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

   Simple linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and dependent(y) variable. Machine learning algorithm that finds the best linear-fit relationship on any given data, between independent and dependent variables. It is mostly done by the Sum of Squared Residuals Method.

   Based on the given data points, we try to plot a line that models the points the best. The line can be modelled based on the linear equation shown below.

   y = mx + c

   The motive of the linear regression algorithm is to find the best values for m and c.

   It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering, and the number of independent variables being used.

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

   Below are the assumptions of linear regression regarding residuals:

   - Zero mean assumption: Residuals have a mean value of zero. There should be no correlation between the residual (error) terms. Absence of this phenomenon is known as Autocorrelation.
   - Normality assumption: The error terms must be normally distributed.
   - Independent error assumption: The independent variables should not be correlated. Absence of this phenomenon is known as multicollinearity.
   - Constant variance assumption: The error terms must have constant variance. This phenomenon is known as homoskedasticity. The presence of non-constant variance is referred to heteroskedasticity.

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

   The correlation coefficient(r) is a measure of the degree of linear association between two continuous variables, i.e. when plotted together, how close is the scatter of points to a straight line. Correlation simply measures the degree to which the two vary together.

   - A positive correlation indicates that as the values of one variable increase the values of the other variable increase.
   - A negative correlation indicates that as the values of one variable increase the values of the other variable decrease.

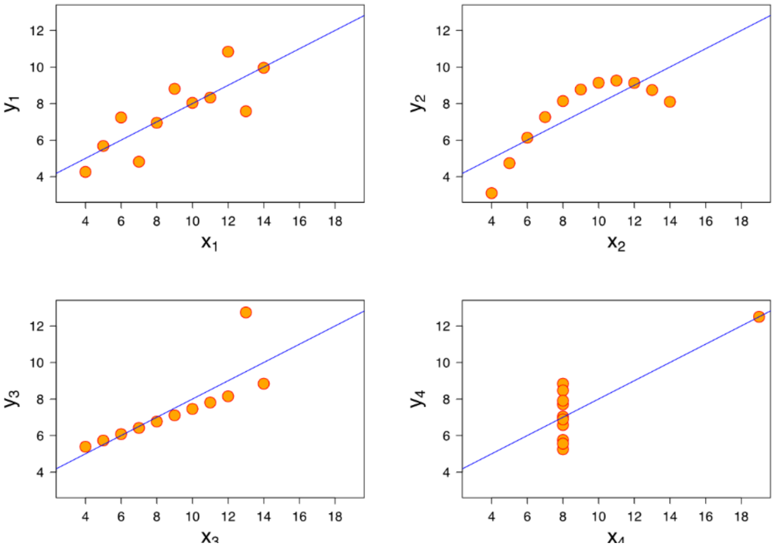
   The range of correlation coefficient is -1 to 1. Closer the value of r to 0, it means the variables are less correlated. Closer the r value to 1 means the variables are strongly correlated.

   The square of the r value is known as the coefficient of determination or r2, describes the proportion of change in the dependent variable Y which is said to be explained by a change in the independent variable X. Range of coefficient of determination is 0 to 1. If two variables have

an r value of 0.40, for example, the coefficient of determination is 0.16 and we state that only 16% of the change in Y can be explained by a change in X.
The larger the correlation coefficient, the larger the coefficient of determination, and the more influence changes in the independent variable have on the dependent variable.

4. Explain the Anscombe's quartet in detail.

Anscombe's quartet comprises four data sets that have nearly identical simple descriptive statistics yet have very different distributions and appear very different when graphed. Each dataset consists of eleven (x,y) points. They were constructed in 1973 by the statistician Francis Anscombe to demonstrate both **the importance of graphing data before analyzing** it and the effect of outliers and other influential observations on statistical properties.



5. What is Pearson's R?

Pearson's *r* correlation is used to assess the relationship between **two continuous variables**. Pearson's r is the most popular correlation test.

- Close to 1, there is a strong relationship between your two variables.
- Close to 0, there is a weak relationship between your two variables.
- Positive (+), as one variable increases in value, the second variable also increases in value. This is called a positive correlation.
- Negative (-), as one variable increases in value, the second variable decreases in value. This is called a negative correlation.

Pearson's Correlation Coefficient formula is as follows,

$$\mathbf{r} = \frac{\mathbf{n}(\sum \mathbf{xy}) - (\sum \mathbf{x})(\sum \mathbf{y})}{\sqrt{[\mathbf{n}\sum \mathbf{x}^2 - (\sum \mathbf{x})^2][\mathbf{n}\sum \mathbf{y}^2 - (\sum \mathbf{y})^2]}}$$

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

   Scaling is a method used to normalize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data preprocessing step. Scaling inputs helps to avoid the situation, when one or several features dominate others in magnitude, as a result, the model hardly picks up the contribution of the smaller scale variables, even if they are strong. But if you scale the target, your mean squared error is automatically scaled.

   Normalized Scaling

   Normalization scales the values of a feature into a range of [0,1].
   Xnew = (X – Xmin) / (Xmax – Xmin)
   Disadvantage of normalization over standardization is that it loses some information in the data, especially about outliers.
   It will be useful when we are sure enough that there are no anomalies (i.e. outliers) with extremely large or small values. For example, in a recommendation system, the ratings made by users are limited to a small finite set like {1, 2, 3, 4, 5}

   Standardized Scaling

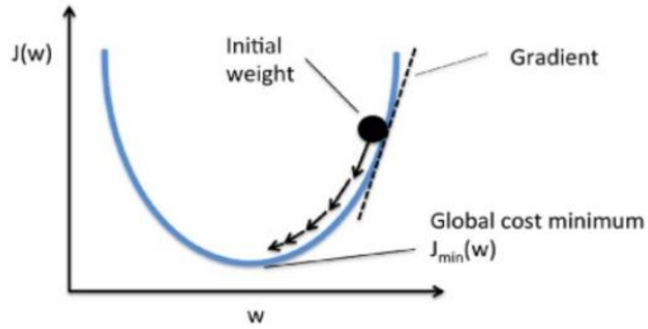
   Standardization refer to the subtraction of the mean (μ) and then dividing by its standard deviation (σ). Standardization transforms your data such that the resulting distribution has a mean of 0 and a standard deviation of 1.
   Xnew = (X - μ) / σ
   For most of the applications, standardization is recommended over normalization

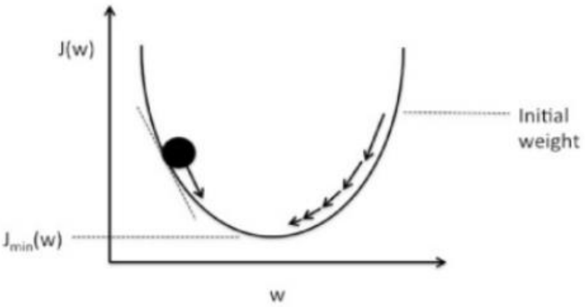
7. **You might have observed that sometimes the value of VIF is infinite. Why does this happen?**
   V.I.F. = 1 / (1 - R^2) - Value of VIF is infinite then it means denominator is 0 which in turn implies R^2 = 1 . This happens if the variables are redundant or in other words feature has a correlation with another variable

8. **What is the Gauss-Markov theorem?**
   The Gauss Markov theorem says that, under certain conditions, the ordinary least squares (OLS) estimator of the coefficients of a linear regression model is the best linear unbiased estimator (BLUE), that is, the estimator that has the smallest variance among those that are unbiased and linear in the observed output variables.

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

Gradient descent is an optimization algorithm used to find the values of parameters (coefficients) of a function (f) that minimizes a cost function (cost).Gradient descent is best used when the parameters cannot be calculated analytically (e.g. using linear algebra) and must be searched for by an optimization algorithm.
Gradient descent works like a ball rolling down a graph (ignoring the inertia). The ball moves along the direction of the greatest gradient and comes to rest at the flat surface (minima).





$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

Gradient Descent starts with a random solution, and then based on the direction of the gradient, the solution is updated to the new value where the cost function has a lower value.

10. **What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.**
    The Q-Q plot or quantile-quantile plot is a graphical tool to help us assess if a set of data plausibly came from some theoretical distribution such as a Normal or exponential. For example, if we run a statistical analysis that assumes our dependent variable is Normally distributed, we can use a Normal Q-Q plot to check that assumption.
    A Q-Q plot is a scatterplot created by plotting two sets of quantiles against one another. If both sets of quantiles came from the same distribution, we should see the points forming a line that’s roughly straight.