1. What is feature engineering, and how does it work? Explain the various aspects of feature engineering in depth.

Ans: **Feature engineering is used to select and transform features in a dataset in order to improve the performance of Machine Learning models. We perform several steps on the raw data and domain expertise is also used to understand the data and its features. In feature engineering, some of the basics steps are data exploration, feature selection, feature transformation, feature extraction and data imputation.**

1. What is feature selection, and how does it work? What is the aim of it? What are the various methods of function selection?

Ans: **feature selection is the process of selecting only those features which are correlated with the target variable and helps to understand the variance in the dependent variable. one of the methods for feature selection is the lasso regression in which the regression coefficient shrinks to zero causing the independent variable to become zero.**

1. Describe the function selection filter and wrapper approaches. State the pros and cons of each approach?

Ans: **in the filter approach, we find the correlation of the independent variable with the dependent variable and based on that we select the variable. In the wrapper approach, we train the model on different subset of features and choose that subset of features which gives the maximum accuracy. Filter is much faster compared to wrapper as it doesn’t require training the model.**

i. Describe the overall feature selection process.

Ans: **the two most common approaches used in feature selection is filter and wrapper. Feature selection process involves selecting only relevant features for model building for which the maximum efficiency of the model can be achieved. in the filter approach, we find the correlation of the independent variable with the dependent variable and based on that we select the variable. In the wrapper approach, we train the model on different subset of features and choose that subset of features which gives the maximum accuracy. Filter is much faster compared to wrapper as it doesn’t require training the model.**

ii. Explain the key underlying principle of feature extraction using an example. What are the most widely used function extraction algorithms?

Ans: **feature extraction involves transforming raw data into numerical features that can be processed while preserving the information in the original data set. It could yield better results than applying ML directly to the raw data. It also helps in reducing the amount of redundant data from the data set which further makes the algorithm to create more generalised model in less time. PCA can be used for feature extraction and reducing the dimensionality of the data. Some other methods for feature extraction are autoencoder, wavelet scattering.**

1. Describe the feature engineering process in the sense of a text categorization issue.

Ans: **for feature engineering, we can use different key words in the text such as count of emojis, type of emjois, frequencies of unique words, number of characters, number of words, number of capital words, count the number of punctuations these all steps can help use create more meaningful features and apply the ML algorithm to learn from the data better.**

6. What makes cosine similarity a good metric for text categorization? A document-term matrix has two rows with values of (2, 3, 2, 0, 2, 3, 3, 0, 1) and (2, 1, 0, 0, 3, 2, 1, 3, 1). Find the resemblance in cosine.

Ans: **Cosine measure the consine angle of two vectors and determines if they are roughly pointing in the same direction. That way we can see if two texts are similar or not. The resemblance between these two rows is 0.69**

7.

1. What is the formula for calculating Hamming distance? Between 10001011 and 11001111, calculate the Hamming gap.

Ans: **The Hamming distance formula is below:**

Text, letter

Description automatically generated

**And the distance between these two binary numbers is 2**

1. Compare the Jaccard index and similarity matching coefficient of two features with values (1, 1, 0, 0, 1, 0, 1, 1) and (1, 1, 0, 0, 0, 1, 1, 1), respectively (1, 0, 0, 1, 1, 0, 0, 1).

Ans:

Graphical user interface, text

Description automatically generated

8. State what is meant by "high-dimensional data set"? Could you offer a few real-life examples? What are the difficulties in using machine learning techniques on a data set with many dimensions? What can be done about it?

Ans: **dimensions are determined by the number of features used in model building. High dimensional data set could have more than 100 features which would need more than 100 dimensions to present the data. due to high number of dimensions, the data become sparse. Algorithms are hard to design in higher dimensions which is also called curse of dimensionality. We can employ different techniques to reduce the number of dimensions and one of them is PCA.**

9. Make a few quick notes on:

PCA is an acronym for Personal Computer Analysis. **PCA stands for principal component analysis which is used for dimensionality reduction.**

2. Use of vectors: **the vectors are used to represent data in the form of coordinates.**

3. Embedded technique: **Embedding technique can be used to transform the non numerical data into numerical data. Machine learning model expects its features to be numbers, not words or emails so that is why it is important to convert a non-numerical data into a numerical data. Embedding method attempts to reduce the dimensionality of data while preserving essential information in the data.**

10. Make a comparison between:

1. Sequential backward exclusion vs. sequential forward selection

Ans: **Sequential forward selection*(*SFS*),* in which features are sequentially added to an empty candidate set until the addition of further features does not decrease the criterion. Sequential backward selection*(*SBS*),* in which features are sequentially removed from a full candidate set until the removal of further features increase the criterion.**

2. Function selection methods: filter vs. wrapper

**in the filter approach, we find the correlation of the independent variable with the dependent variable and based on that we select the variable. In the wrapper approach, we train the model on different subset of features and choose that subset of features which gives the maximum accuracy. Filter is much faster compared to wrapper as it doesn’t require training the model.**

3. SMC vs. Jaccard coefficient:

**the SMC counts both mutual presences (when an attribute is present in both sets) and mutual absence (when an attribute is absent in both sets) as matches and compares it to the total number of attributes in the universe, whereas the Jaccard index only counts mutual presence as matches and compares it to the number of attributes that have been chosen by at least one of the two sets.**