Machine Learning Worksheet – 01

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?
Answer - (A) Least Square Error

2. Which of the following statement is true about outliers in linear regression?
Answer - (A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?
Answer - (B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?
Answer - (B) Correlation

5. Which of the following is the reason for over fitting condition?
Answer - (C) Low bias and high variance

6. If output involves label, then that model is called as:
Answer - (B) Predictive model

7. Lasso and Ridge regression techniques belong to _____?
Answer - (B) Regularization

8. To overcome with imbalance dataset which technique can be used?
Answer - (A) Cross validation

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

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

11. Pick the feature extraction from below:
Answer - (C) Removing stop words

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
Answer - (A) We don't have to choose the learning rate
(B) It becomes slow when number of features is very large

13. Explain the term regularization?
Answer - The term "regularisation" describes methods for calibrating machine learning models to reduce the adjusted loss function and avoid overfitting or underfitting. We can fit our machine learning model on a particular test set using regularisation, which lowers the mistakes in the test set.

Regularization is a method for reducing mistakes by properly fitting the function on the provided training set and avoiding overfitting. Regression model known as LASSO (Least Absolute Shrinkage and Selection Operator) regression uses L1 regularization approach. Ridge regression is a type of L2 regularization regression model. The absolute value of magnitude of the coefficient is added as a penalty term by the Lasso Regression to the loss function.

14. Which particular algorithms are used for regularization?
Answer - The regularisation techniques include -

A) Ridge Regression - A technique for analysing data with multi-collinearity is ridge regression. Ridge regression increases the loss function's penalty (L2 penalty), which is equal to the square of the coefficients' size. The loss function is punished if the coefficients take on large values thanks to the regularisation parameter (λ)

B) LASSO (Least Absolute Shrinkage and Selection Operator) Regression - Ridge regression increases the loss function's penalty (L2 penalty), which is equal to the square of the coefficients' size. A feature selection and regularisation procedure called LASSO is used in regression analysis to improve the model's ability to predict outcomes accurately. In LASSO regression, a penalty (L1 penalty) equal to the coefficients' magnitude is added to the loss function. When the regularisation value is large enough, the penalty in LASSO regression has the effect of pushing some of the coefficient estimates to be exactly equal to zero.

C) Elastic-Net Regression - A regularised regression technique called Elastic-Net, linearly integrates the L1 and L2 penalties of the LASSO and Ridge methods.

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

Answer - The error term, which refers to the sum of the deviations within the regression line and explains the discrepancy between the theoretical value of the model and the actual observed results, denotes the margin of error inside a statistical model. When attempting to ascertain the correlation between one independent variable and one dependent variable, the regression line is utilised as a point of analysis. An error term basically denotes that the model is not entirely precise and yields varying outcomes when used in practical applications. Assume, for instance, that a multiple linear regression function has the following shape -

$Y = \alpha X + \beta \rho + \epsilon$

Where,
α, β = Constant parameters
X, ρ = Independent variables
ϵ = Error term

During an empirical test, if the actual Y is different from the expected or projected Y in the model, the error term does not equal 0, indicating that other factors affect Y.