5 Neural Networks

5.1 Backpropagation

i.

$$\text{Input } \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\text{Hidden Layer output } \mathbf{v} = \sigma(\mathbf{W}^{(1)}\mathbf{x}) = \begin{bmatrix} 4 \\ 0 \\ 11 \end{bmatrix}$$

$$\text{Output Layer output } \mathbf{y}_{predicted} = \mathbf{W}^{(2)}\mathbf{v} = \begin{bmatrix} 18 \end{bmatrix}$$

$$C(\mathbf{y}_{predicted}, \mathbf{y}_{target}) = \|[18] - [3]\|_2^2 = 225$$

ii. Let the input $\mathbf{x} = [x_1, x_2]^\top$ and weights be given by:

$$\mathbf{W}^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} \end{bmatrix} \qquad \mathbf{W}^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} \end{bmatrix}.$$

The output of the hidden layer $\mathbf{v}$ and output layer $\mathbf{y}_{predicted}$ are given by:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} \sigma(w_{11}^{(1)}x_1 + w_{12}^{(1)}x_2) \\ \sigma(w_{21}^{(1)}x_1 + w_{22}^{(1)}x_2) \\ \sigma(w_{31}^{(1)}x_1 + w_{32}^{(1)}x_2) \end{bmatrix}$$

$$\begin{aligned} \mathbf{y}_{predicted} &= \begin{bmatrix} w_{11}^{(2)}v_1 + w_{12}^{(2)}v_2 + w_{13}^{(2)}v_3 \end{bmatrix} \\ &= \begin{bmatrix} w_{11}^{(2)}\sigma(w_{11}^{(1)}x_1 + w_{12}^{(1)}x_2) + w_{12}^{(2)}\sigma(w_{21}^{(1)}x_1 + w_{22}^{(1)}x_2) + w_{13}^{(2)}\sigma(w_{31}^{(1)}x_1 + w_{32}^{(1)}x_2) \end{bmatrix}. \end{aligned}$$

$$\begin{aligned} \frac{\partial C}{\partial w_{11}^{(1)}} &= 2(\mathbf{y}_{predicted} - \mathbf{y}_{target})\frac{\partial \mathbf{y}_{predicted}}{\partial w_{11}^{(1)}} = 2(\mathbf{y}_{predicted} - \mathbf{y}_{target})w_{11}^{(2)}\frac{\partial v_1}{\partial w_{11}^{(1)}} \\ &= 2(\mathbf{y}_{predicted} - \mathbf{y}_{target})\, w_{11}^{(2)}\, \sigma'(w_{11}^{(1)}x_1 + w_{12}^{(1)}x_2)\, x_1 = -30 \end{aligned}$$

iii.

$$\begin{aligned}\frac{\partial C}{\partial w_{11}^{(2)}} &= 2(\mathbf{y}_{predicted} - \mathbf{y}_{target})\frac{\partial \mathbf{y}_{predicted}}{\partial w_{11}^{(2)}}\\ &= 2(\mathbf{y}_{predicted} - \mathbf{y}_{target})v_1 = 120\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{\partial C}{\partial w_{12}^{(2)}} &= 2(\mathbf{y}_{predicted} - \mathbf{y}_{target})v_2 = 0 \qquad \text{and}\\ \frac{\partial C}{\partial w_{13}^{(2)}} &= 2(\mathbf{y}_{predicted} - \mathbf{y}_{target})v_3 = 330.\end{aligned}$$

$$\frac{\partial C}{\partial \mathbf{W}^{(2)}} = \begin{bmatrix}\frac{\partial C}{\partial w_{11}^{(2)}} & \frac{\partial C}{\partial w_{12}^{(2)}} & \frac{\partial C}{\partial w_{13}^{(2)}}\end{bmatrix} = \begin{bmatrix}120 & 0 & 330\end{bmatrix}$$

iv.

$$\begin{aligned}\mathbf{W}_{updated}^{(2)} &= \mathbf{W}^{(2)} - \alpha\frac{\partial C}{\partial \mathbf{W}^{(2)}}\\ &= \begin{bmatrix}-7 & 3 & -14.5\end{bmatrix}\end{aligned}$$