FLIP ROBO

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
    Answer- C) High R-squared value for train-set and Low R-squared value for test-set.

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.
    Answer- B) Decision trees are highly prone to overfitting.

3. Which of the following is an ensemble technique?
    A) SVM   B) Logistic Regression
    C) Random Forest   D) Decision tree
    Answer-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?
    A) Accuracy   B) Sensitivity
    C) Precision   D) None of the above.
    Answer- B) Sensitivity

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
    Answer-B) Model B

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
    Answer- A and D

7. Which of the following is not an example of boosting technique?
    A) Adaboost   B) Decision Tree
    C) Random Forest   D) Xgboost.
    Answer- B and C

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
    Answer- A and C

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

MACHINE LEARNING

D) None of the above
   Answer B)- A tree in the ensemble focuses more on the data points on which the previous tree   was notperforming well

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- Adjusted r2 is the modified version of r2. In r2 the score gets increased as the number of independent variables increases it doesn't take into consideration if the independent variable is important or not. However in adjusted R2 it increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance.

11. Differentiate between Ridge and Lasso Regression.

ANSWER 11 = In ridge regression we penalize the least sum of squared residual line with $\lambda + (slope)^2$, however in lasso regression we penalize the least sum of squared residual with $\lambda + |slope|$.
Ridge regression can shrink the slope close to 0, however Lasso regression can shrink the slope up to 0.
Lasso regression helps in removing the irrelevant variables from the model. Ridge regression is better when there are variables which are important to predict the target variable.

12. What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?
ANSWER 12 = Variance inflation factor (VIF) is a measure of the amount of multicollinearity in a set of multiple regression variables. This is calculated as 1/1-R2. If the R2 score is high then the VIF of that variable will be high and vice versa. A rule of thumb commonly used in practice is if a VIF is > 10, you have high multicollinearity. In our case, with values around 1, we are in good shape, and can proceed with our regression.
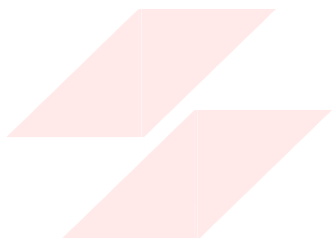
13. Why do we need to scale the data before feeding it to the train the model?

ANSWER 13 = Scaling of data is very important before feeding the data to our model. Machine learning algorithms that use gradient descent for optimisation gets affected with the value of the variables. As our main aim to minimise the loss of the data. If we do not scale our data, we will have multiple (different scaled values) for our input variables and those values will affect the step size of gradient descent. So, by scaling using normal scaling, standard scaling we will make sure that all the variables lie in a same scale which will help the gradient descent to perform well. With scaling the magnitude of the variable remains unchanged. Scaling is also very important for distance based algorithms like Knn, SVM and K-means clustering, distance/similarity matrix.

MACHINE LEARNING

14. What are the different metrics which are used to check the goodness of fit in linear regression?
15. From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy.

Actual/Predicted	True	False
True	1000	50
False	250	1200