1. Feature engineering is the process of selecting, extracting, and transforming features from raw data to improve machine learning model performance. It involves various aspects such as feature selection, feature extraction, and feature transformation.

2. Feature selection is the process of selecting relevant features from a dataset to improve model performance and reduce overfitting. The aim is to reduce the number of features while retaining the most informative ones. Various methods of feature selection include filter methods, wrapper methods, and embedded methods.

3. The filter approach selects features based on statistical measures, such as correlation or mutual information. It is computationally efficient but may not consider the interactions between features. The wrapper approach selects features based on their impact on model performance, considering interactions between features, but is computationally expensive.

4. The feature selection process involves selecting relevant features, preprocessing the data, applying the selected features to the machine learning model, and evaluating the model's performance. Feature extraction involves transforming raw data into meaningful features using algorithms such as PCA, LDA, or t-SNE.

5. In text categorization, feature engineering involves transforming raw text data into numerical features, such as bag-of-words or TF-IDF, to represent the text for machine learning models.

6. Cosine similarity measures the similarity between two vectors' directions, making it a good metric for text categorization because it measures the similarity of text documents based on the direction of word frequencies. The cosine similarity between the two vectors is 0.791.

7. i. The Hamming distance is the number of different bits between two binary strings. The Hamming distance between 10001011 and 11001111 is 3.

ii. The Jaccard index measures the similarity between two sets, while the similarity matching coefficient measures the similarity between two binary vectors. The Jaccard index of the two features is 0.5, and the similarity matching coefficient is 0.625.

8. A high-dimensional dataset is a dataset with a large number of features or dimensions relative to the number of observations. Examples include image recognition, DNA sequencing, and social media data. The difficulties in using machine learning techniques on high-dimensional datasets include overfitting, curse of dimensionality, and computational complexity. Techniques such as dimensionality reduction, feature selection, and regularization can be used to address these issues.

9. PCA stands for Principal Component Analysis and is a technique for reducing the dimensionality of high-dimensional datasets by identifying the most important features. Vectors are used to represent data points in a high-dimensional space. Embedded techniques incorporate feature selection as part of the model training process.

10. Sequential backward exclusion and sequential forward selection are two feature selection methods that differ in the direction of feature selection. Filter and wrapper methods are two approaches to feature selection that differ in how features are evaluated. The Jaccard coefficient and similarity matching coefficient are two similarity measures that differ in how they measure similarity between binary vectors.