Machine Learning Techniques

Introduction

The improvements in technology have made easy for business to collect large amount of data in a variety, velocity, and characteristics [Sinar 2015]. Most organisations are now looking to use their data to create efficiencies, enter new market opportunities and create new products. As a result, many organisations have shifted from using their 'gut instinct-based decision making into adopting data driven strategic and tactful choices.[Sinar 2018]

The availability of low cost of computer processing power and lower processor and cost of memory, it is now practical to analyses large amount of data. As a result, Machine learning algorithms are made possible where large amount of data can be processed to learn and draw patterns and insights from the data.

Machine learning is an application of AI that enables systems to learn and improve from experience without being explicitly programmed. Machine learning focuses on developing computer programs that can access data and use it to learn for themselves.[Selig 2022]

Machine Learning Algorithms can be **broadly** classified into

- **Supervised**
- **Unsupervised Learning**

In **Supervised learning**, the algorithm has a target (labeled) variable which is predicted from a given sets of independent variables. Some examples of Supervised Learning are Regression ,Decision Tree, Random Forrest, K- Nearest Neighbor and Logistical Regression.

In **Unsupervised learning**, there is no target variables to predict and data from a given set is normally clustered into different groups based on certain criteria. Some examples are K-means clustering , Association , Recommendation Systems and Dimensionality Reduction.

Some of the Common Machine Learning Algorithms are

- Linear Regression
- Decision Tree
- K -Nearest Neighbor
- K- Means Clustering

Linear Regression

Linear regression is a supervised learning linear model that assumes a linear relationship between the input variables (x) and the single output variable (y) and that y can be calculated from a linear combination of the input variables (x) [Brownlee 2016] below. You can also use linear regression to provide better insights by uncovering patterns and relationships that your business colleagues might have previously seen and thought they already understood. For example, performing an analysis of sales and purchase data can help you uncover specific purchasing patterns on days or at certain times.

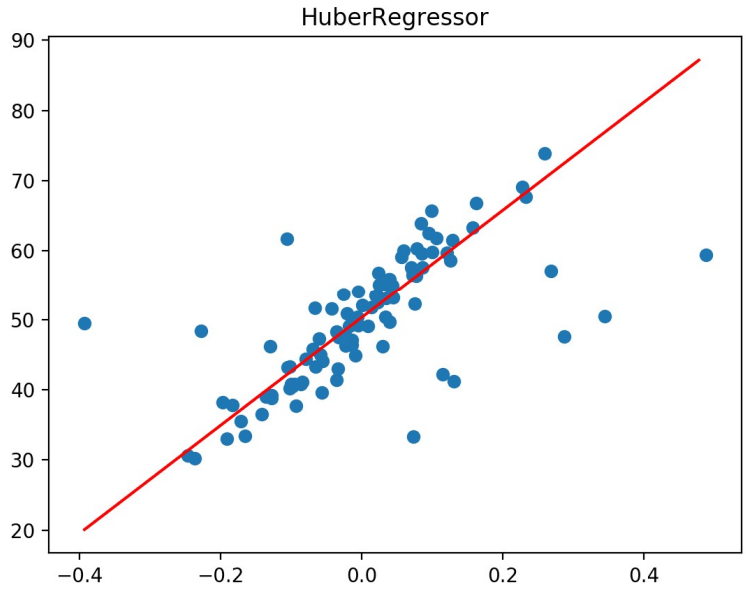

Image taken from Jason Brownlee on March 25, 2016 in Machine Learning Algorithms

Decision Tree Machine Learning

Decision trees are supervised machine learning method that provides solutions to the regression problems using the classification rule (starting from the root to the leaf node)(educba)

It has a tree-like model of decisions that is commonly used in classification and regression models. Each of the internal nodes represents the test on a feature (e.g., whether the age

of person is greater than a set age or not), each leaf node is used to represent the class label(results that need to be computed after taking all the decisions) and the branches represents conjunction of features that lead to the class labels. (educba). Below is an example of Decision Tree model.

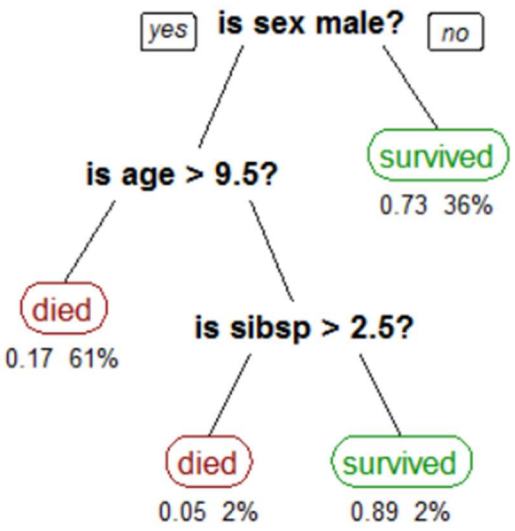

Image taken from wikipedia

K -Means Machine Learning Model

K-Means Clustering is an unsupervised machine learning algorithm that attempts to group observations into different clusters. (Yobero 2018). Kmeans Algorithm is an Iterative algorithm that divides a group of n datasets into k subgroups /clusters based on the similarity and their mean distance from the centroid of that subgroup/ formed.

K, here is the pre-defined number of clusters to be formed by the Algorithm. (Sharma 2021) If K=3, It means the number of clusters to be formed from the dataset is 3 as shown below:

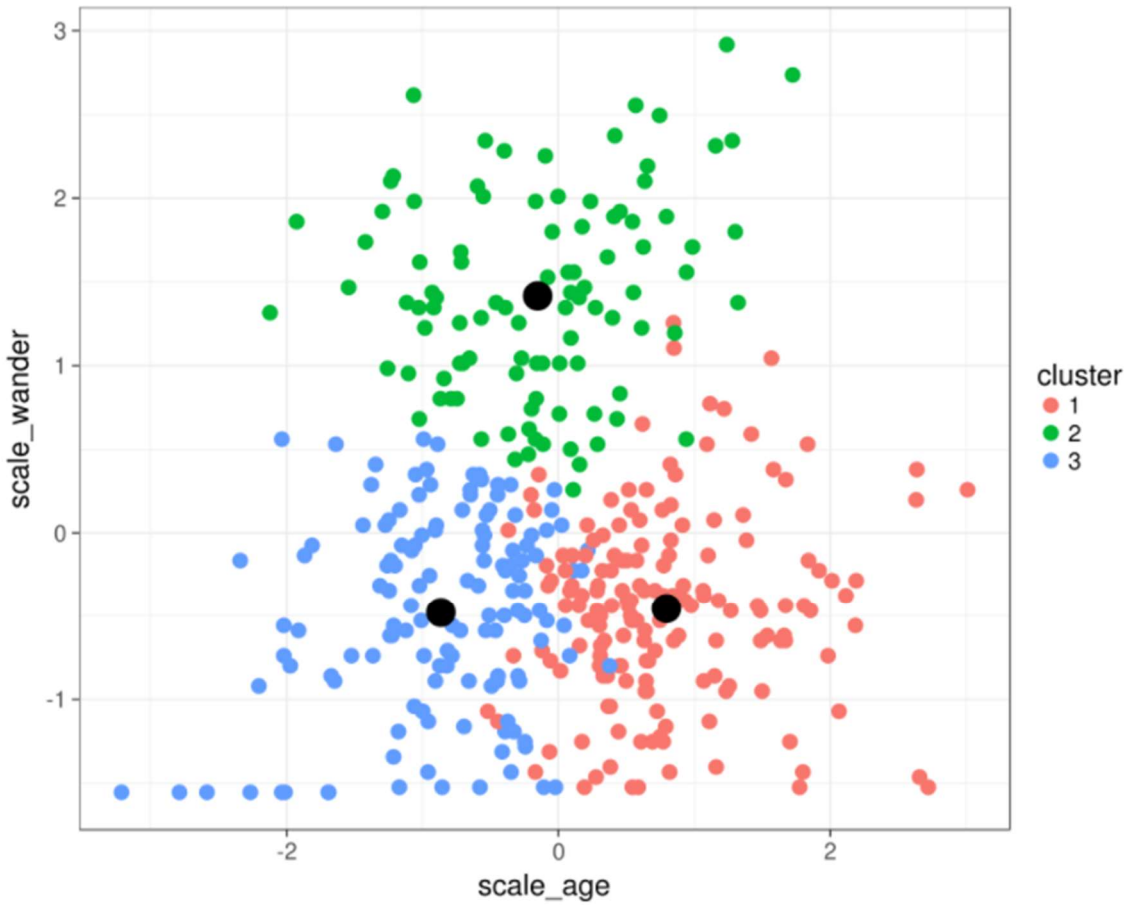

Image from- https://rpubs.com/cyobero/k-means

References

1) Sinar E Data Visualization Book: Big Data at work 2015 : https://www.taylorfrancis.com/chapters/edit/10.4324/9781315780504-11/data-visualization-evan-sinar
2) Sinar E Data Visualization: Get Visual to Drive HR's Impact and Influence 2018
3) Jay Selig - What is Machine Learning March 2022 https://www.expert.ai/blog/machine-learning-definition/
4) Swathi Sree Machine Learning with Python Tutorial November 2018
5) Brownlee J https://machinelearningmastery.com/linear-regression-for-machine-learning/#:~:text=Linear%20regression%20is%20a%20linear,the%20input%20variables%20(x). 2016
6) Decision Tree in Machine Learning | Split creation and Building a Tree (educba.com)
7) https://www.analyticsvidhya.com/blog/2021/04/k-means-clustering-simplified-in-python/
8) https://rpubs.com/cyobero/k-means