**Machine Learning Assignment**

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

Answer. A) Least Square Method

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

Answer. A) Linear regression is sensitive to outliers

1. A line falls from left to right if a slope is \_\_\_\_\_\_?

Answer. B) Negative

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

Answer. B) Correlation

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

Answer. C) Low bias and high variance

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

Answer. B) Predictive Model

1. Lasso and Ridge regression techniques belong to \_\_\_\_\_\_\_\_\_?

Answer. D) Regularization

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

Answer. D) SMOTE

1. 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

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

Answer. B) False

1. Pick the feature extraction from below:

Answer. We haven’t been taught about it yet

1. 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. It is a technique which we can use to reduce the errors by fitting the function appropriately and handle the problems of ‘Underfitting’ & ‘Overfitting’, where the model should learn all the data in proper way & should start giving the right predictions.

A statistical model or a machine learning algorithm is said to have underfitting when it cannot capture the underlying trend of the data. Underfitting destroys the accuracy of our machine learning model.

A statistical model is said to be overfitted when we train it with a lot of data. When a model gets trained with so much data, it starts learning from the noise and inaccurate data entries in our data set. Then the model does not categorize the data correctly, because of too many details and noise.

1. Which particular algorithms are used for regularization?

Answer. Commonly used regularization techniques are:

1. L1 Regularization
2. L2 Regularization
3. ElasticNet

L1 Regularization (Lasso Regression)

This regularization technique performs L1 regularization. It modifies the Residual sum of squares by adding the penalty (shrinkage quantity) equivalent to the sum of the absolute value of coefficients.

L2 Regularization (Ridge Regression)

This regularization technique performs L2 regularization. It modifies the Residual sum of squares by adding the penalty (shrinkage quantity) equivalent to the square of the magnitude of coefficients.

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

Answer. An error term represents the margin of error within a statistical model; it refers to the [sum of the deviations](https://www.investopedia.com/terms/s/sum-of-squares.asp) within the [regression line](https://www.investopedia.com/terms/r/regression.asp), 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.

*Y*=*α*+*βx*+*ϵ*

**where:**

*α*=Intercept

*β=Coefficient/Slope*

*x*=Independent variables

*ϵ*=Error term​