# 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?

ANS-High R-squared value for train-set and Low R-squared value for test-set.

1. Which among the following is a disadvantage of decision trees?

ANS-Decision trees are highly prone to overfitting.

1. is an ensemble technique?

ANS- Random Forest

1. 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- Sensitivity

1. 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-Model B

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

1. Which of the following are the regularization technique in Linear Regression??

ANS-Ridge& Lasso

1. Which of the following is not an example of boosting technique?

Ans-Decision Tree

Which of the techniques are used for regularization of Decision Trees?

ANS- All of the above

1. Which of the following statements is true regarding the Adaboost technique?

ANS-A tree in the ensemble focuses more on the data points on which the previous tree was not performing well

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

1. Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in the model?

ANS- The Adjusted R-squared penalizes the presence of unnecessary predictors by subtracting the effect of each predictor on R-squared.it takes into account the number of predictors in the model and the sample size,so model with many predictors will have lower adjusted R-squared compared to model with fewer predictors.

1. Differentiate between Ridge and Lasso Regression.

ANS- Ridge and lasso regression are used to prevent overfitting in linear regression by adding a regularization term to the cost function.Ridge regression uses L2 regularization and tends to shrink the coefficients towards zero,while lasso regression uses L1 regularization and tend to set some coefficeint to zero.

1. What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?

ANS- VIF-Variance inflation factor is a measure of the amount of multicollinearity in a set of predictor variables in regression model.

Commonly used threshold for the VIF value is 5 or 10,meaning that if a predictor haas VIF greater than 5or 10.the appropriate value of threashold depend upon specific problem & goals of the analysis.

1. 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. Having features on a similar scale can help the gradient descent converge more quickly towards the minima.

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

ANS-There are three error metrics that are commonly used for evaluating and reporting the performance of a regression model; they are: Mean Squared Error (MSE). Root Mean Squared Error (RMSE). Mean Absolute Error

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

|  |  |  |
| --- | --- | --- |
| Actual/Predicted | True | False |
| True | 1000 | 50 |
| False | 250 | 1200 |

ANS-sensitivity-0.8

Specificity-0.96

Precision-0.95

Recall-0.8

Accuracy-0.92