Machine Learning Worksheet

Q1.Which of the following methods do we use to find the best fit line for data in Linear Regression?

Sol: Option A i.e Least Square Error

Q2. Which of the following statement is true about outliers in linear regression?

Sol: Option A i.e Linear regression is sensitive to outliers

Q3. A line falls from left to right if a slope is _____?

Sol: Option B i.e Negative

Q4. Which of the following will have symmetric relation between dependent variable and independent variable?

Sol: Option B i.e Correlation

Q5. Which of the following is the reason for over fitting condition?

Sol: Option C i.e Low bias and high variance

Q6. If output involves label then that model is called as:

Sol: Option B i.e Predictive Model

Q7. Lasso and Ridge regression techniques belong to _____?

Sol: Option D i.e Regularization

Q8. To overcome with imbalance dataset which technique can be used

Sol: Option D i.e SMOTE

Q9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

Sol: Option C i.e Sensitivity and Specificity

Q10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

Sol: Option A i.e True

Q11.Pick the feature extraction from below:

Sol: Option A i.e Construction bag of words from a email

Q12.Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

Sol: Option A, B and D

Q13. Explain the term regularization?

Sol: Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting. Overfitting occurs the model captures the arbitrary data in the training dataset. Such data points that do not have the properties of data make the model noisy. This noise may make the model more flexible, but it can cause low accuracy. In other words, the model has low bias and high variance.

For instance, Consider a classroom of 10 students with an equal number of girls and boys. The overall class grade in the annual examination is 70. The average score of female students is 60, and that of male students is 80. Based on these past scores, we want to predict the students' future scores. Predictions can be made in the following ways:

- Under Fit: The entire class will score 70 marks
- Optimum Fit: This could be a simplistic model that predicts the score of girls as 60 and boys as 80 (same as last time)
- Over Fit: This model may use an unrelated attribute, say the roll number, to predict that the students will score precisely the same marks as last year.

Regularization is a technique used for tuning the function by adding an additional penalty term in the error function. The additional term controls the excessively fluctuating function such that the coefficients don't take extreme values. This technique of keeping a check or reducing the value of error coefficients are called shrinkage methods or weight decay in case of neural networks.

Q14. Which particular algorithms are used for regularization?

Ridge Regression

The Ridge regression technique is used to analyze the model where the variables may be having multicollinearity. It reduces the insignificant independent variables though it does not remove them completely. This type of regularization uses the L2 norm for regularization.

- It uses the L2-norm as the penalty.
- L2 penalty is the square of the magnitudes of beta coefficients.
- It is also known as L2-regularization.
- L2 shrinks the coefficients, however never make them to zero.
- The output of L2 regularization is non-sparse.

Ridge regression is used when it is important to consider all the independent variables in the model or when many interactions are present. That is where collinearity or co-dependency is present amongst the variables.

Lasso Regression

Least Absolute Shrinkage and Selection Operator (or LASSO) Regression penalizes the coefficients to the extent that it becomes zero. It eliminates the insignificant independent variables. It is useful when there are many variables, as this technique can be used as a

feature selection method by itself. This regularization technique uses the L1 norm for regularization.

- It adds L1-norm as the penalty.
- L1 is the absolute value of the beta coefficients.
- It is also known as the L-1 regularization.
- The output of L1 regularization is sparse.

Lasso regression is applied when there are many predictors available and would want the model to make feature selection as well.

Elastic Net Regression

The Elastic Net Regression technique is a combination of the Ridge and Lasso regression technique. It is the linear combination of penalties for both the L1-norm and L2-norm regularization.

The model using elastic net regression allows the learning of the sparse model where some of the points are zero, similar to Lasso regularization, and yet maintains the Ridge regression properties. Therefore, the model is trained on both the L1 and L2 norms. When many variables are present, and determine whether to use Ridge or Lasso regression becomes task, then the Elastic-Net regression is the safest option.

Q15. Explain the term error present in linear regression equation?

Sol: An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis.

The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters *e,* ε, or u.

An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results. The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.

An error term essentially means that the model is not completely accurate and results in differing results during real-world applications. For example, assume there is a multiple linear regression function that takes the following form:

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

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

When the actual Y differs from the expected or predicted Y in the model during an empirical test, then the error term does not equal 0, which means there are other factors that influence Y.