

## MACHINE LEARNING

1. D
2. A
3. C
4. C
5. B
6. C
7. B
8. C
9. .
10. With that said random forests are a strong modelling technique and much more robust. They aggregate many decision tree to limit overfitting as well as error due to bias and therefore yield useful results.
11. This means that you are transforming your data so that it fits within a specific scale like 0-100 or 0-1. You want to scale data when you are using methods based on measures of how far apart data points are,
12. Gradient descent is an optimization algorithm which is commonly used to train machine learning models and neural networks. training data helps these models learn over time and the cost function within gradient descent specifically acts as a barometer gauging its accuracy with each iteration of parameter updates.
13. Accuracy is not a good metric for imbalanced datasets. This model would receive a very good accuracy scores as it predicted correctly for the majority of observations but this hides the true performance of the model which is objectively not good as it only predicts for one class
14. The f score also called the f1 score is a measure of a models accuracy on a dataset. It is used ton evaluate binary classification systems, which classify examples into positive or negative.  
$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$
15. The fit method is used to compute the mean and std dev for a given feature to be used further for scaling. the transform method is used to perform scaling using mean and std dev. Calculating using the. Fit() method. The fit\_transform () method does both fits and transform

## WORKSHEET 7 SQL

- 1- B
- 2- B,C
- 3- C
- 4- C
- 5- C
- 6- C
- 7- C
- 8- B
- 9- B
- 10- A
- 11- A join clause in SQL corresponding to a join operation in a relational algebra combines columns from one or more tables into a new table. Informally a join stitches two tables and puts on the same row records with matching fields .
- 12- Inner join, left join, right join, full join.
- 13- SQL server is a relational database management system developed by Microsoft, as a database server it is a software product with the primary function of storing and retrieving data as requested by other software applications.
- 14- The primary key constraint uniquely identifies each record in a table . primary keys must contain unique values and cannot contain null values. A table can have only one primary key and in the table this primary key can consist of single or multiple columns.
- 15- ETL stands for “extract transform, load”.

## STATISTICS WORKSHEET 7

- 1- B
- 2- D
- 3- B
- 4- D
- 5- A
- 6- .
- 7- C
- 8- B
- 9- D
- 10- A
- 11- A
- 12- C
- 13- C
- 14- B
- 15- B