VDL2

Erik Schwede
Abdalla Arafa
Ehsan Attar
Sai Leela Poduru
Prateek Rathod
Shruti Shrivastava

November 2023

Q2.2 Loss Functions and Optimization

i Derivative of softmax

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{k=1}^{n} e^{z_k}} \tag{1}$$

$$\tag{2}$$

Derivative for condition i = j using quotient rule:

$$\frac{\partial \hat{y}_i}{\partial z_i} = \frac{e^{z_i} \sum e^{z_k} - e^{z_i}.e^{z_i}}{(\sum_k e^{z_k})^2} \tag{3}$$

$$= \frac{e^{z_i}(1 - e^{z_i})}{(\sum_k e^{z_k})^2} \tag{4}$$

$$= \hat{y}_i(1 - \hat{y}_i) \tag{5}$$

Derivative for condition i != j using quotient rule:

$$\frac{\partial \hat{y}_i}{\partial z_j} = \frac{0 - e^{z_i}.e^{z_j}}{(\sum_k e^{z_k})^2} \tag{6}$$

$$= -\hat{y}_i.\hat{y}_j \tag{7}$$

ii Derivative of cross-entropy loss function defined as

$$L(y, \hat{y}) = -\sum_{k=1}^{N} y_k \log \hat{y} \tag{8}$$

$$\frac{\partial L}{\partial \hat{y}_i} = -\sum_i y_i.\frac{1}{\hat{y}_i} \tag{9}$$

Using chain rule:

$$\frac{\partial L}{\partial z_i} = \frac{\partial L}{\partial \hat{y}_i} . \frac{\partial \hat{y}_i}{\partial z_i} \tag{10}$$

$$\begin{aligned}
\frac{\partial L}{\partial z_i} &= -\sum_{i \neq j} y_i . \frac{1}{\hat{y}_i} \frac{\partial \hat{y}_i}{\partial z_i} - y_i . \frac{1}{\hat{y}_i} \frac{\partial \hat{y}_i}{\partial z_i} && (11) \\
&= -\sum_{i \neq j} y_i . \frac{1}{\hat{y}_i} . (-\hat{y}_i \hat{y}_j) - y_i . \frac{1}{\hat{y}_i} . \hat{y}_i (1 - \hat{y}_i) && (12) \\
&= \sum_{i \neq j} y_i . \hat{y}_j + y_i \hat{y}_i - y_i && (13) \\
&= \sum_{i} y_i . \hat{y}_j - y_i && (14) \\
&= \hat{y}_j - y_i && (15)
\end{aligned}$$