Worksheet 5 MACHINE LEARNING

1.(5)

2.(C)

3.(c)

4.(C)

5.

6(a,d)

7.(b)

8.(d)

9.(c.b)

## 10. Adjusted R-Squared:-Adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. The adjusted R-squared increases when the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected. Typically, the adjusted R-squared is positive, not negative. It is always lower than the R-squared.

## The adjusted R-squared compensates for the addition of variables and only increases if the new predictor enhances the model above what would be obtained by probability. Conversely, it will decrease when a predictor improves the model less than what is predicted by chance.

## 11. . Differentiate between Ridge and Lasso Regression.:-

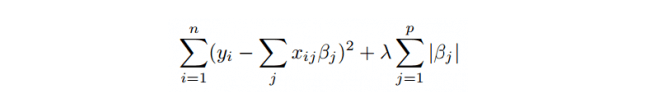
## Lasso Regression

The word “LASSO” denotes Least Absolute Shrinkage and Selection Operator. Lasso regression follows the regularization technique to create prediction. It is given more priority over the other regression methods because it gives an accurate prediction. Lasso regression model uses shrinkage technique. In this technique, the data values are shrunk towards a central point similar to the concept of mean. The lasso regression algorithm suggests a simple, sparse models (i.e. models with fewer parameters), which is well-suited for models or data showing high levels of multicollinearity or when we would like to automate certain parts of model selection, like variable selection or parameter elimination using feature engineering.

Lasso Regression algorithm utilises L1 regularization technique It is taken into consideration when there are more number of features because it automatically performs feature selection.

### Mathematical equation of Lasso Regression Algorithm:

Residual Sum of Squares + λ \* (Sum of the absolute value of the coefficients)  
The equation looks like:



Where,

* λ = the amount of shrinkage.
* If λ = 0 it implies that all the features are considered and now it is equivalent to the linear regression in which only the residual sum of squares is used to build a predictive model.
* If λ = ∞ it implies that no feature is used i.e, as λ gets close to infinity it eliminates more and more features and feature selection is more precise.
* When the bias increases, the value of λ increases
* When the variance increases, the value of λ decreases

## Ridge Regression

Ridge Regression is another type of regression algorithm in data science and is usually considered when there is a high correlation between the independent variables or model parameters. As the value of correlation increases the least square estimates evaluates unbiased values. But if the collinearity in the dataset is very high, there can be some bias value. Therefore, we create a bias matrix in the equation of Ridge Regression algorithm. It is a useful regression method in which the model is less susceptible to overfitting and hence the model works well even if the dataset is very small.

Where λ is the penalty variable. λ given here is denoted by an alpha parameter in the ridge function. Hence, by changing the values of alpha, we are controlling the penalty term. Greater the values of alpha, the higher is the penalty and therefore the magnitude of the coefficients is reduced.

We can conclude that it shrinks the parameters. Therefore, it is used to prevent multicollinearity, it also reduces the model complexity by shrinking the coefficient.

## 12. Variance Inflation Factor (VIF):-

A variance inflation factor (VIF) is a measure of the amount of multicollinearity in regression analysis. [Multicollinearity](https://www.investopedia.com/terms/m/multicollinearity.asp) exists when there is a correlation between multiple independent variables in a multiple regression model. This can adversely affect the [regression](https://www.investopedia.com/terms/r/regression.asp) results. Thus, the variance inflation factor can estimate how much the variance of a regression coefficient is inflated due to multicollinearity.

## Good VIF Value:-

As a rule of thumb, a VIF of three or below is not a cause for concern. As VIF increases, the less reliable your regression results are going to be.

13 Why do we need to scale the data before feeding it to the train the model?

. Answer:-so if the data in any conditions has data points far from each other, scaling is a technique **to make them closer to each other** or in simpler words, we can say that the scaling is used for making data points generalized so that the distance between them will be lower.

14. What are the different metrics which are used to check the goodness of fit in linear regression?

Answer:-

### R-squared

The difference between SST and SSE is the improvement in prediction from the regression model, compared to the mean model. Dividing that difference by SST gives R-squared. It is the proportional improvement in prediction from the regression model, compared to the mean model. It indicates the goodness of fit of the model.

improving the model’s fit. To remedy this, a related statistic, Adjusted R-squared, incorporates the model’s degrees of freedom.

### Adjusted R-squared

Adjusted R-squared will decrease as predictors are added if the increase in model fit does not make up for the loss of degrees of freedom. Likewise, it will increase as predictors are added if the increase in model fit is worthwhile.

Adjusted R-squared should always be used with models with more than one predictor variable. It is interpreted as the proportion of total variance that is explained by the model.

.

### The F-test

The F-test evaluates the null hypothesis that all regression coefficients are equal to zero versus the alternative that at least one is not. An equivalent null hypothesis is that R-squared equals zero.

A significant F-test indicates that the observed R-squared is reliable and is not a spurious result of oddities in the data set. Thus the F-test determines whether the proposed relationship between the response variable and the set of predictors is statistically reliable. It can be useful when the research objective is either prediction or explanation.

### RMSE

The RMSE is the square root of the variance of the residuals. It indicates the absolute fit of the model to the data–how close the observed data points are to the model’s predicted values. Whereas R-squared is a relative measure of fit, RMSE is an absolute measure of fit. As the square root of a variance, RMSE can be interpreted as the standard deviation of the unexplained variance. It has the useful property of being in the same units as the response variable.

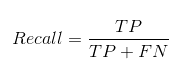
15,

https://cdn.analyticsvidhya.com/wp-content/uploads/2020/04/Equation_Accuracy.png

1000+50\1000+250+50+120=1420.05

Confusion Matrix Precision

1000\1000+250=251



1000\1000+1200=1201