Septiembre 2023

Fundamental concepts of Machine learning

Supervised learning is defined by its use of labeled datasets to train algorithms that classify data, learns to make decisions or predict outcomes accurately. The term "Supervised" refers to the process of providing the algorithm with a supervisor or teacher in the form of labeled data, where each example in the training dataset consists of input-output pairs.

In supervised learning, the algorithm uses a training set to teach models to yield the desire outputs, which allow the model to learn over time. The algorithm measures its accuracy through the loss function, adjusting until the error has been sufficiently minimized.

Supervised learning can be separated into two types of problems when data mining: classification and regression.

Classification uses an algorithm to accurately assign test data into specific categories. It recognizes specific entities within the dataset and attempts to draw some conclusions on how those entities should be labeled or confirmed. Common classification algorithms are linear classifiers, suport vector machines (SVM), decision trees, k-nearest neighbor, and random forest.

Regression is used to understand the relationship between dependent and independent variables. It is commonly used to make projections, such as for sales revenue for a given business. Linear regression, logistical regression and polynomial regression are popular regression algorithms.

Unsupervised learning is a machine learning procedure in which an algorithm is trained on a dataset without labeled output/target values. Unlike supervised learning, where the algorithm learns to make predictions or decisions based on labeled data, unsupervised learning focus on discovering patterns, structures, and relationships within the data without explicit guidance.

It's ability to discover similarities and differences in ~~for [illegible] ideal~~ solution for exploratory data analysis, cross-selling strategies, customer services, and image recognition.

Probabilistic model: Statistical technic to take into account the impact of randoms events or actions in predicting the potential of futures outcomes.

The choice between supervised and unsupervised learning depends on the nature of the data, the problem objectives, and the availability of labeled training data.

The main difference between supervised vs unsupervised learning is the need for labelling training data. Supervised machine learning relies on labelled input and output training data, whereas unsupervised learning processes unlabelled or raw data.

yourself