

Machine learning techniques and definitions

ANNs (Artificial Neural Networks): Computing systems with a set of models and algorithms that are brain-inspired, intended to imitate human learning (Frankenfield, 2018).

ARIMA (Autoregressive integrated moving average model): A prediction technique performed by a statistical analysis model using time series data (Zhang, 2003).

BPNN (Backpropagation neural network): An iterative process of tuning the weights in neural network models from errors provided by a loss function (Chen *et al.*, 2008).

BSM (Black Scholes Model): A mathematical model for calculating price deviation over a period of time for financial instruments (Kenton, 2019a).

CART (Classification and Regression Trees): A binary decision tree model performed through a sequence of questions, the answers to which are input to the next question, and each terminal node is the result of prediction (Breiman *et al.*, 2017; Pradeepkumar and Ravi, 2017).

FCM (Fuzzy c means): A clustering technique where an object can belong to more than one cluster with different probabilities (Zhang, Wang and Zhang, 2007).

GARCH (Generalized Autoregressive Conditional Heteroscedasticity): A statistical model for prediction through analysis of financial time-series data (Francq and Zakoian, 2019).

GAs (Genetic Algorithms): A search technique that imitates the process of natural selection to find optimal solutions (Pal and Wang, 2017).

GMDH (Group method of data handling): A type of inductive algorithm performed by a mathematical model that provides possible automatic discovery of relations in the data (Teng *et al.*, 2017).

GP (Gaussian Processes): A collection of random variables that have a normal distribution (Swain *et al.*, 2016).

GRNN (Generalized Regression Neural Network): A basic neural network-based function estimation

algorithm used for classification, regression, and prediction (Chen and Wang, 2018).

HMM (Hidden Markov Model): A statistical model used for machine learning with hidden states, as opposed to a standard Markov chain where all states are disclosed (Degirmenci, 2014).

Hierarchical clustering: builds a multilevel hierarchy of clusters by creating a cluster tree (Rokach and Maimon, 2005a).

KNN (K Nearest Neighbour): A supervised learning technique use for classification, where an object is considered as a member of a cluster with a plurality voting of its neighbours (Jadhav and Channe, 2016).

k-means: A clustering technique where an object can belong to only one cluster and a cluster must contain at least one object, in which centroids of clusters are means of objects in each cluster (Rokach and Maimon, 2005b).

k-medoids (PAM): A clustering technique where an object can belong to only one cluster and a cluster must contain at least one object, in which centroids of clusters are the most central object in each cluster (Rokach and Maimon, 2005b).

LR (Logistic Regression): A machine learning algorithm for statistical classification, where an outcome is determined by one or more independent variables (Attigeri *et al.*, 2015).

LSTM (Long Short-Term Memory): One of the most successful RNN architectures that introduce the memory cell, a unit of computation that replaces traditional artificial neurons in the hidden layer of the network. These memory cells make networks effectively associate memories with input remote in time (Roondiwala, Patel and Varma, 2017).

MCS (Monte Carlo Simulation): A technique used to model the probability of variant outcomes in hard to predict cases because of the interference of random variables (Morin, 2019).

MLP (Multilayer Perceptron): A type of ANN consisting of at least three layers: input layer, hidden layer, and output layer (Ramchoun *et al.*, 2016).

PSO (Particle Swarm Optimization): A technique performing optimal solution searching, inspired by the behaviour of social creatures in groups (Martínez and Cao, 2018).

QR (Quantile Regression): A regression mechanism that models the relationship between a dependent variable and independent variables at different quantiles (Pradeepkumar and Ravi, 2017).

QRNN (Quantile Regression Neural Network): A QR-based hybrid and a feedforward neural network that can be used to estimate the nonlinear models at different quantiles (Pradeepkumar and Ravi, 2017).

RBF (Radial Basis Function Neural Networks): A type of ANN performed by a linear model (Aljarah *et al.*, 2018).

RF (Random Forest): An ensemble method used to construct a prediction model of both classification and regression problems (Janitza, Tutz and Boulesteix, 2016).

RNN (Recurrent neural networks): A powerful model for processing sequential data such as sound, time-series data or written natural language. Some designs of RNN are used to predict the stock market (Saad, Prokhorov and Wunsch, 1998; Chen, Zhou and Dai, 2015; Lipton, Berkowitz and Elkan, 2015; Rather, Agarwal and Sastry, 2015).

SOM (Self-organizing maps): A type of ANN using unsupervised learning in training to build a two-dimensional map of the problem space (Galutira, Fajardo and Medina, 2019).

SVM (Support Vector Machine): A supervised machine learning technique for classification and regression analysis (Auria and Moro, 2008).

SVR (Support Vector Regression): A supervised machine learning technique for classification and regression analysis (Basak, Pal and Patranabis, 2007).