41) Among the following identify the one in which dimensionality reduction reduces.

a) Performance
b) statistics
c) Entropy
d) Collinearity

Answer: d

Explanation:-

Collinearity refers to the situation where two or more variables in a multiple regression model are highly correlated. Dimensionality reduction techniques, such as Principal Component Analysis (PCA), can be used to address collinearity by transforming the original features into a set of uncorrelated variables (principal components), effectively reducing the dimensionality of the data.

42) Which of the following machine learning algorithm is based upon the idea of bagging?

a) Decision Tree

b) Random Forest

c) Classfication

d) SVM

Answer: b

Explanation:-

Random Forest is a machine learning algorithm based on the idea of bagging (Bootstrap Aggregating). In bagging, multiple instances of a base learning algorithm (in this case, decision trees) are trained on different subsets of the training data, and their predictions are aggregated to produce a final prediction. Random Forest introduces an additional level of randomness by selecting a random subset of features at each split in the decision tree, which helps improve the diversity of the individual trees and the overall performance of the model.

43) Choose a disadvantage of decision trees among the following.

a) Decision tree robust to outliers

b) Factor analysis

c) Decision Tree are prone to overfit

d) all of the above

Answer: c

Explanation:-

Decision trees can create complex structures that perfectly fit the training data but may fail to generalize well to new, unseen data, leading to overfitting. Techniques like pruning and setting limits on tree depth can be employed to mitigate overfitting in decision trees.

44) What is the term known as on which the machine learning algorithms build a model based on sample data?

a) Data Training
b) Sample Data
c) Training data
d) None of the above

Answer: c
Explanation:-
During the training phase, the algorithm learns patterns and relationships from the provided training dataset, which consists of input features and corresponding output labels or responses. The model is then expected to generalize well to new, unseen data based on what it learned during training

45) Which of the following machine learning techniques helps in detecting the outliers in data?

a) Clustering

b) Classification
c) Anamoly detection
d) All of the above

Answer: c
Explanation:-
Anomaly detection involves building models that can distinguish normal patterns from abnormal ones, making it suitable for identifying outliers or unusual instances in a dataset

46) Identify the incorrect numerical functions in the various function representation of machine learning.

a) Support Vector
b) Regression
c) Case based
d) Classification

Answer: a

Explanation:-

Among the given options, Support Vector is not a numerical function in the context of machine learning

47) Analysis of ML algorithm needs

a) Statistical learning theory
b) Computational learning theory
c) None of the above
d) Both a and b

Answer: d

Explanation:

a) Statistical learning theory: Focuses on understanding the properties and behavior of statistical methods for learning from data. It includes concepts like bias-variance tradeoff, model selection, and generalization.

b) Computational learning theory: Deals with the computational aspects of learning algorithms, including their efficiency, complexity, and computational feasibility.

48) Identify the difficulties with the k-nearest neighbor algorithm.

a) Curse of dimensionality
b) Calculate the distance of test case for all training cases
c) Both a and b
d) None

Answer: c
Explanation:-

The difficulties with the k-nearest neighbor algorithm are

a) a) Curse of dimensionality
b) Calculate the distance of test case for all training cases

49) The total types of the layer in radial basis function neural networks is _____
a) 1
b) 2
c) 3
d) 4

Answer: c
Explanation:-
The three types of layers in RBF neural networks are the Input Layer, Radial Basis Function (RBF) Layer, and Output Layer.

50) Which of the following is not a supervised learning
a) PCA
b) Naïve bayes
c) Linear regression
d) KMeans

Answer : d

Explanation:-

Supervised Learning:	
	Definition: In supervised learning, the algorithm is trained on a labeled dataset, where each input is associated with a corresponding output label.
Unsupervised Learning**:**	
	Definition: In unsupervised learning, the algorithm is given unlabeled data and is tasked with finding patterns, relationships, or structure in the data.