Mini Project 2

Personal Report on Machine Learning

Unsupervised learning is a type of self-organized Hebbian learning that helps find previously unknown patterns in data set without pre-existing labels. It is also known as self-organization and allows modeling probability densities of given inputs. It is one of the main three categories of machine learning, along with supervised and reinforcement learning.

Current application of unsupervised learning

Unsupervised learning is more challenging than other strategies, due to the absence of labels. However, they are very significant in machine learning since they can do very complex tasks efficiently.

Clustering

Clustering is the process of grouping the given data into different clusters or groups. Unsupervised learning can be used to do clustering when we don’t know exactly the information about the clusters. Elements in a group or cluster should be as similar as possible and points in different groups should be as dissimilar as possible. Unsupervised learning can be used to do clustering when we don’t know exactly the information about the clusters.

For example, you can go to Walmart or a supermarket and see how different items are grouped and arranged there.

Also, e-commerce websites like Amazon uses clustering algorithms to implement the user specific recommendation system.

Dimensionality Reduction

Dimensionality reduction is the process of reducing the number of random variables under consideration by getting a set of principal variables.

Many machine learning problems contain thousands of features for each training instance. This will make the training slow as well as it will be difficult to obtain a good solution to the problem.

In dimensionality reduction, the objective is to simplify the data without losing too much information. There can be a lot of similar information in your data.

One way to do dimensionality reduction is to merge all those correlated features into one. This method is also called feature extraction.

It is always a good practice to try to reduce the dimensionality of your training data using an algorithm before you feed the data to another machine learning algorithm.

This will make the data less complex, much faster, the data may take up less memory and it will perform better at some times.

## Anomaly Detection

**Anomaly detection is the identification of rare items, events or observations which brings suspicions by differing significantly from the normal data.**

In this case, the system is trained with a lot of normal instances. So, when it sees an unusual instance, it can detect whether it is an anomaly or not.

One important example of this is credit card fraud detection. You might have heard about a lot of events related to credit card frauds.

Visualization

Visualization is the process of creating diagrams, images, graphs, charts, etc., to communicate some information. This method can be applied using unsupervised machine learning.