

I. Pencil-and-Paper

1. Derivative of Softmax

Recall the softmax function:

$$\vec{y}[k] = \frac{e^{\vec{z}[k]}}{\sum_{l=1}^C e^{\vec{z}[l]}} \quad (1)$$

So, consider two cases, $j = k$ and $j \neq k$.

When $j = k$:

$$\frac{\partial \vec{y}[k]}{\partial \vec{z}[j]} = -\frac{e^{\vec{z}[k]} \cdot \sum_{l=1}^C e^{\vec{z}[l]} - e^{\vec{z}[k]} \cdot e^{\vec{z}[k]}}{(\sum_{l=1}^C e^{\vec{z}[l]})^2} = \vec{y}[k] - \vec{y}[k] \cdot \vec{y}[k] \quad (2)$$

When $j \neq k$:

$$\frac{\partial \vec{y}[k]}{\partial \vec{z}[j]} = -\frac{e^{\vec{z}[k]} \cdot e^{\vec{z}[j]}}{(\sum_{l=1}^C e^{\vec{z}[l]})^2} = -\vec{y}[k] \cdot \vec{y}[j] \quad (3)$$

So, we can conclude that:

$$\frac{\partial \vec{y}[k]}{\partial \vec{z}[j]} = \begin{cases} -\vec{y}[k] \cdot \vec{y}[j] & \text{if } k \neq j \\ \vec{y}[k] - \vec{y}[k] \cdot \vec{y}[k] & \text{if } k = j \end{cases} \quad (4)$$

2. Negative Log Likelihood loss for Multi-Class.

Recall the negative log likelihood:

$$L = -\sum_i^N \sum_k^K \mathbf{1}[y_i = k] \cdot \log(\hat{y}_i[k]) \quad (5)$$

In this way, we have:

$$\frac{\partial L}{\partial \hat{y}_i[j]} = -\mathbf{1}[y_i = k] \cdot \frac{\partial \log(\hat{y}_i[j])}{\partial \hat{y}_i[j]} = -\frac{\mathbf{1}[y_i = j]}{\hat{y}_i[j]} \quad (6)$$

3. Avg-pooling (1D)

Recall Avg-pooling (1D) operation with window size W :

$$\vec{y}[i] = \frac{1}{W} \sum_{j=0}^W \vec{x}[i+j] \quad (7)$$

Then, we have:

$$\frac{\partial \vec{y}[i]}{\partial \vec{x}[j]} = \begin{cases} -\frac{1}{W} & \text{if } i \leq j \leq i+W \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

4. Max-pooling (1D)

Recall Max-pooling (1D) operation with window size W :

$$\vec{y}[i] = \max_{j=0}^W \vec{x}[i+j] \quad (9)$$

Then, we have:

$$\frac{\partial \vec{y}[i]}{\partial \vec{x}[j]} = \begin{cases} 1 & \text{if } \max_{k=0}^W \vec{x}[i+k] = \vec{x}[j] \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

5. Convolutional layer (1D)

Recall Convolution (1D) operation, assume \vec{w} is length 3, and zero index at the center:

$$\vec{y}[i] = (\vec{w} * \vec{x})[i] = \sum_{j=-1}^1 \vec{x}[i-j] \cdot \vec{w}[j] \quad (11)$$

From above equation, we have:

$$\frac{\vec{y}[i]}{\vec{x}[j]} = \begin{cases} \vec{w}[i-j] & \text{if } i-1 \leq j \leq i+1 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

$$\frac{\vec{y}[i]}{\vec{w}[j]} = \begin{cases} \vec{x}[i-j] & \text{if } j = -1 \text{ or } 0 \text{ or } 1 \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

II. Code-from-Scratch

1. Methods

2. Results

III. TensorFlow

1. Methods

2. Results