2025-07-24 - Compare and Contrast among activation functions, loss functions, optimizers and regularization for choosing the appropriate method for the given application.

24 July 2025 09:19

$$\hat{y} = W \times X_i + b$$

$$f(x) = \frac{1}{1+e^{-x}}$$

$\hat{y}$

Sigmoid (Logit function)

=> Sigmoid converts linear to non-linear.

=> Logarithmic Loss : BCE

$$L_{BCE}(y_i, \hat{y}_i) = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1-y_i)\cdot\log(1-\hat{y}_i)$$

$$\hat{y}_i \approx \sigma(z) = \frac{1}{1+e^{-z}}$$

$$= \frac{1}{1+e^{-(W \times x_i + b)}}$$

$$\partial/\partial L = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(\hat{y}_i) + (1-y_i)\cdot\log(1-\hat{y}_i)$$

$$= -\frac{1}{N}\sum_{i=1}^{N} y_i \, \partial/\partial L \log(\hat{y}_i) + (1-y_i)\cdot \partial/\partial L \log(1-\hat{y}_i)$$

$$= -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \frac{1}{\hat{y}_i} \, {}^{\partial}/_{\partial L}(\hat{y}_i) + (1-y_i) \cdot \frac{1}{(1-\hat{y}_i)} \cdot {}^{\partial}/_{\partial L}(1-\hat{y}_i)$$

$$= -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \frac{1}{\hat{y}_i} \, {}^{\partial}/_{\partial L}\,\sigma(z) + (1-y_i) \cdot \frac{1}{(1-\hat{y}_i)} \cdot {}^{\partial}/_{\partial L}(1-\sigma(z))$$

$$= -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \frac{1}{\hat{y}_i} \cdot \sigma(z) \cdot (1-\sigma(z)) \cdot {}^{\partial}/_{\partial L}(z) - (1-y_i) \cdot$$

$$\frac{1}{1-\hat{y}_i} \cdot \sigma(z) \cdot (1-\sigma(z)) \, {}^{\partial}/_{\partial L}(z)$$

$$= -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \frac{1}{\hat{y}_i} \cdot (1-\hat{y}_i) \cdot {}^{\partial}/_{\partial L}(z) - (1-y_i) \cdot \frac{1}{1-\hat{y}_i} \cdot \sigma(\hat{y}_i)$$

$$(1-\hat{y}_i) \cdot {}^{\partial}/_{\partial L}(z)$$

$$= -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot (1-\hat{y}_i) \cdot z - (1-y_i) \cdot \hat{y}_i (z)$$

$$= -\frac{1}{N}\sum_{i=1}^{N} (y_i - y_i \cdot \hat{y}_i - \hat{y}_i + y_i \cdot \hat{y}_i) \cdot z$$

$$= -\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i) \cdot z \quad \rightarrow \text{BCE} \approx \text{Log loss}$$

$$\Rightarrow \quad \partial/\partial z \, \sigma(z) = \partial/\partial z \left( \frac{1}{1+e^{-z}} \right)$$

$$= \frac{-(1+e^{-z})}{(1+e^{-z})^2}$$

$$= \frac{e^{-z}}{(1+e^{-z})^2}$$

$$= \left( \frac{1}{1+e^{-z}} \right) \cdot \left( \frac{e^{-z}}{1+e^{-z}} \right)$$

$$= \left( \frac{1}{1+e^{-z}} \right) \cdot \left( \frac{1+e^{-z}}{1+e^{-z}} - \frac{1}{1+e^{-z}} \right)$$

$$= \sigma(z) \cdot \left( \frac{1+e^{-z}}{1+e^{-z}} - \sigma(z) \right)$$

$$= \underline{\sigma(z)(1-\sigma(z))}$$

$$\Rightarrow \quad W = W - \left( \alpha \times \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i) \cdot z \right)$$

$$b = b - \left(\alpha \times \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)\right)$$

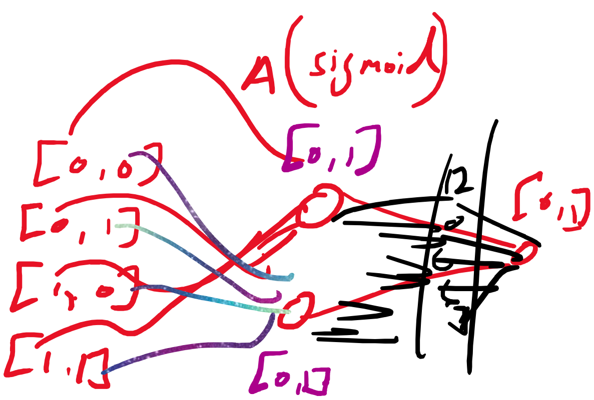
---

XOR

$$\Rightarrow \quad X = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \Rightarrow y$$

X → Train Data

y → Train Label




=> Overfitting : Train Acc ↑ Test Acc ↓

<u>Sol</u> : Train a deeper N/W .

=> Underfitting : Train Acc ↓ Test Acc ↑

=> Underfitting : Train Acc ↓ Test Acc ↑

<u>Sol</u>: Use more data & regulate the epochs.