22MBABA304

USN

Third Semester MBA Degree Examination, June/July 2024
Exploratory Data Analysis for Business

Time: 3 hrs.

Max. Marks: 100

Note: 1. *Answer any FOUR full questions from Q.No.1 to Q.No.7.*
2. *Question No. 8 is compulsory.*
3. *M : Marks , L: Bloom's level , C: Course outcomes.*

			M	L	C
Q.1	a.	What is R and mention its uses in data mining tasks?	3	L1	CO1
	b.	Using the example of email spam filtering, describe the typical steps involved in a data mining process.	7	L2	CO1
	c.	Identify the importance of data visualization in exploratory data analysis and provide examples of tools used for visualizing relationships between two variables.	10	L3	CO1
Q.2	a.	Define Prediction error.	3	L1	CO2
	b.	Outline the holdout method for cross validation and its limitation.	7	L3	CO2
	c.	Evaluate different cross validation techniques to split the sample into multiple training and test data sets.	10	L4	CO2
Q.3	a.	What do you mean by linear regression?	3	L1	CO3
	b.	Explain the difference between point estimates and interval estimates in linear regression.	7	L3	CO3
	c.	Organize the advantages and disadvantages of different variables selection techniques like forward selection and backward selection.	10	L3	CO3
Q.4	a.	Define the concept of shrinkage in regression.	3	L1	CO4
	b.	Illustrate the concept of impurity in decision trees and how it is used to split the data.	7	L2	CO4
	c.	Construct the advantages and disadvantages of using bagging compared to a single decision tree.	10	L3	CO4
Q.5	a.	Define principle components.	3	L1	CO5
	b.	Explain the basic idea behind logistic regression for binary classification.	7	L3	CO5
	c.	Construct how singular value decomposition is used in principal component analysis.	10	L3	CO5
Q.6	a.	List the difference between linear and non-linear support vector machine.	3	L1	CO6
	b.	Apply the basic idea behind using Kernel functions in non-linear support vector machine.	7	L3	CO6
	c.	Choose the concept of support vector in linear SVMS and its role in classification.	10	L3	CO6
Q.7	a.	Define Exploratory data Analysis.	3	L1	CO1
	b.	Explain Training and Testing Error as a function of model complexity.	7	L3	CO2
	c.	Explain the different variable selection methods applicable in context of linear regression.	10	L3	CO3

Q.8		Case Study:			
		Principal Component Analysis (PCA) application in Marketing Research.			
		Case Scenario : ABC corporation is a leading manufacturer of electronic gadgets. The marketing team at ABC Corporation is constantly striving to understand consumer preferences and market trends to develop better products and targeted marketing strategies.			
		As part of their market research efforts, they have collected data on various attributes related to consumer behavior and preferences.			
		Data collected : The data set includes information on the following variables :			
		(i) Age			
		(ii) Gender			
		(iii) Income			
		(iv) Education level			
		(v) Marital status			
		(vi) Bland loyalty score			
		(vii) Purchase frequency			
		(viii) Product preferences			
		The marketing team wants to gain insights into consumer segmentation and identify pattern in consumer behavior that can guide product development and marketing strategies. They have decided to use Principal Component Analysis (PCA) to achieve this objective.			
		Questions :			
	a.	How to perform PCA on the dataset to reduce the dimensionality of the data?	10	L4	CO3
	b.	How can the insights obtained from PCA help in devising effective marketing strategies?	10	L4	CO3

* * * * *