1. d) Collinearity

   Dimensionality reduction techniques, such as Principal Component Analysis (PCA) and feature selection methods, can be used to reduce the collinearity or multicollinearity among variables. Collinearity refers to a high degree of correlation between predictor variables, which can cause issues in statistical models. By reducing collinearity, dimensionality reduction methods can improve the stability and interpretability of statistical models.

2. b) Random Forest

   Random Forest is the machine learning algorithm that is based upon the idea of bagging (Bootstrap Aggregating). Bagging is a technique where multiple models, typically decision trees, are trained on different subsets of the training data and then combined to make predictions. In the case of Random Forest, a collection of decision trees is trained on random subsets of the training data, and the final prediction is determined by aggregating the predictions of individual trees through majority voting (for classification) or averaging (for regression).

3. c) Decision Tree are prone to overfit

   Decision trees have several advantages, such as being easy to understand and interpret, handling both numerical and categorical data, and being able to handle missing values. However, they also have certain disadvantages, and one of them is their tendency to overfit the training data.
   Overfitting occurs when a decision tree becomes too complex and captures the noise or random variations in the training data, rather than the underlying patterns or relationships. This can result in a decision tree that performs well on the training data but generalizes poorly to new, unseen data.

4. c) Training data

   Training data refers to the labeled or annotated data that is used to train a machine learning algorithm. This data consists of input samples (features) and their corresponding known output or target values. During the training phase, the algorithm learns from the training data to create a model that can make predictions or classifications on new, unseen data.

5. c) Anamoly detection

Anomaly detection is the specific machine learning technique that helps in detecting outliers in data. Anomalies, also known as outliers, are data points that significantly deviate from the expected or normal behavior of the majority of the data. Anomaly detection algorithms are designed to identify such unusual patterns or instances in a dataset.

6. c) Case based

Case-based is not a specific numerical function representation in machine learning. It may refer to Case-based Reasoning (CBR), which is a problem-solving methodology that relies on past cases to solve new problems. CBR does not directly represent a numerical function but instead focuses on retrieving and adapting similar cases to solve new instances.

Therefore, the incorrect numerical function representation in machine learning is c) Case-based.

7. d) Both a and b

The analysis of machine learning algorithms requires both statistical learning theory and computational learning theory.

a) Statistical learning theory provides a framework for understanding the behavior and performance of machine learning algorithms. It involves studying the mathematical foundations of learning from data, such as understanding the assumptions, limitations, and guarantees associated with different learning algorithms. Statistical learning theory helps in analyzing the bias-variance tradeoff, generalization error, model complexity, and the convergence properties of machine learning algorithms.

b) Computational learning theory focuses on the computational aspects of machine learning. It deals with the efficiency and scalability of learning algorithms, the complexity of learning problems, and the sample complexity required to achieve a certain level of accuracy. Computational learning theory helps in understanding the time and space complexity of algorithms, optimization

techniques, and the tradeoff between computational resources and model performance.

Therefore, both a) Statistical learning theory and b) Computational learning theory are essential for the analysis of machine learning algorithms.

8. c) Both a and b

The difficulties with the k-nearest neighbor (KNN) algorithm include both the curse of dimensionality and the need to calculate the distance of the test case for all training cases.

a) Curse of dimensionality refers to the challenge that arises when working with high-dimensional data. As the number of dimensions increases, the available data becomes sparse, making it difficult to effectively measure distances and find meaningful nearest neighbors. The curse of dimensionality can lead to increased computational complexity and decreased performance of the KNN algorithm.

b) In the KNN algorithm, the distance between the test case and all the training cases needs to be calculated to determine the nearest neighbors. This can be computationally expensive, especially when dealing with large datasets or high-dimensional feature spaces. Calculating distances for each test case and training case pair can result in longer training and prediction times.

Therefore, the correct answer is c) Both a and b. The KNN algorithm is affected by the curse of dimensionality and requires calculating distances for all training cases when predicting the nearest neighbors.

9. b) 2

Radial Basis Function (RBF) neural networks typically consist of two types of

layers:

Input Layer: This layer receives the input features or variables of the dataset and passes them to the subsequent layers.

RBF Layer: The RBF layer is the main component of RBF neural networks. It consists of a set of neurons, each associated with a radial basis function. These radial basis functions, such as Gaussian or Multiquadric, are responsible for transforming the input data into a higher-dimensional feature space. The RBF layer computes the similarity or distance between each input pattern and the centers of the radial basis functions.

10. d) KMeans

KMeans is not a supervised learning algorithm. It is an unsupervised learning algorithm used for clustering data.

11. c) Neither feature nor number of groups is known

Unsupervised learning is a type of machine learning where the algorithm learns patterns or structures in data without any explicit knowledge of the features or the number of groups/classes in the data. In unsupervised learning, the algorithm is presented with a dataset that does not have labeled or pre-classified examples.

The goal of unsupervised learning is to discover hidden patterns, relationships, or structures in the data. This can include identifying clusters or groups of similar data points, detecting anomalies or outliers, dimensionality reduction, and data visualization.

12. b) SVG

SVG is not a recognized or commonly used machine learning algorithm. It is not an abbreviation commonly associated with any well-known machine learning algorithm or technique.

13. b) Underfitting

Underfitting is the scenario when the model fails to decipher the underlying trend or pattern in the input data. It occurs when a model is too simple or lacks the capacity to capture the complexity of the data.

When a model is underfitting, it typically has high bias and low variance. It fails to capture the important relationships or features in the data and may result in poor performance on both the training and test data.

14. a) Reinforcement learning

Real-time decisions, Game AI, Learning Tasks, Skill acquisition, and Robot Navigation are all applications commonly associated with reinforcement learning.

Reinforcement learning is a type of machine learning where an agent learns to make decisions or take actions in an environment to maximize a cumulative reward. It involves an agent interacting with an environment, receiving feedback in the form of rewards or penalties based on its actions, and adjusting its behavior through a trial-and-error learning process.

15. b) Mean squared error

The average squared difference between the classifier's predicted output and the actual output is indeed called the Mean Squared Error (MSE). The MSE is a commonly used metric in regression tasks to evaluate the performance of a model.

To compute the MSE, the difference between each predicted output and its corresponding actual output is squared, and then the average of these squared differences is calculated. The MSE provides a measure of the overall discrepancy between the predicted values and the true values

16. a) Linear, binary

Logistic regression is a linear regression technique that is used to model data having a binary outcome. It is a supervised learning algorithm primarily used for binary classification tasks, where the target variable or outcome variable has two possible classes or categories.

In logistic regression, the relationship between the input variables (predictors) and the probability of the binary outcome is modeled using a logistic function. The logistic function maps the linear combination of the input variables to a probability value between 0 and 1.

17. A. supervised learning

Classifying reviews of a new Netflix series based on whether they are positive, negative, or neutral is an example of supervised learning. In supervised learning, the algorithm learns from labeled training data where the input data (reviews) is associated with corresponding labels (positive, negative, or neutral). The goal is to train a model that can accurately predict the labels of new, unseen data based on the learned patterns from the labeled training data.

In this case, the algorithm would be trained on a dataset of reviews with their corresponding labels, and it would learn the patterns and features that differentiate positive, negative, and neutral reviews. Once trained, the model can be used to classify the reviews of a new Netflix series by predicting their sentiment based on the learned patterns.

18. C. both a and b

Both Euclidean distance and Manhattan distance are powerful distance metrics used by geometric models.

Euclidean distance, also known as straight-line distance, measures the shortest distance between two points in a straight line. It is calculated using the Pythagorean theorem in two or more dimensions.

Manhattan distance, also known as city block distance or taxicab distance, measures the distance between two points by summing the absolute differences of their coordinates along each dimension. It represents the distance a taxicab would travel when moving horizontally and vertically in a city grid.

19. D. none of these

Let’s review other options:

A. Removing columns which have too many missing values: This option refers to a data preprocessing step that eliminates columns with a high percentage of missing values. While this can be a necessary data cleaning step, it does not directly relate to dimensionality reduction, which aims to reduce the number of features or dimensions while retaining relevant information.

B. Removing columns which have high variance in data: Again, this option does not specifically relate to dimensionality reduction. Removing columns based on high variance may be relevant for certain tasks, such as feature selection, but it does not address the goal of reducing dimensions in the dataset.

C. Removing columns with dissimilar data trends: Similar to the previous options, removing columns based on dissimilar data trends does not directly align with dimensionality reduction techniques. It may be a consideration in some specific analyses or feature selection processes, but it does not encompass the broader goal of dimensionality reduction.

20. C. input attribute.

Both supervised learning and unsupervised clustering require input attributes to perform their respective tasks.

21. (A) SVM allows very low error in classification

In Support Vector Machines (SVM), the term "hard margin" refers to a scenario where the SVM algorithm aims to find a decision boundary that strictly separates the data points of different classes without allowing any misclassifications or errors. In other words, it seeks to achieve perfect classification with no points falling on the wrong side of the decision boundary.

The concept of "margin" in SVM refers to the distance between the decision boundary and the nearest data points from each class. In a hard margin SVM, the goal is to maximize this margin while ensuring that every training sample is correctly classified.

22. (B) Only 2

In Random Forest, increasing the depth of the trees (hyperparameter) can lead to overfitting. The depth of a tree represents the number of levels or splits it can have. When the depth is increased, the trees in the Random Forest become more complex and can potentially fit the training data too closely, resulting in overfitting.

23. (A) -(6/10 log(6/10) + 4/10 log(4/10))

Entropy is a measure of the impurity or uncertainty in a set of data. In the case of a binary target variable with two possible outcomes (0 and 1), the entropy can be calculated using the formula:

Entropy = -(p * log2(p) + q * log2(q))

where p is the proportion of occurrences of one outcome (in this case, 1), and q is the proportion of occurrences of the other outcome (in this case, 0).

Given the actual values of the target variable in the train file as [0, 0, 0, 0, 1, 1, 1, 1, 1, 1], we can calculate the proportions as follows:

p = 6/10 = 0.6 (proportion of 1s) q = 4/10 = 0.4 (proportion of 0s)

Substituting these values into the entropy formula, we get:

Entropy = -(0.6 * log2(0.6) + 0.4 * log2(0.4))

Therefore, the correct answer is (A) -(6/10 log(6/10) + 4/10 log(4/10)). This represents the entropy of the target variable based on the given actual values.

24. (A) weights are regularized with the l1 norm

Lasso (Least Absolute Shrinkage and Selection Operator) is a linear regression technique that incorporates regularization by imposing a penalty on the absolute values of the regression coefficients (weights). This penalty is applied to the sum of the absolute values of the coefficients, also known as the l1 norm.

By using the l1 norm regularization, Lasso encourages sparsity in the coefficient values, effectively shrinking some coefficients to zero and selecting only the most important features. This property makes Lasso useful for feature selection and can provide interpretable models.

25. (D) Perceptron

If the model that was trained on a linearly separable training set is a Perceptron, the learned decision boundary is likely to change if the new labeled data point is added to the training set and retraining is performed.

The Perceptron algorithm updates its decision boundary iteratively, adjusting the weights based on misclassified samples until all samples are correctly classified. Adding a new data point that is far away from the decision boundary may cause the decision boundary to shift or rotate, as the Perceptron tries to accommodate the new point and achieve a better overall classification performance.

26. (D) Either 2 or 3

When faced with multi-collinear features, there are multiple approaches to address the issue:

Option (1) "Both collinear variables should be removed" suggests removing both collinear variables. This can be a valid approach to eliminate redundancy and reduce the risk of overfitting. However, it can also result in the loss of potentially useful information if both variables individually contribute to the target variable.

Option (2) "Instead of deleting both variables, we can simply delete one" suggests removing one of the collinear variables. This approach allows retaining some information while reducing collinearity. By removing only one of the correlated variables, we can potentially preserve the predictive power of the other variable.

Option (3) "Removing correlated variables may result in information loss. We may utilize penalized regression models such as ridge or lasso regression to keep such variables" acknowledges that removing correlated variables can lead to information loss. Instead, using penalized regression models like ridge or lasso regression can be employed. These models can handle collinearity by introducing a penalty term that encourages variable shrinkage, effectively reducing the impact of correlated variables while keeping them in the model.

Therefore, the correct answer is (D) Either 2 or 3.

27. (B) increase by 5 pound

In the given least squares line equation: y = 120 + 5x, the coefficient of x (the height variable) is 5. This means that for every one-unit increase in the height (x), the weight (y) is expected to increase by 5 units.

Therefore, if the height is increased by one inch, the weight should increase by 5 pounds. This is in accordance with the coefficient value of 5 in the least squares line equation.The correct answer is (B) increase by 5 pounds.

28. (D) Minimize the squared distance from the points

The line described by the linear regression equation (OLS) attempts to minimize the squared distance between the predicted values on the line and the actual data points. It aims to find the line that best fits the data by minimizing the sum of the squared residuals, also known as the sum of squared errors.

By minimizing the squared distance from the points, the linear regression line aims to provide the best linear approximation to the relationship between the dependent variable and the independent variable(s) in the given data set. It seeks to minimize the overall prediction error and capture the general trend or pattern in the data.

Therefore, the correct answer is (D) Minimize the squared distance from the points.

29. (B) As the value of one attribute increases the value of the second attribute also increases

A correlation coefficient of 0.85 indicates a strong positive linear relationship between the two real-valued attributes. When the correlation coefficient is positive and close to 1, it suggests that as the value of one attribute increases, the value of the second attribute tends to increase as well. This means that there is a strong tendency for the two attributes to move in the same direction.

30. (B) Convolutional Neural Network

For an image identification problem, such as recognizing a dog in a photo, the most suitable neural network architecture is a Convolutional Neural Network (CNN). CNNs are specifically designed to process and analyze visual data, making them well-suited for image-related tasks.

CNNs excel at capturing and extracting meaningful features from images through the use of convolutional layers. These layers apply filters to the input image, allowing the network to detect patterns, edges, and textures at different scales and orientations. The hierarchical structure of CNNs enables them to learn complex representations and hierarchies of features, making them effective for image classification tasks.