Machine learning 08

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

Ans:- A feature is a distinctive trait or a special attraction. Feature can also mean to give special attention to something. The word feature has several other senses as a noun and a verb. A feature is a unique quality or characteristic that something has.

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

Ans:- This is sometimes done when certain numerical features cannot be directly used in the learning algorithm and are required to be first converted into dummy variables. Thus we first construct a feature from binning and then perform encoding on this newly constructed feature.

3. Describe how nominal variables are encoded.

Ans:- When we have a feature where variables are just names and there is no order or rank to this variable's feature. For example: City of person lives in, Gender of person, Marital Status, etc… In the above example, We do not have any order or rank, or sequence.

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

Ans:- At first thought, converting numeric data to categorical data seems like an easy problem. One simple approach would be to divide the raw source data into equal intervals. For example, for the data in the demo and Figure 2, the range is 78.0 - 60.0 = 18.0.

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. It follows a greedy search approach by evaluating all the possible combinations of features against the evaluation criterion

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

Ans:- One general definition for relevance is that a feature can be regarded as irrelevant if it is conditionally independent of the class labels or it does not influence the class labels; in these cases, it can be discarded.

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

Ans:- In engineering, redundancy is the intentional duplication of critical components or functions of a system with the goal of increasing reliability of the system, usually in the form of a backup or fail-safe, or to improve actual system performance, such as in the case of GNSS receivers, or multi-threaded computer .

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

Ans:- Perhaps four of the most commonly used distance measures in machine learning are as follows: Hamming Distance. Euclidean Distance. Manhattan Distance.

9. State difference between Euclidean and Manhattan distances?

Ans:- Euclidean distance is the shortest path between source and destination which is a straight line as shown in Figure 1.3. but Manhattan distance is sum of all the real distances between source(s) and destination(d) and each distance are always the straight lines

10. Distinguish between feature transformation and feature selection.

Ans:- The main difference:- Feature Extraction transforms an arbitrary data, such as text or images, into numerical features that is understood by machine learning algorithms. Feature Selection on the other hand is a machine learning technique applied on these (numerical) features.