WORKSHEET 6

STATISTICS:

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

are preferred to determine the underlying probability distribution of a data. Box plots on the other hand are more useful when comparing between several data sets.

Histograms is the bar graph gives frequency of data

Box plot gives us the minimun ,max values, frist,second and third quartile

Histogram uses .countplot

Box-plot uses .boxplot

Histogram gives the frequency

Box-plot helps to detect the variance in data as well as outliers in the data set

1. Steps to select metrics in ML:
2. Metrics tries to check accuracy of the model and improves its overall performance.
3. To select a metrics ,it is important to know the kind of problem ,one is dealing with.(classification or regression).
4. For classifications , weuse…confusion matrix(false positive, false negative),precision,accuracy(tp+fn/tp+tn+fp+fn) score,f1 score,recall=tp/tp+tn
5. For regression: MAE- mean absolute error -it represent sum of absolulte errors between absolute and predicted values.

MSE- mean squared error- it is the sum of square absolute differences between absolute and predicted values.

RMSE- It can b used for smaller values..it is square root of MSE.

1. examples :

1.amt of time the car battery lasts.

2. amt of time for which earthquake occurs.

3.amt of time to fully charge a inverter battery.

4.amt of time for which lightning lasts.

1. EXAMPLES:
   * + 1. Selecting which movie to watch ..median vl help which movies a group of people can watch so that all enjoy it.
       2. Grouping based on height: for army selecting ,it is done for grp f people with /above given height and weight.
       3. Selecting a property: it is helpful to select property which s benefical with lesser price.
2. The likelihood function **represents the probability of random variable realizations conditional on particular values of the statistical parameters**.

MACHINE LEARNING:

1. C
2. D
3. C
4. C
5. B
6. B
7. B
8. D
9. A
10. 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. **l**asso regression takes the magnitude of the coefficients, ridge regression takes the square.
12. The Variance Inflation Factor (VIF) measures the severity of multicollinearity in [regression analysis](https://corporatefinanceinstitute.com/resources/knowledge/finance/regression-analysis/). It is a statistical concept that indicates the increase in the variance of a regression coefficient as a result of collinear.

Variance inflation factor (VIF) is used to detect the severity of multicollinearity in the ordinary least square (OLS) regression analysis.

Multicollinearity inflates the variance and type II error. It makes the coefficient of a variable consistent but unreliable.

VIF measures the number of inflated variances caused by multicollinearity