Supervised and Unsupervised Learning – A Comparative Study By

Manjusha Sreekumar
Asst.Professor , T.John College

1. INTRODUCTION

The exponential growth of the data in the current world is generating data that are of various types and structures. Each moment volumes of data are produced from various data sources. Some have a well-defined structure but majority fall under the unstructured category. Data sources include sensor data, videos, audios, image files, textual data etc. These data that burge in from these sources that include are staged, cleaned and stored in warehouses for analytics and visualization. Based on the data, sets, the system predicts. Machine learning is a powerful area and the objective is to comprehend the structure of information and fit that information into models that can be used by individuals. Supervised and Unsupervised Learning are two Machine learning techniques.

2. NEED OF DATA ANALYSIS

The world of big data is so huge that every moment tremendous bytes of data is flowing into web. The data mostly are unstructured like audio, video, mails, messages, images etc that doesn’t have any proper structure or format. Velocity, Veracity, Volume, Value and Variety are the main characteristics of these data that are flowing in from different data sources like web data, enterprise data, sensor data, relational data etc. Any business decision or prediction can only be done with the help of analysis of this data. The data collected are cleaned and staged in warehouses from which they are analysed by various OLAP (Online Analytical Processing) tools and later data is checked for pattern matching and then visualized.

Without analysis or proper study of data, no prediction or decision can be taken which are a part of any business decision. Study says, the more data is been able to analyse, the more accurate will be the decision. Each business organisation has their own way to collect the data and analyse them for future use.

3. SUPERVISED LEARNING

Supervised learning, as the name shows the nearness of an administrator as an instructor. Essentially, it is a learning where we instruct and train the machine utilizing information which is very much labelled and it implies that a few information is as of now tagged with the right answer. The machine is furnished with another arrangement of examples (data) and these data sets produces correct result from theses labelled data. Labelled data implies that a few information is as of now labelled with the right answer and It equals the learning that happens in the presence of a supervisor. This requires some investment and specialized ability from a group of profoundly talented information researchers.

The results of unforeseen data can now be predicted with supervised learning. Machine training algorithms are aptly designed to analyse the data and moreover accurately predict the outcome.

For example, to predict the time to take to your favorite restaurant, the input data can be
1. Traffic -time of the day
2. Weather condition
3. Vacation

Based on these input data, the output that is here, the time that will be taken can be predicted.

Types of Supervised Learning
Classification – It’s a learning technique that will classify data into defined classes. Each data value is studied and is classified into respective group. For each set of data received, the algorithm will train the system to fall into which category of classes.
Example: Depending on salary earned, the customer can be classified into different loan categories.
Regression – For a particular group of datasets, it can predict the product. In the presence of other variables, what is possibility of the presence of another variable. This method is mainly used for forecasting by establishing the relation between target and predicate. Example: Depending on criteria like location of the house, distance from approach road, proximity to essential requirements, the cost or rent of the house can be predicted.

4. UNSUPERVISED LEARNING

Unsupervised is the prediction by the systems without prior training or learning methodologies. These types of systems are presented with unlabelled, uncategorized data and the algorithms designed for prediction act on them without prior learning. It works on its own to discover information and predict. As unlabelled data are used, it is easier to collect from data sources than labelled data that are difficult to be structured. This type of learning will discover all types of patterns in the data than going forward with structured learning algorithms. Learning the data deeply is possible with unsupervised learning. These types of learning will involve the following methods.

Clustering - The data are grouped based on similar patterns-means clustering is a popular algorithm used in this method. A group of data of similar nature is called a cluster. First the data is grouped into clusters based on any similar attribute and then it is labelled. Example – Frequent flyer programs based on flying frequency is an example of this type. People who fly frequently are grouped together and promotions or discounts are given to them.

Association – This describes relations between data that is if X then Y can happen. It mines and finds out all the possible relations that can happen in a data set. Apropri Algorithm is widely used in establishing association rules. Market Analysis is widely done using this. Example – A person purchasing milk can also buy a packet of biscuit. A list of previous transactions is studied and possible combinations of milk are analysed. Artificial Intelligence widely uses Unsupervised Learning.

5 CONCLUSIONS

The invention of different types of learning algorithms have enabled deeper learning of data patterns and thereby better analysis of data. Labelled data will require supervised learning and unlabelled data will need unsupervised learning. Unsupervised learning is a natural process happening to predict data whereas Supervised learning is truly prediction of data in a trained environment. As more and more complex data is getting into the world of web, different training algorithms need to be introduced for better data analysis.

REFERENCES