Cross Entropy loss

$$L = \text{Cross}E(A, P)$$

A = Ground Truth distribution

P = Predicted distribution

$A \in R^{m \times n}$ $\quad m$ = batch size

$P \in R^{m \times n}$ $\quad n$ = Input dim.

$L \in R^{m}$

Forward Propagation

$$L = \text{Cross}E(A, P)$$

$$L_i = -\sum_{j=1}^{n} A_{ij} \log P_{ij}$$

$1 \leq i \leq m$

Derivative

$$L_i = -\sum_{j=1}^{n} A_{ij} \log P_{ij} \qquad \text{---------- (1)}$$

Here P is Predicted distribution

$$\frac{\partial L}{\partial P} \in R^{m \times n}$$

by (1)

$$\left(\frac{\partial L}{\partial P}\right)_{ij} = \frac{\partial L_i}{\partial P_{ij}} = -\frac{A_{ij}}{P_{ij}} \quad , 1 \le i \le m \quad , 1 \le j \le n$$

=) often for batch size one,

$$A = [0, \cdots, 1, \cdots 0]_{1 \times m}$$

$$P = [P_1, \cdots, P_k, \cdots P_m]_{1 \times m}$$

In this case

$$\left(\frac{\partial L}{\partial P}\right)_i = \frac{\partial L_i}{\partial P_i} = \begin{cases} \frac{-1}{P_i} & , \ i = k \\ 0 & , \ \text{else} \end{cases}$$

And Same, we can repeat for each element in batch.