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: 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.

   Ans: B

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

   ANS: 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.
   ANS: B

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: 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
   ANS: 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.
   Ans: 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

   ANS: A and C

MACHINE LEARNING

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
   ANS: 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 the model?

    The adjusted R-squared is a modification of the R-squared metric that adjusts for the number of predictors in the model. It penalizes the presence of unnecessary predictors in the model by adjusting the R-squared value based on the number of predictors used in the model.

    In a regression model, the R-squared metric measures the proportion of the variance in the dependent variable that is explained by the independent variables. However, as we add more predictors to the model, the R-squared value tends to increase, even if the additional predictors do not improve the model's predictive power. This is known as overfitting.

    The adjusted R-squared penalizes the presence of unnecessary predictors by incorporating a penalty term that adjusts the R-squared value based on the number of predictors used in the model. The penalty term increases as the number of predictors increases, which helps to prevent overfitting and ensures that the model is not overly complex.

11. Differentiate between Ridge and Lasso Regression.

    Ridge Regression and Lasso Regression are two popular regularization techniques used in linear regression to prevent overfitting. The main differences between Ridge and Lasso Regression are:

    Penalty term: Ridge regression adds a penalty term equivalent to the square of the magnitude of the coefficients (L2 penalty), whereas Lasso regression adds a penalty term equivalent to the absolute value of the magnitude of the coefficients (L1 penalty).

    Coefficient shrinkage: Ridge regression shrinks the coefficients towards zero but does not set them exactly to zero, whereas Lasso regression can set some of the coefficients to zero, effectively performing feature selection.

    Optimization: Ridge regression can be solved using closed-form solutions, while Lasso regression requires iterative optimization techniques like gradient descent.

    Number of predictors: Ridge regression is effective when dealing with a large number of predictors, whereas Lasso regression is more effective when dealing with a small number of predictors.

    Bias-variance tradeoff: Ridge regression balances the tradeoff between bias and variance by reducing the variance of the model, while Lasso regression can lead to increased bias but reduced variance by reducing the number of predictors.

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

MACHINE LEARNING

modelling?
VIF stands for Variance Inflation Factor, which is a measure of multicollinearity among predictor variables in a regression model. Multicollinearity occurs when two or more independent variables in a regression model are highly correlated with each other, leading to inflated standard errors and unstable coefficients.

The VIF measures the extent to which the variance of an estimated regression coefficient is increased because of collinearity. Specifically, the VIF for a predictor variable is calculated as the ratio of the variance of its coefficient estimate in a model that includes all other predictor variables, to the variance of its coefficient estimate in a model that includes only that predictor variable.

A VIF value of 1 indicates no multicollinearity, while a value greater than 1 indicates the presence of multicollinearity. Generally, a VIF value of 5 or higher indicates high multicollinearity and suggests that the variable may need to be removed from the model. However, the suitable value of VIF for a feature to be included in a regression model may vary depending on the specific context and research question. It is recommended to consult with domain experts and perform sensitivity analysis to determine the appropriate threshold for VIF.

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 machine learning model. Here are some reasons why scaling is necessary:

Helps to improve the convergence speed: Scaling the data helps to improve the convergence speed of optimization algorithms. Many machine learning algorithms, such as gradient descent, are based on the optimization of a cost function. If the features are not scaled, the optimization algorithm may take longer to converge to the optimal solution or may even get stuck in a local minimum.

Helps to prevent domination of features: Features with large scales can dominate the features with small scales during the model training process. This can result in biased coefficients and inaccurate predictions. Scaling the data ensures that all features are equally important in the model fitting process.

Helps to improve the performance of distance-based algorithms: Many machine learning algorithms, such as k-Nearest Neighbors (KNN) and Principal Component Analysis (PCA), are based on distance calculations between data points. If the features are not scaled, the distance calculations may be dominated by features with large scales, resulting in biased results.

Helps to interpret the model coefficients: Scaling the data ensures that the model coefficients are comparable and interpretable. Without scaling, the coefficients may be in different scales and cannot be directly compared to each other.

Overall, scaling the data is an essential step in preparing data for machine learning, as it helps to improve the accuracy, efficiency, and interpretability of the model.

MACHINE LEARNING

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

There are several metrics that can be used to check the goodness of fit in linear regression models. Here are some of the commonly used metrics:

R-squared (R2): R-squared is a measure of how well the regression line fits the data. It is the proportion of the variance in the dependent variable that can be explained by the independent variable(s). An R2 value of 1 indicates a perfect fit, while a value of 0 indicates no relationship between the variables.

Mean Squared Error (MSE): MSE is the average of the squared differences between the predicted and actual values. It measures the average magnitude of the errors in the predictions.

Root Mean Squared Error (RMSE): RMSE is the square root of the MSE. It is a measure of the standard deviation of the residuals, which is the difference between the predicted and actual values.

Mean Absolute Error (MAE): MAE is the average of the absolute differences between the predicted and actual values. It measures the average magnitude of the errors, regardless of their direction.

Residual Standard Error (RSE): RSE is similar to RMSE but normalized by the degrees of freedom. It measures the variability of the residuals around the regression line.

Adjusted R-squared: Adjusted R-squared is a modified version of R-squared that takes into account the number of independent variables in the model. It penalizes the addition of irrelevant variables that do not improve the model's performance.

F-statistic: F-statistic is a test statistic that compares the explained variance by the regression model to the unexplained variance. A high F-statistic indicates that the model is a good fit for the data.

Overall, a combination of these metrics can be used to evaluate the goodness of fit of linear regression models and to compare different models. It is important to choose the appropriate metric(s) depending on the specific context and research question.

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

Actual/Predicted	True	False
True	1000	50
False	250	1200

1. Sensitivity or recall = True Positive / (True Positive + False Negative)
   = 1000 / (1000 + 250)
   = 0.8
2. Specificity = True Negative / (False Positive + True Negative)
   = 1200 / (50 + 1200)
   = 0.96
3. Precision = True Positive / (True Positive + False Positive)
   = 1000 / (1000 + 50)

MACHINE LEARNING

= 0.9524

4. Accuracy = (True Positive + True Negative) / (True Positive + False Positive + False Negative + True Negative)
   = (1000 + 1200) / (1000 + 50 + 250 + 1200)
   = 0.88