Machine Learning Assignment 1:-

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

Least Square error

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

Linear regression is sensitive to outliers

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

Negative no.

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

Correlation

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

Low bias and high variance

If output involves label then that model is called as:

Predictive model.

Lasso and Ridge regression techniques belong to _____?

Cross Validation

To overcome with imbalance dataset which technique can be used?

SMOTE

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

TPR and FPR

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

False

Pick the feature extraction from below:

Apply PCA to project high dimensional data

Truth about Normal Equation used to compute the coefficient of the Linear Regression?

A) We don't have to choose the learning rate.

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

C) We need to iterate.

Explain the term regularization?

Regularization is a process that changes the result answer to be "simpler". It is often used to obtain results for ill-posed problems or to prevent overfitting.

Although regularization procedures can be divided in many ways, one particular delineation is particularly helpful:

- **Explicit regularization** is regularization whenever one explicitly adds a term to the optimization problem. These terms could be priors, penalties, or constraints. Explicit regularization is commonly employed with ill-posed optimization problems. The regularization term, or penalty, imposes a cost on the optimization function to make the optimal solution unique.
- **Implicit regularization** is all other forms of regularization. This includes, for example, early stopping, using a robust loss function, and discarding outliers. Implicit regularization is essentially ubiquitous in modern machine learning approaches, including stochastic gradient descent for training deep neural networks, and ensemble methods (such as random forests and gradient boosted trees).

Which particular algorithms are used for regularization?

- Ridge Regression.
- LASSO (Least Absolute Shrinkage and Selection Operator) Regression.
- Elastic-Net Regression.

Explain the term error present in linear regression equation?

The error term is the difference between the expected price at a particular time and the price that was actually observed.
Linear Regression most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by:

measuring the distance of the observed y-values from the predicted y-values at each value of x;

squaring each of these distances;

calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE