MACHINE LEARNING REPORT

Predictive Analytics & Machine Learning

Name: Drishti Chulani

TY BSc. In Actuarial Science and Quantitative Finance
Roll No: 710
Email : drishti.chulani@gmail.com

Date : 31st March 2022

Table of Contents

Abstract

In this report, we are supposed to give a deep understanding on Machine Learning. What are their techniques, the understanding of their handling, The pros and cons on ML. There will be difference between the type of models and the methodology for using image classification. The mathematics of the models will also be put in this report for clear understanding

Introduction

What is meant by Machine learning?

It is a process how a system or machine improves with experience. It's a type of AI that helps software apps to predict more accurate products/ output without any direct programming or explicitly told so.

It is a study of algorithms that can get more and more accurate results through experience and historical data. It helps find hidden patterns and solve problems, gain insights and make predictions.

Machine Learning can be thought of as the study of a list of sub-problems, meaning: decision making, clustering, classification, forecasting, deep-learning, inductive logic programming, support vector machines, reinforcement learning, similarity and metric learning, genetic algorithms, sparse dictionary learning, etc. There are various type of applications of Machine Learning like web search engine, spam detector, face detector in photos, etc.

The name machine learning was instituted in 1959 by Arthur Samuel. Tom M. Mitchell gave a generally cited, more conventional meaning of the calculations contemplated in the AI field: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

Machine Learning can be categorized into 4 categories

1) Supervised Learning
2) Unsupervised Learning
3) Reinforcement Learning
4) Semi-Supervised Learning

The explanation will be put in more detail in the further report.

Types of Machine Learning

1) Supervised Learning

It is used in prediction models where the output result is specified and the machine is supposed to follow a specific aim. With those specific details and parameters, the best prediction is achieved.

These algorithms work by describing the input data, hypothesized function and the expected results. By repeating the function's execution on the training data along with introduction of control parameters the model improves. It is considered a success if the predictions and the mapping are correct.

We have prediction of real estate prices, image classification, weather for casting, etc can be used as examples for Supervised Learning

2) Unsupervised Learning

It is a type of Machine Learning in which models are trained using unlabeled dataset and are allowed to act on that data without any supervision

This type of learning cannot be directly be put into a regression problem as there no specific output variable given.

Its aim is to basically find the underlying dataset structure that are difficult for humans to process without visualizing in higher dimensions. It works explicitly on unlabelled data with very less user involvement.

This Learning is divided into two categories of algorithm: clustering and association

Few real life examples can be said as finding customer sectors or different target group or reducing the difficult or features of a problem.

3) Reinforcement Learning

In this, machine Learning is basically attempting to find the best way to complete a goal or improve performance on a particular task. The algorithms work by choosing an action and observing the consequences of that action, with this it learns how optimal the result is. This is re-repeated over time and the right strategy is chosen.

Trial error search and delayed reward are important characteristics of reinforcement learning.

Video games, self driving cars, youtube ads are full of reinforcement learning.

4) Semi-supervised Learning

In this type of Machine Learning algorithm, it represents the midway ground between Supervised and Unsupervised Learning Algorithms. It includes both labeled and unlabeled datasets, but it consists more of unlabeled data.

It means the model can be trained to label data without to use as much labeled trained data.

Speech Analysis, text document or web content classification etc are examples for semi supervised learning.

How to choose which learning to use?

There are various questions to look at for decision making

1) if there is labelled expected output? And can we consider supervised learning

2) is the functional relationship between the input and output is clear?

3) if the data is complete and makes sense?

These are few questions that can help in choosing.

Supervised vs Unsupervised

Supervised Learning	Unsupervised Learning
These algorithms are trained using labeled data	These algorithms are trained using unlabeled data
This model takes immediate input for the output	Unsupervised Learning model takes no feedback or criticism
The input and output variables will be given	Only the input data will be given
The use training data to learn the best link between the input and outputs	They do not use output data
They require supervision to train the model	They do not require supervision to train the model
This can be categorized into two: Classification and regression	This can be classified into again two: Clustering and Association
This is highly accurate and trustworthy method	They are less accurate method
This includes algorithms like linear regression, logistic regression, decision trees etc	This includes algorithms like Clustering, K-Means, Hierarchical clustering etc.

Model for Supervised Learning

In this, I will take **Linear Regression**

This is a type of supervised Machine Learning algorithm that helps to get the best approx. linear fit to the points. Linear regression is, at its heart, a linear strategy to detecting the connection between two variables, one of which is dependent and the other independent. The goal is to comprehend how a change in one variable affects the other, resulting in a positive or negative connection.

The regression line is represented by a:
linear equation: Yi = a + bXi , Where a is the intercept and b is the Slope.

This approach is used when the projected output is continuous and has a constant slope, such as when estimating sales based on pricing. Risk assessment for specific assets, for example:
Linear Regression may be extended by including more and more input data to predict the output, but as the number of input variables increases, so does the model's complexity. To fit a multiple linear regression model, a computer programmer is often utilized.

Linear Regression has certain drawbacks. It fails when the dataset has a significant degree of multicollinearity (correlation between predictor variables) and when the data is insufficient. As a result, in such cases, ridge regression is utilized to generate a precise model.

Model for Unsupervised Learning

In this, I will take K-Means algorithm
The K-means algorithm searches for a predetermined number of clusters in an unlabeled multidimensional dataset and concludes by using a simple interpretation of how an optimum cluster may be stated.
The concept would mostly be divided into two phases.
1) To begin, the cluster center is the arithmetic mean (AM) of all data points connected with the cluster.
2) In compared to other cluster centers, each point is next to its cluster center.
The k-means clustering model is built on these two views.

You may think of the center as a data point that defines the cluster's means; however, it may or may not be a member of the dataset. In basic terms, k-means clustering allows us to cluster data into numerous groups by finding unique types of groups in unlabeled datasets without the need for data training. This is a centroid-based technique in which each cluster is connected to a centroid with the goal of minimizing the sum of distances between data points and their respective clusters.
As an input, the method consumes an unlabeled dataset, divides it into k clusters, and iterates the process until the proper clusters are found; the value of k should be preset.

Image Classification

It is a supervised Learning algorithm. It defines a set of target classes which is identified by images, trains them into models and differentiates them into label datasets.
It is basically labeling the images into already pre defined classes

Traffic control, facial recognition, brake light detection, etc are few ways they are used by companies to reduce labor.

Conclusion

There a variety of techniques to apply machine learning in real life. It can be understood that methods come in various flavors yet all share a point practically speaking: they generally comprise of math and stats procedures.
It is vital to comprehend the data before locking in which approach would work with the dataset and according to it, suitable algorithms must be considered to save time and effort while assisting to get more appropriate outputs.

Thank You!!