Submitted By: ANUSHKUMAR K.

1. **Explain the linear regression algorithm in detail.**

   Answer:

   Linear Regression is a supervised machine learning algorithm that performs a regression task. This regression models a target prediction value based on independent variables. It helps in finding out the relationship and forecast them before hand. Linear regression performs the task to predict dependent variable y based on the given independent variable x. Hence this technique finds the linear relationship between x and y. The regression line which best fits the model is

   y = mx + c, where m is slope and c is the constant.

2. **What are the assumptions of linear regression regarding residuals?**

   Answer:

   The below are the assumptions of linear regression,

   1. Linear relationship between x and y
   2. Error terms are normally distributed
   3. Error terms are independent of each other
   4. Error terms have constant variance (homoscedasticity)

3. **What is the coefficient of correlation and the coefficient of determination?**

   Answer:

   Coefficient of correlation(r) is the quantity which measures the strength and direction of linear relationship of two variables. The linear correlation coefficient is sometimes referred as Pearson product moment correlation coefficient. The value of r ranges from -1 to +1.

   Coefficient of determination (r^2) is the square of r shows the percentage variation of one variable with all other variables. This always lies between 0 and 1. Higher the R2-squared value, the better is the model.

4. **Explain the Anscombe's quartet in detail.**

   Answer:

   Anscombe's quartet consists of 4 datasets each containing 11 (x, y) pairs. The essential thing to note is that these datasets share same descriptive statistics, but when we graph them they are changed completely. Each graph gives s different story irrespective of their similar summary statistics. Hence this emphasizes the importance of visualization in statistics.

5. **What is Pearson's R?**

   Answer:

   Coefficient of correlation(r) is the quantity which measures the strength and direction of linear relationship of two variables. The linear correlation coefficient is sometimes referred as Pearson product moment correlation coefficient. The value of r ranges from -1 to +1.

6. **What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?**

   Answer:

   Scaling is the data pre-processing activity which is applied to independent variables or features of the data. Scaling is done to have a normalized form of dataset which can be easily used for analysis.

   Scaling needs to be performed because,
   - Ease of interpretation
   - Faster convergence for gradient descent methods.

   Difference b/w normalized and standardized scaling:

   - Variables are scaled in such a way that their mean is 0 and standard deviation is 1 in standardized scaling
   - Variables are scaled in such a way that all their values lie between 0 and 1 using the minimum and maximum in the dataset.

7. **You might have observed that sometimes the value of VIF is infinite. Why does this happen?**

   Answer:

   The VIF value becomes to infinite when there is a perfect correlation between the variables. This leads to multicollinearity which is explained when two or more independent variables have a highly significant correlation. The accepted VIF value would be anything less than 5.

8. **What is the Gauss-Markov theorem?**

   Answer:

   Gauss-Morkov’s theorem states that in a linear regression model having uncorrelated errors have equal variances and expectation value of 0, the best linear unbiased estimator(BLUE) of the coefficients are given by ordinary least squares(OLS) estimator, provided it exists. The sample distributions are centered to actual population value and are the tightest possible distributions.

9. **Explain the gradient descent algorithm in detail.**

   Answer:

   Gradient descent algorithm is the first order iterative optimization for finding the max and min of a function. To find a local minimum of a function, steps to be taken proportional to the negative of the gradient of the function at the current point. If instead one takes the steps proportional to positive of the gradient local maximum would be achieved which is then called as gradient ascent.

10. **What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.**

    Answer:

    Q-Q plot is a probability plot which is a graphical method for comparing the distributions by plotting their quantiles. If 2 distributions being compared are similar, the points in Q-Q plot will lie approximately on the line y=x. If the distributions are linearly related, the points in the Q-Q plot need not necessarily lie on the line y=x, but will lie on any other line.

Advantages of using the Q-Q plot:

- The sample size does not need to be same or equal.
- Many distributional aspects can be tested simultaneously.