21DS602_2022-23_Quiz3

1. Which of the following techniques is utilized to overcome overfitting in Decision Trees. * (1 Point)

- ( ) Binning
- ( ) PCA
- (•) Prunning
- ( ) Entropy calculation

2. When you perform a correlation analysis between the predictor and the output variables, you find a strong negative correlation. Which of the following statements is correct for this scenario? * (1 Point)

- ( ) Linear regression is used to model relations with close to zero correlation
- ( ) Logistic regression may not be used for classification problems
- ( ) You should not use linear regression since the correlation is negative
- (•) You may use linear regression for this problem

3. First order Markov process refers to which of the following situations? * (1 Point)

- immediate future influences the current decision
- all past and future events influence the current decision
- all past events influence the current decision
- immediate past event influences the current decision

4. A supermarket is collecting data from its customers. All customers are categorized as loyal or neutral based on their monthly grocery purchase behavior. to classify new customers, kNN classifiers are used with training set containing 500 customer data from each class. The k-value set for this classifier is 751. This scenario may be referred to as: * (1 Point)

- Overfit
- Slimfit
- Regularfit
- Underfit

5. Synthetic Minority Oversampling Technique (SMOTE) is used to: * (1 Point)

- bring class balance
- discretize the numeric attribute values
- normalize the attributes with large outliers
- add noise to data for regularization during training

6. A hospital has patient data for all surgeries performed. One of the data fields is if either parents were/are diabetic. For some patients, this field is empty. Which of the following measure may be used to fill the missing values? * (1 Point)

- Entropy
- Median
- Mean
- Mode

7. A test pattern lies between the gutters to one side of the hyperplane. This pattern can not be classified since it lies between the gutters. * (1 Point)

- False
- True

8. Your company sells health drink & supplement products & has collected customer data consisting of height, weight, age & BMI. For a classification problem, you want to use Naive-Bayes classifier (you got wonderful results from NB during your ML term project which has made you a fan of NB). When you present your plan to your manager, she advises you against it without citing a reason. Select your appropriate response from below. * (1 Point)

- You accept the advice because the attributes are related to each other
- You ignore the advice because the data is fairly class balanced
- You accept the advice because you find the data contains numeric values
- You ignore the advice because there are no outliers in the data

9. Explain a scenario where F1 is used for better result analysis in place of accuracy measure. * (1 Point)

F1 is used when the classes are imbalanced.

10. How may AUROC plot be used to identify the best classifier from a multitude of models? * (1 Point)

We can select the best classifier with highest area under the curve.