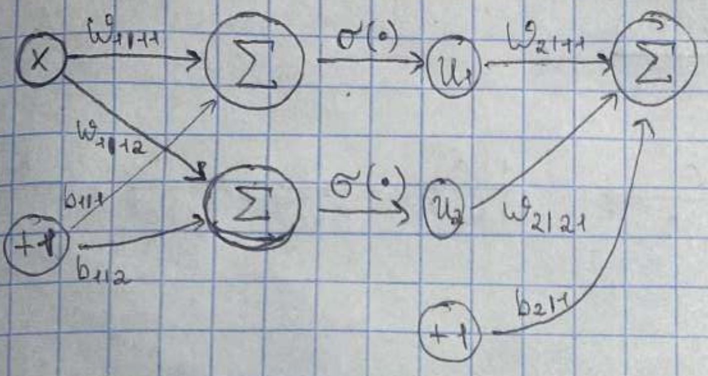
Задача 1.

① Схема нейронной сети




$\#\{\text{параметров}\} = 7 : \; w_{1|1|1}, \; w_{1|1|2}, \; b_{1|1}, \; b_{1|2}$

$w_{2|1|1}, \; w_{2|2|1}, \; b_{2|1}$

② Нахождение оптимальных параметров.

Обучающая выборка: $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$

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

$$\hat{y}(x) = \sum_{h=1}^{2} w_{2h} u_h(x) + b_2$$

2.1 $$\frac{\partial MSE}{\partial \hat{y}(X_i)} = \frac{1}{n}\cdot 2\cdot\left(Y_i - \hat{y}(X_i)\right)\cdot(-1) = -\frac{2}{n}\left(Y_i - \hat{y}(X_i)\right)$$

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

$$\frac{\partial \hat{y}(X_i)}{\partial b_2} = 1$$

$$\frac{\partial \hat{y}(X_i)}{\partial u_h(x)} = w_{2h}$$

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

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

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

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

$$= \sigma(x) - \sigma^2(x)$$

$$\frac{\partial u_h(x_i)}{\partial w_{1h}} = \frac{\partial(\sigma(w_{1h}x_i + b_{1h}))}{\partial w_{1h}} = \left(\sigma(w_{1h}x_i + b_{1h}) - \sigma^2(w_{1h}x_i + b_{1h})\right) \cdot x_i$$

$$\frac{\partial u_h(x_i)}{\partial b_{1h}} = \frac{\partial(\sigma(w_{1h}x_i + b_{1h}))}{\partial b_{1h}} = \sigma(w_{1h}x_i + b_{1h}) - \sigma^2(w_{1h}x_i + b_{1h})$$

Производные MSE по параметрам I слоя.

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

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

(3) Для обновления, в случае большой выборки, можно сделать "батч" из основной выборки $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, и ~~выполняя~~ использовать (SGD) Стохастический Градиентный Спуск:

$$\theta_t = \theta_{t-1} - \eta \cdot \nabla\left(\frac{1}{B} \cdot \sum_{b=1}^{B} MSE\left(y_{\theta_{t-1}}(x_{ib}), y_{ib}\right)\right)$$

формула из Лекции

Вывод: Мы выяснили, что "метод обратного распространения ошибки (backpropagation) можно эффективно вычислять градиент, и в следствие показали возможность ~~(нахождения)~~ ~~(эффектив.)~~ оптимального вычисления параметров $\hat{\theta}$.