

MACHINE LEARNING ASSIGNMENT-1

1. (a) Least square error
2. (a) Linear regression is sensitive to outliers
3. (b) Negative
4. (c) Both of them
5. (c) Low bias and high variance
6. (b) Predictive modal
7. (d) Regularization
8. (a) Cross validation
9. (a) TPR and FPR
10. (b) False
11. (a), (b) and (c)
12. (a) We don't have to choose the learning rate & (c) We need to iterate
13. Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid over fitting. This technique can be used in such a way that it allows maintaining all variables or features in the model by reducing the magnitude of the variables. It mainly reduces the coefficient of features towards zero. The commonly used regularization techniques are:
 - a) L1 regression(LASSO)
 - b) L2 regression(Ridge)
 - c) Elastic- net regression
14. The algorithms which are used in regularization are :
 - a) LASSO Regression: Least Absolute Shrinkage and Selection Operator regression. LASSO is a regression analysis method that performs feature selection and regularization technique in order to enhance the prediction accuracy of the model. LASSO regression adds a penalty (*L1 penalty*) to the loss function that is equivalent to the magnitude of the coefficients. Also called L1 regression.

- b) Ridge regression: It is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions. Ridge regression adds a penalty (L2 penalty) to the loss function that is equivalent to the square of the magnitude of the coefficients. Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as L2 regularization.
- c) Elastic – Net Regression: This method linearly combines the L1 and L2 penalties of the Lasso and ridge methods respectively.

15. Error term is a residual variable produced by a statistical or mathematical model, which is created when model does not fully represent the actual relationship between independent and dependent variables. It is also known as Residual, disturbance etc and is represented by letter e, ϵ or u

Equation: $Y = \alpha X + \beta p + \epsilon$

where: α, β = Constant parameters ; X, p = Independent variables and ϵ = Error term

A regression line always has an error term because in real life also the independent variables are never perfect predictors of the dependent variables. So the larger the error term the less certain the regression line will be.