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

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

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

4.

i. Describe the overall feature selection process.

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

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

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.

7.

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

ii. 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).

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?

9. Make a few quick notes on:

PCA is an acronym for Personal Computer Analysis.

2. Use of vectors

3. Embedded technique

10. Make a comparison between:

1. Sequential backward exclusion vs. sequential forward selection

2. Function selection methods: filter vs. wrapper

3. SMC vs. Jaccard coefficient