MACHINE LEARNING – WORKSHEET 11
(LINEAR REGRESSION)

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

1. What happens to R^2 measure if we add a new feature?
   A) remains same   ➡ B) always increases   C) may or may not increase
             D) always decreases

2. The correct relationship between SST, SSR and SSE is given by:
   A) SSR = SST + SSE
   ➡ B) B) SST = SSR + SSE
   C) C) SSE = SSR – SST   D) None of the above

3. Residuals in regression analysis can be defined as:
   ➡ A) difference between the actual value and the predicted value.
   B) difference between the actual value and the mean of predicted value.
   C) difference between the actual value and mean of dependent variable.
   D) None of the above.

4. In a simple linear regression model, if we change the input variable by 1 unit, then how much output variable will change?
   A) By 1   B) No Change
   ➡ C) By its slope   D) None of the above

5. If the coefficient of determination is equal to 1, then correlation coefficient:
   ➡ A) 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

6. Which of the following plot is best suited for the linear relationship of continuous variables?
   ➡ A) Scatter plot
   B) Histograms
   C) Pie charts
   D) All of the above

7. The ratio of MSR/MSE produces:
   A) t-statistics   ➡ B) f-statistics
   C) z-statistics   D) None of the above.

8. Which of the following regularizations uses only L_2 normalization for its penalty parameter?
   A) Lasso   B) Elastic Nets
   C) Ridge   ➡ D) All of the above

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

9. Which of the following statement/s are true for best fitted line?

A) It shows the causal relationship between dependent and independent variables

→ B) It shows the positive or negative relation between dependent and independent variables

→ C) It always goes through origin

→ D) It is a straight line that is the best approximation of the given data sets

10. Regularizations helps in:
    A) Reducing the training time
    B) Generalizing the test set
→ C) Automatic feature selection
→ D) Grouping the data

11. Linear regression can be implemented through:
→ A) Normal Equation
→ B) Singular Value Decomposition
→ C) Parity checks
    D) nodes

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

12. Explain R^2 and adjusted R^2 metrics?

    R^2 shows how well terms (data points) fit a curve or line. **Adjusted R^2** also indicates how well terms fit a curve or line, but adjusts for the number of terms in a model. ... If you add more useful variables, **adjusted r-squared** will increase. **Adjusted R^2** will always be less than or equal to **R^2**.

13. Explain the cost function of linear regression?

    In the case of gradient descent, the objective is to find a line of best fit for some given *inputs*, or X values, and any number of Y values, or *outputs*. A cost function is defined as:

    a function that maps an event or values of one or more variables onto a real number intuitively representing some "cost" associated with the event.

14. Differentiate SSE, SSR and SST.

    SST is the maximum sum of squares of errors for the data because the minimum information of Y itself was only used for the baseline model. For the regression model, we square all the differences ③ $\hat{Y} - \bar{Y}$ and sum them up, which is called sum of squares due to regression (SSR), $\sum(\hat{Y} - \bar{Y})^2$. SSR is the additional amount of explained variability in Y due to the regression model compared to the baseline model. The difference between SST and SSR is remaining unexplained variability of Y after adopting the regression model, which is called as sum of squares of errors (SSE). SSE can be directly obtained by sum of squares of residual, $\sum(Y - \hat{Y})^2$.

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

The various metrics used to evaluate the results of the prediction are :

1. Mean Squared Error(MSE)

2. Root-Mean-Squared-Error(RMSE).

3. Mean-Absolute-Error(MAE).

4. R² or Coefficient of Determination.

5. Adjusted R²