A Guide to Machine Learning: Understanding the Fundamentals and Techniques

Introduction

Machine Learning is one of the hottest topics in computing as of recent. With the rise of ChatGPT, Artificial Intelligence and Machine Learning are being researched, built, and deployed faster than ever before, and there is good reason for it. Machine Learning is one of the most promising, exciting and controversial fields in the modern era, with algorithms like Linear Regression that can predict a houses value in the future way better than humans and k-Means, which can be used in everything from Banking to Recommendation Engines. The best way to learn Machine Learning is to first learn what it hinges on, which are these algorithms and core concepts.

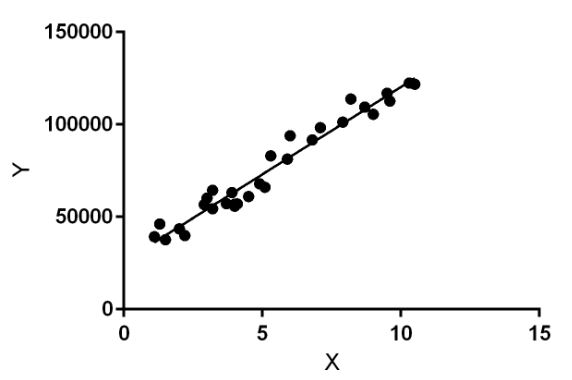
Supervised and Unsupervised Learning

The difference between supervised and unsupervised learning is that supervised learning is where the model is given specially curated data where the Supervisor/Trainer knows what is in the data and what the desired/expected output should be. Unsupervised learning on the other hand is where the model is fed data that doesn’t have labels, with the goal being that the model will uncover and detect the labels itself. Supervised Learning is mainly used for predicting outcomes based on the labeled data, whilst unsupervised learning is used for analysing, exploring, and detecting patterns in data.

Supervised/Unsupervised Learning Algorithms

**Linear Regression**

Linear Regression is a supervised learning algorithm used to predict a ‘target’ with one or more input features in the form of a linear equation. The goal is to find the line of best fit (in simple linear regression) or hyperplane (multiple linear regression) that describes the relationship between the inputs and targets.

Figure 1: https://media.geeksforgeeks.org/wp-content/uploads/linear-regression-plot.jpg

Simple Linear Regression Equation: y = mx + c

‘y’: Dependent variable (target)

‘x’: Independent Variable (feature)

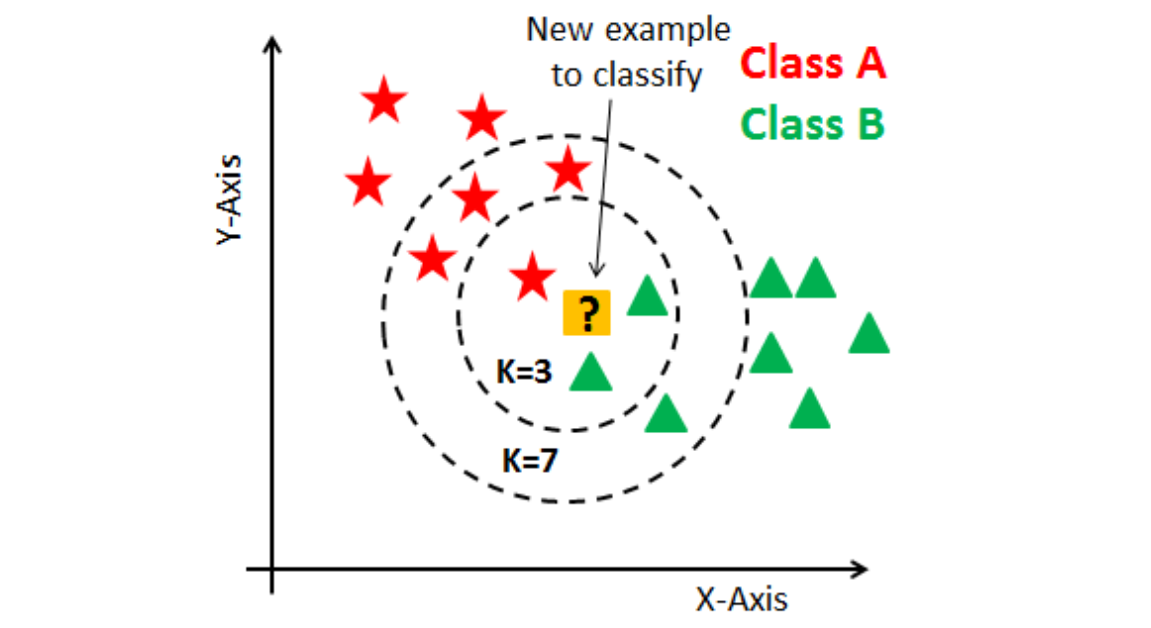
‘m’: Slope of the line

‘c’: y-intercept

Examples: Linear Regression can be applied to areas such as Real-Estate investment, where it can be used to predict a house price based on the size of the house. It is also used in Finance, where traders can see the relationship between a company’s stock price and it’s earnings.

**k-Nearest Neighbours (kNN)**

kNN is a supervised learning algorithm that classifies a data point based on it’s neighboring data points and how they are classified. For any given data point, it looks at the ‘k’ nearest data point and assigns a label to the data point based on the majority class or it computes a mean if it’s a regression task.

Figure 2: https://miro.medium.com/v2/resize:fit:1151/0\*ItVKiyx2F3ZU8zV5

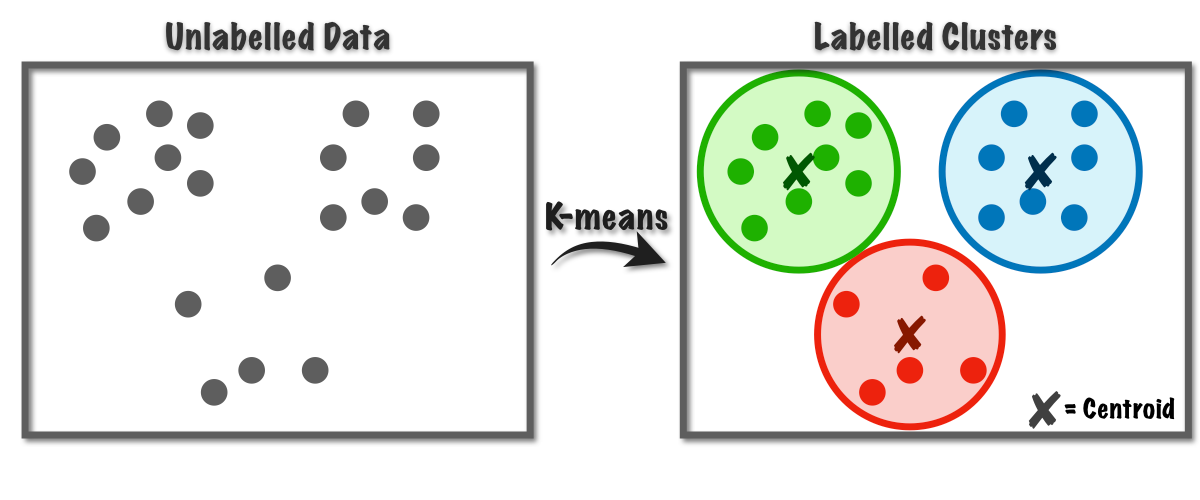
Procedure:

1. Compute the distance between the point to be predicted and all other points in the data set
2. Sort the distances and select the top ‘k’ points
3. Classification: Return the most common class between these top ‘k’ points  
   Regression: Return the average of the ‘k’ points

Examples: kNN can be used for purposes relating to classifying objects based on their similarities to other objects in a data set. An example would be classifying a fruit based on it’s colour, size and shape to other fruit in a dataset.

**K-Means**

k-Means is a unsupervised learning algorithm that is used to partition/divide a dataset into ‘k’ distinct, non-overlapping subsets (named clusters) based on their distance to the center of the cluster, which is commonly known as the ‘Centroid’. It is one of the simplest and most widely used Machine Learning algorithms for unsupervised learning.

Figure 3: https://miro.medium.com/v2/resize:fit:1200/1\*rw8IUza1dbffBhiA4i0GNQ.png

Procedure:

1. Randomly initialize ‘k’ centeroids
2. Assign each data point to the nearest centroid. These sets of points then form clusters
3. Recompute the centeroid of each cluster as the mean of all points in the cluster
4. Repeat the above steps until no significant change in the centeroid occurs

Example: One way that the k-Means algorithm is used is in the sorting of a customer base into different groups based on a behaviour like their purchasing history. It can also be used to detect SPAM emails and predict a score card on an exam.

Conclusion

By understanding the core concepts of Machine Learning, supervised and unsupervised learning, linear regression, kNN, k-Means, it allows the harnessing of the power Machine Learning as a whole provides. Grasping the fundamentals lays a solid foundation for learning the more advanced concepts and techniques, such as Deep Learning, Reinforcement Learning, Natural Language Processing, and Computer Vision

Bibliography

Gupta, M. (2023) *Linear regression in machine learning*, *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/ml-linear-regression/ (Accessed: 12 August 2023).

(LEDU), E.E. (2018) *Understanding K-means clustering in machine learning*, *Medium*. Available at: https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1 (Accessed: 12 August 2023).

Harrison, O. (2019) *Machine learning basics with the K-nearest neighbors algorithm*, *Medium*. Available at: https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761 (Accessed: 12 August 2023).