Machine Learning_assignment - Solutions

Ans. 1 Least Square Error

Ans. 2 Linear regression is sensitive to outliers.

Ans. 3 Line falls from left to right if slope is NEGATIVE.

Ans. 4 Corelation has symmetric relation between dependent and independent variables.

Ans. 5 Reason for Over-fitting is Low bias and high variance.

Ans. 6 Predictive model.

Ans. 7 LASSO & RIDGE regression techniques belong to Regularization

Ans. 8 To overcome imbalanced dataset Synthetic Minority Oversampling Technique (SMOTE) is used.

Ans. 9 TPR & FPR

Ans. 10 False. Higher the AUC better is the model being analyzed.

Ans. 11 Apply PCA to project high dimension data

Ans. 12 A. We don't need to choose learning rate.

B. It slows when number of features is large.

C. We need to iterate.

Ans. 13

Regularization is a method to estimate complexity of machine learning model so that model generalizes, eliminating over-fit and under-fit problem. This is done by reducing degree of freedom of other parameters. It also means that Regularization is a form of regression which reduces the coefficients to 0.

Regularization techniques are

a) L1 regularization:- LASSO (Least Absolute Shrinkage and Selection Operator)

b) L2 regularization:- Ridge regression.

Ans. 14

A) Ridge Regression (L2 regularization): - It penalizes loss function which is equivalent to square of magnitude of coefficients. Regularization parameter λ regularizes the coefficients as to when the coefficients take large values loss function is penalized.

B) LASSO regression (L1 regularization) : - LASSO performs both feature selection and regularization to improve accuracy of the model. It penalizes loss function equal to magnitude of coefficients.

It converts coefficients of less important features to zero, also it reduces coefficients of other features to reduce complexity in the model and avoid over-fitting.

Ans. 15

Error in linear regression equation represents effects of the variables omitted from the equation. It is the residual term which represents the amount at which equation may differ during analysis.