MACHINE LEARNING

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

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

MACHINE LEARNING

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?

The adjusted R-squared is a modified version of R-squared that adjusts for predictors that are not significant in a regression model. Compared to a model with additional input variables, **a lower adjusted R-squared indicates that the additional input variables are not adding value to the model**

11. Differentiate between Ridge and Lasso Regression.

Ridge Regression :

In Ridge regression, we add a penalty term which is equal to the square of the coefficient. The L2 term is equal to the square of the magnitude of the coefficients. We also add a coefficient λ to control that penalty term. In this case if λ is zero then the equation is the basic OLS else if $\lambda > 0$ then it will add a constraint to the coefficient. As we increase the value of λ this constraint causes the value of the coefficient to tend towards zero. This leads to tradeoff of higher bias (dependencies on certain coefficients tend to be 0 and on certain coefficients tend to be very large, making the model less flexible) for lower variance.

$$L_{ridge} = argmin_{\hat{\beta}} \left(\|Y - \beta * X\|^2 + \lambda * \|\beta\|_2^2 \right)$$

where λ is regularization penalty.

Limitation of Ridge Regression:

Ridge regression decreases the complexity of a model but does not reduce the number of variables since it never leads to a coefficient been zero rather only minimizes it. Hence, this model is not good for feature reduction.

Lasso Regression :

Lasso regression stands for Least Absolute Shrinkage and Selection Operator. It adds penalty term to the cost function. This term is the absolute sum of the coefficients. As the value of coefficients increases from 0 this term penalizes, cause model, to decrease the value of coefficients in order to reduce loss. The difference between ridge and lasso regression is that it tends to make coefficients to absolute zero as compared to Ridge which never sets the value of coefficient to absolute zero.

$$L_{lasso} = argmin_{\hat{\beta}} \left(\|Y - \beta * X\|^2 + \lambda * \|\beta\|_1 \right)$$

Limitation of Lasso Regression:

- Lasso sometimes struggles with some types of data. If the number of predictors (p) is greater than the

MACHINE LEARNING

number of observations (n), Lasso will pick at most n predictors as non-zero, even if all predictors are relevant (or may be used in the test set).

- If there are two or more highly collinear variables then LASSO regression select one of them randomly
  which is not good for the interpretation of data

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

A variance inflation factor (VIF) is **a measure of the amount of multicollinearity in regression analysis**. Multicollinearity exists when there is a correlation between multiple independent variables in a multiple regression model. This can adversely affect the regression results.

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

It is performed during the data pre-processing **to handle highly varying magnitudes or values or units**. If feature scaling is not done, then a machine learning algorithm tends to weigh greater values, higher and consider smaller values as the lower values, regardless of the unit of the values

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

1) Mean Absolute Error(MAE)
2) Mean Squared Error(MSE)
3) Root Mean Squared Error(RMSE)
4) Root Mean Squared Log Error(RMSLE)
5) R Squared (R2)
6) Adjusted R Squared

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

Actual/Predicted	True	False
True	1000	50
False	250	1200

1. Accuracy (all **correct** / all) = TP + TN / TP + TN + FP + FN=1000+1200/1000+1200+50+250=.88
2. Precision (**true** positives / **predicted** positives) = TP / TP + FP=1000/1000+50=.952
3. Sensitivity or Recall (**true** positives / all **actual** positives) = TP / TP + FN=1000/1000+250=.8

4 Specificity (**true** negatives / all **actual** negatives) =TN / TN + FP= 1200/1200+50=.96