Machine Learning Algorithms

By Reda Ouzidane

April 3, 2025

Classification Algorithms

These algorithms are used to predict a categorical label or class.

- **Logistic Regression**
  Purpose: Predict binary outcomes (yes/no, true/false) for classification tasks.
  Type: Supervised, Classification.
- **K-Nearest Neighbors (KNN)**
  Purpose: Classifies data points based on the majority class of their k nearest neighbors.
  Type: Supervised, Classification.
- **Support Vector Machine (SVM)**
  Purpose: Finds the hyperplane that best separates different classes in high-dimensional space.
  Type: Supervised, Classification.
- **Decision Trees**
  Purpose: Builds a tree-like model of decisions based on features, classifying data by traversing the tree.
  Type: Supervised, Classification.
- **Random Forest**
  Purpose: An ensemble of decision trees that improves accuracy by averaging predictions over multiple trees.
  Type: Supervised, Classification.
- **Naive Bayes**
  Purpose: Classifies data based on Bayes' theorem, assuming feature independence.
  Type: Supervised, Classification.
- **Gradient Boosting (XGBoost, LightGBM, CatBoost)**
  Purpose: Uses an ensemble of weak learners (typically decision trees) to improve accuracy by focusing on errors of previous models.
  Type: Supervised, Classification.

- **K-Means (when used for classification)**
  Purpose: Uses centroid-based clustering to predict labels for data points.
  Type: Supervised/Unsupervised, Classification.

Regression Algorithms

These algorithms are used to predict continuous numerical outcomes.

- **Linear Regression**
  Purpose: Predicts a continuous value based on the linear relationship between input features and the target variable.
  Type: Supervised, Regression.
- **Ridge Regression (L2 Regularization)**
  Purpose: A linear regression model with regularization to prevent overfitting.
  Type: Supervised, Regression.
- **Lasso Regression (L1 Regularization)**
  Purpose: A variation of linear regression that uses L1 regularization to perform feature selection.
  Type: Supervised, Regression.
- **Polynomial Regression**
  Purpose: Extends linear regression to model nonlinear relationships by fitting polynomial functions.
  Type: Supervised, Regression.
- **Support Vector Regression (SVR)**
  Purpose: Uses the principles of SVM to fit a regression model, aiming to minimize error within a certain margin.
  Type: Supervised, Regression.
- **Decision Trees (when used for regression)**
  Purpose: Predicts continuous outcomes by splitting the feature space based on certain decision rules.
  Type: Supervised, Regression.
- **Random Forest (when used for regression)**
  Purpose: An ensemble of decision trees used to predict continuous outcomes by averaging results over multiple trees.
  Type: Supervised, Regression.
- **Gradient Boosting Regression (XGBoost, LightGBM, CatBoost)**
  Purpose: Uses an ensemble of decision trees to predict continuous values, focusing on the errors of previous models.
  Type: Supervised, Regression.

Clustering Algorithms

These algorithms are used to group similar data points into clusters without predefined labels.

- **K-Means Clustering**
  Purpose: Groups data into k clusters based on feature similarity, minimizing within-cluster variance.
  Type: Unsupervised, Clustering.
- **Hierarchical Clustering**
  Purpose: Builds a tree-like structure (dendrogram) of nested clusters by either agglomerating or dividing clusters.
  Type: Unsupervised, Clustering.
- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**
  Purpose: Clusters data based on density, identifying areas of high density and separating noise (outliers).
  Type: Unsupervised, Clustering.
- **Gaussian Mixture Models (GMM)**
  Purpose: Models data as a mixture of multiple Gaussian distributions, useful for soft clustering.
  Type: Unsupervised, Clustering.

Dimensionality Reduction Algorithms

These algorithms are used to reduce the number of features or dimensions in data while retaining most of the variance.

- **Principal Component Analysis (PCA)**
  Purpose: Reduces dimensionality by projecting data onto the directions of maximum variance (principal components).
  Type: Unsupervised, Dimensionality Reduction.
- **t-Distributed Stochastic Neighbor Embedding (t-SNE)**
  Purpose: Reduces dimensions for visualization, preserving local structures in data.
  Type: Unsupervised, Dimensionality Reduction.
- **Linear Discriminant Analysis (LDA)**
  Purpose: Reduces dimensionality while maximizing class separability.
  Type: Supervised, Dimensionality Reduction.

Reinforcement Learning Algorithms

These algorithms are used in environments where agents learn by interacting with the environment and receiving rewards.

- **Q-Learning**
  Purpose: An off-policy reinforcement learning algorithm that learns the value of actions to maximize long-term rewards.
  Type: Reinforcement Learning.
- **Deep Q-Network (DQN)**
  Purpose: Combines Q-learning with deep learning for high-dimensional state spaces, like images.
  Type: Reinforcement Learning.
- **Policy Gradient Methods**
  Purpose: Directly optimize the policy (action selection) in reinforcement learning, useful for continuous action spaces.
  Type: Reinforcement Learning.
- **Actor-Critic Methods**
  Purpose: Combines value-based and policy-based reinforcement learning, using two models: an actor for choosing actions and a critic for evaluating them.
  Type: Reinforcement Learning.

Anomaly Detection Algorithms

These algorithms are used to detect unusual patterns or outliers in the data.

- **Isolation Forest**
  Purpose: Detects anomalies by isolating observations in the data.
  Type: Unsupervised, Anomaly Detection.
- **One-Class SVM**
  Purpose: A variation of SVM used to detect outliers in a dataset by learning a boundary around the normal data points.
  Type: Unsupervised, Anomaly Detection.
- **Autoencoders (for anomaly detection)**
  Purpose: Neural networks trained to reconstruct input data, where high reconstruction error indicates an anomaly.
  Type: Unsupervised, Anomaly Detection.

Association Algorithms

These algorithms are used to find relationships or patterns between variables.

- **Apriori Algorithm**
  Purpose: Finds frequent item sets in transaction data and generates association rules between them.
  Type: Unsupervised, Association.

- **Eclat Algorithm**
  Purpose: Similar to Apriori but uses a different method (intersection of item sets) to find frequent item sets.
  Type: Unsupervised, Association.