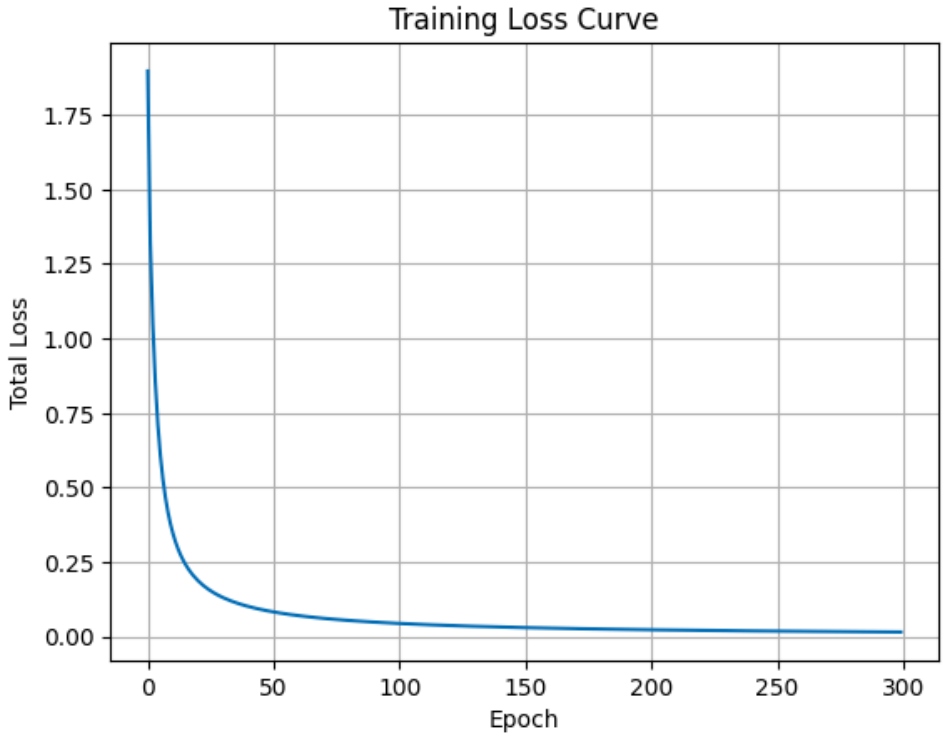
LAB7_inclass

Results:

```
Epoch [50/300] - Loss: 0.0848
Epoch [100/300] - Loss: 0.0451
Epoch [150/300] - Loss: 0.0309
Epoch [200/300] - Loss: 0.0236
Epoch [250/300] - Loss: 0.0191
Epoch [300/300] - Loss: 0.0160
```




```
=== Testing ===
Sentence: ['The', 'movie', 'is', 'bad'] → Prediction: 0
Sentence: ['The', 'movie', 'is', 'good'] → Prediction: 1
```

QA:

What does the hidden state □ represent at each time step when processing a sequence of □ words?

At each time step **t**, the hidden state $\mathbf{s_t}$ represents the **summarized information** of the input sequence **up to the**

current word.
It captures both the **current input word** and the **context from all previous words** in the sequence.
In other words, $\mathbf{s_t}$ **acts like the memory** of the RNN that accumulates and updates information as it reads each word in order.

Why are RNNs hard to train?

RNNs are hard to train mainly because of the problems of **vanishing gradients** and **exploding gradients** during backpropagation through time (BPTT).

Vanishing gradients make the model unable to learn long-range dependencies, since the gradients shrink exponentially and updates become too small.

Exploding gradients cause the gradients to become very large, making the model unstable or the loss diverge.