ML Assignment – 3

1. what are ensemble techniques in machine learning?

Ensemble techniques in machine learning involve combining multiple models to improve performance. One common ensemble technique is bagging, which uses bootstrap sampling to create multiple datasets from the original data and trains a model on each dataset.

2. Explain bagging and how it works in ensemble techniques.

Bagging is also known as bootstrap aggregration, is the ensemble learning method that is commonly used to reduce variance within a noisy data set. It uses the random sample of the data in a training set is selected with replacement meaning that the individual data points can be chosen more than once. it involves training multiple models independently and combining their predictions through averaging or voting.

3. What is the purpose of bootstrapping in bagging?

The purpose of bootstrapping in bagging is to create multiple diverse subsets of the original dataset by sampling with replacement. This technique helps to reduce variance and improve the stability and accuracy of machine learning models by averaging the predictions from multiple models trained on these different subsets.

4. Describe the random forest algorithm.

Random forest is a commonly -used machine learning algorithm, trademarks by Leo Breiman and Adele cutler, that combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems.

5. How does randomization reduce overfitting in random forests?

Randomization reduces overfitting in random forests by introducing two key elements:

Random Feature Selection: At each split in the decision trees, only a random subset of features is considered, which ensures that no single feature dominates the model and helps capture different aspects of the data.

Bootstrap Aggregation (Bagging): Multiple decision trees are trained on different random samples of the data, and their predictions are averaged, which reduces the model's variance and prevents it from fitting noise in the training data.

6. Explain the concept of feature bagging in random forests.

The random forest algorithm is a bagging algorithm, we draw random bootstrap samples from your training set. however, for the bootstrap samples, we also draw random sets of features for training the individual trees.

7. What is the role of decision trees in gradient boosting?

Gradient boosting decision trees are popular for solving prediction problems in classification and regression domains. The approach improves the learning process by simplifying the objective and reducing the number of iterations to reach a sufficiently optimal solution.

8. Differentiate between bagging and boosting.

Bagging:

-the simplest way of combining predictions that belong to the same type.

-aim to decrease variance, not bias.

-each model receives equal weight.

-each model is built independently.

-different training data subsets are randomly drawn with replacements from the entire training dataset.

Boosting:

-A way of combining predictions that belong to the different types.

-Aim to decrease bias, not variance.

-Models are weighted according to their performance.

-New models are influenced by the performance of previously built models.

-Every new subset contains the elements that were misclassified by previous models.

9. What is the AdaBoost algorithm, and how does it work?

Adaboost is also known as adaptive boosting is a machine learning approach that is utilised as an ensemble method. It is most commonly used estimator is decision trees with one level, which is decision trees with just one split. these trees were often referred to as decision stumps.

10.Explain the concept of weak learners in boosting algorithms.

Weak learners have low prediction accuracy, similar to random guessing. They are prone to overfitting- that is they can't classify data that varies too much from their original dataset.

11. Describe the process of adaptive boosting.

Adaptive boosting follows the metadata algorithms of machine learning.it is used for weak learners and is adaptive in the sense.it is weak for the instances that are misclassified by the previous classifiers. The properties of AdaBoost are sensitive to data that are noisy and outliers.

12. How does AdaBoost adjust weights for misclassified data points?

Adaboost is a boosting technique that iteratively corrects the errors of previous models by assigning higher weights to misclassified data points. By sequentially training weak learners on these weighted data points, adaboost constructs a strong ensemble model capable of making accurate predictions.

13. Discuss the XGBoost algorithm and its advantages over traditional gradient boosting.

XGBoost stands for Extreme Gradient Boosting.it is an advanced implementation of gradient boosting designed for speed and performance. It introduces several key enhancements over traditional gradient boosting methods. It is a machine learning algorithm under ensemble learning. It is a trend for supervised learning tasks, such as regression and classification. XGBoost builds a predictive model by combining the predictions of multiple individual models, often decision trees, in an iterative manner.

14. Explain the concept of organization in XGBoost.

An organization is a body built for a collection of individuals who join together to achieve some common goals and objectives bounded by legal entities. Organizations are often referred to as a company, institutions, associations, government bodies, etc. they follow certain legal procedures like business registration, tax identification, and maintaining corporate book records.

15. What are the different types of ensemble techniques?

The different types of ensemble techniques:

-bagging: bagging is the short form of bootstrap aggregating, and is mainly applied in classification and regression.

-Boosting: Boosting is an ensemble technique that learns previous predictor mistakes to make better predictions in the future.
-stacking: first in last out

16. Compare and contrast bagging and boosting.

Bagging:

-the simplest way of combining predictions that belong to the same type.

-aim to decrease variance, not bias.

-each model receives equal weight.

-each model is built independently.

-different training data subsets are randomly drawn with replacements from the entire training dataset.

Boosting:

-A way of combining predictions that belong to the different types.

-Aim to decrease bias, not variance.

-Models are weighted according to their performance.

-New models are influenced by the performance of previously built models.

-Every new subset contains the elements that were misclassified by previous models.

17. Discuss the concept of ensemble diversity.

Diversity is an important concept in ensemble learning and it refers to the idea that the individual models prediction in an ensemble should be as different from each other as possible. This is because different models are likely to make different types of errors.

18. How do ensemble techniques improve predictive performance?

Improving accuracy with Ensemble techniques Ensemble models can improve predictive accuracy by reducing variance, bias, or stacking. By training the same algorithm on different data subsets and then averaging. bagging decreases variance caused by data sampling.

19. Explain the concept of ensemble variance and bias.

Bias is simply defined as the inability of the model because of that there is some difference or error occurring between the model's predicted value and the actual value. These differences between actual or expected values and the predicted values are known as error or bias error or error due to bias. Bias is a systematic error that occurs due to wrong assumptions in the machine learning process

Variance is the measure of spread in data from its mean position. In machine learning variance is the amount by which the performance of a predictive model changes when it is trained on different subsets of the training data. More specifically, variance is the variability of the model that how much it is sensitive to another subset of the training dataset. i.e. how much it can adjust on the new subset of the training dataset.

20. Discuss the trade-off between bias and variance in ensemble learning.

Bias refers to the error introduced by approximating a real-world problem, which may be complex, by a much simpler model. High bias can cause the model to miss relevant relations between features and target outputs, leading to underfitting.

Variance refers to the error introduced by the model's sensitivity to small fluctuations in the training set. High variance can cause the model to model the random noise in the training data rather than the intended outputs, leading to overfitting.

21. What are some common applications of ensemble techniques?

Some common applications of ensemble techniques include Random forests, Boosting tree-models, gradient boosted tree-models and stacking.

22. How does ensemble learning contribute to model interpretability?

Ensemble learning helps improve machine learning results by combining several models and allows the production of better predictive performance compared to a single model.

23. Describe the process of stacking in ensemble learning.

Stacking is one of the most popular ensemble machine-learning techniques used to predict multiple nodes to build a new model and improve model performance. Stacking enables us to train multiple models to solve similar problems, and based on their combined output, it builds a new model with improved performance. We can make better predictions in the future. In stacking, an algorithm takes the outputs of sub-models as input and attempts to learn how to best combine the input predictions to make a better output prediction.

24. Discuss the role of meta-learners in stacking.

Stacking is a combination of several models of different types using the concept of a meta-learner. The first training data is used to train base learners at the first layer to generate a new dataset that is reused to train the second layer meta-learner. It is a process that helps models learn new and unseen tasks on their own. the meta-learning model gets better at solving a new and unseen task.

25. what are some challenges associated with ensemble techniques?

Some challenges associated with ensemble techniques include computational complexity, training time, computational cost, overfitting, model interpretability, data quality, parameter pruning, data preparation, Scalability, integration, implementation, storage requirements, and deployment complexity.

26. What is boosting, and how does it differ from bagging?

Bagging is the simplest way to combining predicitons that belong to the same type while boosting is a way of sombining predictions that belong to the different types. Bagging aims to decrease variance, not bias while boosting aims to decrease bias, not variance.

27. Explain the function behind boosting.

Boosting is a special type of ensemble learning technique that works by combining several weak learners into a strong learner. This works by each model paying attention to its predecessor's mistakes.

28. Describe the concept of sequential training in boosting.

In boosting, a random sample of data is selected, fitted with a model, and then trained sequentially that is each model tries to compensate for the weakness of its predecessor. With each iteration, the weak rules from each individual classifier are combined to form one, strong prediction rule.

29. How does boosting handle misclassified data points?

Boosting is a powerful machine learning technique where you combine multiple weak learners to create a strong learner. it focuses on the misclassified data points during each iteration, giving them some weight, and subsequently improves the model's accuracy.

30. Discuss the role of weights in boosting algorithms.

The boosting algorithm assesses model predictions and increases the weight of samples with a more significant error. It also assigns a weight based on model performance. A model that outputs excellent predictions will have a high amount of influence over the final decision.

31. what are the differences between boosting and AdaBoost?

Boosting:

- boosting is an ensemble modeling technique that attempts to build a strong classifier from the number of weak classifiers.

-boosting can improve the accuracy of the model by combining several weak models' accuracies and averaging them for regression over them for classification to increase the accuracy of the final model.

- boosting can reduce the risk of overfitting by reweighting the inputs that are classified wrongly.

Adaboost:

-AdaBoost is an ensemble method that trains and deploys trees in series. It implements boosting, wherein a set of weak classifiers is connected in series such that each weak classifier tries to improve the classification of samples that were misclassified by the previous weak classifier.

-Adaboost is best used to boost the performance of decision trees on binary classification problems.

-AdaBoost was originally called AdaBoost .it can be used to boost the performance of any machine learning algorithm.it is the best weak learners.

32. How does Adaboost adjust weights for misclassified samples?

Adaboost is also known as adaptive boosting as the weights are re-assigned to each instance, with higher weights assigned to incorrectly classified instances. It builds a model and gives equal weights to all the data points. It then assigns higher weights to points that are wrongly classified.

33. explain the concept of weak learners in boosting algorithm.

Weak learners in boosting algorithms have low prediction accuracy, similar to random guessing. They are prone to overfitting that is, they can’t classify data that varies too much from their original dataset.

34. discuss the concept of gradient boosting.

Gradient boosting is a type of machine learning boosting.it relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. The key idea is to set the target outcomes for this next model in order to minimize the error.

35. what is the purpose of gradient descent in gradient boosting?

The purpose of gradient descent in gradient boosting is to minimize the loss function by iteratively adding weak learners that correct the errors of the ensemble's previous iterations. Each new learner is trained to predict the residual errors of the combined model from the previous step, effectively reducing the overall prediction error.

36. describe the role of the learning rate in gradient boosting.

The learning rate in gradient boosting controls the contribution of each new weak learner to the overall model. It scales the corrections made by each subsequent learner, with lower learning rates making the model updates smaller and more gradual. This helps in preventing overfitting and allows for more refined convergence towards the optimal solution, although it may require more iterations to reach the same performance level.

37. how does gradient boosting handle overfitting?

Regularization techniques are used to reduce the overfitting effect, eliminating the degradation byn ensuring the fitting procedure is constrained. One popular regularization parameter is M, which denotes the number of iterations of gradient boosting.

38. discuss the difference between gradient boosting and XGBoost.

Gradient boosting:

-gradient boosting handles the categorical variables it may require preprocessing like on-hot encoding.

-scalability

-it memory usage.

-it has no regularization

- it has no parallel processing.

XG Boost:

- XGBoost doesn't handle the categorical variables.

-it allows the regularizations.

-it has accepted the parallel processing.

-it supports GPU.

39. explain the concept of regularized boosting.

Regularized boosting incorporates additional constraints and penalties into the boosting process to prevent overfitting. This typically involves techniques like L1 or L2 regularization, which penalize the complexity of the model, and constraints on tree structures, such as limiting depth or the number of leaves, ensuring the model remains generalizable.

40. what are the advantages of using XGBoost over traditional gradient boosting?

The advantages of using XGBoost over traditional gradient boosting.

- Flexibility

-regularization.

-Handling missing data

-speed and efficiency

-scalability

41. describe the process of early stopping in the gradient algorithm.

Early stopping is a form of regularization used in training iterative algorithms like gradient descent.it involves halting the training process when the validation error minimizes, thereby preventing themodel from learning the noise and idiosyncrasies in the training data.

42. How does early stopping prevent overfitting in boosting?

Early stopping is a technique in gradient boosting that allows us to find the optimal number of iterations required to build a model that regularizes well to unseen data and avoids overfitting.

43. Discuss the role of hyperparameters in boosting algorithms.

Hyperparameters govern the learning process of a GBM impacting its complexity, training time, and generalizability. Finetuning these parameters is crucial for optimal performance. We shall now use the tuning methods on the Titanic dataset.

44. what are some common challenges associated with boosting?

Some common challenges associated with boosting include:

-vulnerability to outlier data

-Real-time implementation.

45. explain the concept of boosting convergence.

Boosting convergence refers to the process by which a boosting algorithm iteratively improves its predictions and reduces errors over successive iterations. Each new weak learner added to the ensemble focuses on correcting the residual errors (or gradients) of the combined model from previous iterations. As the iterations progress, the overall model converges towards a more accurate solution, ideally achieving lower bias and variance, and improving predictive performance on the training and validation data.

46. How does boosting improve the performance of weak learners?

Boosting improves the performance of weak learners by sequentially adding them to the ensemble, where each new learner focuses on correcting the errors made by the previous ones. This iterative process enhances the overall model's accuracy by combining the strengths of multiple weak learners.

47. Discuss the impact of data imbalance on boosting algorithms.

Data imbalance can lead boosting algorithms to focus excessively on the majority class, resulting in poor performance on the minority class. This can cause the model to be biased and less effective at accurately predicting minority class instances.

48. what are some real-world applications of boosting?

Some real-world applications of boosting include: healthcare, IT, and finance.

49. Describe the process of ensemble selection in boosting.

Ensemble selection in boosting involves iteratively adding weak learners to the model based on their ability to correct the residual errors of the current ensemble. Each learner is selected and weighted to minimize the overall loss function, enhancing the combined model's performance.

50. How does boosting contribute to model interpretability?

Interpretability is used in the model while some ensemble methods can be like black box boosting algorithms often provide a more interpretable view of the decision-making process by combining simpler models.

51. Explain the curse of dimensionality and its impact on KNN.

KNN is susceptible to overfitting due to the curse of dimensionality. curse of dimensionality also describes the phenomenon where the feature space becomes increasingly sparse for an increasing number of dimensions of a fixed-size of training dataset.

52. what are the applications of KNN in real-world scenarios?

The applications of KNN in real-world scenarios include:

-forecasting the stock market.

-currency exchange rate.

-understanding and financial risks.

-trading futures

-credit rating

-loan management

-bank customer profiling.

-money laundering analysis.

53. Discuss the concept of weighted KNN.

The K-nearest neighbor modifies weighted KNN. One of the many issues that affect the KNN algorithm's performance is the hyperparameter's k choice. if k is too small, the algorithm would be more sensitive to outliers. If k is too large. Then the neighborhood may include too many points from other classes. Another issue is the approach to combining the class labels. The simplest method is to take the majority vote, but this can be a problem if the nearest neighbors vary widely in their distance and the closest neighbors more

Reliably indicates the class of the object.

54. How do you handle missing values in KNN?

In KNN, missing values can be handled by imputing them before applying the algorithm. Common imputation methods include filling missing values with the mean, median, or mode of the feature, or using more sophisticated techniques like k-nearest neighbors imputation, where the missing value is predicted based on the values of the nearest neighbors.

55. Explain the difference between lazy learning and eager learning algorithms, and where does KNN fit in.

A lazy learning algorithm takes a shorter time for training and a longer time for predicting. The eager training algorithm processes the data during the training phase only. An eager learning algorithm is faster than the lazy learning algorithm for predicting data observations.

56. what are some methods to improve the performance of KNN?

Some method to improve the performance of KNN include:

-feature scaling

-choosing optimal k.

-dimensionality reduction.

-weighted KNN

-Data cleaning

57. Can KNN be used for regression tasks? If yes, how?

The KNN algorithm can be used for both classification and regression. The KNN algorithm uses “feature similarity” to predict the values of any new data points. This means that the new points is assigned a value of based on how closely it resembles the points in the training dataset.

58. Describe the boundary decision made by the KNN algorithm.

It is a surface or hypersurface that determines the class label assigned to a data point based on the majority class of its k-nearest neighbors. The decision boundary essentially defines the region in the feature space where one class is favoured over others.

59. How do you choose the optimal value of K in KNN?

To choose the optimal value of K in KNN, use techniques like cross-validation to evaluate the model's performance for different values of K. Select the K that results in the lowest prediction error or highest accuracy on the validation set.

60. Discuss the trade-offs between using a small and large value of K in KNN.

A small value of K can make the KNN model sensitive to noise and outliers, leading to high variance. In contrast, a large value of K provides a smoother decision boundary and reduces variance but may lead to high bias, causing the model to overlook subtle patterns in the data.

61. Explain the process of feature scaling in the context of KNN.

Feature scaling in the context of KNN involves standardizing or normalizing the features so that they all have a similar scale. This ensures that no single feature disproportionately influences the distance calculations, leading to more accurate neighbor identification and improved model performance.

62. Compare and contract KNN with other classification algorithms like SVM and decision trees.

KNN is a simple, instance-based algorithm that relies on the distance between data points, making it sensitive to feature scaling and large datasets. SVM, on the other hand, constructs hyperplanes to separate classes and is effective for high-dimensional data. Decision trees build a model by splitting data based on feature values, providing interpretable results but can overfit if not pruned.

63. How does the choice of distance metric affect the performance of KNN?

The choice of distance metric significantly impacts KNN's performance by determining how distances between data points are calculated. Metrics like Euclidean distance are sensitive to feature scaling, while Manhattan distance is less affected. Choosing the appropriate metric depends on the dataset's characteristics and the underlying problem's requirements for accuracy and robustness.

64. What are some techniques to deal with imbalanced datasets in KNN?

Some techniques to deal with imbalanced datasets in KNN include:

- class weights

-resampling

-distance metrics

-ensemble methods.

65. Explain the concept of cross-validation in the context of tuning KNN parameters.

Cross-validation in the context of tuning KNN parameters involves dividing the dataset into multiple subsets (folds). It iteratively uses each fold as a validation set while using the rest for training. This process allows for evaluating the KNN model's performance across different parameter values (such as K) and helps select the optimal parameter setting that generalizes well to unseen data, reducing the risk of overfitting.

66. what is the difference between uniform and distance-weighted voting in KNN?

Uniform voting:

- uniform voting is that each neighbor contributes equally to the decision-making process regardless of its distance from the query point.

-this approach treats all neighbors equally in determining the class label of the query point.

Distance-weighted voting:

-the distance weighted voting is the neighbor's contribute to the decision-making process in proportion to their distance from the query point.

- closer neighbors have a greater influence, while more distant neighbors have less influence on the predicted class label.

67. Discuss the computational complexity of KNN.

The computational complexity of the KNN algorithm mainly depends on the size of the dataset, the number of features, and the value of k. The time complexity of the KNN algorithm for a single query point is O(nd), where n is the number of training examples and d is the number of features.

68. How does the choice of distance metric impact the sensitivity of KNN to outliers?

The choice of distance metric can impact how sensitive KNN is to outliers. Metrics like Euclidean distance are sensitive to outliers because they consider the overall spread of data. Using robust distance metrics, like Manhatten distance, can reduce sensitivity to outliers.

69. Explain the process of selecting an appropriate value for k using the elbow method.

This method is a visual technique used to determine the best k value for a k-means clustering algorithm. In this method, a graph known as the elbow graph plots the within-cluster sum of square values against various k values. The optimal k value is identified at the point where the graph bends like an elbow.

70. can KNN be used for the next classification tasks? If yes, how?

The KNN algorithm can be used for both the classification and regression problems. The KNN algorithm uses ‘feature similarity’ to predict the values of any new data points. This means that the new point is assigned a value based on how closely it resembles the points in the training set.

71. How do you decide the number of principal components to retain in PCA?

The sum of the eigenvalues turns out to be equal to the trace of the covariance matrix; therefore, the mean eigenvalue is equal to the trace divided by p. One procedure for deciding on the number of PCs to retain is to retain those for which the eigenvalue is greater than average that is grater than the negative of the lambda.

72. Explain the reconstruction error in the context of PCA.

The difference between the new matrix and the original data matrix is compared at each model run and gives an error. Calculated reconstruction error is examined by anomaly detection method, to expose at which level of error there is abnormal or not.

73. what are the applications of PCA in real-world scenarios?

The applications of principal component analysis in real-world scenarios include:

-data compression.

-feature extraction

-noise reduction

-data visualization

-anomaly detection.

74. Discuss the limitations of PCA.

-PCA assumes that the relationship between variables is linear. If the data is embedded on a nonlinear manifold, PCA produce wrong results.

-PCA is also sensitive to outliers, such data inputs could produce results that are very much of the correct projection of the data.

-interpretation

-partially interpret the model.

75. what is singular value decomposition? And how is it related to PCA?

Singular Value Decomposition (SVD) is a matrix factorization technique that decomposes a matrix into three components: U, Sigma, and V^T. It is related to Principal Component Analysis because PCA uses SVD to identify the principal components by decomposing the data covariance matrix, allowing for dimensionality reduction and feature extraction.

76. explain the concept of latent semantic analysis and its application in natural language processing.

Latent semantic analysis is used in natural language processing and information retrieval to anlayze word relationships in a large text corpus. It is a method for discovering the underlying structure of meaning within a collection of documents.

77. what are some alternatives to PCA for dimensionality reduction?

Some alternatives to PCA for dimensionality reduction include:

Linear discriminant analysis

Kernel PCA

Factor analysis

Independent component analysis

Autoencoders

Projection

Feature selection

Multidimensional scaling

78. Describe t-distributed Stochastic Neighbor Embedding and its advantages over PCA.

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a nonlinear dimensionality reduction technique that visualizes high-dimensional data by converting similarities between data points into joint probabilities and minimizing the Kullback-Leibler divergence between these probabilities in low-dimensional space. Compared to PCA, t-SNE better preserves local structures and is more effective for visualizing complex patterns in high-dimensional data.

79. How does t-SNE preserve local structure compared to PCA?

Local relationship preservation T-SNE excels at keeping nearby data points together in both high and low-dimensional spaces. Making it ideal for visualizing complex data.

80. Discuss the limitations of t-SNE.

Limitations of T-SNE include:

-computational complex

-non-deterministic

-requires hyperparameter tuning

-noisy patterns

81. what is the difference between PCA and independent component analysis?

PCA-

-Principal component analysis is a dimensionality reduction and machine learning method used to simplify a large data set into a smaller set while still maintaining significant patterns and trends

-it seeks to identify the largest contributors to variance that is magnitude of impact.

-‘signals’ maximize sum of variances of returns of each security within the universe

Independent component analysis:

- the independent analysis is a computational method for separating a multivariate signal into additive subcomponents. This is done by assuming that at most one subcomponent is gaussian and that the subcomponents are statistically independent from each other.

-it seeks to identify contributions to market behavior that are meaningful

-‘signals’ maximize independence, non-normality and /or complexity.

82. Explain the concept of manifold learning and its significance in dimensionality reduction.

Nonlinear dimensionality reduction are also known as manifold learning, is any of various related techniques that aim to protect high-dimensional data onto lower-dimensional latent manifolds, eith the goal of either visualizing the data in the low-dimensional space or learning the mapping.

83. what are autoencoders, and how are they used for dimensionality reduction?

An autoencoder is defined as an artificial neural network used for unsupervised learning of efiicient codings. The autoencoders are specific type of deep learning architecture used for learning representation of data, typically for the purpose of dimensionality reduction. An autoencoder reduces dimensionality by forcing the encoders to compress the input data into a smaller latent space. This means that the latent space has to capture the most important features or patterns in the data.while discarding the noise or redundancy.

84. discuss the challenges of using the nonlinear dimensionality reduction technique.

Nonlinear dimensionality reduction, also known as manifold learning, is any of various related techniques that aim to project high-dimensional data onto lower-dimensional latent manifolds, with the goal of either visualizing the data in the low-dimensional space or learning the mapping either from the high-dimensional space to the low-dimensional embedding or vice versa itself. The techniques described below can be understood as generalizations of linear decomposition methods used for dimensionality reduction, such as singular value decomposition and principal component analysis.

85. how does the choice of distance metric impact the performance of dimensionality reduction techniques?

The choice of distance metric in dimensionality reduction techniques impacts how similarities and dissimilarities between data points are measured, affecting the preservation of data structure. Appropriate metrics can enhance the algorithm's ability to capture relevant patterns, while inappropriate metrics may distort the data representation, leading to suboptimal reduction and visualization results.

86. what are some techniques to visualize high-dimensional data after dimensionality reduction?

Some techniques to visualize high-dimensionality data after dimensionality reduction:

-scatter plots

-heatmaps

-parallel coordinates

-biplots

87. explain the concept of feature hashing and its role in dimensionality reduction.

Feature hashing represents text documents of variable length as numeric feature vectors of equal length to reduce dimensionality. If you tried to use the text column for training as is it would be treated as a categorical feature column with many distinct values.

88. what is the difference between global and local feature extraction methods?

Global feature extraction methods:

-global features represent characteristics of an entire image as a whole.

-they capture information that describes the overall properties of the image.

-these features are typically computed by considering the entire image without focusing on specific regions.

Local feature extraction methods:

-local features capture information from specific regions of parts of an image.

- they describe distinctive patterns or structures within localized areas of the image.

-local features are often used for tasks like object detection, recognition, and matching.

89. how does feature sparsity affect the performance of dimensionality reduction techniques?

Sparsity denotes the no-value features that can be ignored while training a model. Moreover, auch featurea occur redundantly in the given dataset and poses issues while clustering similar features. To address this curse of dimensionality, dimensionality reduction is resorted.

90. discuss the impact of outliers on dimensionality reduction algorithms.

That can help improve the performance and interpretability of machine learning models, as well as reduce computational costs and storage requirements. However, dimensionality reduction methods can also be sensitive to outliers, which are data points that deviate significantly from the rest of the distribution.