# Machine Learning Assignment-6

1. C
2. B
3. C
4. B
5. B
6. A,D
7. B,C
8. D
9. A,B
10. The 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 only if 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 by chance.
11. Lasso regression stands for Least Absolute Shrinkage and Selection Operator. It adds penalty term to the cost function. In Ridge regression, we add a penalty term which is equal to the square of the coefficient. The L2 term is equal to the square of the magnitude of the coefficients.
12. Variance inflation factor (VIF) is a measure of the amount of multicollinearity in a set of multiple regression variables. Mathematically, the VIF for a regression model variable is equal to the ratio of the overall model variance to the variance of a model that includes only that single independent variable. This ratio is calculated for each independent variable.
13. Scaling, which is not as painful as it sounds, is a way to maintain a cleaner mouth and prevent future plaque build-up. Though it's not anyone's favourite past-time to go to the dentist to have this procedure performed, it will help you maintain a healthy mouth for longer.
14. Mean Squared Error (MSE). Root Mean Squared Error (RMSE). Mean Absolute Error (MAE) and R2 score/ adjusted R2 score.
15. Accuracy: 88%, Recall: 95%, Precision: 80%

# Sql\_worksheet\_6

1. A
2. B
3. B
4. B
5. D
6. B
7. A
8. C
9. D
10. A
11. Denormalization is a strategy used on a previously-normalized database to increase performance. In computing, denormalization is the process of trying to improve the read performance of a database, at the expense of losing some write performance, by adding redundant copies of data or by grouping data.
12. A database cursor can be thought of as a pointer to a specific row within a query result.
13. Navigational Search Queries, Informational Search Queries, Transactional Search Queries.
14. Constraints are used to limit the type of data that can go into a table. This ensures the accuracy and reliability of the data in the table.
15. Auto-increment allows a unique number to be generated automatically when a new record is inserted into a table. Often this is the primary key field that we would like to be created automatically every time a new record is inserted.

# Statistics Worksheet -6

1. D
2. A
3. D
4. C
5. C
6. A
7. B
8. B
9. A
10. Histograms are better in determining the underlying distribution of the data, box plots allow you to compare multiple data sets better than histograms as they are less detailed and take up less space. It is recommended that you plot your data graphically before proceeding with further statistical analysis.
11. The five easy steps listed below will enable them to systematically arrive at the appropriate metrics.

**Step 1** Why is the measurement required?  
**Step 2** What needs to be measured?  
**Step 3** What is the precision of measurement required?  
**Step 4** How will it be measured?  
**Step 5** What use will the measurement be put to? By whom?

1. Statistical significance can be accessed using hypothesis testing: – Stating a null hypothesis which is usually the opposite of what we wish to test (classifiers A and B perform equivalently, Treatment A is equal of treatment B)
2. Don’t know
3. If you have a skewed distribution, the median is often the best measure of central tendency.
4. Likelihood is the hypothetical probability that an event that has already occurred would yield a specific outcome. The concept differs from that of a probability in that a probability refers to the occurrence of future events, while a likelihood refers to past events with known outcomes.