Homework 3 - Group 076

Aprendizagem 2021/2022

1 Pen and Paper

1) Let $b^{[l]}$, $net^{[l]}$ and $a^{[l]} = \phi(net^{[l]})$ denote the vector of biases, net values and activations of the l-th layer, respectively (with $\phi_i(net^{[l]}) = \tanh(net_i^{[l]})$ being the activation function). Let $W^{[l]} = [w_{ij}]$ be the matrix of weights w_{ij} that connect the j-th activation of layer $l-1$ to the i-th net of layer l.

- **Forward Propagation**
Given that $a^{[l]} = \phi(net^{[l]}) = \phi(W^{[l]}a^{[l-1]} + b^{[l]})$ $(i \in \{1,2,3\})$, considering that $a^{[0]} = \mathbf{x}$:

$$\begin{aligned}
a^{[1]} = \phi(W^{[1]}a^{[0]} + b^{[1]}) &= \phi\left(\begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1.0 \\ 1.0 \\ 1.0 \end{bmatrix}\right) \\
&= \phi\left(\begin{bmatrix} 6.0 \\ 1.0 \\ 6.0 \end{bmatrix}\right) = \begin{bmatrix} \tanh(6.0) \\ \tanh(1.0) \\ \tanh(6.0) \end{bmatrix} = \begin{bmatrix} 0.99999 \\ 0.76159 \\ 0.99999 \end{bmatrix}
\end{aligned}$$

$$a^{[2]} = \phi(W^{[2]}a^{[1]} + b^{[2]}) = \phi\left(\begin{bmatrix} 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \end{bmatrix} \begin{bmatrix} 0.99999 \\ 0.76159 \\ 0.99999 \end{bmatrix} + \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}\right) = \begin{bmatrix} 0.99892 \\ 0.99892 \end{bmatrix}$$

$$a^{[3]} = \phi(W^{[3]}a^{[2]} + b^{[3]}) = \phi\left(\begin{bmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \end{bmatrix} \begin{bmatrix} 0.99892 \\ 0.99892 \end{bmatrix} + \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix}\right) = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix}$$

- **Backpropagation**
Consider the squared error loss $E = \frac{1}{2}\sum_{i=1}^{2}(\mathbf{z}_i - \hat{\mathbf{z}}_i)^2 = \frac{1}{2}\sum_{i=1}^{2}(\mathbf{z}_i - a_i^{[l]})^2$ and define $\delta^{[l]} = \nabla_{net^{[l]}}E$. By the chain rule of derivation, we have:

$$\begin{aligned}
\delta^{[l]} &= \nabla_{net^{[l]}}E = \nabla_{net^{[l]}}a^{[l]}\nabla_{a^{[l]}}net^{[l+1]}\nabla_{net^{[l+1]}}E \\
&= \mathrm{diag}(\tanh'(net^{[1]}), ..., \tanh'(net^{[n_l]}))(W^{[l+1]})^T\delta^{[l+1]} \\
&= [\tanh'(net^{[1]}), ..., \tanh'(net^{[n_l]})]^T \circ \left((W^{[l+1]})^T\delta^{[l+1]}\right)
\end{aligned}$$

where n_l is the number of units in the l-th layer