

*ML Algorithms support parallel, distributed, in-database execution for performance and scalability, improved memory utilization, partitioned models, and automatic mining of text columns*

## Predict Categories → Classification

*Target variable contains 2 or more distinct category values*

Decision Tree	Generates human-interpretable rules, can be used for segmentation
Random Forest*	High accuracy predictions, avoids overfitting
Naïve Bayes	Yields interpretable probabilities, assumes predictor independence
Support Vector Machine	High accuracy; linear or Gaussian kernel; IPM (non-linear , linear ) and SGD (linear, wide data) solvers; sparsity optimizations
Logistic Regression	Narrow, wide , sparse data; QR, Cholesky, and SGD solvers; enables ridge, feature selection/generation, sparsity optimization
Neural Network*	Well-suited to noisy and complex data, supports many hidden layers, binary classification only

## Predict Numeric Values → Regression

*Target variable contains number (integer, real) values*

Linear Model	Same results as R's lm(); global statistics, coefficient statistics, and row diagnostics
Generalized Linear Model	Narrow, wide , sparse data; QR, Cholesky, and SGD solvers; enables ridge, feature selection/generation, sparsity optimization
Support Vector Machine	High accuracy; linear or Gaussian kernel; IPM (non-linear , linear ) and SGD (linear, wide data) solvers; sparsity optimizations
Stepwise Regression*	Selects "best" set of predictors for linear model; supports forward, backward, both, alternate, and none direction
Neural Network*	Well-suited to noisy and complex data, supports many hidden layers

## Rank Predictors → Attribute Importance

*Supervised and unsupervised ranking of variables*

Minimum Description Length	Select most important variables for classification and regression
Expectation Maximization	Supports unsupervised attribute ranking and pairwise dependency estimates

## Group or Segment Cases → Clustering

*All Input variables considered for grouping rows (cases) into clusters*

K-Means	Produces k clusters in hierarchy; Euclidean and cosine distance, generates probabilities, rules, and statistics; sparsity optimizations
Orthogonal Partitioning	Discovers natural clusters up to maximum specified; density-based, generates probabilities, rules, and statistics
Expectation Maximization	Automated model search; protection against overfitting; numeric and multinomial distributions; high quality probability estimates; generates cluster hierarchy, rules, and other statistics

## Derive New Values → Feature Extraction

*All Input variables considered to generate reduced set of variables*

Non-negative Matrix Factorization	Derives features based on non-negative linear combinations which makes features more interpretable
Singular Value Decomposition	Narrow data via Tall and Skinny solvers; wide data via stochastic solvers; eigensolvers (faster, sparsity) or SVD (more stable)
Principal Component Analysis	Uses SVD to obtain a set of uncorrelated variables that contain the maximum amount of variance from dataset
Explicit Semantic Analysis	Text categorization with human-readable topic labels derived from corpus; semantic similarity estimates among documents

## Identify Unusual Cases → Anomaly Detection

*Flag cases as normal or anomalous by learning pattern of normal data*

One-Class SVM	High accuracy; linear or Gaussian kernel; IPM (non-linear , linear ) and SGD (linear, wide data) solvers; sparsity optimizations
---------------	--

## Predict Sequential Numeric Data → Time Series

*Target variable contains number (integer, real) values*

Exponential Smoothing*	Smooths time series – single and double; enables forecasting
------------------------	--

## Market Basket Analysis → Association Rules

*Transactional or 2D data representation to extract frequently occurring patterns*

Apriori	Finds frequent itemsets and generates human-interpretable rules; can compute aggregate measures associated with rules
---------	---

\* Oracle R Enterprise only

*...plus R extensible algorithm support in Oracle Data Mining, and open source R packages in combination with Oracle R Enterprise embedded R data-parallel and task-parallel execution*