Lesson 12

Learning with different Loss Functions and Their Derivatives

Two Commonly used Loss Functions are

- Mean Square Error – Standard Loss Function for Regression
- Cross Entropy Loss - Standard Loss Function for Classification

Cross Entropy Loss

Cross-entropy is a measure of the difference between two probability distributions for a given random variable or set of events. If p and q are two probability distributions drown from a random variable X, cross entropy is defined as

$$CE = -\sum_{x \epsilon X}^{n} p(x) \log q(x)$$

Cross Entropy Loss

Cross-entropy is a measure of the difference between two probability distributions for a given random variable or set of events. If p and q are two probability distributions drown from a random variable X, the distance of p from q i.e., cross entropy is defined as

$$CE = -\sum_{x \epsilon X}^{n} q(x) \log p(x)$$

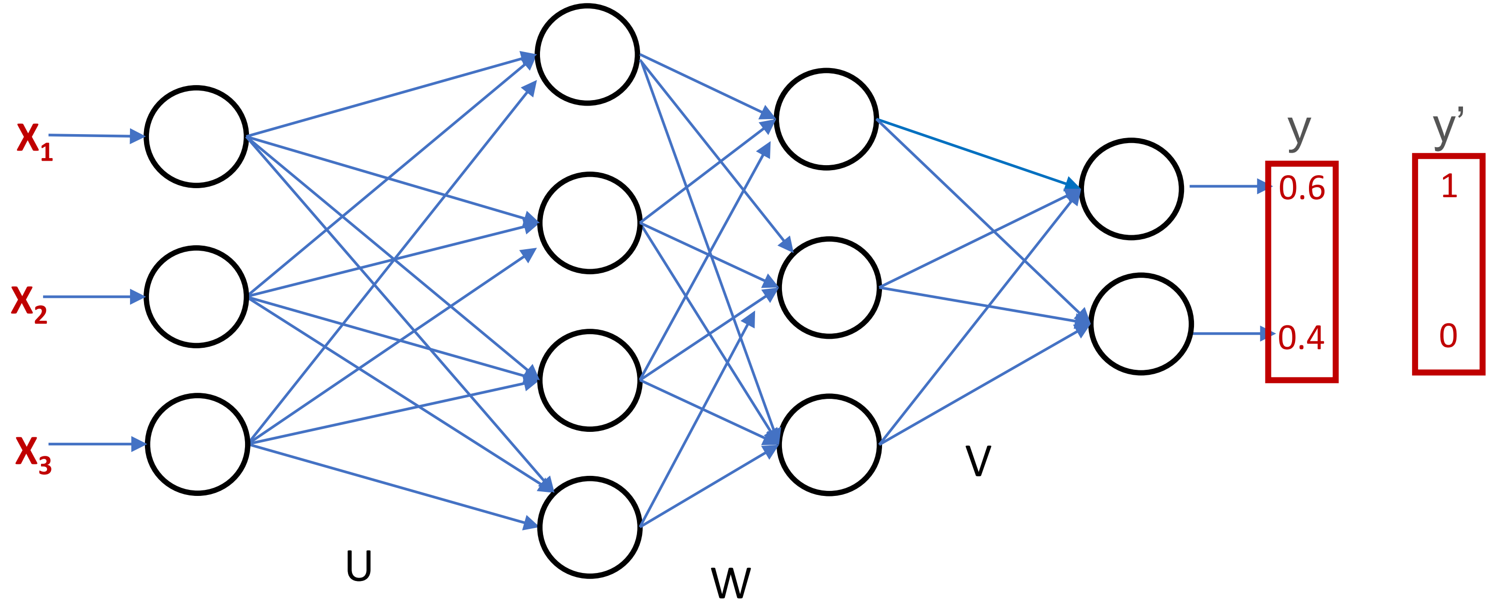
X = {H, T}, p = {0.9, 0.1} and q = {0.6,0.4}

How different p from q?

CE = - 0.6 log(0.9) - 0.4 log(0.1) = 1.42

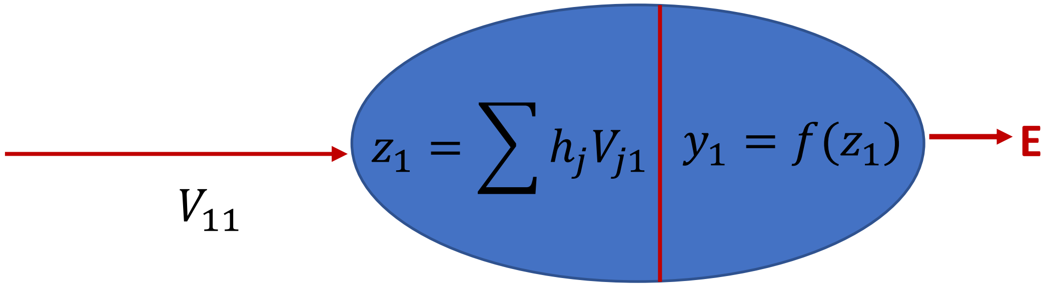
Cross Entropy Loss

$$CE = -\sum_{x \epsilon X}^{n} q(x) \log p(x)$$



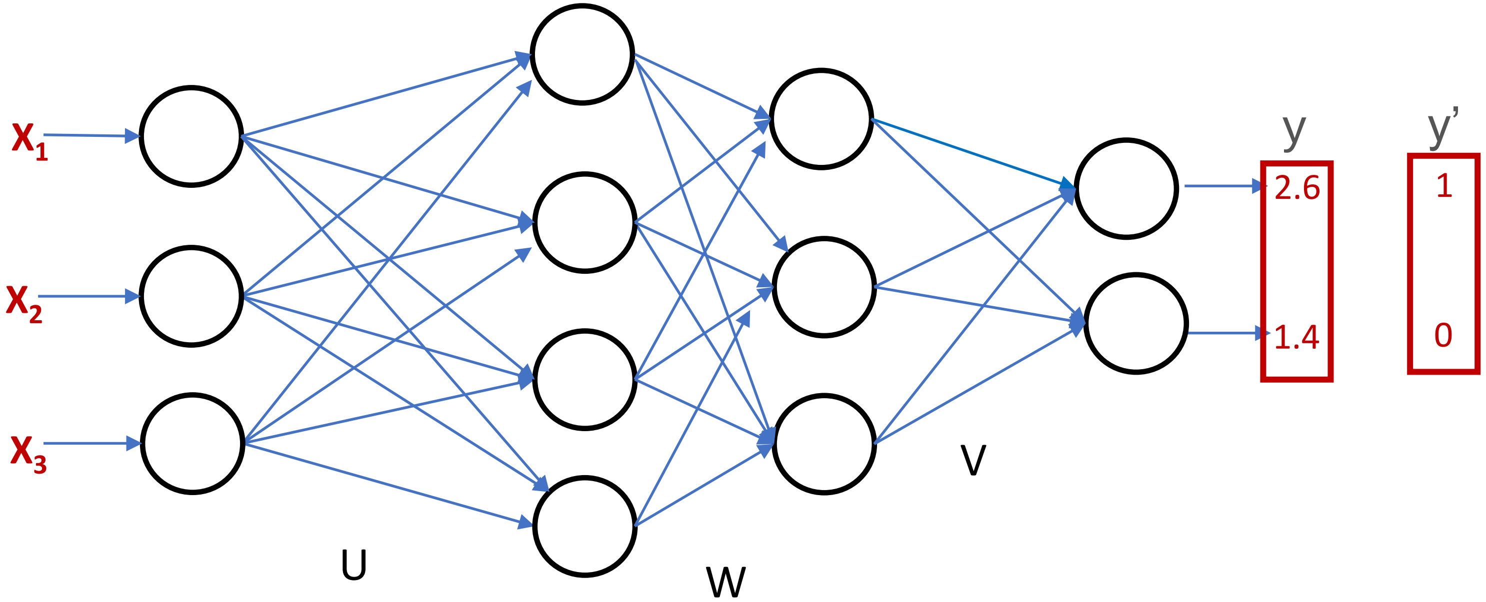
$$E = -\sum_{i}^{C} \acute{y}_i \log(y_i)$$

$$\frac{\delta E}{\delta y_i} = -\sum_{i}^{C} \frac{\acute{y}_i}{y_i}$$

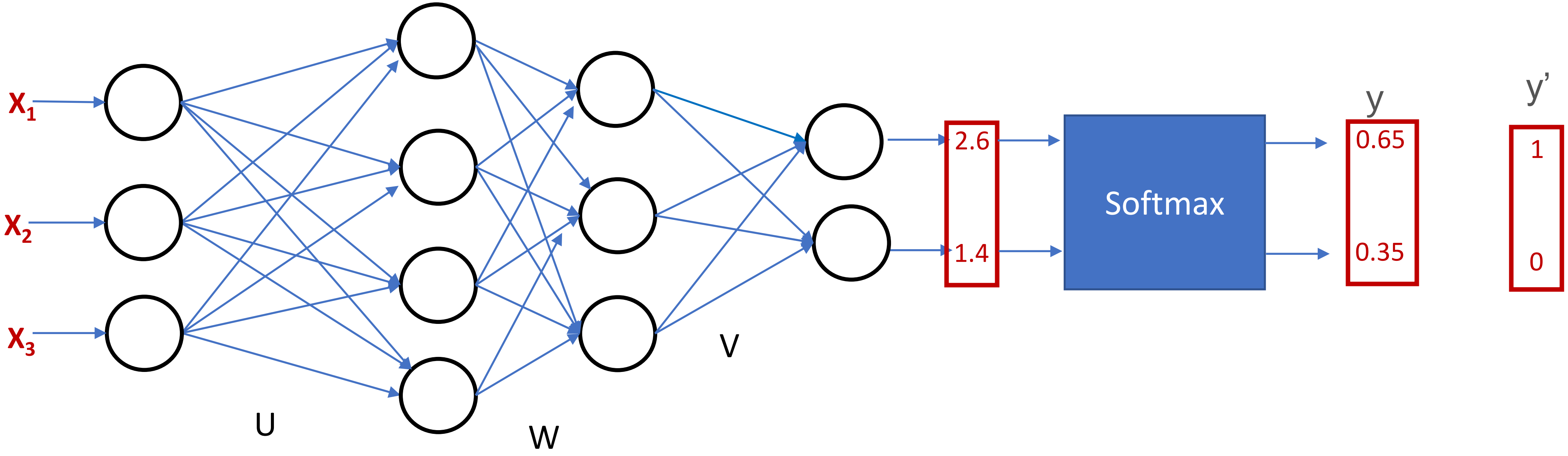


$$\frac{\delta E}{\delta V_{11}} = \frac{\delta z_1}{\delta V_{11}} \times \frac{\delta y_1}{\delta z_1} \times \frac{\delta E}{\delta y_1}$$

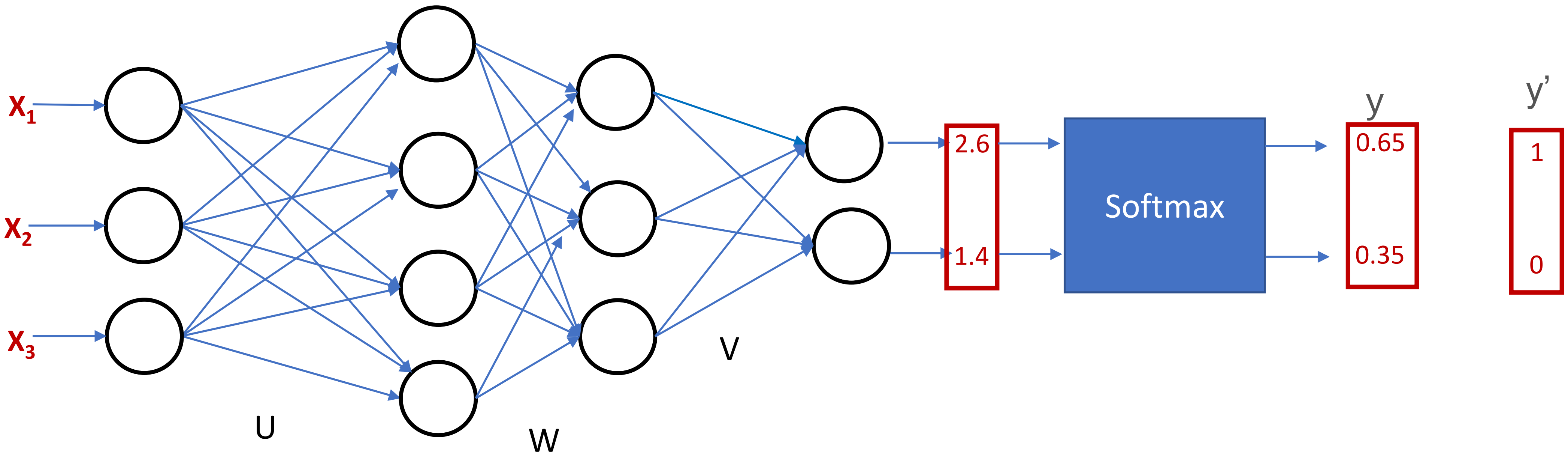
Cross Entropy Loss



Cross Entropy Loss



Backpropagation



Softmax

softmax: $\mathbb{R}^n \rightarrow \mathbb{R}^n$

z= $\{z_1, z_2, z_3, \ldots., z_n\}$

Z = {1.1, 2.2, 0.2, -1.7}

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

Z = {0.224, 0.672, 0.091, 0.013}

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

Z = {1.1, 2.2, 0.2, -1.7}

$$\left\{ \frac{e^{z_1}}{\sum_{j=1}^{n} e^{z_j}}, \quad \frac{e^{z_2}}{\sum_{j=1}^{n} e^{z_j}}, \quad \frac{e^{z_3}}{\sum_{j=1}^{n} e^{z_j}} \ \ldots, \quad \frac{e^{z_n}}{\sum_{j=1}^{n} e^{z_j}} \right\}$$

$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$

Z = {1.1, 2.2, 0.2, -1.7}

$\{ \quad \frac{e^{z_1}}{\sum_{j=1}^{n} e^{z_j}}, \quad \frac{e^{z_2}}{\sum_{j=1}^{n} e^{z_j}}, \quad \frac{e^{z_3}}{\sum_{j=1}^{n} e^{z_j}} \ldots, \quad \frac{e^{z_n}}{\sum_{j=1}^{n} e^{z_j}} \}$

If i = k

$$\frac{\delta \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}}{\delta z_k} = \frac{e^{z_i} \sum_{j=1}^{n} e^{z_j} - e^{z_k} e^{z_i}}{(\sum_{j=1}^{n} e^{z_j})^2} = \frac{e^{z_i}(\sum_{j=1}^{n} e^{z_j} - e^{z_k})}{(\sum_{j=1}^{n} e^{z_j})^2} = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \times \frac{\sum_{j=1}^{n} e^{z_j} - e^{z_k}}{\sum_{j=1}^{n} e^{z_j}} = p_i(1 - p_i)$$

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

Z = {1.1, 2.2, 0.2, -1.7}

$$\left\{ \frac{e^{z_1}}{\sum_{j=1}^{n} e^{z_j}}, \quad \frac{e^{z_2}}{\sum_{j=1}^{n} e^{z_j}}, \quad \frac{e^{z_3}}{\sum_{j=1}^{n} e^{z_j}} \ldots, \quad \frac{e^{z_n}}{\sum_{j=1}^{n} e^{z_j}} \right\}$$

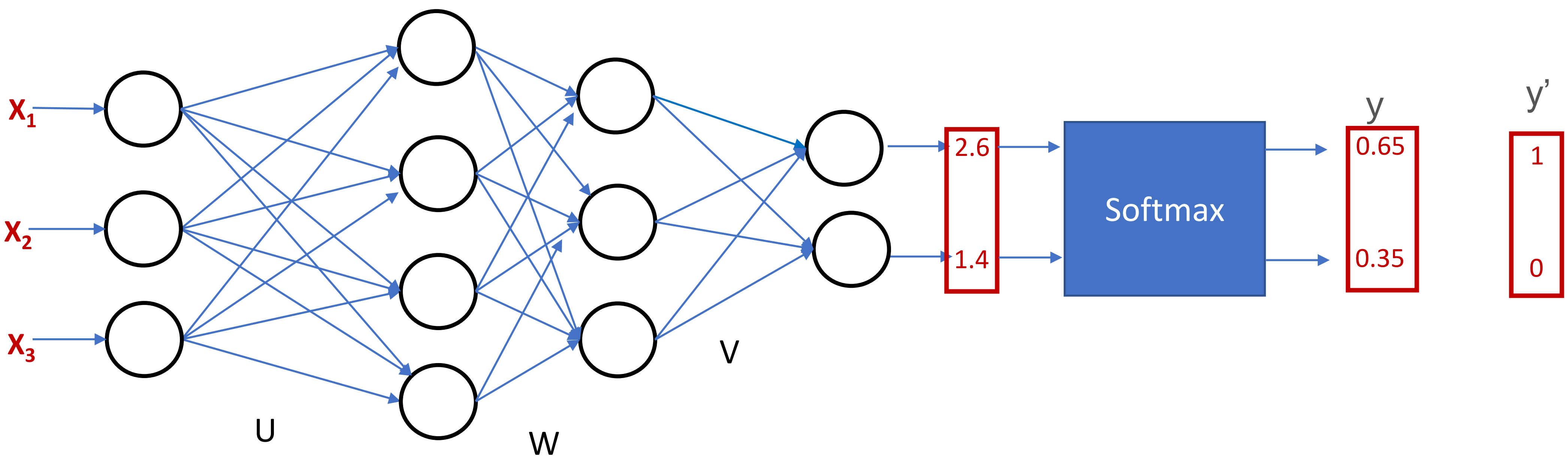
If i = k

$$\frac{\delta \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}}{\delta z_k} = \frac{e^{z_i} \sum_{j=1}^{n} e^{z_j} - e^{z_k} e^{z_i}}{(\sum_{j=1}^{n} e^{z_j})^2} = \frac{e^{z_i}(\sum_{j=1}^{n} e^{z_j} - e^{z_k})}{(\sum_{j=1}^{n} e^{z_j})^2} = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \times \frac{\sum_{j=1}^{n} e^{z_j} - e^{z_k}}{\sum_{j=1}^{n} e^{z_j}} = p_i(1 - p_i)$$

If i != k

$$\frac{\delta \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}}{\delta z_k} = \frac{0 - e^{z_k} e^{z_i}}{(\sum_{j=1}^{n} e^{z_j})^2} = \frac{-e^{z_k}}{\sum_{j=1}^{n} e^{z_j}} \times \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} = -p_i p_k$$

Backpropagation



Summary

- Cross Entropy Loss Function and Softmax and their gradients.