MACHINE LEARNING

In Q1 to Q5, only one option is correct, Choose the correct option:

1. In which of the following you can say that the model is overfitting?
   A) High R-squared value for train-set and High R-squared value for test-set.
   B) Low R-squared value for train-set and High R-squared value for test-set.
   C) High R-squared value for train-set and Low R-squared value for test-set.
   D) None of the above
2. Which among the following is a disadvantage of decision trees?
   A) Decision trees are prone to outliers.
   B) Decision trees are highly prone to overfitting.
   C) Decision trees are not easy to interpret
   D) None of the above.
3. Which of the following is an ensemble technique?
   A) SVM B) Logistic Regression
   C) Random Forest 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?
   A) Accuracy B) Sensitivity
   C) Precision D) None of the above.
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?
   A) Model A B) Model B
   C) both are performing equal D) Data Insufficient

In Q6 to Q9, more than one options are correct, Choose all the correct options:

6. Which of the following are the regularization technique in Linear Regression??
   A) Ridge B) R-squared
   C) MSE D) Lasso
7. Which of the following is not an example of boosting technique?
   A) Adaboost B) Decision Tree
   C) Random Forest D) Xgboost.
8. Which of the techniques are used for regularization of Decision Trees?
   A) Pruning B) L2 regularization
   C) Restricting the max depth of the tree D) All of the above
9. Which of the following statements is true regarding the Adaboost technique?
   A) We initialize the probabilities of the distribution as 1/n, where n is the number of data-points
   B) A tree in the ensemble focuses more on the data points on which the previous tree was not performing well
   C) It is example of bagging technique
   D) None of the above

Q10 to Q15 are subjective answer type questions, Answer them briefly.

10. Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in the model?
    Answer: The R-squared has been changed and the number of predictors in the model has been taken into account in the adjusted R-squared. It is employed to clarify how much input variables (predictor variables) influence the variation of output variables (predicted variables).

It calculates the percentage of variance that can be explained by just the independent variables that have a significant impact on the explanation of the dependent variable. If you include independent variables that have no bearing on predicting the dependent variable, you will be penalised.

Only when the additional term enhances the model more than would be predicted by chance does the adjusted R-squared rise. It falls off when a predictor boosts the model by a smaller amount than would be predicted by chance. Although it's uncommon, the adjusted R-squared can be negative. It is consistently less than R-Squared

11. Differentiate between Ridge and Lasso Regression.
    Answer: The distinction between lasso regression and ridge regression is that lasso regression frequently sets coefficient values to zero, whereas ridge regression never does.

12. What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?
    Answer: A set of multiple regression variables' variance inflation factor (VIF) is a gauge of how multicollinear they are. The VIF for a regression model variable is mathematically equivalent to the ratio of the variance of the entire model to the variance of a model with only that one independent variable.
    A general rule of thumb in practise is that strong multicollinearity indicates a VIF > 10. A high VIF indicates a considerable collinearity between the independent variable and the other variables in the model. One is deemed to be able to move forward with the regression method if the VIF value is greater than 1.

13. Why do we need to scale the data before feeding it to the train the model?
    Answer: The act of scaling entails putting values on the same scale or range such that no variable may be controlled by another. It is a method for standardising the various independent features seen in data within a given range. It helps deal with highly variable magnitudes, values, or units during the pre-processing of data.
    A machine learning algorithm would choose to weight larger values as higher and evaluate smaller values as lower, regardless of the unit of measurement, if feature scaling was not done. We scale the data before feeding it to the model to make sure that the gradient descent progresses smoothly towards the minima and that the steps for gradient descent are updated at the same pace for all the features.

14. What are the different metrics which are used to check the goodness of fit in linear regression?
    Answer:

    a. Root Squared Mean Error
    b. Mean Squared Error
    c. Mean Absolute Error

15. From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy.

Actual/Predicted	True	False
True	1000	50
False	250	1200

    Answer:

    Sensitivity (true positives / all actual positives) = TP / (TP + FN)
    = 1000/ (1000+250)
    = 1000/ 1250
    = ***0.8 or 80% Sensitivity***
    Specificity (true negatives / all actual negatives) = TN / (TN + FP)

= 1200/ (1200+50)
= 1200/1250
= 0.96 or 96% Specificity

Precision (true positives / predicted positives) = TP / (TP + FP)
= 1000/ (1000+50)
= 1000/ 1050
= ***0.9523 or 95.23% Precision***

Recall (true positives / all actual positives) = TP / (TP + FN)
= 1000/ (1000+250)
= 1000/ 1250
= ***0.8 or 80% Recall***

Accuracy (all correct / all) = (TP + TN) / (TP + TN + FP + FN)
= (1000+1200) /(1000+1200+50+250)
= 2200 / 2500
= 0.88 or 88% Accuracy