MACHINE LEARNING

1. In which of the following you can say that the model is overfitting?
Ans:

2. Which among the following is a disadvantage of decision trees?
Ans: B) Decision trees are highly prone to overfitting.

3. Which of the following is an ensemble technique?
Ans: D) Decision tree

4. Suppose you are building a classification model for detection of a fatal disease where detection of the disease is most important. In this case which of the following metrics you would focus on?
Ans: A) Accuracy

5. The value of AUC (Area under Curve) value for ROC curve of model A is 0.70 and of model B is 0.85. Which of these two models is doing better job in classification?
Ans: B) Model B

6. Which of the following are the regularization technique in Linear Regression??
Ans: A) Ridge
    D) Lasso

7. Which of the following is not an example of boosting technique?
Ans: B) Decision Tree

8. Which of the techniques are used for regularization of Decision Trees?
Ans: A) Pruning
    B) L2 regularization

9. Which of the following statements is true regarding the Adaboost technique?
Ans: D) None of the above

10. Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in the model?
Ans: The adjusted R-squared is a modified version of R-squared that adjusts for predictors that are not significant in a regression model. Compared to a model with additional input variables, a lower adjusted R-squared indicates that the additional input variables are not adding value to the model.

11. Differentiate between Ridge and Lasso Regression.

Ans: Ridge Regression:ridge regression puts a similar constraint on the coefficients by introducing a penalty factor
Lasso Regression:lasso regression takes the magnitude of the coefficients, ridge regression takes the square.

12. What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?
Ans: VIF:A variance inflation factor (VIF) is a measure of the amount of multicollinearity in regression analysis.
As a rule of thumb, a VIF of three or below is not a cause for concern. As VIF increases, the less reliable your regression results are going to be.

13. Why do we need to scale the data before feeding it to the train the model?
Ans:To ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features, we scale the data before feeding it to the model.

14. What are the different metrics which are used to check the goodness of fit in linear regression?
Ans: Three statistics are used in Ordinary Least Squares (OLS) regression to evaluate model fit: R-squared, the overall F-test, and the Root Mean Square Error (RMSE).

15. From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy.
Ans:
Sensitivity = TP/TP+FN
$\quad\quad\quad$ = 1000/1000+250
$\quad\quad\quad$ = 0.8
Specificity = TN/FP+TN
$\quad\quad\quad$ = 1200/1200+50
$\quad\quad\quad$ = 0.96
Precision = TP/TP+FP
$\quad\quad\quad$ = 1000/1000+50
$\quad\quad\quad$ = 0.95
Recall $\quad$ = TP/TP+FN
$\quad\quad\quad$ = 1000/1000+250
$\quad\quad\quad$ = 0.8
Accuracy = TP+TN/TP+TN+FP+FN
$\quad\quad\quad$ = 1000+1200/1000+1200+50+250
$\quad\quad\quad$ = 0.88