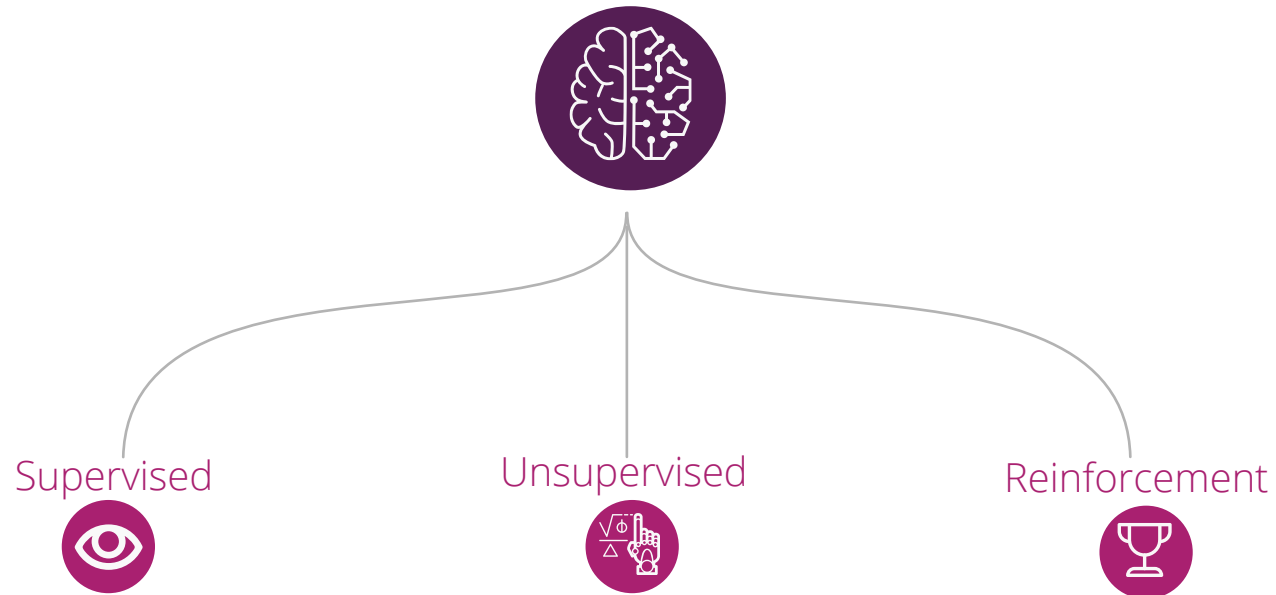# Machine Learning

MACHINE LEARNING

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.
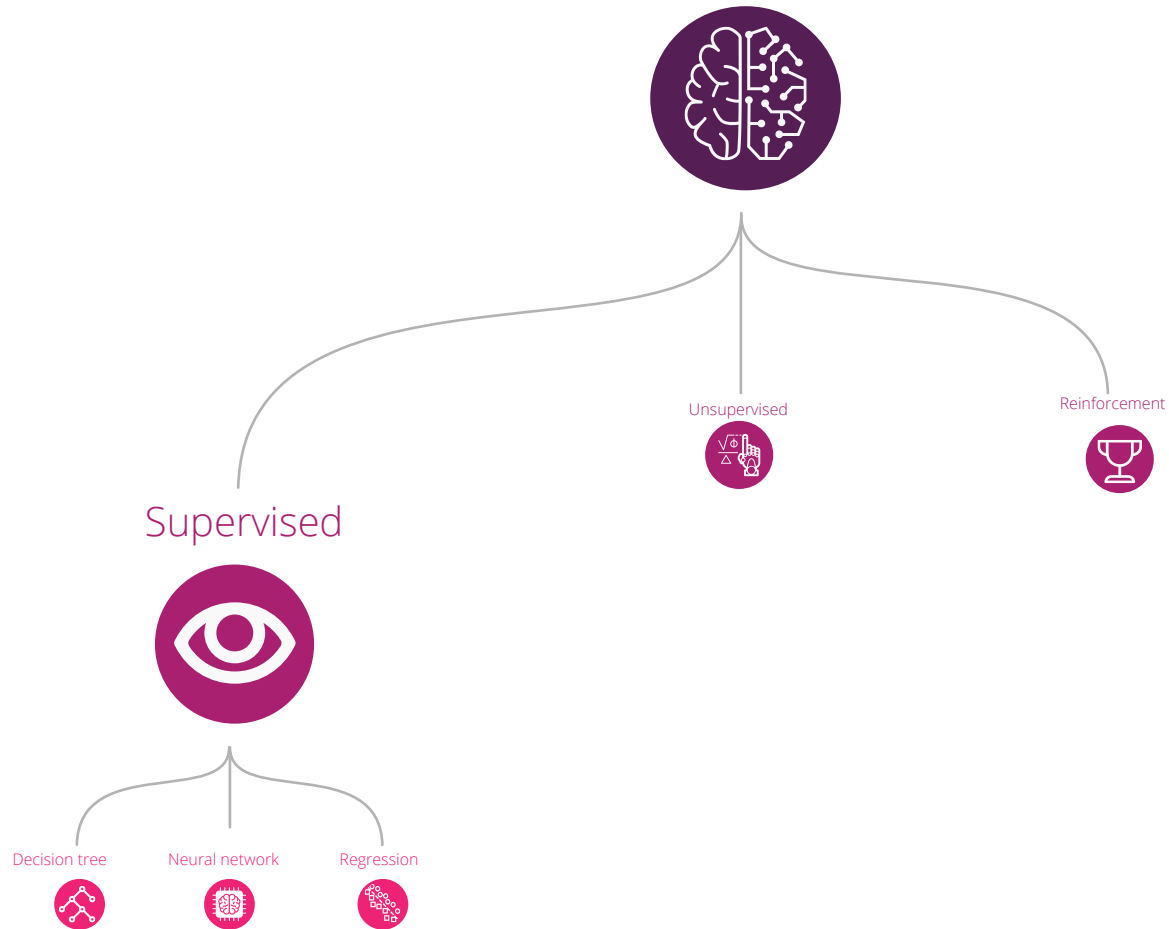
Machine Learning

THREE LEARNING METHODS

Supervised

Unsupervised

Reinforcement

Supervised

Unsupervised

Reinforcement

Machine Learning

Supervised

Unsupervised

Reinforcement

Decision tree

Neural network

Regression

SUPERVISED LEARNING

Supervised learning algorithms are trained using labeled examples, such as an input where the desired output is known. For example, a piece of equipment could have data points labeled either "F" (failed) or "R" (runs). The learning algorithm receives a set of inputs along with the corresponding correct outputs, and the algorithm learns by comparing its actual output with correct outputs to find errors. It then modifies the model accordingly. Through methods like classification, regression, prediction and gradient boosting, supervised learning uses patterns to predict the values of the label on additional unlabeled data. Supervised learning is commonly used in applications where historical data predicts likely future events. For example, it can anticipate when credit card transactions are likely to be fraudulent or which insurance customer is likely to file a claim.

Machine Learning

Supervised

Unsupervised

Reinforcement

Cluster Analysis

Support vector machine
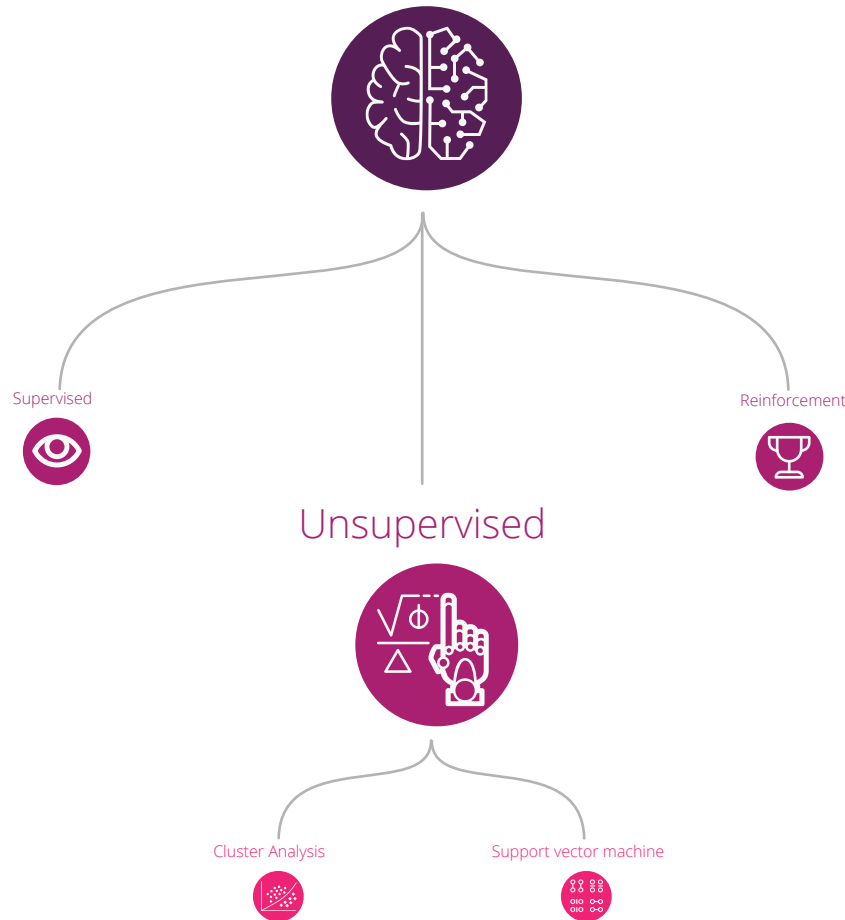
UNSUPERVISED LEARNING

Unsupervised learning is used against data that has no historical labels. The system is not told the "right answer." The algorithm must figure out what is being shown. The goal is to explore the data and find some structure within. Unsupervised learning works well on transactional data. For example, it can identify segments of customers with similar attributes who can then be treated similarly in marketing campaigns. Or it can find the main attributes that separate customer segments from each other. Popular techniques include self-organizing maps, nearest-neighbor mapping, k-means clustering and singular value decomposition. These algorithms are also used to segment text topics, recommend items and identify data outliers.
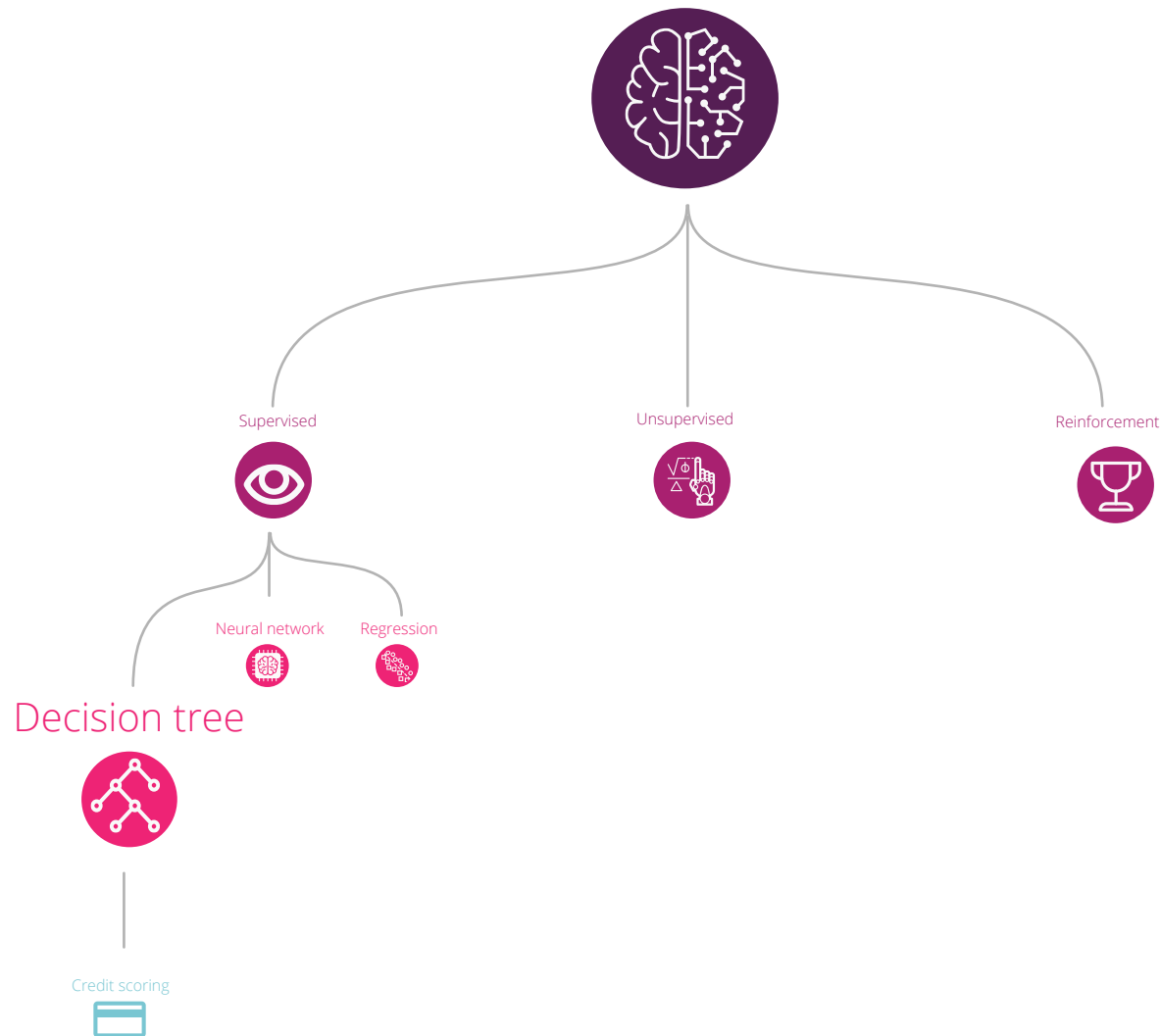
Machine Learning

Supervised

Unsupervised

Reinforcement
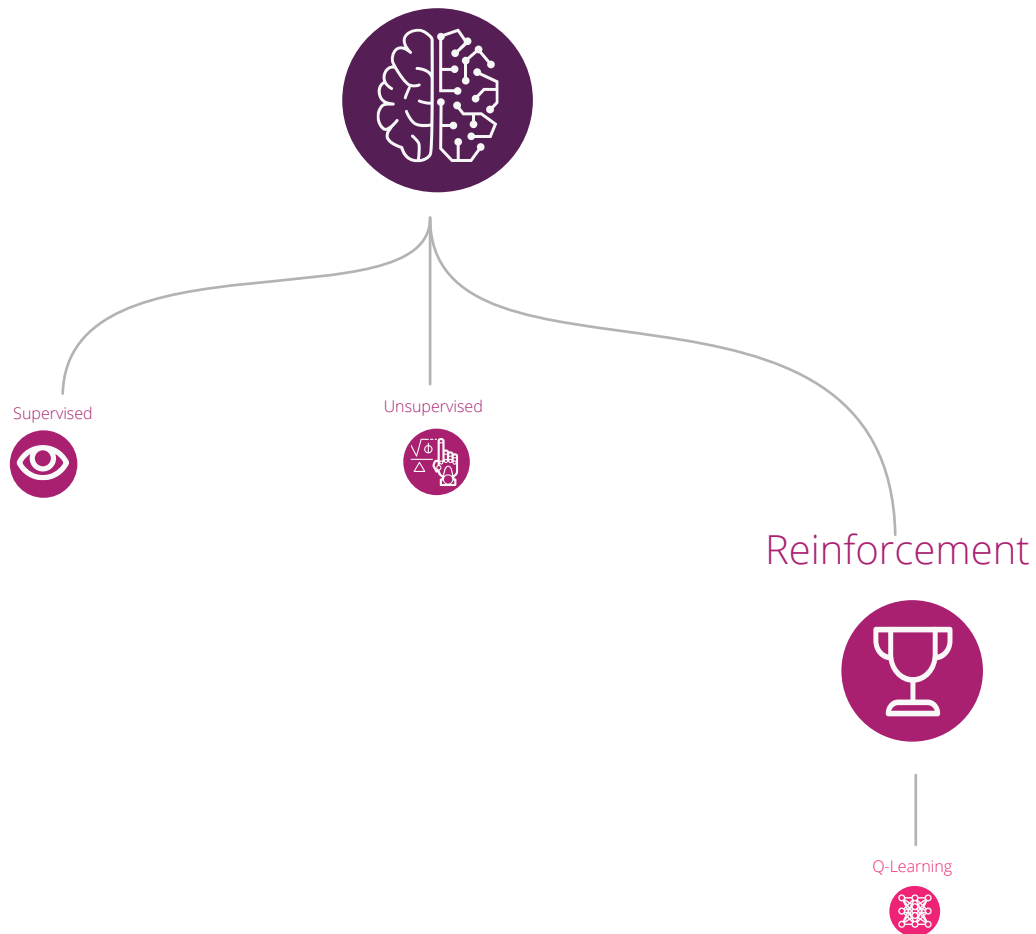
Q-Learning

REINFORCEMENT LEARNING

Reinforcement learning is often used for robotics, gaming and navigation. With reinforcement learning, the algorithm discovers through trial and error which actions yield the greatest rewards. This type of learning has three primary components: the agent (the learner or decision maker), the environment (everything the agent interacts with) and actions (what the agent can do). The objective is for the agent to choose actions that maximize the expected reward over a given amount of time. The agent will reach the goal much faster by following a good policy. So the goal in reinforcement learning is to learn the best policy.

Machine Learning

Supervised

Unsupervised

Reinforcement

Neural network

Regression

Decision tree

Credit scoring

DECISION TREES

Decision tree learning uses a decision tree to go from observations about an item to conclusions about the item's target value. It is one of the predictive modeling approaches used in statistics, data mining and machine learning.
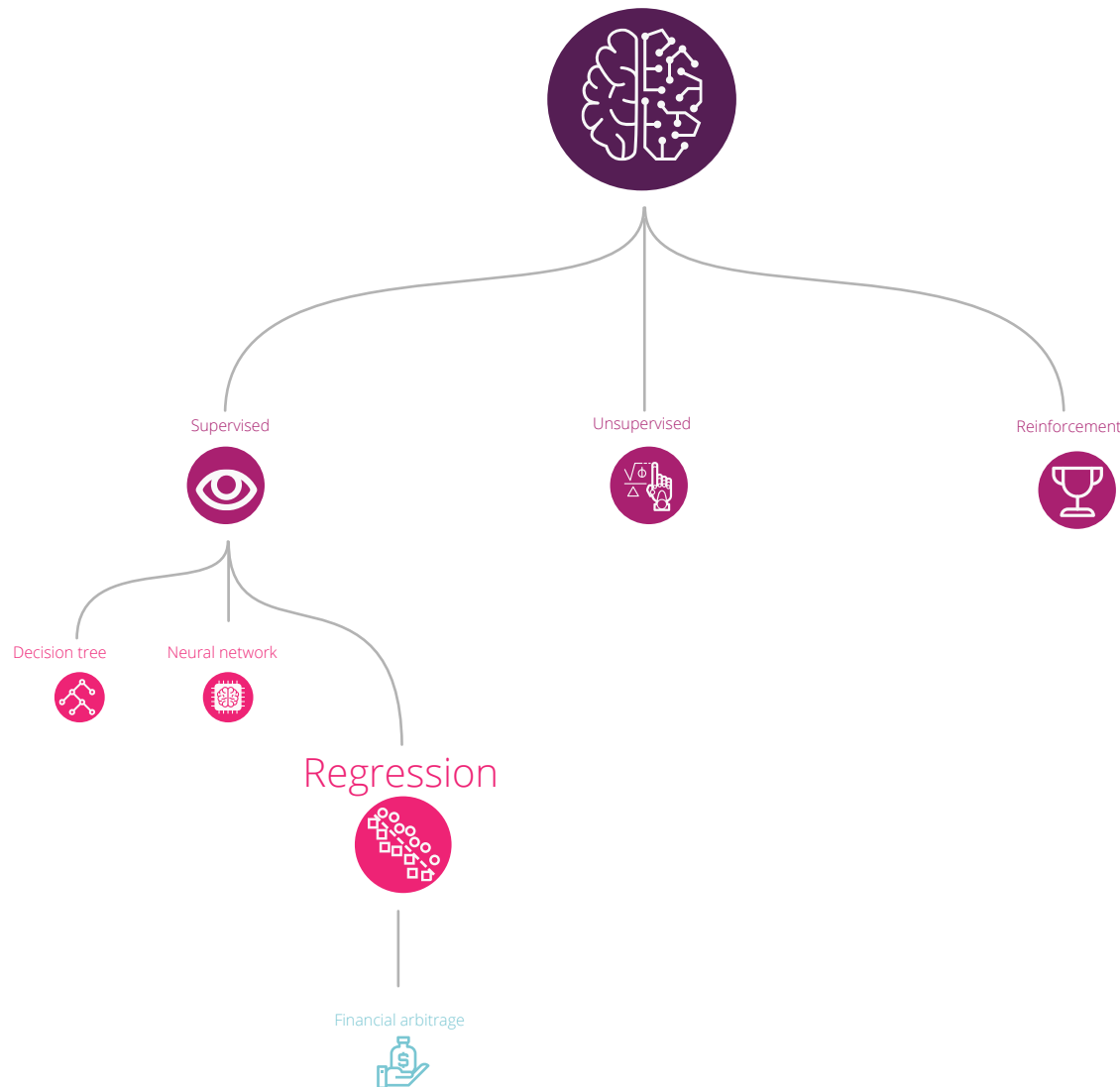
Machine Learning

Supervised

Unsupervised

Reinforcement

Decision tree

Regression

Neural network

Speech recognition
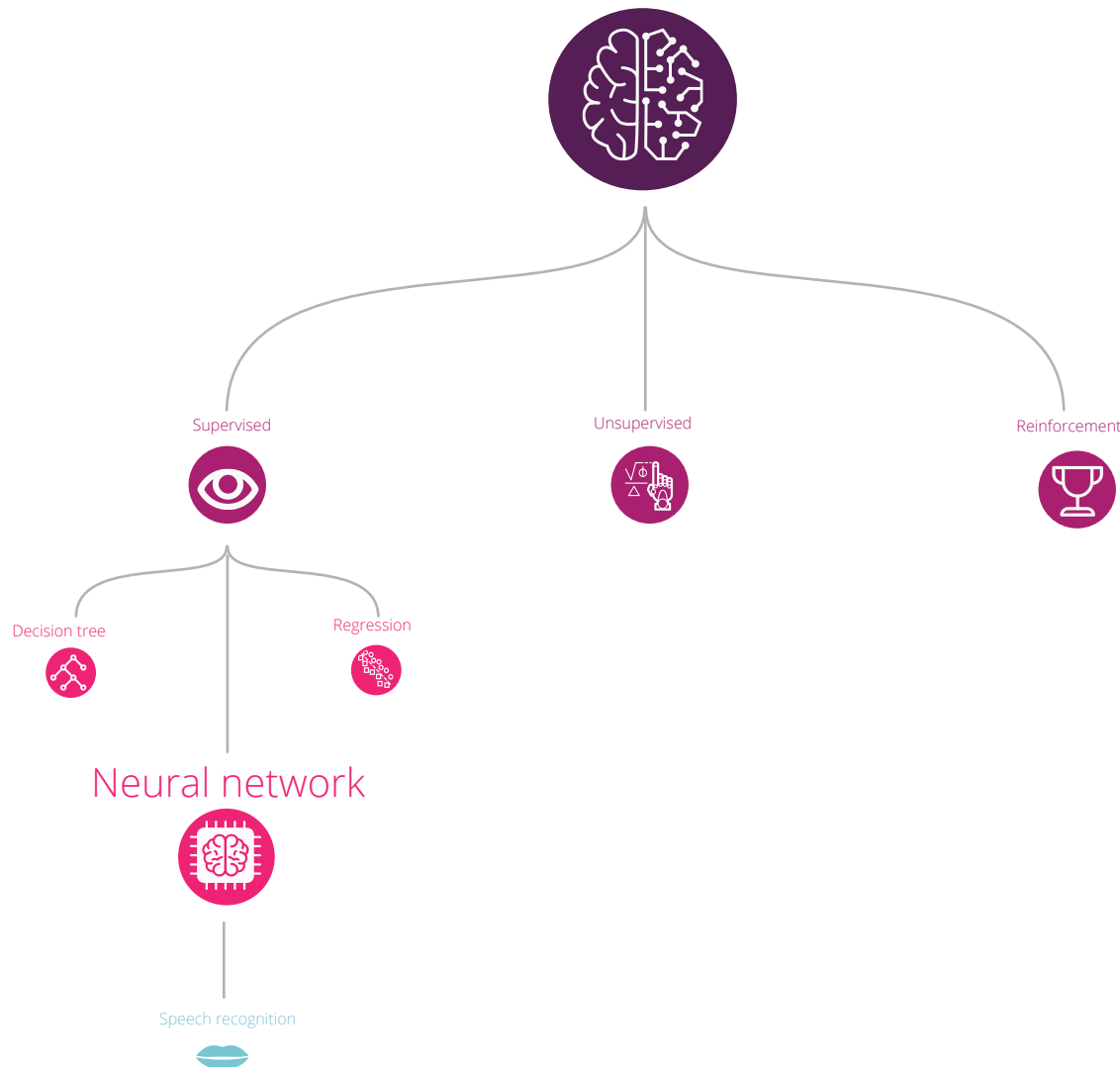
NEURAL NETWORK

The term neural network was traditionally used to refer to a network or circuit of neurons.[1] The modern usage of the term often refers to artificial neural networks, which are composed of artificial neurons or nodes. Thus the term may refer to either biological neural networks, made up of real biological neurons, or artificial neural networks, for solving artificial intelligence (AI) problems. The connections of the biological neuron are modeled as weights. A positive weight reflects an excitatory connection, while negative values mean inhibitory connections. All inputs are modified by a weight and summed. This activity is referred as a linear combination. Finally, an activation function controls the amplitude of the output. For example, an acceptable range of output is usually between 0 and 1, or it could be −1 and 1.
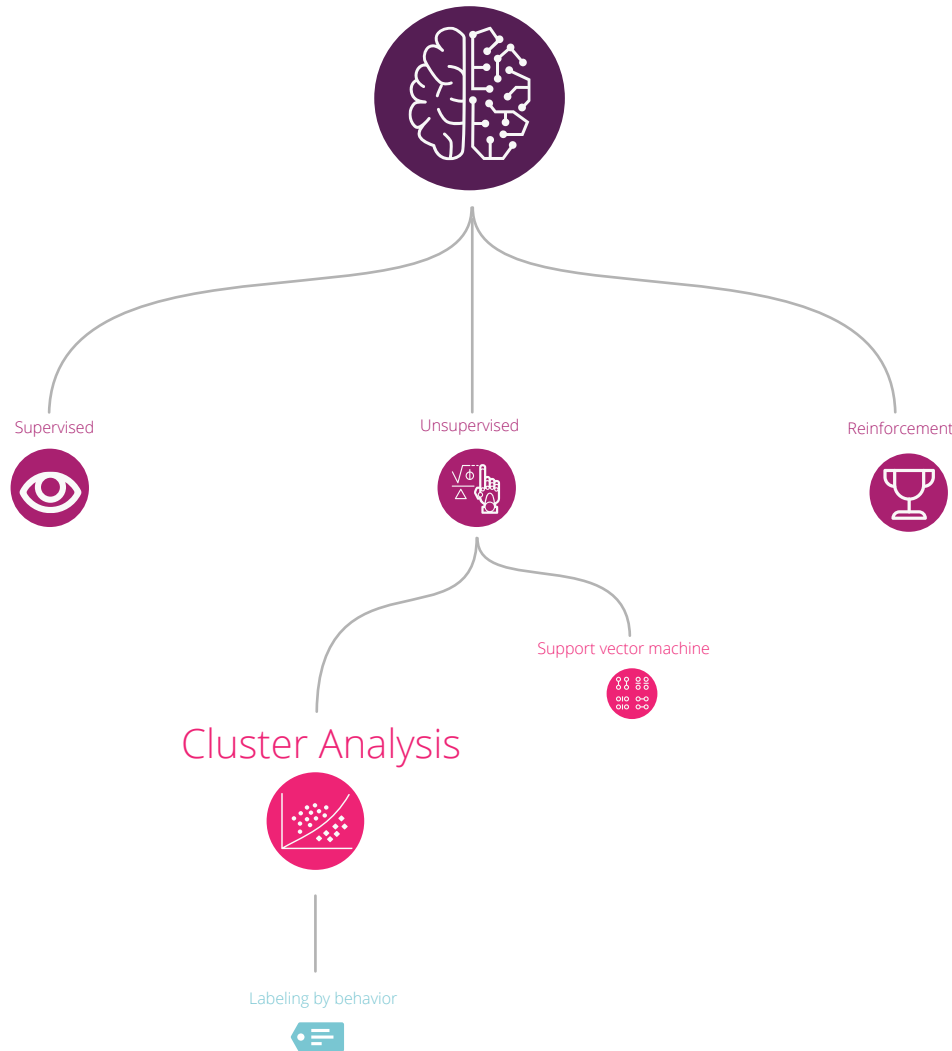
Machine Learning

Supervised

Unsupervised

Reinforcement

Decision tree

Neural network

Regression

Financial arbitrage

REGRESSION

In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression.[1] This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

Machine Learning

Supervised

Unsupervised

Reinforcement

Support vector machine

Cluster Analysis

Labeling by behavior

CLUSTER ANALYSIS

In machine learning and statistics, classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. Examples are assigning a given email to the "spam" or "non-spam" class, and assigning a diagnosis to a given patient based on observed characteristics of the patient (sex, blood pressure, presence or absence of certain symptoms, etc.). Classification is an example of pattern recognition.
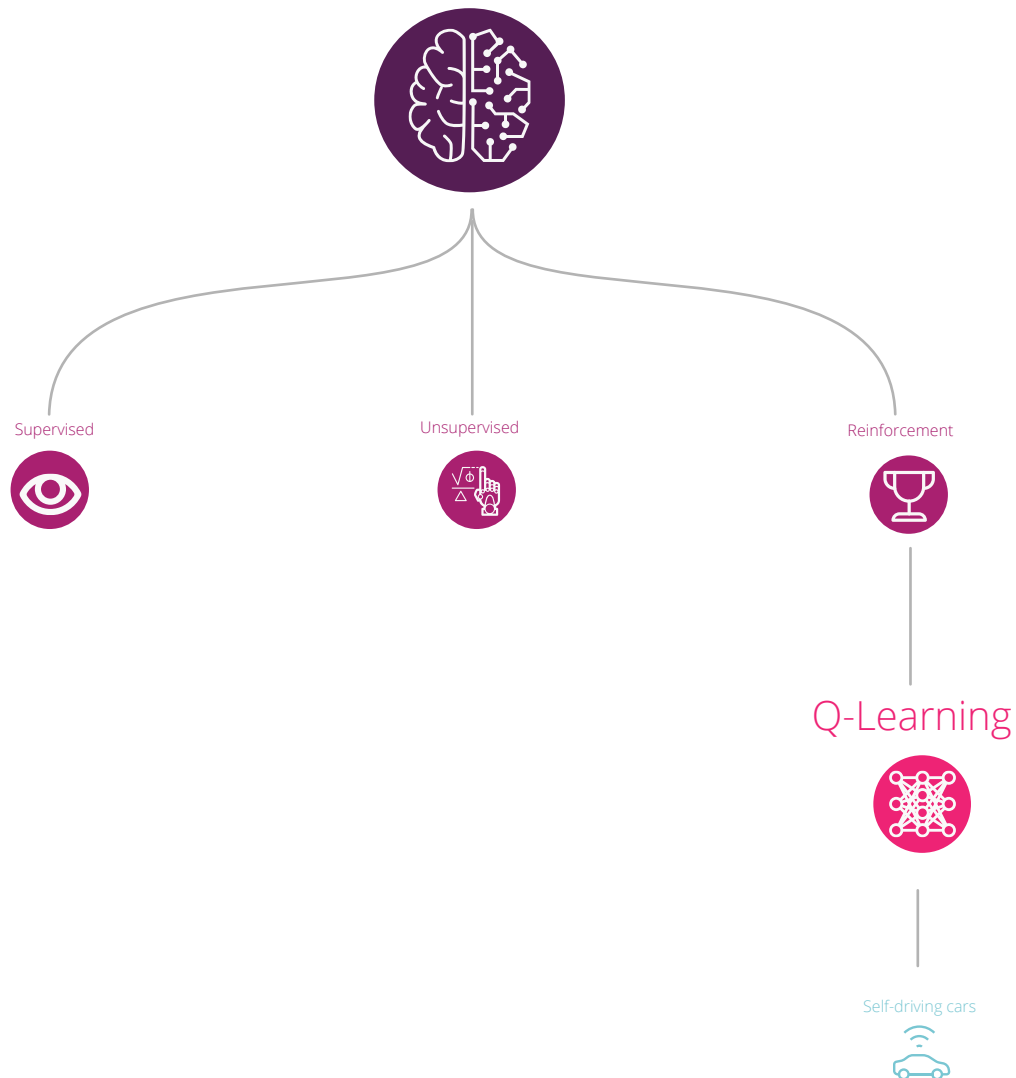
Machine Learning

Supervised

Unsupervised

Reinforcement

Cluster Analysis

Anomaly detection

Detecting intruders
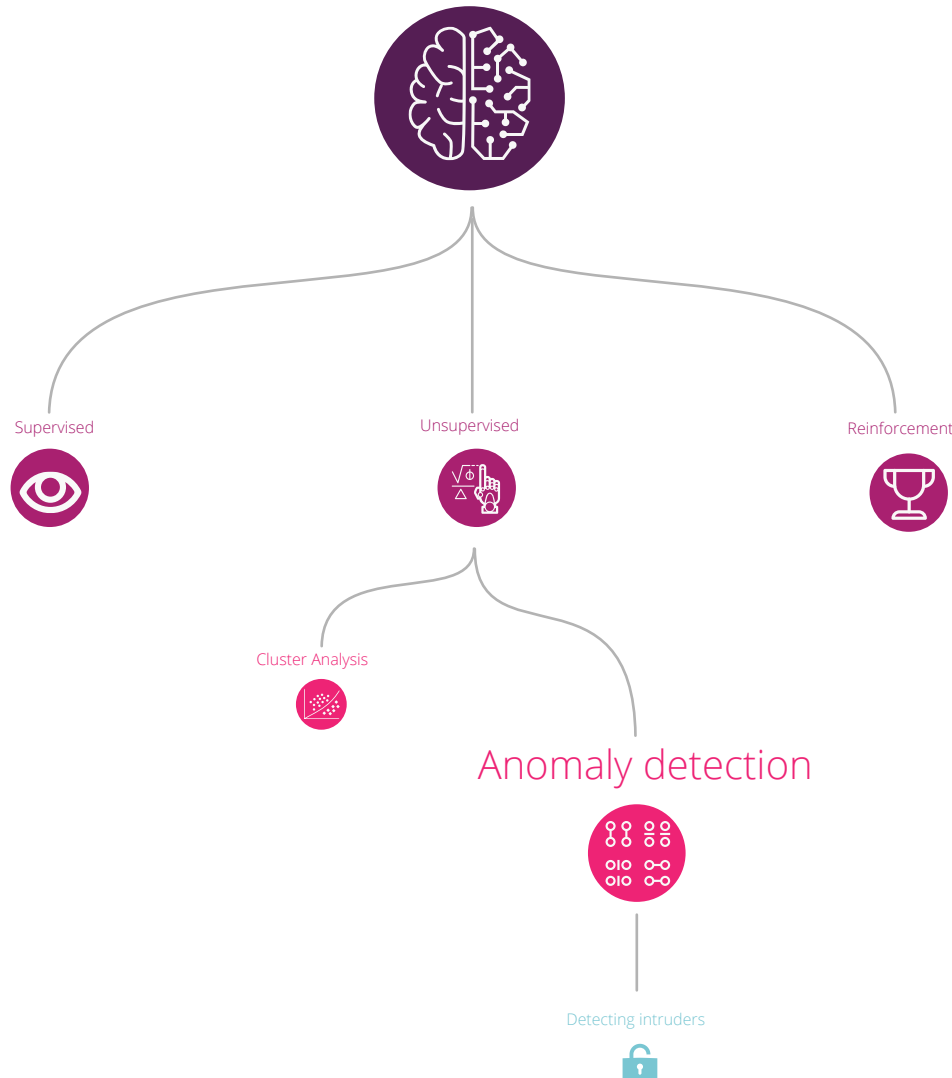
ANOMALY DETECTION

In data mining, anomaly detection (also outlier detection[1]) is the identification of rare items, events or observations which raise suspicions by differing significantly from the majority of the data.[1] Typically the anomalous items will translate to some kind of problem such as bank fraud, a structural defect, medical problems or errors in a text. Anomalies are also referred to as outliers, novelties, noise, deviations and exceptions. Three broad categories of anomaly detection techniques exist.[4] Unsupervised anomaly detection techniques detect anomalies in an unlabeled test data set under the assumption that the majority of the instances in the data set are normal by looking for instances that seem to fit least to the remainder of the data set.

Machine Learning

Supervised

Unsupervised

Reinforcement

Q-Learning

Self-driving cars

Q-LEARNING

Q-learning is a reinforcement learning technique used in machine learning. The goal of Q-Learning is to learn a policy, which tells an agent what action to take under what circumstances. It does not require a model of the environment and can handle problems with stochastic transitions and rewards, without requiring adaptations.

Machine Learning

Supervised

Unsupervised

Reinforcement

Types of learning

Decision tree

Neural network

Regression

Cluster Analysis

Anomaly detection

Q-Learning

Learning algorithms

Credit scoring

Speech recognition

Financial arbitrage

Labeling by behavior

Detecting intruders

Self-driving cars

Real-world uses

About

This website aims to give the interested reader a brief, but comprehensive overview of machine learning. Sources are from SAS, a statistical software company that sets the standard for statistical analytics, and from Wikipedia. Special thanks to the statistician Logan Lossing.