Homework 2

Yunian Pan

February 21, 2019

1 Problem 1: Backpropagation

1.1 case 1

Sol:

Given $x_i = g(y_j) = \dfrac{1}{1+e^{-\sum_j w_{ji} y_j}}$, and $\sum_i \dfrac{\partial E}{\partial x_i} = -\sum_i (\dfrac{t_i}{x_i} - \dfrac{1-t_i}{1-x_i})$ apply the chain rule, first take x_i and y_j as fixed to compute the gradient for w_{ji}, then take z_k and y_j as fixed to compute the gradient for w_{kj} we have

$$\begin{aligned}
\frac{\partial E}{\partial w_{ji}} &= \frac{\partial E}{\partial x_i}\frac{\partial x_i}{\partial w_{ji}} \\
&= (\frac{1-t_i}{1-x_i} - \frac{t_i}{x_i})\frac{\partial g(w_{ji}, y_j)}{\partial w_{ji}} \\
&= (\frac{1-t_i}{1-x_i} - \frac{t_i}{x_i})x_i^2 y_j e^{-\sum_j w_{ji} y_j} \\
&= (\frac{1-t_i}{1-x_i} - \frac{t_i}{x_i})x_i y_j (1-x_i) \\
&= (x_i - t_i) y_j
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial E}{\partial w_{kj}} &= \sum_i (\frac{\partial E}{\partial x_i}\frac{\partial x_i}{\partial y_j})\frac{\partial y_j}{\partial w_{kj}} \\
&= \sum_i \frac{\partial E}{\partial x_i}(-w_{ji})x_i(1-x_i)(-z_k)y_j(1-y_j) \\
&= \sum_i \frac{x_i - x_i t_i - t_i + x_i t_i}{(1-x_i)x_i}(-(-w_{ji})x_i(1-x_i))(-(-z_k)y_j(1-y_j)) \\
&= \sum_i (x_i - t_i)w_{ji}y_j(1-y_j)z_k
\end{aligned}
$$

Thus, we have $\quad \sum_i \frac{\partial E}{\partial w_{ji}} = \sum_i \delta_j^i y_j, \qquad \sum_j \frac{\partial E}{\partial w_{kj}} = \sum_j \delta_k^j z_k.$

1.2 case 2

Sol:

Given the cross-entropy $E = -\sum_i t_i \log(x_i)$ and softmax activation function $x_i = \frac{e^{\sum_j w_{ji}y_j}}{\sum_i e^{\sum_j w_{ji}y_j}} = f(w_{11}, \ldots, w_{j1}, \ldots, w_{ji}, \ldots, w_{jm}, \ldots, y_j, \ldots)$, we have

$$
\frac{\partial E}{\partial x_i} = -\frac{t_i}{x_i}
$$

$$
\begin{aligned}
\frac{\partial x_m}{\partial w_{ji}} &= \frac{y_j(\delta(i-m)e^{\sum_j w_{jm}y_j}(\sum_i e^{\sum_j w_{ji}y_j}) - e^{\sum_j w_{jm}y_j}e^{\sum_j w_{ji}y_j})}{(\sum_i e^{\sum_j w_{ji}y_j})^2} \\
&= y_j x_m(\delta(i-m) - x_i)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial E}{\partial w_{ji}} &= \sum_m \frac{\partial E}{\partial x_m}\frac{\partial x_m}{\partial w_{ji}} \\
&= \sum_m y_j(-\frac{t_m}{x_m})x_m(\delta(i-m) - x_i) \\
&= y_j(\sum_m t_m \cdot x_i - t_i)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial E}{\partial w_{kj}} &= \sum_i (\sum_m (\frac{\partial E}{\partial x_m} \frac{\partial x_m}{\partial y_j}) \frac{\partial y_j}{\partial w_{kj}}) \\
&= \sum_i (w_{ji} (\sum_m t_m x_i - t_i)) y_j (1 - y_j) z_k \\
&= \sum_i (\sum_m t_m x_i - t_i)(w_{ji} y_j (1 - y_j)) z_k
\end{aligned}
$$

Here $\delta_j^i = \sum_m t_m \cdot x_i - t_i, \quad \delta_k^j = \sum_i (\sum_m t_m x_i - t_i)(w_{ji} y_j (1 - y_j)).$