DS-GA 1008 HW1 - Part 1

Nikhil Supekar

February 18, 2020

1 Affine Module

1.1 Question 1a)

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

$$\begin{aligned}
\frac{\partial C}{\partial \mathbf{W}} &= \begin{pmatrix} \frac{\partial C}{\partial w_{11}} & \frac{\partial C}{\partial w_{12}} \\ \frac{\partial C}{\partial w_{21}} & \frac{\partial C}{\partial w_{22}} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial C}{\partial y_1} \cdot \frac{\partial y_1}{\partial w_{11}} + \frac{\partial C}{\partial y_2} \cdot \frac{\partial y_2}{\partial w_{11}} & \frac{\partial C}{\partial y_1} \cdot \frac{\partial y_1}{\partial w_{12}} + \frac{\partial C}{\partial y_2} \cdot \frac{\partial y_2}{\partial w_{12}} \\ \frac{\partial C}{\partial y_1} \cdot \frac{\partial y_1}{\partial w_{21}} + \frac{\partial C}{\partial y_2} \frac{\partial y_2}{\partial w_{21}} & \frac{\partial C}{\partial y_1} \cdot \frac{\partial y_1}{\partial w_{22}} + \frac{\partial C}{\partial y_2} \cdot \frac{\partial y_2}{\partial w_{22}} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial C}{\partial y_1} \cdot x_1 + \frac{\partial C}{\partial y_2} \cdot 0 & \frac{\partial C}{\partial y_1} \cdot x_2 + \frac{\partial C}{\partial y_2} \cdot 0 \\ \frac{\partial C}{\partial y_1} \cdot 0 + \frac{\partial C}{\partial y_2} \cdot x_1 & \frac{\partial C}{\partial y_1} \cdot 0 + \frac{\partial C}{\partial y_2} \cdot x_2 \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial C}{\partial y_1} x_1 & \frac{\partial C}{\partial y_1} x_2 \\ \frac{\partial C}{\partial y_2} x_1 & \frac{\partial C}{\partial y_2} x_2 \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial C}{\partial y_1} \\ \frac{\partial C}{\partial y_2} \end{pmatrix} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \\
&= \frac{\partial C}{\partial \mathbf{y}} \mathbf{x}^\top
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial C}{\partial \mathbf{b}} &= \begin{pmatrix} \frac{\partial C}{\partial b_1} & \frac{\partial C}{\partial b_2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial C}{\partial y_1}\frac{\partial y_1}{\partial b_1} + \frac{\partial C}{\partial y_2}\frac{\partial y_2}{\partial b_1} & \frac{\partial C}{\partial y_1} \cdot \frac{\partial y_1}{\partial b_2} + \frac{\partial C}{\partial y_2}\frac{\partial y_2}{\partial b_2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial C}{\partial y_1} \cdot 1 + \frac{\partial C}{\partial y_2} \cdot 0 & \frac{\partial C}{\partial y_1} \cdot 0 + \frac{\partial C}{\partial y_2} \cdot 1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial C}{\partial y_1} & \frac{\partial C}{\partial y_2} \end{pmatrix} \\
&= \left(\frac{\partial C}{\partial \mathbf{y}} \right)^T
\end{aligned}
$$

1.2 Question 1b)

$$
\begin{aligned}
\frac{\partial C_2}{\partial \mathbf{W}} &= \begin{pmatrix} \frac{\partial C_2}{\partial w_{11}} & \frac{\partial C_2}{\partial w_{12}} \\ \frac{\partial C_2}{\partial w_{21}} & \frac{\partial C_2}{\partial w_{22}} \end{pmatrix} \\
&= \begin{pmatrix} 3\frac{\partial C}{\partial w_{11}} & 3\frac{\partial C}{\partial w_{12}} \\ 3\frac{\partial C}{\partial w_{21}} & 3\frac{\partial C}{\partial w_{22}} \end{pmatrix} \\
&= 3 \begin{pmatrix} \frac{\partial C}{\partial w_{11}} & \frac{\partial C}{\partial w_{12}} \\ \frac{\partial C}{\partial w_{21}} & \frac{\partial C}{\partial w_{22}} \end{pmatrix} \\
&= 3\frac{\partial C}{\partial \mathbf{y}}\mathbf{x}^\top
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial C_2}{\partial b} &= \begin{pmatrix} \frac{\partial C_2}{\partial b_1} & \frac{\partial C_2}{\partial b_2} \end{pmatrix} \\
&= 3 \begin{pmatrix} \frac{\partial C}{\partial b_1} & \frac{\partial C}{\partial b_2} \end{pmatrix} \\
&= 3 \left(\frac{\partial C}{\partial \mathbf{y}} \right)^\top
\end{aligned}
$$

2 Softmax Module

2.1 Question 2

$$y_i = \frac{\exp(\beta x_i)}{\sum_{n=1}^{K} \exp(\beta x_n)}$$

$$\begin{aligned}
\frac{\partial y_i}{\partial x_i} &= \frac{\beta \exp(\beta x_i) \sum_{n=1}^{K} \exp(\beta x_n) - \beta \exp(\beta x_i) \exp(\beta x_i)}{\left(\sum_{n=1}^{K} \exp(\beta x_n)\right)^2} \\
&= \beta \exp(\beta x_i) \frac{\sum_{n=1, n \neq i}^{K} \exp(\beta x_n)}{\left(\sum_{n=1}^{K} \exp(\beta x_n)\right)^2}
\end{aligned}$$

For $i \neq j$,

$$\begin{aligned}
\frac{\partial y_i}{\partial x_j} &= \frac{-\exp(\beta x_i)}{\left(\sum_{n=1}^{K} \exp(\beta x_n)\right)^2} \beta \exp(\beta x_j) \\
&= \frac{-\beta \exp(\beta (x_i + x_j))}{\left(\sum_{n=1}^{K} \exp(\beta x_n)\right)^2}
\end{aligned}$$