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

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 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.

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

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?
11. Differentiate between Ridge and Lasso Regression.
12. What is VIF? What is the suitable value of a VIF for a feature to be included in a regressionmodelling?
13. Why do we need to scale the data before feeding it to the train the model?

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

10. Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in themodel?

R-squared is a statistical measure that represents the proportion of variance in the dependent variable that is explained by the independent variables in a linear regression model. However, when additional predictors are added to the model, the R-squared value will always increase, even if the new predictors do not improve the accuracy of the model. Therefore, adjusted R-squared is used as a modified version of R-squared that penalizes the presence of unnecessary predictors in the model.

Adjusted R-squared is calculated by adjusting R-squared based on the number of predictors in the model and the sample size. It takes into account the number of predictors used in the model and the degree of freedom. It is given by the formula:

Adjusted R-squared = 1 - [(1 - R-squared) * (n - 1) / (n - k - 1)]

where,

- **R-squared is the original R-squared value**
- **n is the sample size**
- **k is the number of predictors in the model**

As we can see from the formula, adjusted R-squared reduces the R-squared value by adding a penalty term, which is a function of the number of predictors in the model. The penalty term increases as the number of predictors increases, which helps to adjust the R-squared value for the overfitting problem caused by adding unnecessary predictors. The adjusted R-squared value will only increase if the additional predictor improves the accuracy of the model, and will decrease if the additional predictor does not improve the accuracy of the model.

Therefore, adjusted R-squared penalizes the presence of unnecessary predictors in the model by reducing the value of R-squared when additional predictors do not improve the accuracy of the model. It helps in selecting the best subset of predictors for the model by avoiding overfitting and improving the generalization performance of the model.

MACHINE LEARNING

11. Differentiate between Ridge and Lasso Regression.

Ridge Regression and Lasso Regression are two popular linear regression techniques used for regularization, which helps to prevent overfitting and improve the generalization performance of the model. However, they have different approaches to regularization and differ in their properties.

Here are some of the key differences between Ridge and Lasso Regression:

Regularization approach: Ridge Regression uses L2 regularization, while Lasso Regression uses L1 regularization. L2 regularization adds a penalty term proportional to the square of the coefficients, whereas L1 regularization adds a penalty term proportional to the absolute value of the coefficients.

Shrinkage effect: Both Ridge and Lasso Regression use shrinkage, which shrinks the coefficients towards zero. However, the degree of shrinkage is greater in Lasso Regression compared to Ridge Regression. This means that Lasso Regression tends to drive some coefficients to exactly zero, while Ridge Regression only reduces the magnitude of the coefficients.

Feature selection: Ridge Regression can reduce the coefficients of less important features close to zero, but it cannot eliminate them completely. On the other hand, Lasso Regression can perform feature selection by eliminating the coefficients of less important features, making it useful when there are a large number of features.

Number of variables: When the number of variables is greater than the number of observations, Ridge Regression can be used, while Lasso Regression may fail to converge. This is because Lasso Regression may end up selecting more variables than necessary, leading to overfitting.

Computation: Lasso Regression involves solving a non-linear optimization problem, while Ridge Regression has a closed-form solution that can be solved analytically. However, both methods can be computationally expensive for large datasets with a large number of features.

In summary, Ridge Regression and Lasso Regression are two regularization techniques that differ in their approach to regularization, shrinkage effect, feature selection, number of variables, and computation. The choice of the regularization technique depends on the specific requirements of the problem and the properties of the dataset.

MACHINE LEARNING

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

VIF (Variance Inflation Factor) is a measure of multicollinearity in a regression analysis. It quantifies how much the variance of the estimated regression coefficient is inflated due to the correlation between the predictor variables.

The VIF of a predictor variable is calculated as the ratio of the variance of the model with all the predictor variables to the variance of the model without that predictor variable. A high VIF value indicates that the predictor variable is highly correlated with the other predictor variables in the model and may lead to instability in the regression coefficients and reduced predictive accuracy.

A commonly used rule of thumb for VIF is that a value greater than 5 or 10 indicates high multicollinearity, which means that the predictor variable may not be suitable for inclusion in the regression model. However, the specific threshold for VIF may depend on the particular application and the nature of the data. In general, it is desirable to keep the VIF value as low as possible to minimize multicollinearity and increase the stability and reliability of the regression coefficients.

It is a good practice to calculate the VIF values for all the predictor variables in the regression model before finalizing the model. If a predictor variable has a high VIF value, it may be necessary to remove it from the model or combine it with other predictor variables to reduce multicollinearity and improve the performance of the model.

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

Scaling the data is an important preprocessing step before training a model, especially for machine learning algorithms that use distance-based metrics or rely on gradient-based optimization algorithms. There are several reasons why we need to scale the data:

Avoid bias: When the range of features in the dataset is large, features with larger values will dominate over features with smaller values. This can result in a biased model, where the dominant features are given more importance and may not accurately represent the underlying relationships in the data.

Improve convergence: Gradient-based optimization algorithms like gradient descent converge faster when the features are normalized to a similar scale. Otherwise, features with large values may lead to large steps in the weight space, which can make convergence slower or unstable.

Better performance: Many machine learning algorithms assume that the features are normally distributed and have similar variances. Scaling the features can help to achieve this assumption, which can improve the performance of the model.

Interpretation: Scaling the features can make it easier to interpret the coefficients of the model. Without scaling, it can be difficult to compare the importance of different features.

Overall, scaling the data is an important step in the data preprocessing pipeline, and it can improve the performance, stability, and interpretability of machine learning models.

MACHINE LEARNING

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

In linear regression, there are several metrics that can be used to evaluate the goodness of fit of the model. Here are some of the commonly used metrics:

R-squared (R^2): It is the most commonly used metric to evaluate the goodness of fit of the linear regression model. It measures the proportion of variance in the dependent variable that is explained by the independent variables. The value of R-squared ranges between 0 and 1, with higher values indicating a better fit.

Adjusted R-squared: It is a modification of the R-squared that takes into account the number of predictors in the model. It penalizes the presence of unnecessary predictors and provides a more accurate estimate of the goodness of fit of the model.

Mean squared error (MSE): It measures the average squared difference between the actual and predicted values of the dependent variable. A lower value of MSE indicates a better fit of the model.

Root mean squared error (RMSE): It is the square root of the MSE and provides a measure of the typical distance between the actual and predicted values. A lower value of RMSE indicates a better fit of the model.

Mean absolute error (MAE): It measures the average absolute difference between the actual and predicted values of the dependent variable. Like MSE, a lower value of MAE indicates a better fit of the model.

These metrics provide different aspects of the goodness of fit of the linear regression model and can be used together to get a better understanding of the model performance.

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

Actual/Predicted	True	False
True	1000	50
False	250	1200

Ans:

Using the given confusion matrix, we can calculate the following metrics:

True Positive (TP) = 1000
False Positive (FP) = 50
False Negative (FN) = 250
True Negative (TN) = 1200

Sensitivity (Recall) = TP / (TP + FN) = 1000 / (1000 + 250) = 0.80

Specificity = TN / (TN + FP) = 1200 / (1200 + 50) = 0.96

Precision = TP / (TP + FP) = 1000 / (1000 + 50) = 0.95

MACHINE LEARNING

Accuracy = (TP + TN) / (TP + TN + FP + FN) = (1000 + 1200) / (1000 + 1200 + 50 + 250) = 0.90

Therefore, the sensitivity is 0.80, specificity is 0.96, precision is 0.95, recall is 0.80 and accuracy is 0.90.