1) D) All of Mentioned Above
2) A) Discrete
3) A) PDF
4) C) Mean
5) C) Empirical Mean
6) A) Variance
7) C) 0 to 1
8) B) Bootstrap
9) B) Summaried

13)

Count data: Data that represent the number of occurrences of a certain event, such as the number of cars passing through a toll booth in a given hour, or the number of customers in a store during a certain period. Count data often follow a Poisson or a negative binomial distribution, rather than a Gaussian or a log-normal distribution.

Categorical data: Data that are divided into categories or classes, such as the type of fruit in a basket or the color of a car. Categorical data are discrete and cannot be modeled using a Gaussian or a log-normal distribution.

Skewed data: Data that are not symmetric and have a long tail on one side, such as income distribution, which is often skewed to the right. Skewed data can be modeled using distributions such as the gamma, Weibull, or Pareto distributions.

14)

The median is a better measure than the mean in situations where the data is skewed, In such cases, the mean can be heavily influenced by extreme values or outliers

12

We perform a hypothesis test, which involves setting up a null hypothesis and an alternative hypothesis, and calculating a test statistic based on the observed data. null hypothesis mean no significant difference or relationship between the variables of interest, while the alternative hypothesis mean there is a significant difference or relationship

10

Histograms are useful for visualizing the shape of a distribution, such as whether it is symmetric, skewed, it is used to identify the outliers, it is used for continuous variable

boxplots are more useful for comparing distributions across groups or categories, it is also used for detection of outliers

11

Metrics shall be selected based on the type of problem we are solving,

If we are dealing with regression problem, we use Mean absolute error, mean squared error and root mean squared error

If we are dealing with classification problem we can use confusion matrix which involve precision, f1 score, recall and accuracy metric

15

Likelihood is a fundamental concept of statistics, it is the probability of happening a particular event like flipping coin here are 50% chance of getting head and 50% of chance of getting tail

STATS

1) B) Low R-squared value for train-set and High R-squared value for test-set
2) A) Decision trees are prone to outliers
3) C) Random Forest
4) A) Accuracy
5) B) model B
6) Ridge and Lasso
7) Random Forest and decision tree
8) Pruning and restricting the max depth of decision tree
9) A tree in the ensemble focuses more on the data points on which the previous tree was not performing well
10) The adjustment factor in the formula is (n-1)/(n-k-1), which increases as the number of predictors in the model increases, which penalizes the inclusion of unnecessary predictors

because it reduces the adjusted R-squared value as the number of predictors increases without a significant improvement.

11) Both Ridge and Lasso used forregularization techniques used in linear regression models to prevent overfitting and improve the model's predictive accuracy.
    Ridge regression does not perform feature selection
    Lasso regression performs feature selection by setting the coefficients of some features to zero
    Ridge regression balances the bias-variance tradeoff by reducing the variance of the model

12) VIF stands for Variance Inflation Factor, which is a measure of multicollinearity in a regression model. Multicollinearity occurs when there is a high correlation between two or more predictor

    it is best to keep the VIF value below 5 to avoid issues of multicollinearity

13) Scaling the features puts them on the same scale, which can help avoid bias in the model towards variables with higher magnitudes
    Scaling the features can reduce the computational complexity of some algorithms

14) 1) R-squared
    2) Root Mean Squared Error
    3) Mean Absolute Error
    4) Residual plots

15) Specificity = TN / (TN + FP) = 1200 / (1200 + 250) = 0.8276

    Precision = TP / (TP + FP) = 1000 / (1000 + 250) = 0.8

    Accuracy = (TP + TN) / (TP + FP + TN + FN) = (1000 + 1200) / (1000 + 250 + 1200 + 50) = 0.8889