Machine Learning

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

Ans. d) none of above

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

Ans. a) Decision trees are prone to outliers.

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 disease is most important. In this case which of the following metrics you would focus on?

Ans. c) Precision

5. The value of AUC 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 jo in classification?

Ans. b) Model B

6. Which of 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. c) Random Forest

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

Ans. a) Pruning

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

Ans. c) A tree in the ensemble focuses more on the data points on which the previous tree was not performing well.

10.

11. Lasso tends to do well if there are small number of significant parameters.

Ridge works well if there are many large parameters of about the same value.

Lasso does as that lasso check the relation between each feature and label and if any feature has not any kind of relation to label then lasso will neglect that feature, give zero importance and consider only those features that contributes.

RIDGE -- it will give a little a very little importance also to those features which don't have relation to label.

12. Variance inflation factor (VIF) is a measure of amount of multicollinearity during model building. The suitable value of VIF is less than 5.

13. It is mandatory to standardize the data before feeding it to model because data present in dataset is of different scale. Then, in that situation, what will model do, we have to pass a single standard unit data to model. So that model can train in data easily. After standardize the data, it becomes very easy for model to predict better and error became low as possible.

14. R-square, adjusted r-square, mean absolute error, mean squared error, root mean squared error.

15. ---- In python notebook