Web Data Mining

Assignment 2

Amal Joy

INT. MCA S9, Roll no. 7

Q: Compare and explain supervised and unsupervised learning with example.

A: Supervised learning and unsupervised learning are the 2 techniques of machine learning. But, both the techniques are used in different scenarios and with different data set.

Supervised learning

Supervised learning is the learning of the model with the input variable (x) and an output variable (y) and an algorithm to map the input to the output.

ie, $y = f(x)$

The basic aim is to appropriate the mapping function so well that when there is a new input data (x) then the corresponding output variable can be predicted.

Example-:

Suppose there is a basket which is filled with some fresh fruits, the task is to arrange the same type of fruits at one place. Also, suppose that the fruits are apple, banana, cherry, grape. Suppose one already knowns from their previous work that the shape of each and every fruit present in the basket. So, it is easy for them to arrange the same type of fruits in one place.

Here, the previous work is called as training data in data mining technology.

So, it learns the things, from the training data.

This is because, it has a response variable which says y that if some fruit

and so featured then it is grape, and similarly for each and every fruit.

This type of information is deciphered from the data that is used to train the model. This type of learning is called supervised learning. Such problems are listed under classical classification tasks.

Unsupervised learning

Unsupervised learning is where only the input data is present and no corresponding output variable is there. The main aim of unsupervised learning is to model the distribution in the data inorder to learn more about the data. It is called so, because there is no correct answer and there is no such teacher. Algorithms are left to their own devises to discover and present the interesting structure in the data.

Example -:

Again, Suppose there is a basket and it is filled with some fresh fruits, the task is to arrange the same type of fruits at one place.

This time there is no information about those fruits before hand, its the first time the fruits are being seen or discovered

Red color group: Bananas and grapes

So, now, take another physical character say, SIZE, so now the groups will be something like this

Red color and big size: apple

Red color and small size: cherry

Green color and big size: Bananas

Green color and small size: Grapes

The job is done.

- Explain an algorithm for each learning:

supervised learning algorithm

Random forest Algorithm

Step 1: Select random k data points from the training set.

Step 2: Build the decision trees associated with the selected data points

Step 3: Choose the number N for decision trees that you want to build

Step 4: Repeat step 1 and step 2

Step 5: For new data points, find the predictions of each decision tree and assign the new data points to the category that wins the majority votes.

Unsupervised learning algorithm

K-Means clustering

Step 1: Select the number k to decide the number of cluster

Step 2: Select random k points on centroids

Step 3: Assign each data points to their closed centeroid, which will form the predefined k clusters.

Step 4: Calculate the variance and place a new centeroid of each cluster.

Step 5: Repeat the third steps, which means reassign each data points to the new closest centroid of each cluster.

Step 6: If any reassignment occurs, then go to Step 4. else go to FINISH

Step 7 : The model is ready.