WORKSHEET

MACHINE LEARNING – WORKSHEET 11 (LINEAR REGRESSION)

# In Q1 to Q8, only one option is correct, Choose the correct option:

1. What happens to R2 measure if we add a new feature?
   1. remains same B) always increases

C) may or may not increase D) always decreases

**ANS- B**

1. The correct relationship between SST, SSR and SSE is given by:
   1. SSR = SST + SSE B) SST = SSR + SSE

C) SSE = SSR – SST D) None of the above

**ANS-B**

1. Residuals in regression analysis can be defined as:
   1. difference between the actual value and the predicted value.
   2. difference between the actual value and the mean of predicted value.
   3. difference between the actual value and mean of dependent variable.
   4. None of the above.

**ANS- A**

1. In a simple linear regression model, if we change the input variable by 1 unit, then how much output variable will change?
   1. By 1 B) No Change

C) By its slope D) None of the above

**ANS- C**

1. If the coefficient of determination is equal to 1, then correlation coefficient:
   1. must also be equal to 1 B) can be either -1 or 1

C) can be any value between -1 to 1 D) must be -1

**ANS- B**

1. Which of the following plot is best suited for the linear relationship of continuous variables?
   1. Scatter plot B) Histograms

C) Pie charts D) All of the above

**ANS-A**

1. The ratio of MSR/MSE produces:
   1. t-statistics B) f-statistics

C) z-statistics D) None of the above.

**ANS-B**

1. Which of the following regularizations uses only L2 normalization for its penalty parameter?
   1. Lasso B) Elastic Nets

C) Ridge D) All of the above

**ANS-C**

# In Q9 to Q11, more than one options are correct, Choose all the correct options:

1. Which of the following statement/s are true for best fitted line?
   1. It shows the causal relationship between dependent and independent variables
   2. It shows the positive or negative relation between dependent and independent variables
   3. It always goes through origin
   4. It is a straight line that is the best approximation of the given data sets

**ANS-B, D**

1. Regularizations helps in:
   1. Reducing the training time B) Generalizing the test set

C) Automatic feature selection D) Grouping the data

ANS-

1. Linear regression can be implemented through:
   1. Normal Equation B) Singular Value Decomposition

C) Parity checks D) nodes

ANS-

# Q12 to Q15 are subjective answer type questions, Answer them briefly.

1. Explain R2 and adjusted R2 metrics?

Adjusted R2 and R2 both define that how well the model fits in the data point.

Adjusted R2 penalizes the model for using more features. In case we increase the number of features in training data the R2 will increase but adjusted R2 will only increase if the new feature adds value to the model.

For this reason, adjusted R2 is considered as a better evaluation metric than R2.

Adjusted R2 is always less than or equal to R2.

The formula to calculate adjusted R2 is as follows:



Where,

n = number of data points in the dataset

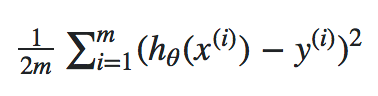
K = Number of features in the dataset excluding the constant term

1. Explain the cost function of linear regression?

The cost function may be defined as the difference between the actual values and the predicted values. It generates a real value.

It is a measure to define the ability of the model to draw a relationship between X and Y and helps the learner to correct and minimize the cost function.

Cost function is gives as;



Where;

m= no of samples

h0(x(i))= hypothetical value of x at i

1. Differentiate SSE, SSR and SST.

SSE stands for Sum of Squares Error

SSR stands for Sum of Squares Regression

SST stands for Sum of Squares Total

**SST** is the sum of squared difference between the observed values and the mean . It shows the variability of the dataset

**SSR** is the sum of difference between the predicted and mean values, If the values of SSR and SST becomes same we can say the model has captured all the points.

**SSE** is the sum of predicted and observed value it can also expressed as the difference between SST and SSR

|  |
| --- |
| y  SSE  Yi    SST  SSR    Ymean    x |

1. What are the various evaluation metrics for linear regression?

The evaluation matrices of Linear Regression are as follows;

R2 Score

Mean Squared Error

Root mean squared Error

Adjusted R2

Mean Absolute Error