Vertica Machine Learning Overview

### Naive Bayes

Naive Bayes is a probabilistic classification method based on Bayes' theorem. It assumes that the features are conditionally independent given the class. Naive Bayes classifiers are simple and effective for many classification tasks.

### Support Vector Machines (SVM)

Support Vector Machines (SVMs) are supervised learning models used for classification and regression analysis. They are used in pattern recognition and data mining. SVMs maximize the margin between different classes.

### Random Forests

Random Forests is an ensemble learning method for classification, regression, and other tasks. It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

### Linear Regression

Linear Regression is a simple and popular statistical model used for predictive analysis. It models the relationship between a dependent variable and one or more independent variables. The model is given by the equation $y = mx + b$,

### Logistic Regression

Logistic Regression is a statistical model that uses a logistic function to model a binary dependent variable. It is used for classification problems where the outcome is binary (0 or 1).

### K-Means Clustering

K-Means Clustering is an unsupervised learning algorithm used for data segmentation and classification. It works by partitioning a dataset into a specified number of clusters.

### Database-Supported Machine Learning

Vertica is the first database to support machine learning on columnar data, allowing users to perform machine learning on their data without moving it to the host. It supports a variety of machine learning algorithms, including regression, classification, and clustering.

### Supervised Learning Functions

Supervised learning functions in Vertica allow users to perform regression and classification tasks directly on their data. These functions include:

- Linear Regression
- Logistic Regression
- Support Vector Machines (SVM)
- Random Forests

### Unsupervised Learning Functions

Unsupervised learning functions are used for data analysis and exploration when labels are not available. These functions include:

- Clustering
- Normalization
- Anomaly Detection

### Vertica Machine Learning Overview Diagram

The diagram illustrates the processes and algorithms involved in Vertica's machine learning capabilities, including data preparation, training and testing, and model evaluation.

---

### Resources

- [Vertica Machine Learning Overview](https://www.vertica.com/documentation)
- [Example code and data sets](https://github.com/vertica/Machine-Learning-Examples)