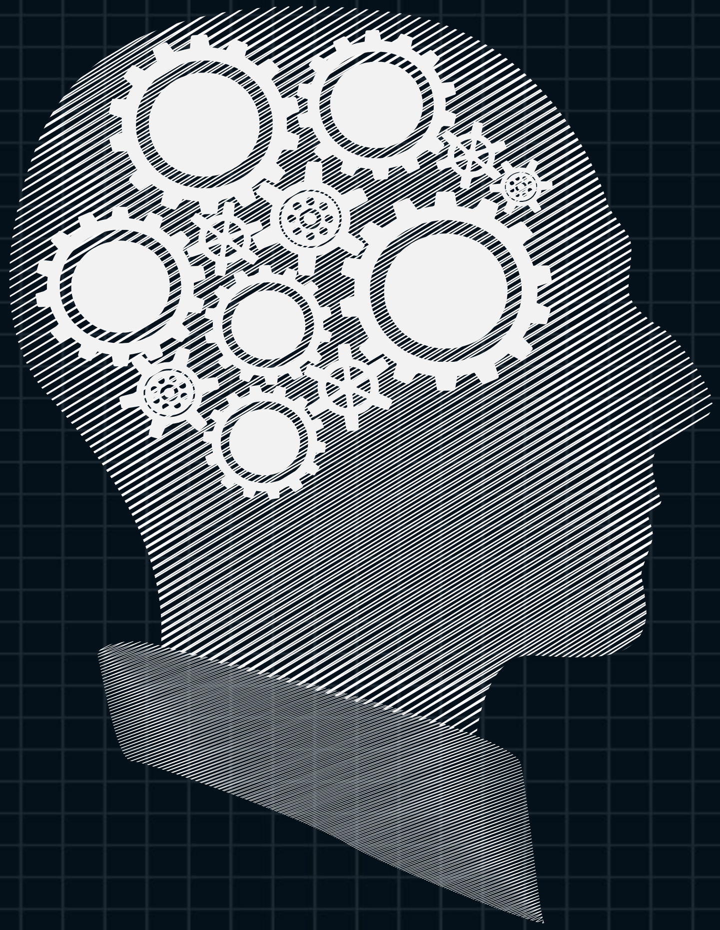


# Forward Propagation of Neural Networks

## Lecture.4
## Loss Functions

## - Cartesian Product for Predictions/Labels

$$\mathbb{R}$$

$$\mathbb{B} = \{0, 1\}$$

$$\mathbb{C} = \{c_1, c_2, \dots, c_K\}$$

$$\mathbb{P} = \{x \mid 0 \leq x \leq 1\}$$

$$\mathbb{B}_1^n = \left\{ (b_1, b_2, \dots, b_n)^T \;\middle|\; \forall b_i \in \mathbb{B}, \; \sum_{i=1}^{n} b_i = 1 \right\}$$

Lecture.4
Loss Functions

- Mean Squared Error

## Dataset for Regression

$$\left(\vec{x}^{(1)}\right) = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \end{pmatrix} \quad y^{(1)} \in \mathbb{R}$$

$$\left(\vec{x}^{(2)}\right) = \begin{pmatrix} x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \end{pmatrix} \quad y^{(2)} \in \mathbb{R}$$

$$\vdots$$

$$\left(\vec{x}^{(N)}\right) = \begin{pmatrix} x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix} \quad y^{(N)} \in \mathbb{R}$$

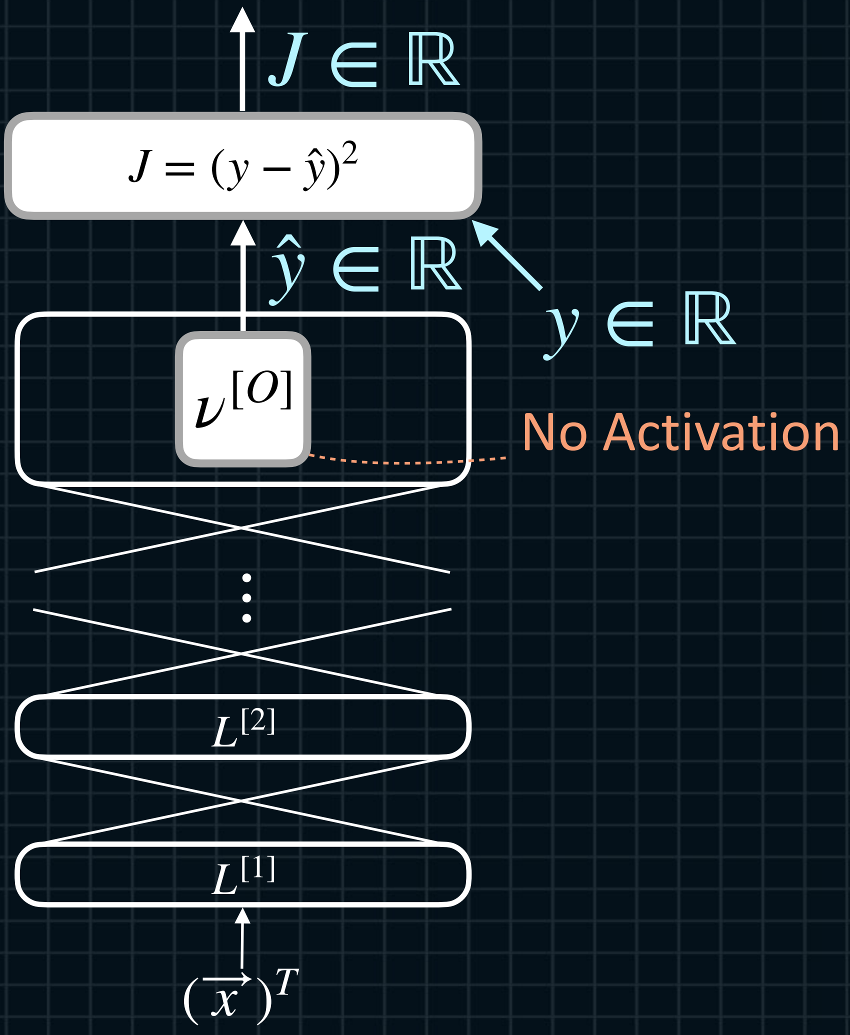
$$\longrightarrow$$

$$X^T = \begin{pmatrix} \longleftarrow & \vec{x}^{(1)} & \longrightarrow \\ \longleftarrow & \vec{x}^{(2)} & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{x}^{(N)} & \longrightarrow \end{pmatrix} \in \mathbb{R}^{N \times l_I}$$

$$Y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{pmatrix} \in \mathbb{R}^{N \times 1}$$

# - Mean Squared Error

Squared Error

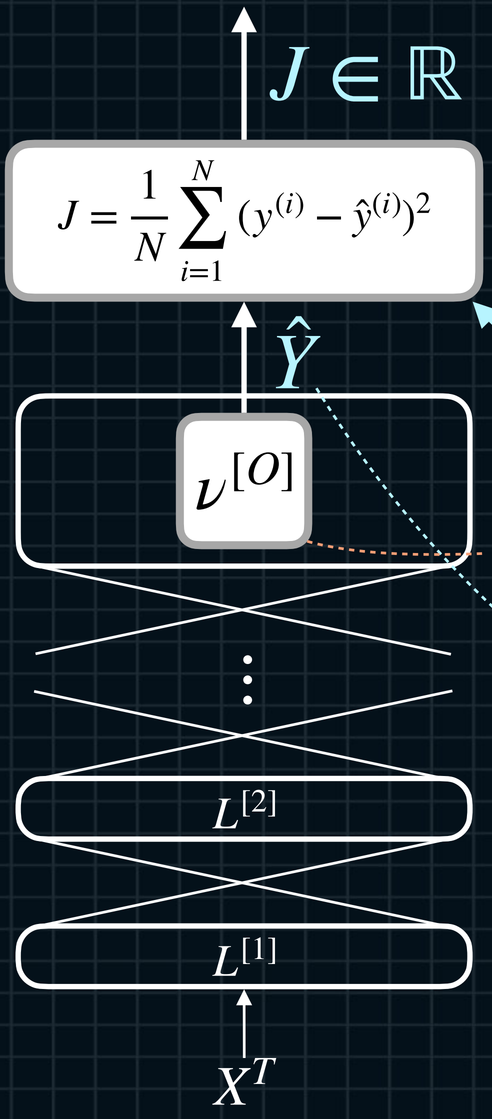



$$J = (y - \hat{y})^2$$

# Lecture.4
# Loss Functions

## - Mean Squared Error

$$J = \frac{1}{N}\sum_{i=1}^{N}(y^{(i)} - \hat{y}^{(i)})^2$$




$Y$

No Activation

$$Y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{pmatrix} \in \mathbb{R}^{N\times 1}$$

$$\hat{Y} = \begin{pmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(N)} \end{pmatrix} \in \mathbb{R}^{N\times 1}$$

- Binary Cross Entropy

# Dataset for Binary Classification

$$\left(\vec{x}^{(1)}\right) = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \end{pmatrix} \quad y^{(1)} \in \mathbb{B}$$

$$\left(\vec{x}^{(2)}\right) = \begin{pmatrix} x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \end{pmatrix} \quad y^{(2)} \in \mathbb{B}$$

$$\vdots$$

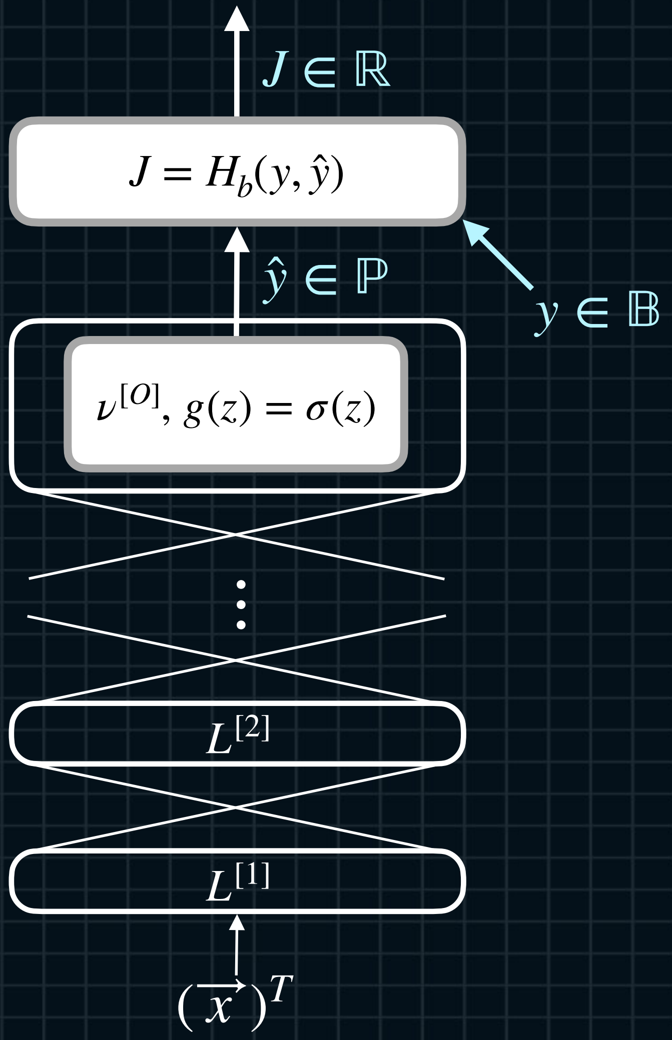
$$\left(\vec{x}^{(N)}\right) = \begin{pmatrix} x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix} \quad y^{(N)} \in \mathbb{B}$$

$$\longrightarrow$$

$$X^T = \begin{pmatrix} \longleftarrow & \vec{x}^{(1)} & \longrightarrow \\ \longleftarrow & \vec{x}^{(2)} & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{x}^{(N)} & \longrightarrow \end{pmatrix} \in \mathbb{R}^{N \times l_I}$$

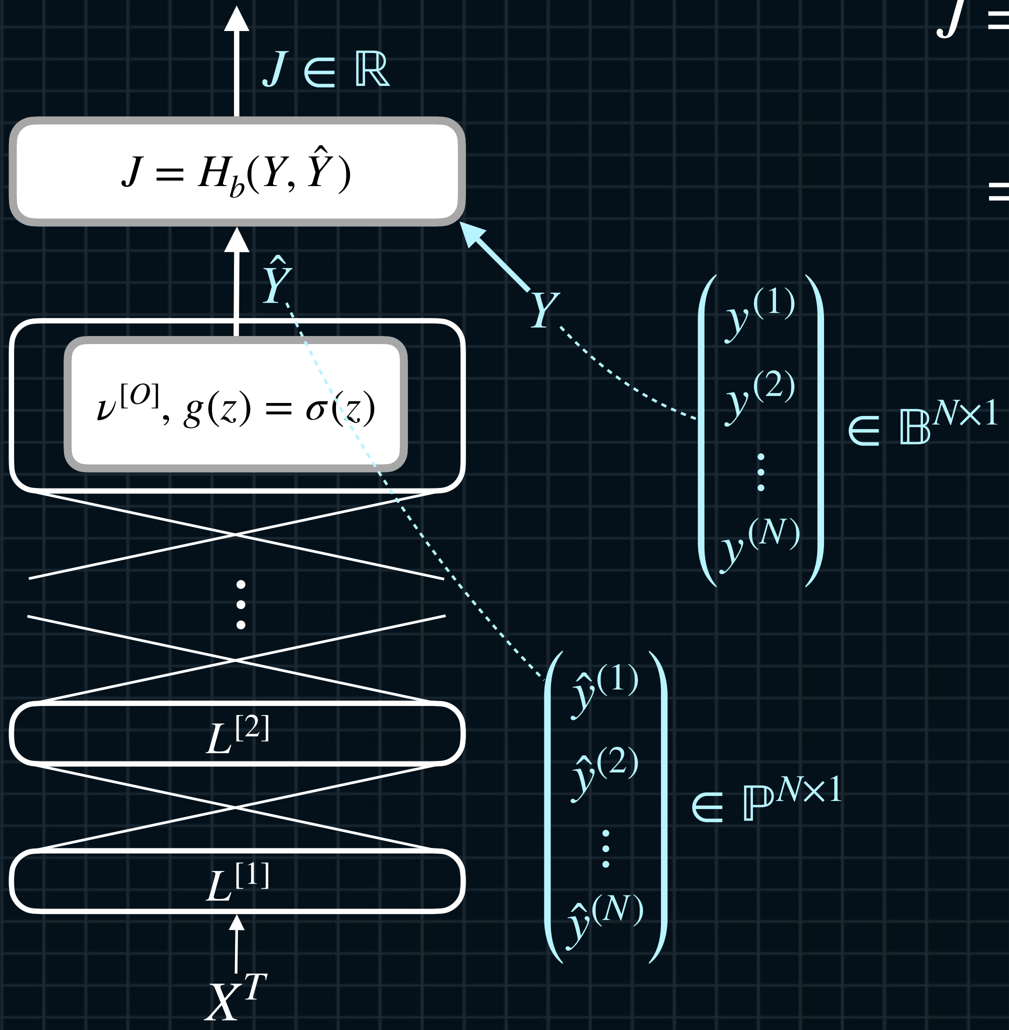
$$Y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{pmatrix} \in \mathbb{B}^{N \times 1}$$

## - Binary Cross Entropy




$$H_b(y, \hat{y}) = -\left[y log(\hat{y}) + (1 - y) log(1 - \hat{y})\right]$$

## - Binary Cross Entropy




$$J = H_b(Y, \hat{Y})$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\left[y^{(i)}log(\hat{y}^{(i)}) + (1 - y^{(i)})log(1 - \hat{y}^{(i)})\right]$$

## - Categorical Cross Entropy

# Dataset for Multi-class Classification

$$\mathbb{C} = \{c_1, c_2, \dots, c_K\}$$

$$\left(\vec{x}^{(1)}\right) = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \end{pmatrix} \quad y^{(1)} \in \mathbb{C}$$

$$\left(\vec{x}^{(2)}\right) = \begin{pmatrix} x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \end{pmatrix} \quad y^{(2)} \in \mathbb{C}$$

$$\vdots$$

$$\left(\vec{x}^{(N)}\right) = \begin{pmatrix} x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix} \quad y^{(N)} \in \mathbb{C}$$

$$\longrightarrow$$

$$X^T = \begin{pmatrix} \longleftarrow & \vec{x}^{(1)} & \longrightarrow \\ \longleftarrow & \vec{x}^{(2)} & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{x}^{(N)} & \longrightarrow \end{pmatrix} \in \mathbb{R}^{N \times l_I}$$

$$Y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{pmatrix} \in \mathbb{C}^{N \times 1}$$

# Dataset for Multi-class Classification + One-hot Encoding

$$\left(\vec{y}^{(i)}\right)^T \in \mathbb{B}_1^{1\times K}$$

$$\left(\vec{y}^{(1)}\right)^T = (1 \quad 0 \quad \dots \quad 0),\ \alpha = 1$$
$$\left(\vec{y}^{(2)}\right)^T = (0 \quad 1 \quad \dots \quad 0),\ \alpha = 2$$
$$\vdots$$
$$\left(\vec{y}^{(K)}\right)^T = (0 \quad 0 \quad \dots \quad 1),\ \alpha = K$$

$$\left(\vec{y}^{(1)}\right)^T \in \mathbb{B}_1^{1\times K}$$
$$\left(\vec{y}^{(2)}\right)^T \in \mathbb{B}_1^{1\times K}$$
$$\vdots$$
$$\left(\vec{y}^{(N)}\right)^T \in \mathbb{B}_1^{1\times K}$$

$$\longrightarrow \quad Y^T = \begin{pmatrix} \longleftarrow & \left(\vec{y}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{y}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{y}^{(N)}\right)^T & \longrightarrow \end{pmatrix} \in \mathbb{B}_1^{N\times K}$$

# Lecture.4
# Loss Functions

## - Categorical Cross Entropy

$$H(\overrightarrow{y}, \overrightarrow{\hat{y}}) = -\sum_{i=1}^{K} y_i log(\hat{y}_i)$$

$J$

$$J = H(\overrightarrow{y}, \overrightarrow{\hat{y}})$$

$$(\overrightarrow{y})^T = (y_1 \quad y_2 \quad \dots \quad y_K)$$

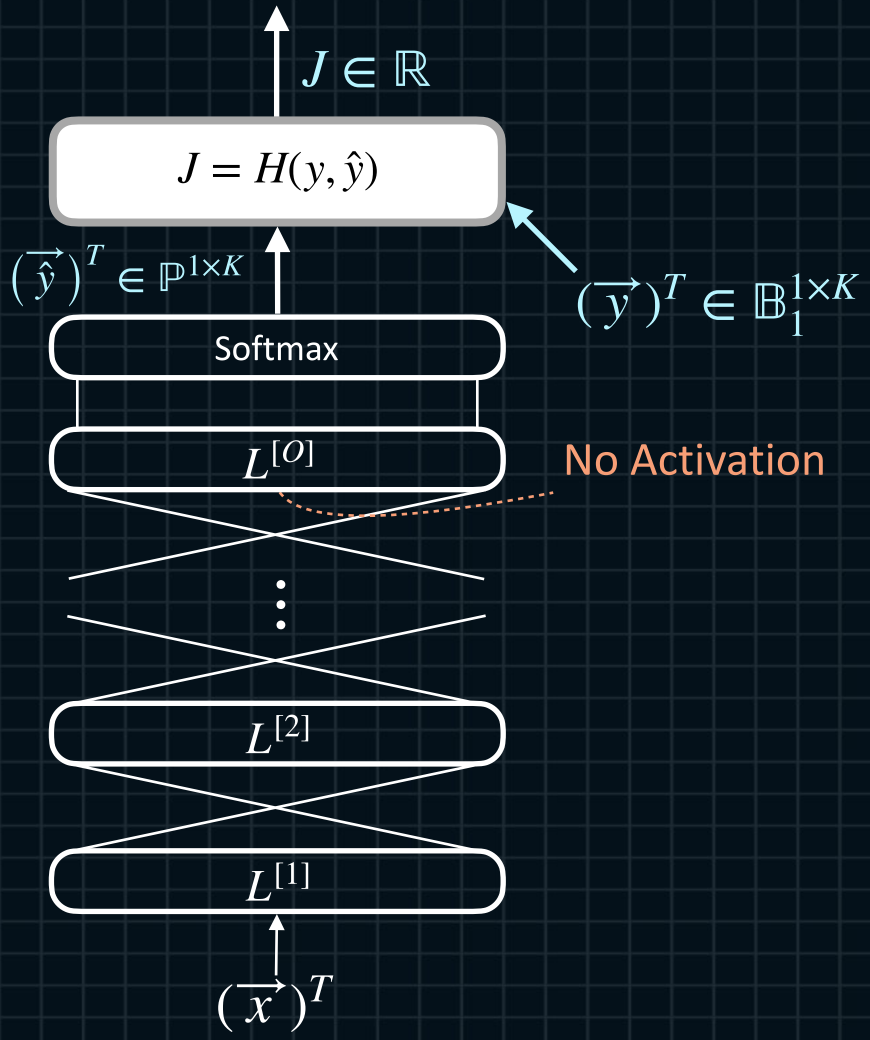
$$y_i \in \{0, 1\}$$

$$(\overrightarrow{\hat{y}})^T = \left(\hat{y}_1 \quad \hat{y}_2 \quad \dots \quad \hat{y}_K\right)$$

$$0 \leq \hat{y}_i \leq 1$$

Lecture.4
Loss Functions

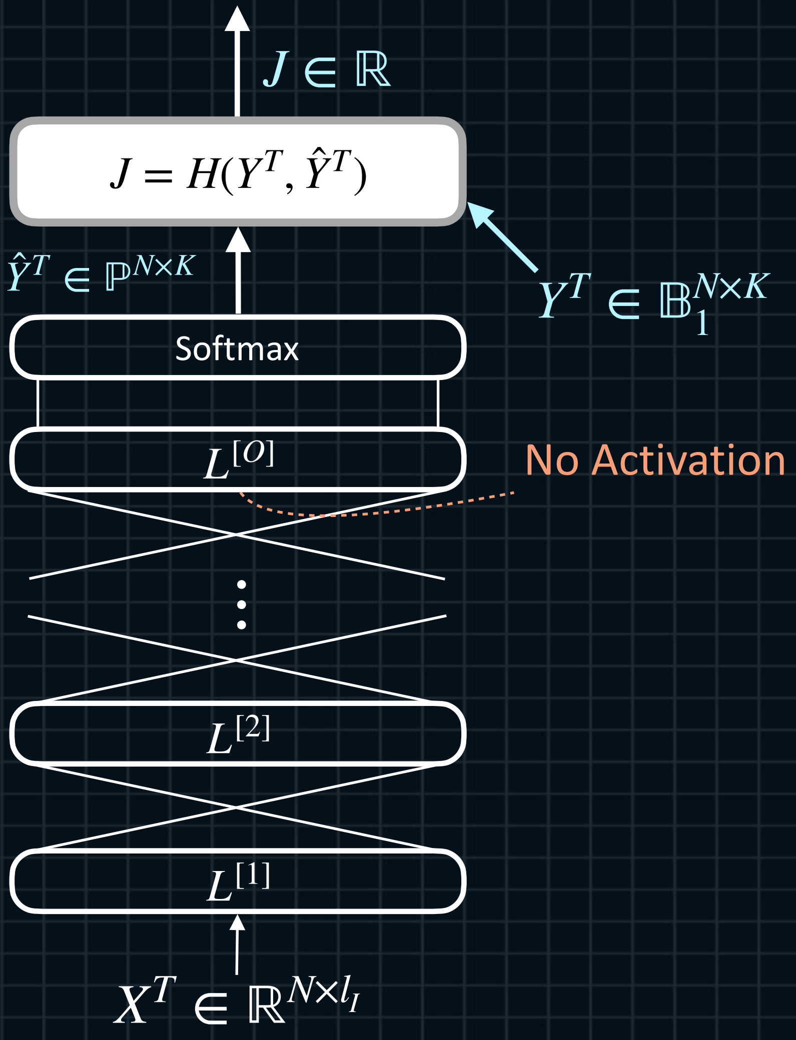
## - Categorical Cross Entropy




$$H(\vec{y}, \vec{\hat{y}}) = -\sum_{i=1}^{K} y_i log(\hat{y}_i)$$

# Lecture.4
# Loss Functions

## - Categorical Cross Entropy

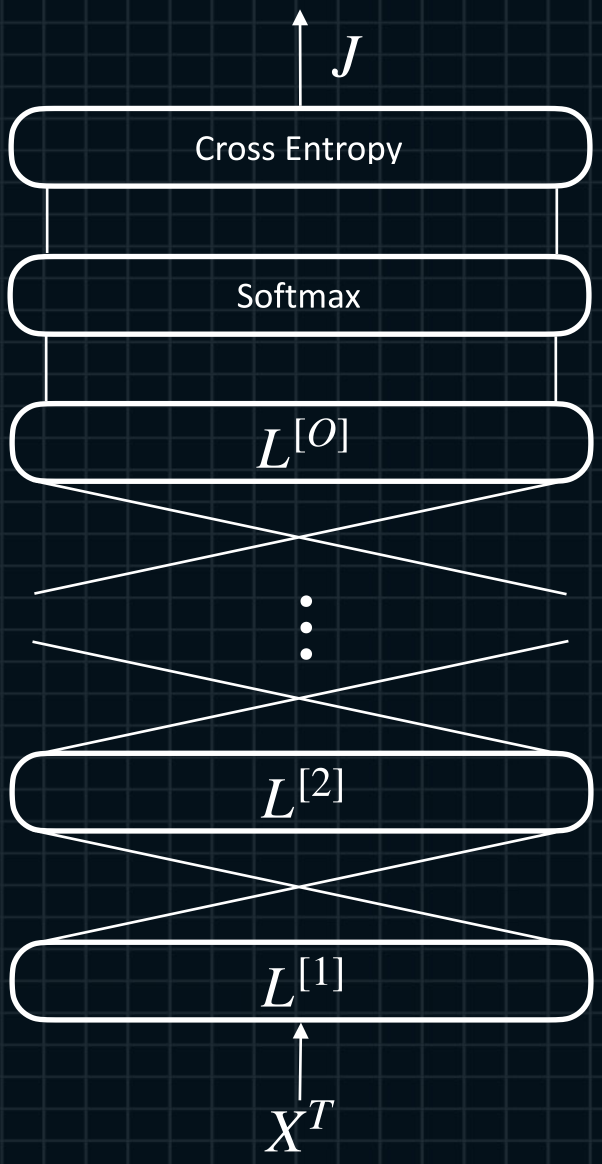



$$H(Y^T, \hat{Y}^T) = \frac{1}{N}\sum_{i=1}^{N} H(\overrightarrow{y}^{(i)}, \overrightarrow{\hat{y}}^{(i)})$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} y_j^{(i)} log(\hat{y}_j^{(i)})$$

# Lecture.4
# Loss Functions

## - from Feature to Loss



# Forward Propagation
# of
# Neural Networks

## Lecture.4
## Loss Functions