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

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

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

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:- C

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:- C

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:- 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

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:- D

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

MACHINE LEARNING

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
Answer:- A&B

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?**

Answer:- 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**.

Answer:- 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.

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

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.

The shrinkage factor is lambda times the sum of squares of regression coefficients (The last element in the above equation).

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.

MACHINE LEARNING

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

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

Unlike Ridge Regression, Lasso uses |β| to penalize the high coefficients.

The shrinkage factor is lambda times the sum of Regression coefficients (The last factor in the above equation).

Without further delay, let us look into an example of how we can build Ridge and Lasso Models!
Let us Consider the dataset from an Ad Agency! that advertised their Ad through different forums such as TV, Radio, and Newspapers and recorded their sales against it!

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

    Answer:- Variance Inflation Factors (VIFs) measure the correlation among independent variables in least squares regression models. Statisticians refer to this type of correlation as multicollinearity. Excessive multicollinearity can cause problems for regression models. Most research papers consider a VIF (Variance Inflation Factor) > 10 as an indicator of multicollinearity, but some choose a more conservative threshold of 5 or even 2.5.

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

Answer:- Scaling of the data comes under the set of steps of data pre-processing when we are performing machine learning algorithms in the data set. As we know most of the supervised and unsupervised learning methods make decisions according to the data sets applied to them and often the algorithms calculate the distance between the data points to make better inferences out of the data.

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

Answer:- There are three error metrics that are commonly used for evaluating and reporting the performance of a regression model; they are: **Mean Squared Error (MSE). Root Mean Squared Error (RMSE). Mean Absolute Error (MAE).**

1. R Square/Adjusted R Square

R Square measures how much variability in dependent variable can be explained by the model. It is the square of the Correlation Coefficient(R) and that is why it is called R Square.

MACHINE LEARNING

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

R Square is calculated by the sum of squared of prediction error divided by the total sum of the square which replaces the calculated prediction with mean. R Square value is between 0 to 1 and a bigger value indicates a better fit between prediction and actual value.

2. Mean Square Error (MSE)/Root Mean Square Error(RMSE)

While R Square is a relative measure of how well the model fits dependent variables, Mean Square Error is an absolute measure of the goodness for the fit.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Mean Square Error formula

MSE is calculated by the sum of square of prediction error which is real output minus predicted output and then divide by the number of data points. It gives you an absolute number on how much your predicted results deviate from the actual number. You cannot interpret many insights from one single result but it gives you a real number to compare against other model results and help you select the best regression model.

3. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is similar to Mean Square Error (MSE). However, instead of the sum of square of error in MSE, MAE is taking the sum of the absolute value of error.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

Mean Absolute Error formula

MACHINE LEARNING

Compare to MSE or RMSE, MAE is a more direct representation of sum of error terms. **MSE gives larger penalization to big prediction error by square it while MAE treats all errors the same**.

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

Actual/Predicted	True	False
True	1000	50
False	250	1200

Answer:- The overall accuracy is calculated by **summing the number of correctly classified values and dividing by the total number of values**. The correctly classified values are located along the upper-left to lower-right diagonal of the confusion matrix.

1. Sensitivity: - 0.90
2. Specificity:- 0.576
3. Precision :- 0.91
4. recall and accuracy.:- 0.88 & 0.80

MACHINE LEARNING