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

   Ans: A) High R-squared value for train-set and High R-squared value for test-set

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: C) Random Forest

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 C) Random Forest

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

   Ans: A) Pruning C) Restricting the max depth of the tree.

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

Ans: A) We initialize the probabilities of the distribution as 1/n, where n is the number of data-points

C) It is example of bagging technique

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

The adjusted R-squared compensates for the addition of variables and only increases if the new predictor enhances the model above what would be obtained by probability. Conversely, it will decrease when a predictor improves the model less than what is predicted by chance.

11. Differentiate between Ridge and Lasso Regression.

Lasso Regression: It puts constraints to the coefficient by introducing penalty factor,It takes the magnitude of the coefficient.It tends to make the coefficient value to absolute zero .

Ridge Regression: It also puts constraints to the coefficient by taking the squares of the values,It never tends to make the coeeficient value to absolute zero.

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

VIF means Variance Inflation Factor, As the name suggest ,it gives the values of how multicollinearity is varied from one Independent variables to other .

In particular, the variance inflation factor for the j^{th} predictor is:
VIFJ=1/1-R2j

where R$_j$2 is the R^2-value obtained by regressing the j^{th} predictor on the remaining predictors.

As a rule we should not exceed than 4,if the value of VIF is like 10 then we have serious multicollinearity requiring correction.

13. Why do we need to scale the data before feeding it to the train the model?

Machine learning algorithm just takes the numbers of the value,if the number are in highest range, the model automatically thinks that they have the superiority a lot, that's y feature scaling is important for every feature to bring in the same foot.

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

- R-square.
- Adjusted R-square.
- Root mean squared error (RMSE)

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

Actual/Predicted	**True**	**False**
True	**1000**	**50**
False	**250**	**1200**

Sensitivity: TP/TP+FN=1000/1000+250=0.8

Specificity: TN/TN+FP=1200/1200+50=1200/1250=0.96

Precision: TP/TP+FP=1000/1000+50=1000/1050=0.95

Recall: TP/TP+FN=1000/1000+250=1000/1250=0.8

Accuracy=TP+TN/TP+FP+FN+TN=1000+1200/1000+1200+50+250

2200/2500=0.88.