

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
B) always increases
2. The correct relationship between SST, SSR and SSE is given by:
A) $SSR = SST + SSE$ B) $SST = SSR + SSE$ D) None of the above
C) $SSE = SSR - SST$
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.
A) difference between the actual value and the predicted value.
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
C) By its slope
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
A) must also be equal to 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
A) Scatter plot
7. The ratio of MSR/MSE produces:
A) t-statistics B) f-statistics C) z-statistics D) None of the above.
B) f-statistics
8. Which of the following regularizations uses only L2 normalization for its penalty parameter?
A) Lasso B) Elastic Nets C) Ridge D) All of the above
A) Lasso

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
- B) It shows the positive or negative relation between dependent and independent variables***
C) It always goes through origin
10. Regularizations helps in:
- A) Reducing the training time
 - B) Generalizing the test set
 - C) Automatic feature selection
 - D) Grouping the data
- B) Generalizing the test set***
C) Automatic feature selection
11. Linear regression can be implemented through:
- A) Normal Equation
 - B) Singular Value Decomposition
 - C) Parity checks
 - D) nodes
- A) Normal Equation***
B) Singular Value Decomposition

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

12. Explain R² and adjusted R² metrics?
- R-squared statistics or coefficient of determination is a scale invariant statistics that gives the proportion of variation in target variable explained by the linear regression model. The Adjusted R-squared takes into account the number of independent variables used for predicting the target variable/ In doing so, we can determine wheather adding new variables to the model actually increases the model fit.***
13. Explain the cost function of linear regression?
- a cost function is a measure of how wrong the model is in terms of its ability to estimate the relationship between x and y or in simple words it helps the learner to correct / change behaviour to minimize mistakes.***
14. Differentiate SSE, SSR and SST.
- The sum of squares total, denoted as SST , is the squared difference between the observed dependent variable and its mean.***
The second term is the sum of squares due to regression, or SSR. It is the sumof the mean of the dependent variable.
The last term is the sum of squares error, or SSE. The error is the difference between the observed value and the predicted value.
15. What are the various evaluation metrics for linear regression?
- There are multiple metrics:***
Mean squared error
Mean absolute error
Root mean squared error
r² score