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

ANS- 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.
   ANS- Decision trees are not easy to interpret

3. Which of the following is an ensemble technique?
   A) SVM
   B) Logistic Regression
   C) Random Forest
   D) Decision tree

   ANS- Decision tree

4. Suppose you are building a classification model for detection of a fatal disease where detection ofthe 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.
   ANS- 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?
   A) Model A
   B) Model B
   C) both are performing equal
   D) Data Insufficient
   ANS- both are performing equal

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
   ANS- Ridge

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

ANS- Pruning

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

MACHINE LEARNING

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
E) ANS- None of the above

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 themodel?

ANS- 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.

R-squared tends to reward you for including too many independent variables in a regression model, and it doesn't provide any incentive to stop adding more. Adjusted R-squared and predicted R-squared use different approaches to help you fight that impulse to add too many. The protection that adjusted R-squared and predicted R-squared provide is critical because too many terms in a model can produce results that you can't trust. These statistics help you include the correct number of independent variables in your regression model.

11. Differentiate between Ridge and Lasso Regression.

ANS- **Ridge regression** -Ridge regression is a technique used to analyze multi-linear regression (multicollinear), also known as L2 regularization. It is Applied when predicted values are greater than the observed values.

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2=\mathrm{RSS}+\lambda\sum_{j=1}^{p}\beta_j^2$$

Above equation represents the formula for Ridge Regression! where,

Lambda (λ) in the equation is tuning parameter which is selected using cross-validation technique which makes the fit small by making squares small (β2) by adding shrinkage factor.

Lasso Regression:

Lasso stands for – Least Absolute Shrinkage and Selection Operator. It is a technique where data points are shrunk towards a central point, like the mean. Lasso is also known as L1 regularization.

It is applied when the model is overfitted or facing computational challenges.

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j|\ =\mathrm{RSS}+\lambda\sum_{j=1}^{p}|\beta_j|.$$

MACHINE LEARNING

The above equation represents the formula for Lasso Regression! where, Lambda (λ) is a tuning parameter selected using the before Cross-validation technique.

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

ANS- 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.
What Is a Good VIF Value? As a rule of thumb, a VIF of three or below is not a cause for concern. As VIF increases, the less reliable your regression results are going to be

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

ANS- Feature scaling is essential for machine learning algorithms that **calculate distances between data**. If not scale, the feature with a higher value range starts dominating when calculating distances, as explained intuitively in the “why?” section.

MACHINE LEARNING

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

    ANS- The **adjusted R-square statistic** is generally the best indicator of the fit quality when you add additional coefficients to your model. The adjusted R-square statistic can take on any value less than or equal to 1, with a value closer to 1 indicating a better fit. A RMSE value closer to 0 indicates a better fit.

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

Actual/Predicted	True	False
True	1000	50
False	250	1200

ANS- **Recall**

$$Recall = \frac{TP}{TP + FN}$$

Recall [Image 7] (Image courtesy: My Photoshopped Collection)

The above equation can be explained by saying, from all the positive classes, how many we predicted correctly.

Recall should be high as possible.

Precision

$$Precision = \frac{TP}{TP + FP}$$

Precision [Image 8] (Image courtesy: My Photoshopped Collection)

The above equation can be explained by saying, from all the classes we have predicted as positive, how many are actually positive.

Precision should be high as possible.

and

Accuracy

From all the classes (positive and negative), how many of them we have predicted correctly. In this case, it will be 4/7.

Accuracy should be high as possible.