Fuzzy Logic & Neural Networks (CS-514)

Dr. Sudeep Sharma

IIIT Surat

sudeep.sharma@iiitsurat.ac.in

Loss Function

Categorical Cross-Entropy Loss

- It is used for multi-class classification tasks where the labels are one-hot encoded.
- The Loss function measures the difference

$$\mathrm{L}oss = -\sum_{i=1}^{K} y_i \log(S_i)$$

where K is the number of classes, y_i is the true label (1 for the correct class, 0 otherwise), and S_i is the predicted probability for class i.

- This loss value indicates how well the model predicted the correct class. The lower the loss, the better the prediction.

Loss Function

Categorical Cross-Entropy Loss

- **Example Case:** Consider a 3-class classification problem where the true label is class 3 (one-hot encoded as [0,0,1]), and the model predicts probabilities [0.1,0.3,0.6].

$$y = [0,\ 0,\ 1]$$

$$S = [0.1,\ 0.3,\ 0.6]$$

$$\mathrm{L}oss = -\sum_{i=1}^{3} y_i \log(S_i)$$

$$= -y_1 \log(S_1) - y_2 \log(S_2) - y_3 \log(S_3) = 0.511$$

Loss Function Derivative

Categorical Cross-Entropy Loss Derivative

$$\mathrm{L}oss = -\sum_{i=1}^{K} y_i \log(S_i)$$

$$\frac{\partial}{\partial S_i}\mathrm{L}oss = \frac{\partial}{\partial S_i}\left(-\sum_{i=1}^{K} y_i \log(S_i)\right)$$

$$\frac{\partial}{\partial S_i}\mathrm{L}oss = -\frac{y_i}{S_i}$$

Loss Function Derivative

Categorical Cross-Entropy Loss

$$\mathrm{L}oss = -\sum_{i=1}^{K} y_i \log(S_i)$$

$$\frac{\partial \mathrm{L}oss}{\partial z_m} = \sum_{i=1}^{K} \frac{\partial \mathrm{L}oss}{\partial S_i} \frac{\partial S_i}{\partial z_m}$$

The summation is used to solve the Jacobian matrix terms involved in the above product.

$$\frac{\partial \mathrm{L}oss}{\partial z_i} = S_i - y_i \ ; \ when \ i = m$$

$$\frac{\partial \mathrm{L}oss}{\partial z_m} = y_i S_m \ ; \ when \ i \neq m$$

Loss Function Derivative

Categorical Cross-Entropy Loss Derivative

- Example case to understand it properly: Let

$$S = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix}$$

$$\frac{\partial S}{\partial z} = \begin{bmatrix} \frac{\partial S_1}{\partial z_1} & \frac{\partial S_1}{\partial z_2} & \frac{\partial S_1}{\partial z_3} \\ \frac{\partial S_2}{\partial z_1} & \frac{\partial S_2}{\partial z_2} & \frac{\partial S_2}{\partial z_3} \\ \frac{\partial S_3}{\partial z_1} & \frac{\partial S_3}{\partial z_2} & \frac{\partial S_3}{\partial z_3} \end{bmatrix}$$

Loss Function Derivative

Categorical Cross-Entropy Loss Derivative

$$\frac{\partial \mathrm{L}oss}{\partial z_m} = \sum_{i=1}^{K} \frac{\partial \mathrm{L}oss}{\partial S_i} \frac{\partial S_i}{\partial z_m}$$

$$S = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix}$$

$$\frac{\partial S}{\partial z_1} = \begin{bmatrix} \frac{\partial S_1}{\partial z_1} \\ \frac{\partial S_2}{\partial z_1} \\ \frac{\partial S_3}{\partial z_1} \end{bmatrix} \qquad \frac{\partial Loss}{\partial S} = \begin{bmatrix} \frac{\partial Loss}{\partial S_1} \\ \frac{\partial Loss}{\partial S_2} \\ \frac{\partial Loss}{\partial S_3} \end{bmatrix}$$

Loss Function Derivative

Categorical Cross-Entropy Loss Derivative

$$\frac{\partial \mathrm{L}oss}{\partial z_1} = \sum_{i=1}^{K} \frac{\partial \mathrm{L}oss}{\partial S_i} \frac{\partial S_i}{\partial z_1} =$$

$$\begin{bmatrix} \frac{\partial Loss}{\partial S_1} & \frac{\partial Loss}{\partial S_2} & \frac{\partial Loss}{\partial S_3} \end{bmatrix} \begin{bmatrix} \frac{\partial S_1}{\partial z_1} \\ \frac{\partial S_2}{\partial z_1} \\ \frac{\partial S_3}{\partial z_1} \end{bmatrix}$$

Loss Function Derivative

Categorical Cross-Entropy Loss Derivative

$$\frac{\partial \mathrm{L}oss}{\partial z_1} = \sum_{i=1}^{K} \frac{\partial \mathrm{L}oss}{\partial S_i} \frac{\partial S_i}{\partial z_1} =$$

$$\begin{bmatrix} \frac{\partial Loss}{\partial S_1} & \frac{\partial Loss}{\partial S_2} & \frac{\partial Loss}{\partial S_3} \end{bmatrix} \begin{bmatrix} \frac{\partial S_1}{\partial z_1} \\ \frac{\partial S_2}{\partial z_1} \\ \frac{\partial S_3}{\partial z_1} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{y_1}{S_1} & -\frac{y_2}{S_2} & -\frac{y_3}{S_3} \end{bmatrix} \begin{bmatrix} S_1(1-S_1) \\ -S_2 S_1 \\ -S_3 S_1 \end{bmatrix}$$

Loss Function Derivative

Categorical Cross-Entropy Loss Derivative

- **Example Case:** Consider a 3-class classification problem where the true label is class 3 (one-hot encoded as [0,0,1]), and the model predicts probabilities [S_1, S_2, S_3].

$$\frac{\partial Loss}{\partial z_1} = \begin{bmatrix} -\frac{y_1}{S_1} & -\frac{y_2}{S_2} & -\frac{y_3}{S_3} \end{bmatrix} \begin{bmatrix} S_1(1-S_1) \\ -S_2 S_1 \\ -S_3 S_1 \end{bmatrix}$$

$$= y_1(S_1 - 1) + y_2 S_1 + y_3 S_1$$

$$as\ y_1 = 0,\ y_2 = 0,\ y_3 = 1$$

$$\frac{\partial Loss}{\partial z_1} = S_1 = S_1 - y_1$$

Loss Function Derivative

Categorical Cross-Entropy Loss Derivative

➢ **Example Case:** Consider a 3-class classification problem where the true label is class 3 (one-hot encoded as [0,0,1]), and the model predicts probabilities [S_1, S_2, S_3].

$$\frac{\partial Loss}{\partial z_2} = \begin{bmatrix} -\frac{y_1}{S_1} & -\frac{y_2}{S_2} & -\frac{y_3}{S_3} \end{bmatrix} \begin{bmatrix} -S_1 S_2 \\ S_2(1-S_2) \\ -S_3 S_2 \end{bmatrix}$$

$$= y_1 S_2 + y_2(S_2 - 1) + y_3 S_2$$

$$as \;\; y_1 = 0, \;\; y_2 = 0, \;\; y_3 = 1$$

$$\frac{\partial Loss}{\partial z_2} = S_2 = S_2 - y_2$$

Loss Function Derivative

Categorical Cross-Entropy Loss Derivative

- **Example Case:** Consider a 3-class classification problem where the true label is class 3 (one-hot encoded as [0,0,1]), and the model predicts probabilities [S_1, S_2, S_3].

$$\frac{\partial Loss}{\partial z_3} = \begin{bmatrix} -\frac{y_1}{S_1} & -\frac{y_2}{S_2} & -\frac{y_3}{S_3} \end{bmatrix} \begin{bmatrix} -S_1 S_3 \\ -S_2 S_3 \\ S_3(1-S_3) \end{bmatrix}$$

$$= y_1 S_3 + y_2 S_3 + y_3 (S_3 - 1)$$

$$as\ \ y_1 = 0,\ \ y_2 = 0,\ \ y_3 = 1$$

$$\frac{\partial Loss}{\partial z_3} = S_3 - 1 = S_3 - y_3$$

Loss Function Derivative

Categorical Cross-Entropy Loss Derivative

$$\frac{\partial \mathrm{L}oss}{\partial z_i} = S_i - y_i \ ; \ when \ i = m$$

$$\frac{\partial \mathrm{L}oss}{\partial z_m} = y_i S_m \ ; \ when \ i \neq m$$

$$the \ \ \mathrm{sec}ond \ term \ y_i S_m = 0 \ ; \ when \ i \neq m$$

$$therefore \ \frac{\partial \mathrm{L}oss}{\partial z_i} = S_i - y_i$$

Loss Function Derivative

Categorical Cross-Entropy Loss Derivative

$$\frac{\partial \mathrm{L}oss}{\partial z_i} = S_i - y_i \;;\; when\ i = m$$

$$\frac{\partial \mathrm{L}oss}{\partial z_m} = y_i S_m \;;\; when\ i \neq m$$

$$the\ \ \mathrm{sec}ond\ term\ y_i S_m = 0 \;;\; when\ i \neq m$$

$$therefore\ \frac{\partial \mathrm{L}oss}{\partial z_i} = S_i - y_i$$

Backpropagation

- The backpropagation algorithm involves 4 steps:

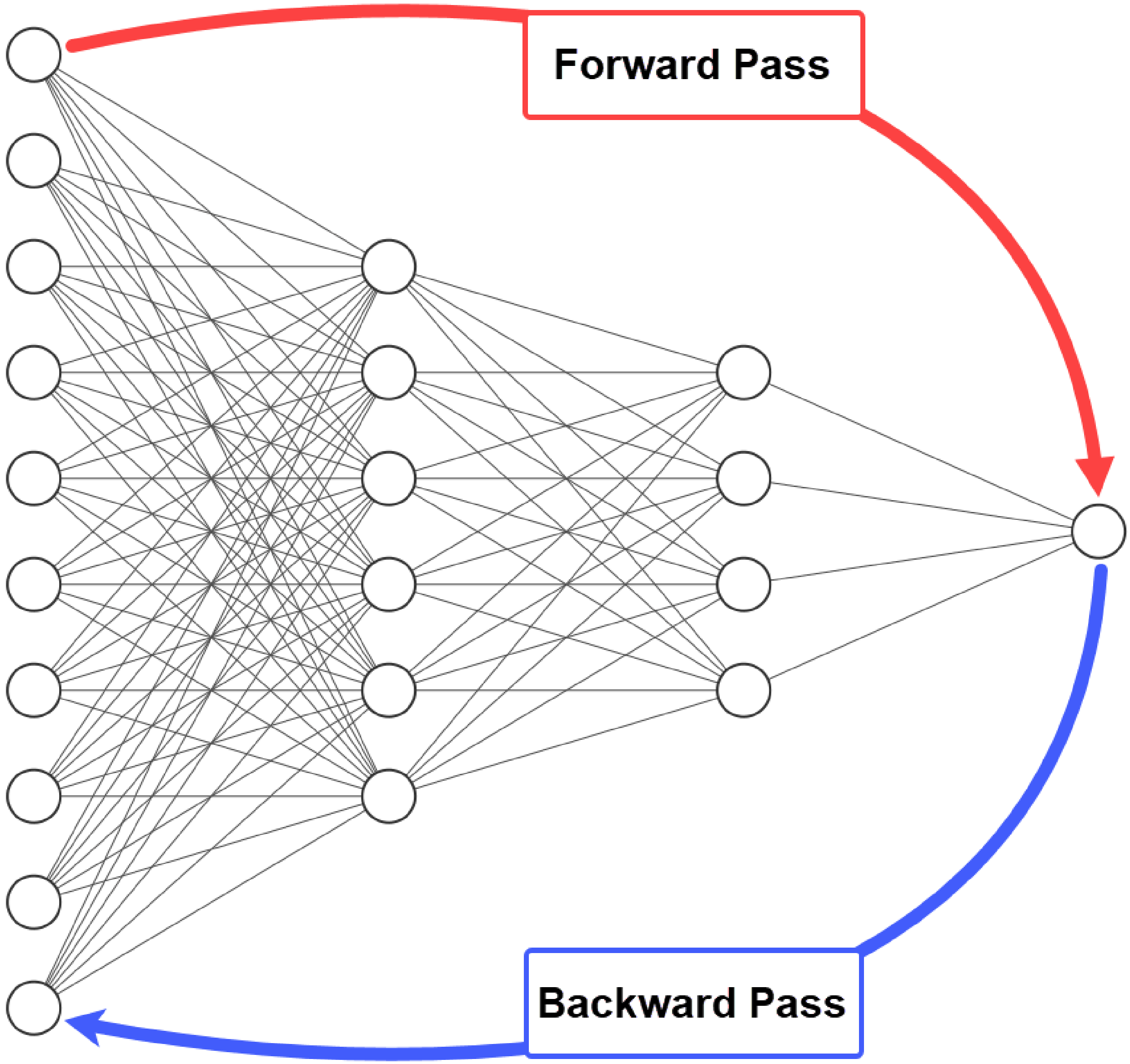
1. **Forward Pass:** Calculate output for the applied input using current weights and biases.
2. **Loss Computation:** Measure error between predicted and target values to define Loss Function.
3. **Backward Pass:** Calculate gradients of weights and biases.
4. **Weight Update:** Adjust weights and biases to reduce future errors.

Backpropagation

- The advantages of the backpropagation algorithm:

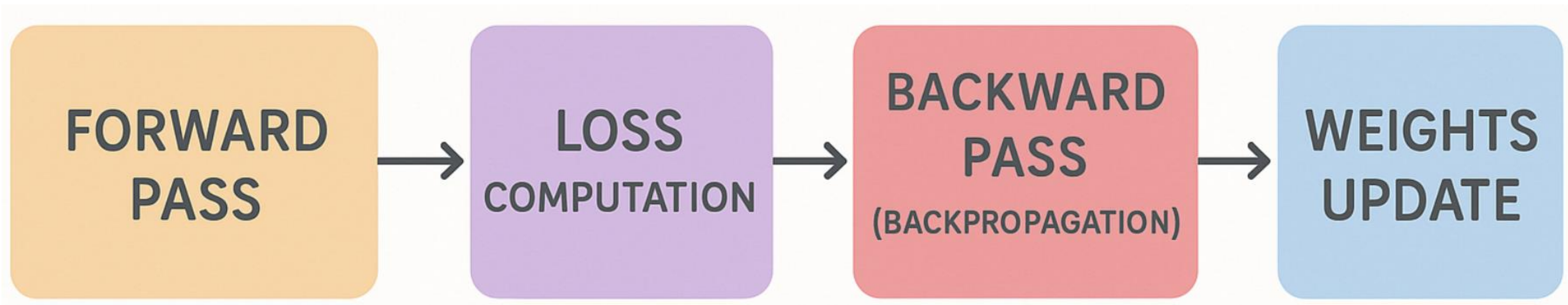
- It’s memory-efficient in calculating the derivatives.
- The backpropagation algorithm is fast, especially for small and medium-sized networks.
- This algorithm is good enough to work with different network architectures.
- There are no parameters to tune the backpropagation algorithm.

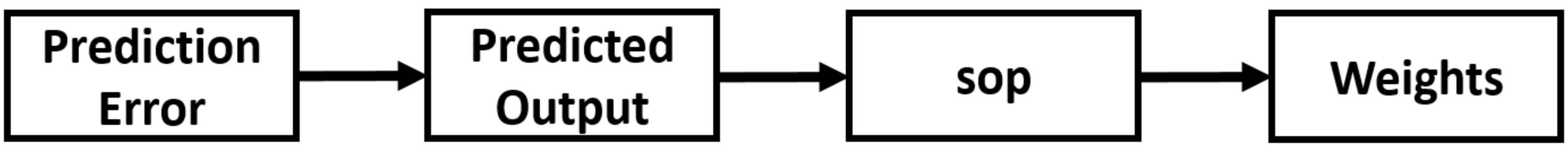
Backpropagation



Source: Internet

Backpropagation





$E = \frac{1}{2}(desired - predicted)^2$

$$\frac{\partial E}{\partial Predicted}$$

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

$$\frac{\partial Predicted}{\partial s}$$

$s = X_1 * W_1 + X_2 * W_2 + b$

$$\frac{\partial s}{\partial W_1}$$

W_1, W_2

$$\frac{\partial s}{\partial W_2}$$

Source: Internet