DS-GA 1008: Deep Learning - Homework 1 (Spring 2020)

Alfred Rajakumar

February 18, 2020

1 Problem - 1 Backpropagation:

1.1 Affine Module:

Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$, affine transformation $\boldsymbol{W} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \in \mathbb{R}^{2\times 2}$ and $\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \in \mathbb{R}^2$. Now, the output $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in \mathbb{R}^2$ is given by:

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{b} \tag{1}$$

a.

$$\begin{aligned}
\frac{\partial C}{\partial \boldsymbol{W}} &= \frac{\partial C}{\partial \boldsymbol{y}} \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{W}} \\
\implies \begin{bmatrix} \frac{\partial C}{\partial w_{11}} & \frac{\partial C}{\partial w_{12}} \\ \frac{\partial C}{\partial w_{21}} & \frac{\partial C}{\partial w_{22}} \end{bmatrix} &= \begin{bmatrix} \frac{\partial C}{\partial y_1} & \frac{\partial C}{\partial y_2} \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial w_{11}} & \frac{\partial y_1}{\partial w_{12}} \\ \frac{\partial y_1}{\partial w_{21}} & \frac{\partial y_1}{\partial w_{22}} \end{bmatrix}, \begin{bmatrix} \frac{\partial y_2}{\partial w_{11}} & \frac{\partial y_2}{\partial w_{12}} \\ \frac{\partial y_2}{\partial w_{21}} & \frac{\partial y_2}{\partial w_{22}} \end{bmatrix} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial C}{\partial y_1} & \frac{\partial C}{\partial y_2} \end{bmatrix} \begin{bmatrix} \begin{bmatrix} x_1 & x_2 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ x_1 & x_2 \end{bmatrix} \end{bmatrix} \\
\therefore \frac{\partial C}{\partial \boldsymbol{W}} &= \begin{bmatrix} x_1 \frac{\partial C}{\partial y_1} & x_2 \frac{\partial C}{\partial y_1} \\ x_1 \frac{\partial C}{\partial y_2} & x_2 \frac{\partial C}{\partial y_2} \end{bmatrix}
\end{aligned} \tag{2}$$

Similarly,

$$\begin{aligned}
\frac{\partial C}{\partial \boldsymbol{b}} &= \frac{\partial C}{\partial \boldsymbol{y}} \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{b}} \\
\implies \begin{bmatrix} \frac{\partial C}{\partial b_1} \\ \frac{\partial C}{\partial b_2} \end{bmatrix} &= \begin{bmatrix} \frac{\partial C}{\partial y_1} & \frac{\partial C}{\partial y_2} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial b_1} & \frac{\partial y_2}{\partial b_1} \\ \frac{\partial y_1}{\partial b_2} & \frac{\partial y_2}{\partial b_2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial C}{\partial y_1} & \frac{\partial C}{\partial y_2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
\therefore \frac{\partial C}{\partial \boldsymbol{b}} &= \begin{bmatrix} \frac{\partial C}{\partial y_1} \\ \frac{\partial C}{\partial y_2} \end{bmatrix} = \frac{\partial C}{\partial \boldsymbol{y}}
\end{aligned} \tag{3}$$

b. Now, $C_2(\boldsymbol{y}) = 3C(\boldsymbol{y})$. So, in this case,

$$\frac{\partial C_2}{\partial \boldsymbol{W}} = \frac{\partial C}{\partial C_2} \frac{\partial C}{\partial \boldsymbol{W}} = 3\frac{\partial C}{\partial \boldsymbol{W}} \tag{4}$$

Similarly,

$$\frac{\partial C_2}{\partial \boldsymbol{b}} = \frac{\partial C}{\partial C_2} \frac{\partial C}{\partial \boldsymbol{b}} = 3\frac{\partial C}{\partial \boldsymbol{b}} \tag{5}$$

where $\frac{\partial C}{\partial \boldsymbol{W}}$ and $\frac{\partial C}{\partial \boldsymbol{b}}$ are same as given in eqns. (2) and (3) respectively.

1.2 Softmax Module:

We have, $\boldsymbol{y} = softmax_\beta(\boldsymbol{x})$, i.e. $y_k = \frac{\exp(\beta x_k)}{\sum_n \exp(\beta x_n)}$, where $\sum_k y_k = 1$, $y_k \geq 0$ and $\boldsymbol{x} \in \mathbb{R}^k$

Differentiating y_i w.r.t. each component of $\boldsymbol{x}$, gives,

$$
\begin{aligned}
\frac{\partial y_i}{\partial x_j} &= \begin{cases} \frac{\left[\sum_n \exp(\beta x_n)\right] \beta \exp(\beta x_i) - \beta \exp(\beta x_i) \exp(\beta x_i)}{\left[\sum_n \exp(\beta x_n)\right]^2}, & \text{when } i = j \\ \frac{\left[\sum_n \exp(\beta x_n)\right] 0 - \beta \exp(\beta x_i) \exp(\beta x_j)}{\left[\sum_n \exp(\beta x_n)\right]^2}, & \text{when } i \neq j \end{cases} \\
\implies \frac{\partial y_i}{\partial x_j} &= \begin{cases} \frac{\beta \sum_{i \neq n} \exp(\beta(x_i + x_n))}{\left[\sum_n \exp(\beta x_n)\right]^2}, & \text{when } i = j \\ \frac{-\beta \exp(\beta x_i) \exp(\beta x_j)}{\left[\sum_n \exp(\beta x_n)\right]^2}, & \text{when } i \neq j \end{cases}
\end{aligned} \tag{6}
$$