

## Course Conclusion

Congratulations! You've successfully completed the course and are now equipped with valuable insights into various aspects of Natural Language Processing (NLP) and AI model development. At this point, you know that:

- One-hot encoding converts categorical data into feature vectors.
- The bag-of-words representation portrays a document as the aggregate or average of one-hot encoded vectors.
- A document classifier seamlessly categorizes articles by analyzing the text content.
- A neural network is a mathematical function consisting of a sequence of matrix multiplications with a variety of other functions.
- The Argmax function identifies the index of the highest logit value, corresponding to the most likely class.
- Hyperparameters are externally set configurations of a neural network.
- The prediction function works on real text that starts by taking in tokenized text. It processes the text through the pipeline, and the model predicts the category.
- A neural network functions via matrix and vector operations, called learnable parameters.
- In neural network training, learnable parameters are fine-tuned to enhance model performance. This process is steered by the loss function, which serves as a measure of accuracy.
- Cross-entropy is used to find the best parameters.
- For unknown distribution, estimate it by averaging the function applied to a set of samples. This technique is known as Monte Carlo sampling.
- Optimization is used to minimize the loss.
- The training data is split into training and validation, and then data loaders are set up for training, validation, and testing.
- Batch size specifies the sample count for gradient approximation, and shuffling the data promotes better optimization.
- 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.
- 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.
- Word2Vec is the short form for "word to vector." It is a group of models that produce word embeddings or vectors, which are numerical representations capturing the essence of words.
- A neural network model consists of an input layer, an embedding layer, and an output layer. Words are fed into an embedding layer, which interacts with an output layer to predict context words.
- The continuous bag of words, or CBOW model, utilizes context words to predict a target word and generate its embedding.
- Skip-gram model predicts surrounding context words from a specific target word. It operates in contrast to the CBOW model.
- Sequence-to-sequence models within generative AI are used in machine translation, such as converting English phrases into French.
- Sequence-to-label tasks take multiple inputs to produce a single label, useful in document classification.
- Label-to-sequence tasks generate a full sequence from a single input, as seen in generative models for image creation.
- RNN is a type of artificial neural network that uses sequential or time series data.
- RNNs can be used to create sequence-to-sequence models that receive one sequence as input and generate another sequence as output.

- The encoder-decoder architectures are introduced so that the sequences do not necessarily require to be of the same length.
- Perplexity is a metric and a precious tool for evaluating the efficiency of LLMs and GenAI models. It is calculated as an exponent of the loss obtained from the model.
- F1-Score is a harmonic mean of precision and recall that is used to judge the performance of a model based on them.