ORACLE Machine Learning in Oracle Database – What do you want to do?

Oracle Machine Learning enables building AI applications and dashboards, delivering powerful in-database ML algorithms, automatic ML functionality, and integration with open source Python and R. OML algorithms support parallel execution for performance and scalability with improved memory utilization, and support for partitioned models and automatic mining of text columns

Classification

Predict target variable containing 2 (binary) or more (multi-class) category values

Decision Tree	Generates human-interpretable rules, can be used for segmentation
Explicit Semantic Analysis	Text categorization suitable for large text corpora
Extreme Gradient Boosting	Scalable implementation of popular XGBoost algorithm; supports tree and linear models
Logistic Regression / Generalized Linear Model	Predict binary (0/1, Yes/No) target attributes with attribute coefficients and model statistics; narrow, wide, sparse data; enables ridge, feature selection/generation; row diagnostics
Naïve Bayes	Computes conditional probabilities and yields interpretable probabilities; assumes predictor attribute independence
Neural Network	Well-suited to noisy and complex data, supports many hidden layers
Random Forest	Tree-based ensemble method that relies on bagging and feature randomness
Support Vector Machine	Solves linear and non-linear problems; multiple solvers; sparsity optimizations; supports multi-target classification (a list of targets per row)

Regression

Predict numeric target variable

Extreme Gradient Boosting	Scalable implementation of popular XGBoost algorithm; supports tree and linear models
Generalized Linear Model	Predict numeric target attributes with attribute coefficients and model statistics; narrow, wide, sparse data; enables ridge, feature selection/generation; row diagnostics
Neural Network	Well-suited to noisy and complex data, supports many hidden layers
Stepwise Regression	Selects “best” set of predictors for linear model; supports forward, backward, both, and alternate direction
Support Vector Machine	Solves linear and non-linear problems; multiple solvers; sparsity optimizations

Attribute Importance

Supervised and unsupervised ranking of variables to improve model quality

CUR Decomposition	Supports a low-rank SVD-based approach for ranking attribute importance as unsupervised method
Expectation Maximization	Supports unsupervised variable ranking and pairwise dependency estimates
Minimum Description Length	Select most important variables for classification and regression;

Ranking

Supervised prediction probability of one item ranking over other items

Extreme Gradient Boosting	Supports pairwise and list-wise ranking

Clustering

Group or segment cases into hierarchical clusters producing probabilities, rules, and statistics

Expectation Maximization	Automated model search; protection against overfitting; numeric and multinomial distributions; high quality probability estimates
K-Means	Produces specified number, k, of clusters; Euclidean and cosine distance functions; sparsity optimizations
Orthogonal Partitioning	Discovers natural clusters up to maximum number specified; density-based

Feature Extraction

Derive new values where all Input variables considered to generate reduced set of variables

Explicit Semantic Analysis	Text categorization with human-readable topic labels derived from corpus; semantic similarity estimates among documents
Non-negative Matrix Factorization	Derives features based on non-negative linear combinations for greater feature interpretability
Principal Component Analysis	Uses SVD to obtain a set of uncorrelated variables that contain the maximum amount of variance from dataset
Singular Value Decomposition	Narrow data via tall and skinny solvers; wide data via stochastic solvers

Anomaly Detection

Identify cases as normal or anomalous by learning patterns of normal data

One-Class SVM	Special case of SVM classification that does not use a target; Solves linear and non-linear problems; multiple solvers; sparsity optimizations
MSET-SPRT	Process monitoring to detect anomalies with non-linear, non-parametric patterns in IoT sensor data; “Multivariate State Estimation Technique”

Time Series

Forecast or predict sequential numeric data using series order column with either Number or Date/Timestamp types

Exponential Smoothing	Single, double, and triple exponential smoothing for regular and irregular series, with and without trend and seasonality; multiple methods supported, including Holt-Winters

Association Rules

Market basic analysis using transactional or 2D data representation to extract frequently occurring patterns and rules

Apriori	Finds frequent itemsets and generates human-interpretable rules; computes support, confidence, lift, and aggregate measures associated with rules

Row Importance

Unsupervised ranking of rows

CUR Decomposition	Supports low-rank SVD-based approach for ranking row importance as unsupervised method