

MACHINE LEARNING – WORKSHEET 11

1. B
2. B
3. A
4. C
5. B
6. A
7. B
8. C
9. B,C
10. B,D
11. A,C

12. R-Squared : R-squared, also known as the coefficient determination, defines the degree to which the variance in the dependent variable (or target) can be explained by the independent variable (features).

Adjusted R-Squared: Similar to R-squared, the Adjusted R-squared measures the variation in the dependent variable (or target), explained by only the features which are helpful in making predictions. Unlike R-squared, the Adjusted R-squared would penalize you for adding features which are not useful for predicting the target.

13. In ML, cost functions are used to estimate how badly models are performing. Put simply, 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. This is typically expressed as a difference or distance between the predicted value and the actual value. The cost function (you may also see this referred to as loss or error.) can be estimated by iteratively running the model to compare estimated predictions against “ground truth” — the known values of y.

14. SST: The sum of squares total, denoted SST, is the squared differences between the observed dependent variable and its mean.

SSR: The second term is the sum of squares due to regression, or SSR. It is the sum of the differences between the predicted value and the mean of the dependent variable. Think of it as a measure that describes how well our line fits the data. If this value of SSR is equal to the sum of squares total, it means our regression model captures all the observed variability and is perfect. Once again, we have to mention that another common notation is ESS or explained sum of squares.

SSE: The last term is the sum of squares error, or SSE. The error is the difference between the observed value and the predicted value. We usually want to minimize the error. The smaller the error, the better the estimation power of the regression. Finally, I should add that it is also known as RSS or residual sum of squares. Residual as in: remaining or unexplained.

15. 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. R^2 or Coefficient of Determination.
5. Adjusted R^2