



Week 7 Glossary

Activation function

A function which transforms linear inputs into nonlinear forms to help the network learn complex patterns in the data.

Convolutional neural network (CNN)

A deep-learning neural network that makes use of convolutional layers. It learns convolutional filters that process the input image to a particular layer and produce a new image (with many channels). CNNs are particularly useful for processing image data but can also be used for audio signals or videos.

Cosine similarity

A function that computes the similarity between two vectors (or sequences of numbers).

Count vectorizer

A scikit-learn library tool in Python used to convert a collection of text into a vector. It can tokenize a collection of text documents and build a vocabulary of known words.

Data preprocessing

The manipulation or dropping of data before it is used to ensure or enhance performance.

Deep averaging network

A neural network consisting of two components: a word embedding and a traditional neural network (sometimes even a linear classifier, which is a neural network without hidden layers).

Deep neural network

A neural network with many more than two hidden layers, each with many more than two nodes. Traditional networks have a few hidden layers, whereas deep learning models can have hundreds of hidden layers. Empirically deep neural networks perform better on certain machine learning tasks, such as image recognition.

Feedforward neural network

A type of artificial neural network in which nodes' connections do not form a loop. A feedforward neural network typically takes one word or sentence in a text sequence. Still, it loses any context of what was previously input.



Gradient descent

An optimization algorithm that searches for the minimum of a loss function by slightly changing the parameters of a classifier in the direction of the greatest negative gradient of the loss. It is often too hard to calculate the gradient of the loss for all training examples; for this reason, we usually use stochastic gradient descent instead.

Hidden layer

A layer of a neural network located between the input and output layers which applies a linear transformation followed by a nonlinear transformation, called the activation function, to its input values. All neural networks have at least one hidden layer; deep neural networks have many hidden layers.

Hyperparameters

The “knobs” that you tweak during successive runs of training a model; they help guide the learning process. They are parameters in the model that are not learned but set before learning. Hyperparameters often trade off the complexity vs. simplicity of models.

Machine learning (ML)

A broad class of methods and algorithms for building predictive models from data without prescribing the specific form of relationships between inputs and outputs. Machine learning is considered a subfield in the larger field of artificial intelligence but also straddles the world of data science, which is an amalgamation of human insight and automated inference.

N-grams

Combinations of individual word tokens.

Natural language processing (NLP)

A branch of artificial intelligence (AI) that enables machines to understand the human language. NLP interprets raw, arbitrary written text and transforms it into something a computer can understand.

Neural network

A supervised learning algorithm designed to solve complex, real-world problems. It can recognize complex patterns and nonlinear relationships between features and labels. Neural networks are often used in the NLP field.

Output layer

In neural networks, the last layer that outputs a prediction.



Recurrent neural networks (RNN)

A special type of neural network designed so that a given node's output flows back into the same node. In RNN, the information cycles through a loop. When it makes a decision, it considers the current input and what it has learned from the inputs it has previously received.

Scikit-learn pipeline

A scikit-learn utility that orchestrates the flow of data into and out of a machine learning model to automate the machine learning workflow.

Stochastic gradient descent

An approximation of gradient descent. The gradient of the loss function is applied to a subset of all the training examples instead of the whole set, which is much faster to compute.

Stop word

A token that appears very frequently in different examples of text but also adds very little predictive value.

Stop word removal

A data preprocessing step that removes the words that commonly occur. Stop words are usually removed to reduce data size and to speed up computation.

TF-IDF vectorizer

Also known as “term frequency-inverse document frequency”; a process for encoding text that captures the relative importance of a word to a given document.

Text classification

A machine learning technique that assigns a set of predefined categories to open-ended text, categorizing text into organized groups.

Tokenization

The process of parsing text to remove certain words. Tokenization allows you to use textual data for predictive modeling and map every word in training data to a future position.

Vectorization

The simple process of representing a word's binary presence or frequency in a given text example. Vectorization is a common strategy for mapping individual word tokens to a number.



Word embedding

A type of word representation which allows words with similar meanings to have an equal representation. In word embedding, each word can be represented by a k -dimensional vector. Those factors are commonly pre-trained and available as a lookup table.

