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

Answer:Feature engineering in ML consists of four main steps: Feature Creation, Transformations, Feature Extraction, and Feature Selection. Feature engineering consists of creation, transformation, extraction, and selection of features, also known as variables, that are most conducive to creating an accurate ML algorithm.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.Feature engineering is the process that takes raw data and transforms it into features that can be used to create a predictive model using machine learning or statistical modeling, such as deep learning.Feature can also mean to give special attention to something. The word feature has several other senses as a noun and a verb. A feature is a unique quality or characteristic that something has. Real-life examples: Elaborately colored tail feathers are peacocks' most well-known feature.

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

Answer: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.In the machine learning process, feature selection is used to make the process more accurate. It also increases the prediction power of the algorithms by selecting the most critical variables and eliminating the redundant and irrelevant ones. This is why feature selection is important.Feature selection is the process of isolating the most consistent, non-redundant, and relevant features to use in model construction. Methodically reducing the size of datasets is important as the size and variety of datasets continue to grow.

Saving...

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

Answer: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.They take into consideration the interaction of features like wrapper methods do. They are faster like filter methods. They are more accurate than filter methods. They find the feature subset for the algorithm being trained.A wrapper function is a function (another word for a subroutine) in a software library or a computer program whose main purpose is to call a second subroutine or a system call with little or no additional computation.

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?

Answer: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.Why is Feature Selection important? In the machine learning process, feature selection is used to make the process more accurate. It also increases the prediction power of the algorithms by selecting the most critical variables and eliminating the redundant and irrelevant ones.There are mainly two types of Feature Selection techniques, which are: Supervised Feature Selection technique. Supervised Feature selection techniques consider the target variable and can be used for the labelled dataset. Unsupervised Feature Selection technique. What are common feature selection methods? It can be used for feature selection by evaluating the Information gain of each variable in the context of the target variable. Chi-square Test. ... Fisher's Score. ... Correlation Coefficient. ... Dispersion ratio. ... Backward Feature Elimination. ... Recursive Feature Elimination. ... Random Forest Importance.

ii. Explain the key underlying principle of feature extraction using an example. What are the most widely used function extraction algorithms?
Answer:-->Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better resultsThough PCA is a very useful technique to extract only the important features but should be avoided for supervised algorithms as it completely hampers the data. If we still wish to go for Feature Extraction Technique then we should go for LDA insteadDenoising Autoencoder. Variational Autoencoder. Convolutional Autoencoder. Sparse Autoencoder. than applying machine learning directly to the raw data. The main difference between them is that feature selection is about selecting the subset of the original feature set, whereas feature extraction creates new features. Feature selection is a way of reducing the input variable for the model by using only relevant data in order to reduce overfitting in the model

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

Answer:Feature engineering is one of the most important steps in machine learning. It is the process of using domain knowledge of the data to create features that make machine learning algorithms work.Feature engineering in ML consists of four main steps: Feature Creation, Transformations, Feature Extraction, and Feature Selection. Feature engineering consists of creation, transformation, extraction, and selection of features, also known as variables, that are most conducive to creating an accurate ML algorithm.After getting to know your data through data summaries and visualizations, you might want to transform your variables further to make them more meaningful. This is known as feature processing. For example, say you have a variable that captures the date and time at which an event occurred.Feature engineering is the creation of features from raw data. Feature engineering includes: Determining required features for ML mode. Analysis for understanding statistics, distribution, implementing one hot encoding and imputation, and more.

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.

Cosine similarity measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. It is often used to measure document similarity in text analysis.

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).

Answer:-->1...While comparing two binary strings of equal length, Hamming distance is the number of bit positions in which the two bits are different. Hamming distance can also be calculated by taking XOR of two code words 11011001 ⊕ 10011101 = 01000100. Since, this contains two 1s, the Hamming distance, d(11011001, 10011101) = 2

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?

Answer:Data on health status of patients can be high-dimensional (100+ measured/recorded values for thousands of genes which is “high dimensional” data set. 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. High dimensional data is common in healthcare datasets where the number of features for a given individual can be massive (i.e. blood pressure, resting heart rate, immune system status, surgery history, height, weight, existing conditions, etc.)So high-dimensional data is generally going to be big data as well. The converse is not true - big data does not need many dimensions for you to learn from it. But if you are only working with a few dimensions, it's probably not as necessary to collect a large number of data points to do your analysis. 2..In machine learning classification problems, there are often too many factors on the basis of which the final classification is done. These factors are basically variables called features. The higher the number of features, the harder it gets to visualize the training set and then work on it.Dimensionality reduction is a machine learning (ML) or statistical technique of reducing the amount of random variables in a problem by obtaining a set of principal variables.

9. Make a few quick notes on:

PCA is an acronym for Personal Computer Analysis.

2. Use of vectors
3. Embedded technique Answer: 1.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.Principal component analysis (PCA) is the process of computing the principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest.PCA is an unsupervised machine learning algorithm that attempts to reduce the dimensionality (number of features) within a dataset while still retaining as much information as possible.

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.Navigating by air and by boat is generally done using vectors. Planes are given a vector to travel, and they use their speed to determine how far they need to go before turning or landing. Flight plans are made using a series of vectors. Sports instructions are based on using vectors.Vectors are used in science to describe anything that has both a direction and a magnitude. They are usually drawn as pointed arrows, the length of which represents the vector's magnitude

3.Emdedded Technique--->Embedded methods combine the qualities' of filter and wrapper methods. It's implemented by algorithms that have their own built-in feature selection methods. Some of the most popular examples of these methods are LASSO and RIDGE regression which have inbuilt penalization functions to reduce overfitting. Advantages of Embedded Methods: They take into consideration the interaction of features like wrapper methods do. They are faster like filter methods. They are more accurate than filter methods. They find the feature subset for the algorithm being trained. They are much less prone to over-fitting.