1. **Explain the linear regression algorithm in detail.**
   Linear regression is a method of finding the best straight-line fitting to the given data, i.e. finding the best linear relationship between the independent and dependent variables. In technical terms, linear regression is a 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.

2. **What are the assumptions of linear regression regarding residuals?**
   - Normality assumption: It is assumed that the error terms are normally distributed.
   - Zero mean assumption: It is assumed that the residuals have a mean value of zero, i.e., the error terms are normally distributed around zero.
   - Constant variance assumption: It is assumed that the residual terms have the same (but unknown) variance. This assumption is also known as the assumption of homogeneity or homoscedasticity.
   - Independent error assumption: It is assumed that the residual terms are independent of each other, i.e. their pair-wise covariance is zero.

3. **What is the coefficient of correlation and the coefficient of determination?**
   Coefficient of Correlation is the R Value of a regression output. R square is called as the coefficient of determination. Multiply R times R to get the R square value. In other words, Coefficient of Determination is the square of the Coefficient of correlation.
   R square or coefficient of determination shows percentage variation in y which is explained by all the x variables together. Higher the better. It is always between 0 and 1. It can never be negative – since it is a squared value.

4. **Explain the Anscombe's quartet in detail.**
   Anscombe's Quartet was developed by statistician Francis Anscombe. It comprises four datasets, each containing eleven (x,y) pairs. The essential thing to note about these datasets is that they share the same descriptive statistics. But things change completely when they are graphed. Each graph tells a different story irrespective of their similar summary statistics.

   Let's see the Quartets Summary Stats and Their Graphs:

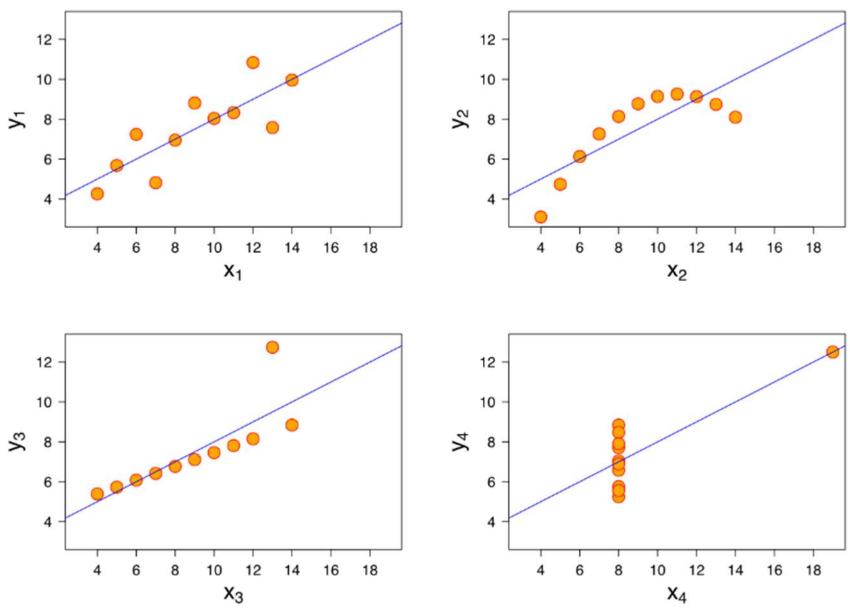
	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

Quartet's Summary Stats

The summary statistics show that the means and the variances were identical for x and y across the groups:

- Mean of x is 9 and mean of y is 7.50 for each dataset.
- Similarly, the variance of x is 11 and variance of y is 4.13 for each dataset
- The correlation coefficient (how strong a relationship is between two variables) between x and y is 0.816 for each dataset

When we plot these four datasets on an x/y coordinate plane, we can observe that they show the same regression lines as well but each dataset is telling a different story:



- Dataset I appear to have clean and well-fitting linear models.
- Dataset II is not distributed normally.
- In Dataset III the distribution is linear, but the calculated regression is thrown off by an outlier.
- Dataset IV shows that one outlier is enough to produce a high correlation coefficient.

This quartet emphasizes the importance of visualization in Data Analysis. Looking at the data reveals a lot of the structure and a clear picture of the dataset.

5. **What is Pearson's R?**
   Pearson's r is a numerical summary of the strength of the linear association between the variables. If the variables tend to go up and down together, the correlation coefficient will be positive. If the variables tend to go up and down in opposition with low values of one variable associated with high values of the other, the correlation coefficient will be negative.

6. **What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?**
   Feature 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 values, 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). If there is perfect correlation, then VIF = infinity

8. **What is the Gauss-Markov theorem?**
   The Gauss-Markov theorem states that if your linear regression model satisfies the first six classical assumptions, then ordinary least squares (OLS) regression produces unbiased estimates that have the smallest variance of all possible linear estimators.
   The critical point is that when you satisfy the classical assumptions, you can be confident that you are obtaining the best possible coefficient estimates. The Gauss-Markov theorem does not state that these are just the best possible estimates for the OLS procedure, but the best possible estimates for *any* linear model estimator.

9. **Explain the gradient descent algorithm in detail.**
   Gradient Descent is one of the most used algorithms in Machine Learning and Deep Learning. It is an optimization algorithm used in training a model. In simple words, Gradient Descent finds the parameters that minimize the cost function (error in prediction). Gradient Descent does this by iteratively moves toward a set of parameter values that minimize the function, taking steps in the opposite direction of the gradient.
   A gradient is a vector-valued function that represents the slope of the tangent of the graph of the function, pointing the direction of the greatest rate of increase of the function. It is a derivative that indicates the incline or the slope of the cost function.
   Let's explain this with an example:
   imagine you are in the top of a mountain. Your goal is to reach the bottom field, but there is a problem: you are blind. How can you come with a solution? Well, you will have to take small

steps around and move towards the direction of the higher incline. You do this iteratively, moving one step at a time until finally reach the bottom of the mountain.
That is exactly what Gradient Descent does. Its goal is to reach the lowest point of the mountain. The mountain is the data plotted in a space, the size of the step you move is the learning rate, feeling the incline around you and decide which is higher is calculating the gradient of a set of parameter values, which is done iteratively. The chosen direction is where the cost function reduces (the opposite direction of the gradient). The lowest point in the mountain is the value -or weights- where the cost of the function reached its minimum (the parameters where our model presents more accuracy).

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. A quantile is a fraction where certain values fall below that quantile. For example, the median is a quantile where 50% of the data fall below that point and 50% lie above it. 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.
In linear regression, the Q – Q plot tells the following:
A residual value is a measure of how much a regression line vertically misses a data point. ... You can think of the lines as averages; a few data points will fit the line and others will miss. A Q-Q plot has the Residual Values on the vertical axis; the horizontal axis displays the independent variable