**QUESTION 1**

1. Which of the following statements is true concerning linear discriminant analysis?

|  |  |  |
| --- | --- | --- |
|  | A. | For predictors, either a continuous or discrete type variables can be used. |
|  | B. | Equal variances are not required across groups. |
|  | C. | Predictor variables are assumed to be normally distributed. |
|  | D. | None of the above. |

**2 points**

**QUESTION 2**

1. Which of the following statements is true concerning classification and regression tree (CART)?

|  |  |  |
| --- | --- | --- |
|  | A. | For classifiers (predictors), either continuous or discrete variable can be used. |
|  | B. | Best attribute for a tree model is the one that yields the largest tree. |
|  | C. | It grows each tree on an independent bootstrap sample from the training data. |
|  | D. | None of the above. |

**2 points**

**QUESTION 3**

1. If you run a linear discriminant analysis with ‘age’, ‘duration’ and ‘previous’, what is the coefficient of linear discriminant for ‘previous’?

|  |  |  |
| --- | --- | --- |
|  | A. | [2.1, 2.2] |
|  | B. | [1.3, 1.4] |
|  | C. | [1.1, 1.2] |
|  | D. | None of the above |

**2 points**

**QUESTION 4**

1. When you build a linear discriminant model using the training dataset and predict using the validation dataset, what is the predictive hit rate (accuracy) and the Jaccard similarity index (accuracy measure that discards true negatives)?

|  |  |  |
| --- | --- | --- |
|  | A. | [0.88, 0.90] and [0.25, 0.26] |
|  | B. | [0.87, 0.88] and [0.35, 0.36] |
|  | C. | [0.25, 0.26] and [0.88, 0.90] |
|  | D. | None of the above |

**2 points**

**QUESTION 5**

1. If you consider a Naïve Bayes classifier, what is the a priori likelihood of classification membership (i.e., prior proportion of membership before any data information is added)?

|  |  |  |
| --- | --- | --- |
|  | A. | [0.87, 0.88] and [0.11, 0.12] |
|  | B. | [0.88, 0.89] and [0.11, 0.12] |
|  | C. | [0.88, 0.89] and [0.10, 0.11] |
|  | D. | None of the above |

**2 points**

**QUESTION 6**

1. When you build a Naïve Bayes Classifier with ‘age’, ‘duration’ and ‘previous’ on the training dataset and predict using the validation dataset, what is the predictive hit rate (accuracy) and the Jaccard similarity index?

|  |  |  |
| --- | --- | --- |
|  |  | [0.88, 0.90] and [0.25, 0.26] |
|  |  | [0.88, 0.90] and [0.28, 0.30] |
|  |  | [0.25, 0.26] and [0.88, 0.90] |
|  |  | None of the above |

**2 points**

**QUESTION 7**

1. If you run a CART model with ‘age’, ‘duration’, and ‘previous’ on the training dataset, what will be the most important variable?

|  |  |  |
| --- | --- | --- |
|  | A. | Age |
|  | B. | Previous |
|  | C. | Duration |
|  | D. | None of the above |

**2 points**

**QUESTION 8**

1. What is the first node-split criteria for a CART model with ‘age’, ‘duration’, and ‘previous’ on the training dataset?

|  |  |  |
| --- | --- | --- |
|  | A. | Duration < 524.5 |
|  | B. | Previous = 1 |
|  | C. | Age > 40.5 |
|  | D. | None of the above |

**2 points**

**QUESTION 9**

1. When you build a CART model with ‘age’, ‘duration’, and ‘previous’ on the training dataset and predict using the validation dataset, what is the predictive hit rate and the Jaccard similarity index?

|  |  |  |
| --- | --- | --- |
|  | A. | [0.92, 0.93] and [0.23, 0.24] |
|  | B. | [0.87, 0.88] and [0.28, 0.30] |
|  | C. | [0.89, 0.91] and [0.23, 0.24] |
|  | D. | None of the above |

**2 points**

**QUESTION 10**

**#Rando**

1. If you run a Random Forest model with ‘age’, ‘marital’, ‘housing’, ‘duration’, and ‘previous’ using training dataset by creating 1000 trees, what is the Jaccard similarity index upon prediction?

|  |  |  |
| --- | --- | --- |
|  | A. | [0.15, 0.2] |
|  | B. | [0.23, 0.28] |
|  | C. | [0.30, 0.35] |
|  | D. | None of the above |

**2 points**

**QUESTION 11**

1. If you run a Random Forest model with ‘age’, ‘duration’, and ‘previous’ using training dataset by creating 500 trees, what is the Jaccard similarity index upon prediction?

|  |  |  |
| --- | --- | --- |
|  | A. | [0.14, 0.18] |
|  | B. | [0.22, 0.28] |
|  | C. | [0.20, 0.21] |
|  | D. | None of the above |

**2 points**

**QUESTION 12\_reg**

1. If you use factor analysis to try and reduce dimensions (without oblique rotation) for ratings on Q1 – Q22, what proportion of the variance can be explained by a three factor model?

|  |  |  |
| --- | --- | --- |
|  | A. | [88, 90] |
|  | B. | [65, 67] |
|  | C. | [68, 71] |
|  | D. | None of the above |

**2 points**

**QUESTION 13**

1. In a three factor model for Q1 – Q22 (without oblique rotation) what are the three questions that load onto the 3rd factor that likely captures “Tangibles” dimension?

|  |  |  |
| --- | --- | --- |
|  | A. | Q11, Q12, Q14 |
|  | B. | Q11, Q12, Q1 |
|  | C. | Q11, Q1, Q3 |
|  | D. | None of the above |

**2 points**

**QUESTION 14**

1. In a three factor model for Q1 - Q22 (without oblique rotation), what can the 2nd factor be representative of in terms of the five potential dimensions of the survey?

|  |  |  |
| --- | --- | --- |
|  | A. | Assurance |
|  | B. | Reliability |
|  | C. | Empathy |
|  | D. | None of the above |

**2 points**

**QUESTION 15**

1. In a three factor model with oblique rotation for Q1 – Q22, what percentage of the total variation can be explained by the three factors?

|  |  |  |
| --- | --- | --- |
|  | A. | [0.60, 0.63] |
|  | B. | [0.70, 0.73] |
|  | C. | [0.65, 0.68] |
|  | D. | None of the above |

**2 points**

**QUESTION 16**

1. In a three factor model with oblique rotation for Q1 – Q22, which two factors have the greatest correlation?

|  |  |  |
| --- | --- | --- |
|  | A. | F1 and F2 |
|  | B. | F1 and F3 |
|  | C. | F2 and F3 |
|  | D. | None of the above |

**2 points**

**QUESTION 17**

1. What can you say about the regression of servqual (dependent variable) on the three orthogonal (i.e. not correlated) factors obtained using a three factor model?

|  |  |  |
| --- | --- | --- |
|  | A. | All factors have a negative and significant coefficient |
|  | B. | Not all factors are statistically significant |
|  | C. | All factors have a positive and significant coefficient |
|  | D. | None of the above |

**2 points**

**QUESTION 18**

1. What can you say about the regression of servqual (dependent variable) on the three correlated factor scores (with oblique rotation) obtained using a three factor model?

|  |  |  |
| --- | --- | --- |
|  | A. | All factors have a negative and significant coefficient. |
|  | B. | Not all factors are statistically significant. |
|  | C. | All factors have a positive and significant coefficient. |
|  | D. | None of the above |

**2 points**

**QUESTION 19**

1. Which of the following is true about factor analysis?

|  |  |  |
| --- | --- | --- |
|  | A. | It can be used to extract uncorrelated dimensions from a large number of highly correlated dimensions. |
|  | B. | Factor analysis allows for theory based extraction of factors by allowing for some degree of correlations among the factors. |
|  | C. | Both A and B |
|  | D. | None of the above |

**2 points**

**QUESTION 20**

1. Which of the following is a procedure for dimension reduction that allows some domain-specific knowledge to select final model?

|  |  |  |
| --- | --- | --- |
|  | A. | Principal Component Analysis |
|  | B. | Factor Analysis |
|  | C. | Multi-dimensional Scaling |
|  | D. | None of the above |