Module 2: Fundamentals of Machine Learning

Glossary

Central Limit Theorem

A theorem proving that the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger.

Conditional Probabilities

The likelihood of an event or outcome occurring, based on the occurrence of a previous event or outcome.

Continuous Distribution

A type of probability distribution that shows the probabilities of the possible values of a continuous random variable. A continuous random variable is a random variable with a set of possible values (known as the range) that is infinite and uncountable.

Correlation

"The degree to which the needs, demands, goals, objectives, and/or structures of one component are consistent with those of the other."
Source: Nadler and Tushman

Congruence Hypothesis

The strength and direction with which two variables move in coordination with one another.

Covariance

The direction of the linear relationship between variables.

Covariance Matrix

A square matrix giving the covariance between each pair of elements in a given set of variables.

Discrete Distribution

A type of probability distribution that shows the probability of occurrence of each value of a discrete random variable. A discrete random variable is a random variable that has countable values, such as a list of non-negative integers.

Gaussian Distribution

A type of distribution, also known as a Normal Distribution, whose probability distribution is symmetric about the mean. In graph form, a Gaussian distribution appears as a bell curve.

Law of Large Numbers

A theorem proving that as a sample size grows, its mean gets closer to the average of the whole population.

Mean

The sum of all values in a dataset divided by the total number of values.

Measures of Central Tendency

A descriptive summary of a dataset that takes the form of a single value and reflects the center of the data distribution. These values are mean, median, and mode.

Median

The middle value in a dataset that is arranged in ascending order (from the smallest value to the largest value). If a dataset contains an even number of values, the median of the dataset is the mean of the two middle values.

Multivariate Distribution

A type of probability distribution, also known as a Gaussian Distribution, that is symmetric about the mean. In graph form, a normal distribution appears as a bell curve.

Normal Distribution

A type of probability distribution, also known as a Gaussian Distribution, that is symmetric about the mean. In graph form, a normal distribution appears as a bell curve.

Pair Plot

A type of plot that is used to understand the best set of features with which to explain a relationship between two variables or to form the most separated clusters.

Scatterplot

A type of plot in which the values of two variables are plotted along two axes, and the pattern of the resulting points reveal any correlation present.

Standard Deviation

A descriptive statistic that measures how spread out the numbers within a data set are. It is calculated as the square root of the variance.

Uniform Distribution

A type of probability distribution in which all outcomes are equally likely.