

## Summary and Highlights

Congratulations! You have completed this lesson. At this point in the course, you know that:

- A bi-gram model is a conditional probability model with context size one, that is, you consider only the immediate previous word to predict the next one.
- A trigram model is also a conditional probability function and can improve on the bigram model's limitations by increasing the context size to two.
- The concept of a trigram can be generalized to an N-gram model, which allows for an arbitrary context size.
- In the realm of neural networks, the context vector is generally defined as the product of your context size and the size of your vocabulary. Typically, this vector is not computed directly but is constructed by concatenating the embedding vectors.
- An N-gram model allows for an arbitrary context size.
- In Pytorch, the n-gram language model is essentially a classification model, using the context vector and an extra hidden layer to enhance performance.
- The n-gram model predicts words surrounding a target by incrementally shifting as a sliding window.
- In training the model, prioritize the loss over accuracy as your key performance indicator or KPI.