ECE5984 Homework3

소프트웨어학과 2022710836 김민근

1. Backpropagation

(a).

By definition,

$$y_k = \sum_{n=1}^{d} W_{kn}\, x_n + b_k$$

And by using partial derivative and Dirac delta function,

$$\frac{\partial W_{kn}}{\partial W_{ij}} = \begin{cases} 1 \dots (k = i, n = j) \\ 0 \dots . (otherwise) \end{cases} = \delta_{ik}\delta_{jn}$$

$$\therefore \frac{\partial y_k}{\partial W_{ij}} = \sum_{n=1}^{d} \delta_{ik}\delta_{jn}\, x_n$$

$$\therefore \frac{\partial J}{\partial W_{ij}} = \sum_{k=1}^{m}\sum_{n=1}^{d} \frac{\partial J}{\partial y_k}\delta_{ik}\delta_{jn}\, x_n$$

$$= \sum_{k=1}^{m} \delta_{ik}\frac{\partial J}{\partial y_k} \sum_{n=1}^{d} \delta_{jn} x_n$$

$In \sum_{k=1}^{m} \delta_{ik}\frac{\partial J}{\partial y_k}$,

$$\sum_{k=1}^{m} \delta_{ik}\frac{\partial J}{\partial y_k} = 0_1 + 0_2 + \cdots + \frac{\partial J}{\partial y_i} + \cdots + 0_m = \frac{\partial J}{\partial y_i}$$

$$\left(Because\ \ \delta_{ik} = \begin{cases} 1 \dots (i = k) \\ 0 \dots . (otherwise) \end{cases}\right)$$

$In \sum_{n=1}^{d} \delta_{jn} x_n,$

$$\sum_{n=1}^{d} \delta_{jn} x_n = 0_1 + 0_2 + \cdots + x_j + \cdots + 0_d = x_j$$

$$\left(Because\ \ \delta_{jn} = \begin{cases} 1 \dots (j = n) \\ 0 \dots . (otherwise) \end{cases}\right)$$

$$\therefore \frac{\partial J}{\partial W_{ij}} = \frac{\partial J}{\partial y_i} x_j$$

(b).

$$x = (x_1, x_2, x_3, \dots, x_d)^T$$

$$\frac{\partial J}{\partial y} = \left(\frac{\partial J}{\partial y_1}, \frac{\partial J}{\partial y_2}, \frac{\partial J}{\partial y_3}, \dots, \frac{\partial J}{\partial y_m}\right)^T$$

$$By\ \frac{\partial J}{\partial W_{ij}} = \frac{\partial J}{\partial y_i} x_j\ \ in\ \ problem\ 1\text{-}(a),$$

$$\frac{\partial J}{\partial W_{ij}} = \frac{\partial J}{\partial y_i} x_j$$

$$And\ x^T \in \mathbb{R}^{d\times 1}, \qquad \frac{\partial J}{\partial y} \in \mathbb{R}^{1\times m}$$

$$\therefore \frac{\partial J}{\partial W} = \frac{\partial J}{\partial y} x^T$$

(c).

By Chain rule,

$$\frac{\partial J}{\partial x_j} = \sum_{k=1}^{m} \frac{\partial y_k}{\partial x_j} \cdot \frac{\partial J}{\partial y_k}$$

And by definition,

$$y_k = \sum_{n=1}^{d} W_{kn} x_n + b_k$$

Using partial derivative.

$$\frac{\partial x_n}{\partial x_j} = \begin{cases} 1 \dots. & (n = j) \\ 0 \dots. & (otherwise) \end{cases}$$

$$\therefore \frac{\partial y_k}{\partial x_j} = \sum_{n=1}^{d} \frac{\partial W_{kn} x_n}{\partial x_j} = 0_1 + 0_2 + \dots + W_{kj} + \dots + 0_d = W_{kj}$$

$$\therefore \frac{\partial J}{\partial x_j} = \sum_{k=1}^{m} W_{kj} \frac{\partial J}{\partial y_k} = \sum_{k=1}^{m} {W^T}_{jk} \frac{\partial J}{\partial y_k}$$

$$\left(Because\quad W_{kj} = {W^T}_{jk}\right)$$

$$\therefore \frac{\partial J}{\partial x} = W^T \frac{\partial J}{\partial y}$$