**Assignment\_8**

1. What exactly is a feature? Give an example to illustrate your point.

**Ans: Features are independent variables in ML. What is required to be learned in any specific machine learning problem is a set of there features , coefficients of these features, and parameters for coming with model.**

* **Ex - A model for predicting the risk of cardiac disease may have features such as the following:**
  + **Age**
  + **Gender**
  + **Weight**
  + **Whether the person smokes**
  + **Whether the person is suffering from diabetic disease, etc.**

2. What are the various circumstances in which feature construction is required?

**Ans: A machine learning models require feature to give the accurate output based on its learning. Thus, this is where feature construction is important part of machine learning.**

3.Describe how nominal variables are encoded.

Ans:

4. Describe how numeric features are converted to categorical features.

**Ans: Converting categorical feature into numeric feature using domain knowledge. For example, we are given a list of countries and say we know the distance to these countries from India then we can replace it with distance from India. So, every country can be represented as its distance from India.**

5. Describe the feature selection wrapper approach. State the advantages and disadvantages of this approach?

**Ans: In wrapper methods, the feature selection process is based on a specific machine learning algorithm that we are trying to fit on a given dataset.**

**Advantage: A fast algorithm for feature ranking is used to select relevant features. The performances of the model are evaluated and validated using the classification accuracy to compare existing approaches in the literature.**

6. When is a feature considered irrelevant? What can be said to quantify it?

7. When is a function considered redundant? What criteria are used to identify features that could be redundant?

**Ans: Duplicated features or information, that adds as a precaution against failure or error.**

**For ex – if two feature {X1, X2} are highly correlated, then the two features become redundant features since they have same information in tearms of correlation measure.**

8. What are the various distance measurements used to determine feature similarity?

**Ans: The various various distance measurements used to determine feature similarity are : Hamming distance, Euclidean Distance, Manhattan Distance**.

**Euclidean distance is the shortest path between source and destination which is a straight line.**

**Manhattan distance is sum of all the real distances between sources and destination and each are always the straight lines.**

9. State difference between Euclidean and Manhattan distances?

**Ans: Euclidean distance is the shortest path between source and destination which is a straight line.**

**Manhattan distance is sum of all the real distances between sources and destination and each are always the straight lines.**

10. Distinguish between feature transformation and feature selection.

**Ans: Feature transformation of data to improve the accuracy of the algorithm.**

**Feature selection is removing unnecessary features**

11. Make brief notes on any two of the following:

1.SVD (Standard Variable Diameter Diameter)

1. Collection of features using a hybrid approach

**Ans: It is classification in small sample size data sets. The filter step is based on instance learning taking advantage of small sample size of data.**

1. The width of the silhouette :

**The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters. The silhouette ranges from -1 to +1, where a high value indicates the object is well matched to its own cluster and poorly matched to neighboring clusters.**