

Рассмотрим очень маленькую сеть:
Вход \rightarrow один скрытый нейрон \rightarrow выход

Архитектура:

- вход: $x = 1.0$
- веса:
 - $w_1 = 0.4$ (вход \rightarrow скрытый)
 - $w_2 = 0.3$ (скрытый \rightarrow выход)
- смещения: $b_1 = 0.1, b_2 = 0.2$
- активация: сигмоида
- истинный ответ: $y = 1$

1. Прямой проход

1) Скрытый нейрон

$$z_1 = x \cdot w_1 + b_1 = 1 \cdot 0.4 + 0.1 = 0.5$$
$$h = \sigma(z_1) \approx 0.622$$

2) Выходной нейрон

$$z_2 = h \cdot w_2 + b_2 = 0.622 \cdot 0.3 + 0.2 \approx 0.3866$$
$$\hat{y} = \sigma(z_2) \approx 0.595$$

2. Ошибка

Используем MSE:

$$L = \frac{1}{2}(y - \hat{y})^2 \approx \frac{1}{2}(1 - 0.595)^2 \approx 0.082$$

3. Обратное распространение

Шаг 1. Ошибка на выходе

Для сигмoиды: $\sigma'(z) = \sigma(z) \cdot (1 - \sigma(z))$

$$\frac{\partial L}{\partial \hat{y}} = \hat{y} - y = -0.405$$
$$\frac{\partial \hat{y}}{\partial z_2} = \hat{y}(1 - \hat{y}) \approx 0.595 \cdot 0.405 \approx 0.241$$
$$\delta_2 = \frac{\partial L}{\partial z_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_2} = -0.405 \cdot 0.241 \approx -0.0976$$

Шаг 2. Градиенты весов и смещения второго слоя:

$$\frac{\partial L}{\partial w_2} = \delta_2 \cdot h \approx -0.0976 \cdot 0.622 \approx -0.0607$$
$$\frac{\partial L}{\partial b_2} = \delta_2 = -0.0976$$

Шаг 3. Ошибка на скрытом слое

$$\begin{aligned}\frac{\partial z_2}{\partial h} &= w_2 = 0.3 \\ \frac{\partial L}{\partial h} &= \delta_2 \cdot w_2 \approx -0.0976 \cdot 0.3 \approx -0.0293 \\ \frac{\partial h}{\partial z_1} &= h(1-h) \approx 0.622 \cdot 0.378 \approx 0.235 \\ \delta_1 &= \frac{\partial L}{\partial z_1} = \frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial z_1} \approx -0.0293 \cdot 0.235 \approx -0.00688\end{aligned}$$

Шаг 4. Градиенты для первого слоя:

$$\begin{aligned}\frac{\partial L}{\partial w_1} &= \delta_1 \cdot x \approx -0.00688 \cdot 1 \approx -0.00688 \\ \frac{\partial L}{\partial b_1} &= \delta_1 = -0.00688\end{aligned}$$

4. Обновление весов (например, $\eta = 0.1$)

$$\begin{aligned}w_2^{\text{нов}} &= w_2 - \eta \cdot \frac{\partial L}{\partial w_2} = 0.3 - 0.1 \cdot (-0.0607) = 0.30607 \\ b_2^{\text{нов}} &= b_2 - \eta \cdot \frac{\partial L}{\partial b_2} = 0.2 - 0.1 \cdot (-0.0976) = 0.20976 \\ w_1^{\text{нов}} &= w_1 - \eta \cdot \frac{\partial L}{\partial w_1} = 0.4 - 0.1 \cdot (-0.00688) = 0.400688 \\ b_1^{\text{нов}} &= b_1 - \eta \cdot \frac{\partial L}{\partial b_1} = 0.1 - 0.1 \cdot (-0.00688) = 0.100688\end{aligned}$$