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

In machine learning and pattern recognition, a feature is an individual measurable property or characteristic of a phenomenon. Choosing informative, discriminating and independent features is a crucial element of effective algorithms in pattern recognition, classification and regression.

The definition of a feature is a part of the face, a quality, a special attraction, article or a major film showing in the theatre. An example of feature is a nose. An example of feature is freckles. Feature is defined as to give or bring special attention to someone or something.

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

Top reasons to use feature selection are: It enables the machine learning algorithm to train faster. It reduces the complexity of a model and makes it easier to interpret. It improves the accuracy of a model if the right subset is chosen.

Feature subset selection is the process of identifying and removing as much of the irrelevant and redundant information as possible. This reduces the dimensionality of the data and allows learning algorithms to operate faster and more effectively.

3. Describe how nominal variables are encoded.

A column with nominal data has values that cannot be ordered in any meaningful way. Nominal data is most often one-hot (aka dummy) encoded, but there are many options that might perform better for machine learning. Rank. In contrast, ordinal data can be rank ordered.

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

-

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

n 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.

The filter method has the fastest running time; however, it does not consider feature dependencies and tends to each feature separately when univariate techniques are used . The wrapper method has the advantages of better generalization and robust interaction with the classifier used for feature selection

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

Feature Selection is the process where you automatically or manually select those features which contribute most to your prediction variable or output in which you are interested in. Having irrelevant features in your data can decrease the accuracy of the models and make your model learn based on irrelevant features.

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

If the same product gets entered twice by mistake, data redundancy takes place. The same retailer may keep customer files in a file storage system. If a customer purchases from the company more than once, their name may be entered multiple times. Duplicate entries of the customer name is considered redundant data.

Redundant features add no relevant information to your other features, because they are correlated or because they can be obtained by [linear] combination of other features. Having them on your set will not add anything, but it won't hurt either, information-wise.

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

Hamming Distance. Euclidean Distance. Manhattan Distance, Minkowski Distance

9. State difference between Euclidean and Manhattan distances?

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

but Manhattan distance is sum of all the real distances between source(s) and destination(d) and each distance are always the straight lines. Also The distance between two points measured along axes at right angles

10. Distinguish between feature transformation and feature selection.

feature transformation: transformation of data to improve the accuracy of the algorithm.

feature selection: removing unnecessary features

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

1.SVD (Standard Variable Diameter Diameter)

2. Collection of features using a hybrid approach

3. The width of the silhouette

4. Receiver operating characteristic curve