Bidirectional LSTM (BiLSTM)

Overview:

- BiLSTM = Bidirectional Long Short-Term Memory

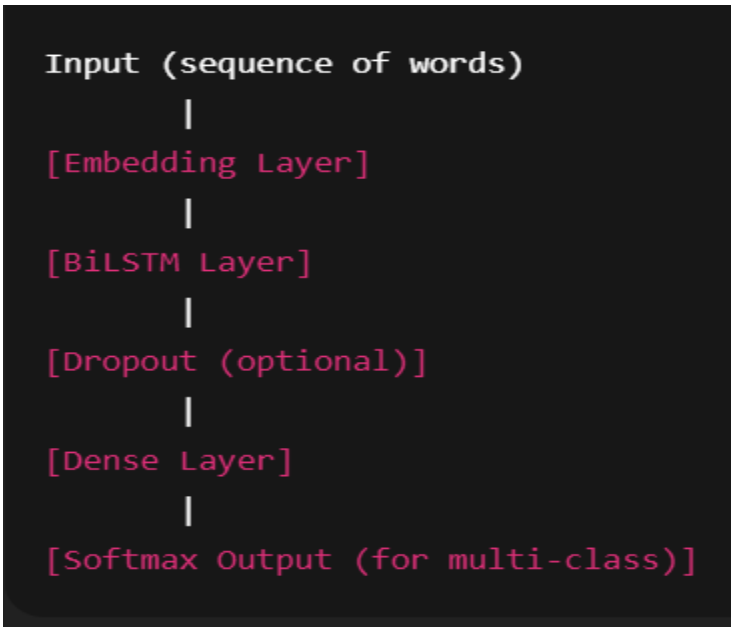
It is a type of RNN (Recurrent Neural Network) designed to learn long-term dependencies in sequence data (like text). BiLSTM processes the sequence from both directions:

- Forward (left-right)
- Backward (right-left)

Bidirectional Long Short-Term Memory (BiLSTM) is an advanced Recurrent Neural Network (RNN) architecture that processes sequential data in both forward and backward directions. It is widely used in natural language processing (NLP) tasks, time series prediction, and speech recognition.

Key Advantage: BiLSTM captures past (previous context) and future (next context) simultaneously, improving sequence understanding over standard unidirectional LSTMs.

Architecture:



- **Input Layer** → Sequence of tokens (words).
- **Embedding Layer** → Converts words into dense vectors.
- **Forward LSTM** → Reads the sequence left to right.
- **Backward LSTM** → Reads the sequence right to left.
- **Concatenation** → Combines both hidden states (forward + backward).
- **Dense Layer** → Fully connected layer for classification.
- **Output Layer** → Prediction (Softmax for multi-class, Sigmoid for binary)

1. **LSTM Unit:**

   Handles sequential data using gates:

   - Forget gate: Decides what information to discard
   - Input gate: Decides what new information to store
   - Output gate: Decides what to output

2. **Bidirectional Structure:**

   Two LSTM layers:

   - Forward LSTM: Processes sequence from t=0 to t=T
   - Backward LSTM: Processes sequence from t=T to t=0

   Outputs are concatenated (or summed) to produce final sequence representation.

Diagram:

Embedding Layer (for NLP tasks)

- Purpose: Converts discrete tokens (word indices) into dense, continuous vector representations.
- Function: Captures semantic meaning of words, where similar words have similar vector representations.
- Notes: Pre-trained embeddings like GloVe or Word2Vec can improve performance, but embeddings can also be learned during training.

Bidirectional LSTM Layer

- Purpose: Processes the input sequence in both forward and backward directions.
- Function:
    - Forward LSTM: Reads the sequence from start to end.
    - Backward LSTM: Reads the sequence from end to start.
    - Outputs are combined (usually concatenated) to form a rich representation for each timestep.
- Key Benefit: Captures context from past and future simultaneously, improving sequence understanding

Dropout Layer

- Purpose: Regularization to prevent overfitting.
- Function: Randomly sets a fraction of inputs to zero during training.
- Notes: Helps the network generalize better to unseen data, especially when the dataset is small.

Dense (Fully Connected) Layer

- Purpose: Maps the features extracted by the BiLSTM to the output space.
- Function: Combines information from all hidden units to predict the final output.
- Notes: Each neuron in this layer connects to all inputs from the previous layer.

Activation Layer (Softmax for multi-class classification)

- Purpose: Converts the Dense layer's output into probabilities.
- Function: Ensures that the sum of all predicted class probabilities equals 1.

- Notes:
    - For rating prediction (1–5), each neuron corresponds to a rating class.
    - The class with the highest probability is the model's prediction.

Why BiLSTM?

- Standard LSTM only reads input forward, so it misses future context.
- BiLSTM reads the input twice:
    - Once forward
    - Once backward
- Useful in NLP tasks where context on both sides matters (e.g., sentiment analysis, POS tagging, etc.)

Advantages of BiLSTM

- **Contextual Awareness:** Captures both past and future context.
- **Improved Accuracy:** Better performance in tasks like sentiment analysis, text classification, and speech recognition.
- **Flexible:** Can be combined with embedding layers, attention mechanisms, or other neural architectures.