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

Feature engineering is a machine learning technique that leverages data to create new variables that aren't in the training set. It can produce new features for both supervised and unsupervised learning, with the goal of simplifying and speeding up data transformations while also enhancing model accuracy.

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

Feature selection is the process of reducing the number of input variables when developing a predictive model. It is desirable to reduce the number of input variables to both reduce the computational cost of modeling and, in some cases, to improve the performance of the model

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

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.

Filter methods measure the relevance of features by their correlation with dependent variable while wrapper methods measure the usefulness of a subset of feature by actually training a model on it. Filter methods are much faster compared to wrapper methods as they do not involve training the models.

The main differences between the filter and wrapper methods for feature selection are: Filter methods measure the relevance of features by their correlation with dependent variable while wrapper methods measure the usefulness of a subset of feature by actually training a model on it.

4.

1. Describe the overall feature selection process.

Feature Selection is the method of reducing the input variable to your model by using only relevant data and getting rid of noise in data. It is the process of automatically choosing relevant features for your machine learning model based on the type of problem you are trying to solve.

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

Denoising Autoencoder.

Variational Autoencoder.

Convolutional Autoencoder.

Sparse Autoencoder.

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

Feature engineering is the process of selecting, manipulating, and transforming raw data into features that can be used in supervised learning. In order to make machine learning work well on new tasks, it might be necessary to design and train better features.

1. 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.

The cosine similarity is beneficial because even if the two similar data objects are far apart by the Euclidean distance because of the size, they could still have a smaller angle between them. Smaller the angle, higher the similarity.

7.

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

To calculate the Hamming distance, you simply count the number of bits where two same-length messages differ. An example of Hamming distance 1 is the distance between 1101 and 1001 . If you increase the distance to 2 , we can give as an example 1001 and 1010 .

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?

High-dimensional data are defined as data in which the number of features (variables observed), p, are close to or larger than the number of observations (or data points), n. The opposite is low-dimensional data in which the number of observations, n, far outnumbers the number of features, p.

A data with more than 100 features

The curse of dimensionality basically means that the error increases with the increase in the number of features. It refers to the fact that algorithms are harder to design in high dimensions and often have a running time exponential in the dimensions.

One of the ways to reduce the impact of high dimensions is to use a different measure of distance in a space vector. One could explore the use of cosine similarity to replace Euclidean distance. Cosine similarity can have a lesser impact on data with higher dimensions.

9. Make a few quick notes on:

PCA is an acronym for Personal Computer Analysis.

Principal component analysis, or PCA, is a statistical procedure that allows you to summarize the information content in large data tables by means of a smaller set of “summary indices” that can be more easily visualized and analyzed.

2. Use of vectors

Most commonly in physics, vectors are used to represent displacement, velocity, and acceleration. Vectors are a combination of magnitude and direction and are drawn as arrows. The length represents the magnitude and the direction of that quantity is the direction in which the vector is pointing.

3. Embedded technique

The word embedding techniques are used to represent words mathematically. One Hot Encoding, TF-IDF, Word2Vec, FastText are frequently used Word Embedding methods. One of these techniques (in some cases several) is preferred and used according to the status, size and purpose of processing the data.

10. Make a comparison between:

1. Sequential backward exclusion vs. sequential forward selection

Forward selection starts with a (usually empty) set of variables and adds variables to it, until some stop- ping criterion is met. Similarly, backward selection starts with a (usually complete) set of variables and then excludes variables from that set, again, until some stopping criterion is met.

1. Function selection methods: filter vs. wrapper

Filter methods measure the relevance of features by their correlation with dependent variable while wrapper methods measure the usefulness of a subset of feature by actually training a model on it.

1. 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.