CSDS 440: Assignment 5

Shaochen (Henry) ZHONG, `sxz517`

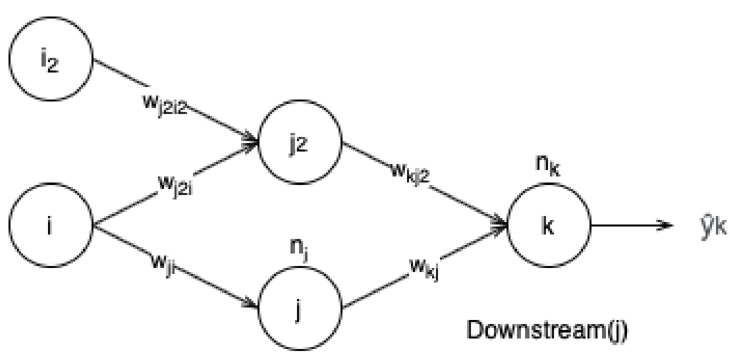
Mingyang TIE, `mxt497`

Due on 10/09/2020, submitted early on 10/02/2020
Fall 2020, Dr. Ray

Problem 20

We denote $h(u)$ to be the activation function for u and we do not make assumption of which activation function is used (so we won't have the sigmoid simplication of $\frac{\partial h}{\partial u} = h(u)(1 - h(u))$). For traning example α, $h(\alpha) = o_\alpha$ (representing the prediction output), and we denotes $y = y_\alpha$ for the output label.

For the ease of understanding, we denote our symbols with respect to the following NN, which use almost identical notation to the one taught in *Lecture 10*.



Hidden-to-Output

$$
\begin{aligned}
\frac{\partial L}{\partial w_{ji}} &= \frac{\partial L}{\partial n_j}\frac{\partial n_j}{w_{ji}} \\
&= \frac{\partial L}{\partial n_j} \cdot \frac{\partial (w_{ji} \cdot x_{ji})}{w_{ji}} \\
&= \frac{\partial L}{\partial n_j} \cdot x_{ji} \\
&= \frac{\partial L}{\partial h(n_j)}\frac{\partial h(n_j)}{\partial n_j} \cdot x_{ji} \\
&= \frac{\partial}{\partial h(n_j)}(-y\log(h(n_j)) - (1-y)\log(1-h(n_j))) \cdot \frac{\partial h(n_j)}{\partial n_j} \cdot x_{ji} \\
&= (-\frac{-y}{h(n_j)} + \frac{1-y}{1-h(n_j)}) \cdot \frac{\partial h(n_j)}{\partial n_j} \cdot x_{ji} \\
&= (\frac{1-y}{1-h(n_j)} - \frac{-y}{h(n_j)}) \cdot \frac{\partial h(n_j)}{\partial n_j} \cdot x_{ji}
\end{aligned}
$$

Base on the calucation of the output layer, we may tall the backpropagation weight updates for hidden-to-output layer is:

$$
\frac{\partial L}{\partial w_{ji}} = (\frac{1-y}{1-h(n_j)} - \frac{-y}{h(n_j)}) \cdot \frac{\partial h(n_j)}{\partial n_j} \cdot x_{ji}
$$

Input-to-Hidden

Since j affects the $\hat{y}$ outputs only through $Downstream(j)$, we have:

$$
\begin{aligned}
\frac{\partial L}{\partial w_{ji}} &= \frac{\partial L}{\partial n_j} \cdot \frac{\partial n_j}{\partial w_{ji}} \\
&= \frac{\partial L}{\partial n_j} \cdot \frac{\partial (w_{ji} \cdot x_{ji})}{w_{ji}} \\
&= \frac{\partial L}{\partial n_j} \ \cdot x_{ji} \\
&= \sum_{k \in Downstream(j)} \frac{\partial L}{\partial n_k} \cdot \frac{\partial n_k}{\partial n_j} \ \cdot x_{ji} \\
&= \sum_{k \in Downstream(j)} \frac{\partial L}{\partial n_k} \cdot \frac{\partial (w_{kj} h(n_j))}{\partial n_j} \ \cdot x_{ji} \\
&= \sum_{k \in Downstream(j)} \frac{\partial L}{\partial n_k} \cdot w_{kj} \frac{\partial h(n_j)}{\partial n_j} \ \cdot x_{ji} \\
&= \frac{\partial h(n_j)}{\partial n_j} \cdot \sum_{k \in Downstream(j)} \frac{\partial L}{\partial n_k} \cdot w_{kj} \ \cdot x_{ji} \\
&= \frac{\partial h(n_j)}{\partial n_j} x_{ji} \cdot \sum_{k \in Downstream(j)} \frac{\partial L}{\partial w_{kj}} \frac{w_{kj}}{x_{kj}}
\end{aligned}
$$

Base on the calucation of the hidden layer, we may tall the backpropagation weight updates for input-to-output layer is:

$$
\frac{\partial L}{\partial w_{ji}} = \frac{\partial h(n_j)}{\partial n_j} x_{ji} \cdot \sum_{k \in Downstream(j)} \frac{\partial L}{\partial w_{kj}} \frac{w_{kj}}{x_{kj}}
$$

I don't know if it is required (as it is now showed on the slide), but if you'd expect us to show an input-to-hidden equation with loss function plugged in, it will be:

$$
\frac{\partial L}{\partial w_{ji}} = \frac{\partial h(n_j)}{\partial n_j} x_{ji} \cdot \sum_{k \in Downstream(j)} \left(\frac{1-y}{1-h(n_k)} - \frac{-y}{h(n_k)}\right) \cdot \frac{\partial h(n_k)}{\partial n_k} \ \cdot x_{kj} \ \cdot \frac{w_{kj}}{x_{kj}}
$$