Module 1: Introduction to Machine Learning

Glossary

‘Black Box’

This is a metaphor for understanding machine learning.

Continuous Distribution

This is a distribution in which data can take on any value within a specified range (which may be infinite). Temperature, for instance, can be on a continuum from absolute zero to the surface of the sun, but also, there can be infinite points of data between 12ºC and 13ºC, depending on the quality of the thermometer.

DataFrame

This is a Python object of rows and columns.

Discrete Distribution

This is a distribution in which data can only take on certain values, for example, integers. The recording of the outcomes of a six-sided dice roll would generate a discrete distribution because there are only six possibilities for an outcome.

Expected Value

This is the result you can expect from some action. It can be calculated by summing the values of a random variable with each value multiplied by its probability of occurrence.

Expected Variance

This is a measure of how far a set of numbers is spread out from their average value.

Histogram

The histogram divides a numeric variable into multiple bins, and these bins calculate the observations that fall into each bin. This columnar representation of binned counts gives you an instant sense of the distribution of values in a variable.

Mean

The mean is the average of a set of given numbers. For instance, in the given number set [0,1,2,5,6,15,20], the mean would be the sum of all the numbers divided by the count of numbers in the set. The sum of all numbers is 49. There are 7 numbers in the set, so the mean would be 49/7, or 7.

Median

The median is the middle number in a set of given numbers. For instance, in the given number set [0,1,2,5,6,15,20], the median would be the middle number, or 5.

Outliers

These are explained or unexplained values that lie outside the normal distribution.