial u_h}$$

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

2.3)

$$\frac{\partial u_n(x_i)}{\partial w_{1n}} = \sigma'(w_{1n}x_i + b_{1n}) \cdot x_i$$

$$\frac{\partial u_n(x_i)}{\partial b_{1n}} = \sigma'(w_{1n}x_i + b_{1n}) \cdot 1$$

$$\hookrightarrow \frac{\partial MSE}{\partial w_{1n}} = \sum_{i=1}^{n} \frac{\partial MSE}{\partial u_n} \cdot \frac{\partial u_n}{\partial w_{1n}} =$$

$$= \sum_{i=1}^{n} \frac{\partial MSE}{\partial \hat{y}(x_i)} \cdot \frac{\partial \hat{y}(x_i)}{\partial u_n(x_i)} \cdot \frac{\partial u_n(x_i)}{\partial w_{1n}}$$

$$\hookrightarrow \frac{\partial MSE}{\partial b_{1n}} = \sum_{i=1}^{n} \frac{\partial MSE}{\partial \hat{y}(x_i)} \cdot \frac{\partial \hat{y}(x_i)}{\partial u_n(x_i)} \cdot \frac{\partial u_n(x_i)}{\partial b_{1n}}$$

3) Использовать SGD:

$$\theta_t = \theta_{t-1} - \eta \cdot \sum_{j=1}^{k} x_{i_j}(x_{i_j}\theta_t - y_{i_j}),$$

где $i_1, i_2, \dots, i_k \sim U\{1, \dots, n\}$ с повторами

(выбираем индексы на каждом шаге