

MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

Ans = A) Least Square Error

2. Which of the following statement is true about outliers in linear regression?

Ans = A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?

Ans = B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?

Ans = C) Both of them

5. Which of the following is the reason for over fitting condition?

Ans = C) Low bias and high variance

6. If output involves label then that model is called as:

Ans = B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

Ans = D) Regularization

8. To overcome with imbalance dataset which technique can be used?

Ans = D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

Ans = A) TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

Ans = B) False

11. Pick the feature extraction from below:

Ans = B) Apply PCA to project high dimensional data

In Q12, more than one options are correct, choose all the correct options:

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

Ans = B) It becomes slow when number of features is very large.

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Ans = when we use regression models to train some data. There is a good chance that the model will overfit the given training data set. regularization helps sort this overfitting problem by restricting the degression of freedom of a given equation simily reducing the number of degression function by reducing their corresponding weights.

In a linear equation we do not huge weight coefficient as a chance in weight can a large different for the dependent variable (y) so regularization constrints the weight of such feature to avoid overfitting

To regularization the models a shrinkage penalty is added to be cost function lets see different type of regularization in regression

The words regularize means to make things regular or acceptable this is exactly why we use it for Explain the term regularization are techniques used to reduce the error by fitting a function appropriately on the giving training set and avoid overfitting

The weights are evenly distributed this can be achieved by doing Explain the term regularization there are two types of regularization as follows L1 regularization or LASSO regularization L2 regularization or RIDGE regularization

Regularization is a technique which makes slight modifications to the learning algorithm such that the model generalizes better this in turn improves the model's performance on the unseen data as well.

14. Which particular algorithms are used for regularization?

Ans = 1) LASSO :- (Least Absolute Shrinkage and Selection Operator) lasso regression penalizes the models based on the sum of magnitude of a coefficient the regularization term is given by LASSO

Lasso tends to do well if there are a small number of significant parameters and the others are close to zero (ergo when only a few predictors actually influence the response) RIDGE works well if there are many large parameters of about the same values

Lasso was introduced in order to improve the prediction accuracy and interpretability of regression models it reduces the number of known covariates for use in a model.

2) RIDGE :- ridge regression (L2 form) ridge regression penalizes the models based on the sum of square of magnitude of the coefficient the regularization term is given by regularization

Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization when the issue of multicollinearity occurs least squares are unbiased and variances are large this results in predicted values being far away from the actual values

Ridge regression example for example ridge regression can be used for the analysis of prostate specific antigen and clinical measure among people who were about to have prostates removed the performance of ridge regression is good when there is a subset which are small or even zero

15. Explain the term error present in linear regression equation?

Ans = Linear regression is one of the most fundamental algorithms in the machine world it is door to the magical world ahead but before proceeding with the algorithm let's first discuss the lifecycle of any machine learning model this diagram example the creation of a machine learning model further with hyperparameters tuning to increase its accuracy deciding the deployment strategies for that models once deployment strategies for that models monitoring setting regression analysis is an important tools for analysis and modelling data here we fit a curve line to the data point in such a manner that different between distance of the actual data point plotted curved le minimum

Linear regression most often uses mean square error (MSE) to calculate the error of the models MSE is calculated by measuring the distance of the observed y-values from predict y-values at each values x:squaring each of there distance calculating the mean of each of the square distance

Error of prediction are define as the difference between the observed values of the dependent variable and the predict values for the variable obtained using a given regression equation and the values of the independent variable

Residual variance (also called unexplained variance or error variance) is the variance of any (residual) the exact definition depend on what type analysis You performing for example in regression analysis random fluctuation cause variance around the True regression line