1. Explain the linear regression algorithm in detail.

   Its a method of finding the best straight line fitting to the given data, I.e finding the best linear relationship between dependent and independent variable. It is mostly done by the sum of square residual method.

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

   The assumption of linear regression are:
   1. It is assumed that there is a linear relationship between the dependent and independent variables. It is known as linearity assumption.
   2. Assumption about the residuals:
      1. It is assumed that the error terms are normally distributed.
      2. It is assumed that the residuals have a mean value of zero, I.e the error terms are normally distributed around zero.
      3. It is assumed that the residual terms has the same variance.
      4. It is assumed that the residual terms are independent of each other.
      5. It is assumed that the independent variables are measured without error.
      6. There is no multicollinearity in the data.

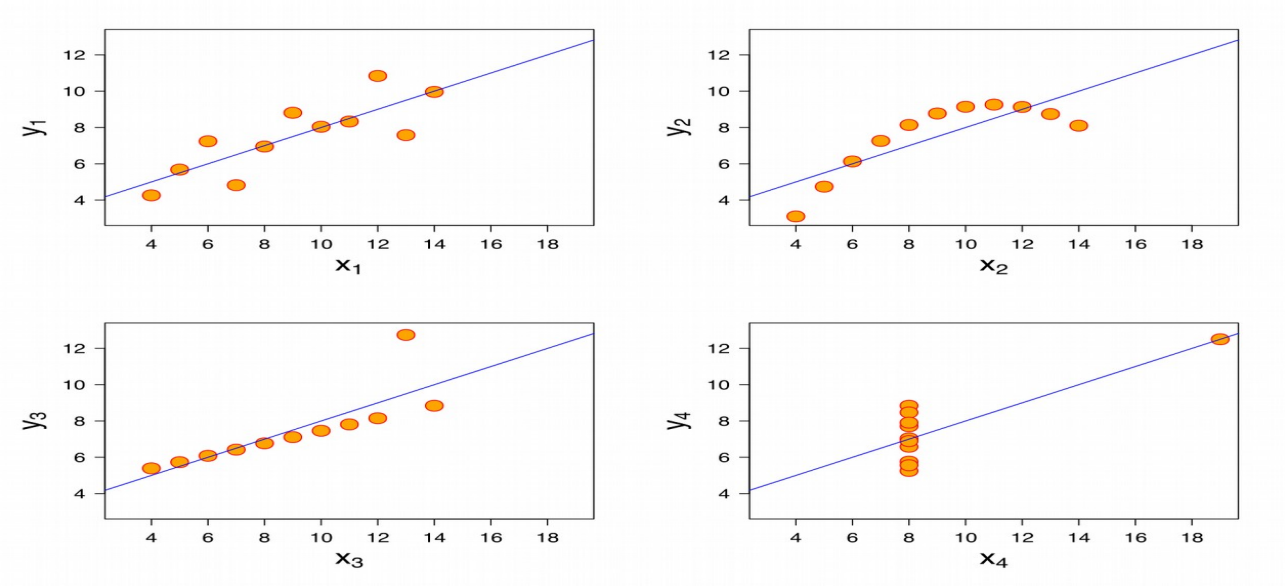
3. What is the coefficient of correlation and the coefficient of determination?

   A correlation coefficient is a numerical measure of some type of correlation, meaning a statistical relationship between two variables. The variables may be two columns of a given data set of observations, often called a sample, or two components of a multivariate random variable with a known distribution

   In statistics, the coefficient of determination, denoted R^2 or r^2 and pronounced "R squared", is the proportion of the variance in the dependent variable that is predictable from the independent 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.



5. What is Pearson's R?

   It is a measure of linear correlation between two variables X and Y. It has a value between -1 to +1. Where +1 is a positive correlation and means non linear and -1 means negative correlation.

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

   Scaling is a technique to standardize the independent features present in the data in a fixed range.
   If feature scaling is not done, then a machine learning algorithm tends to weigh greater value, higher and consider smaller values as the lower values, regardless of the unit of the values.
   Normalization rescales the values into a range of [0,1]. This might be useful in some cases where all parameters need to have the same positive scale. However, the outliers from the data set are lost. Standardization rescales data to have a mean (μ) of 0 and standard deviation (σ) of 1 (unit variance).

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

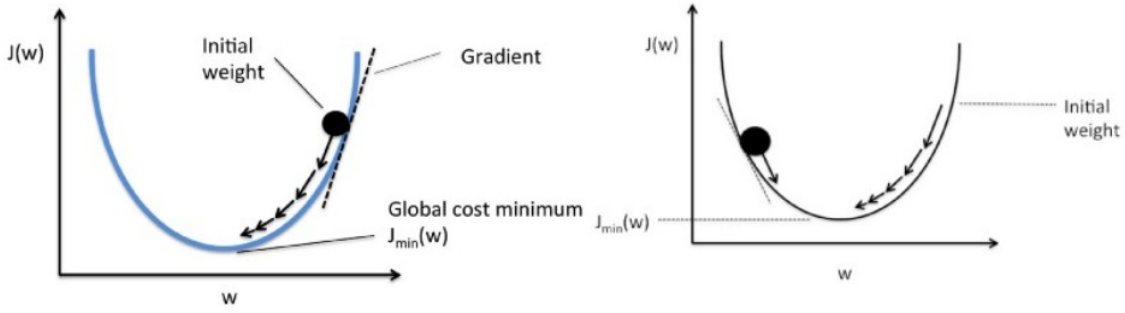
   An infinite VIF value indicates that the corresponding variable may be expressed exactly by a linear combination of other variables (which show an infinite VIF as well).

8. What is the Gauss-Markov theorem?

   In statistics, the Gauss–Markov theorem states that in a linear regression model in which the errors are uncorrelated, have equal variances and expectation value of zero, the best linear unbiased estimator of the coefficients is given by the ordinary least squares estimator, provided it exists

9. Explain the gradient descent algorithm in detail.
   Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model. Parameters refer to coefficients in Linear Regression.

   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).



Gradient Descent

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.

$$\theta_j = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}).\, x_j^{(i)} \quad \text{for j = 1,2,...,n}$$

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

    Q Q Plots (Quantile-Quantile plots) are plots of two quantiles against each other. ... The purpose of Q Q plots is to find out if two sets of data come from the same distribution. A 45 degree angle is plotted on the Q Q plot; if the two data sets come from a common distribution, the points will fall on that reference line