

ASSIGNMENT-6

TITLE : SUPERVISED LEARNING REGRESSION

AIM : Generate a proper 2D set of N points, split dataset into train & test sets.

- i. Perform linear regression with least square
- ii. Plot graphs for MSE of train & test sets. Comment on curve fitting.
- iii. Verify effect of dataset size & bias-var. tradeoff
- iv. Apply cv & plot graph errors
- v. Apply subset selection method & plot graphs for errors
- vi. Describe finding in each case.

THEORY :

- LEAST SQUARE METHOD FOR REGRESSION :

i. LINEAR REGRESSION :

Linear approach to modelling the relationship between dependant & independant variable/s

$$y = mx + c \rightarrow y\text{-intercept}$$

\swarrow dependant var. \searrow independant var.
 \downarrow slope

Challenge is to find m & c for which error is minimised.

FINDING ERROR :

Loss function simply measures how far predicted value is different than actual value

$$L(x) = \sum_{i=1}^n (y_i - p_i)^2$$

y = actual

p = predicted.

iii. LEAST SQUARES METHOD:

To minimise loss function, we find partial derivative of L & equate to 0.

$$m = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$c = \bar{y} - m \cdot \bar{x}$$

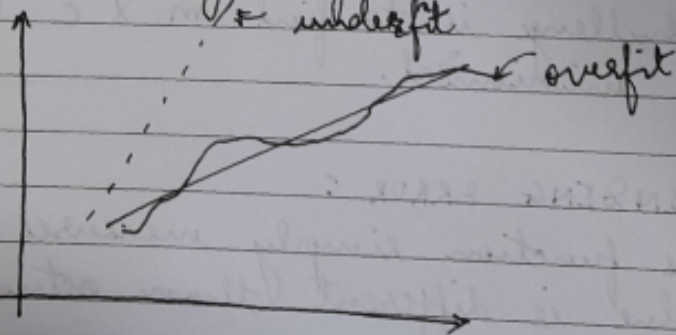
• MEAN SQUARED ERROR:

MSE of estimated measures average of the squares of errors, i.e., avg. squared difference b/w them.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

• CURVE FITTING & GENERALISATION ERROR:

It's a process of constructing a curve or mathematical function that has best fit in series of data points, possibly subjected to constraints.



Techniques to reduce generalisation error:

- i. Simplify model
- ii. Get more data
- iii. Regularization.

- EFFECT OF DATASET:

More the data, better the fit of linear regression & lesser generalization error.

- BIAS VARIANCE TRADEOFF:

Property of model that the variance of parameter estimates across examples can be reduced in bias in estimated parameters.

- CROSS-VALIDATION:

- i. LOOCV
- ii. Stratified cv
- iii. Repeated cv
- iv. Nested cv

- CONCLUSION:

Linear regression was studied & implemented successfully.