1. Explain the linear regression algorithm in detail.

Ans : **Linear Regression** is one of the most widely used models which is used for linear relationship between the dependent and independent variables. We use Linear regression for **predictive analysis** and model building.

It is used where the output variable to be predicted is a continuous variable. Hence, it is a **regression** kind of Machine learning algorithm which comes under **supervised** ML. We have labels here which are continuous. For **example**: We can use it for predicting the scores of the students in PA course.

We first take the data set and divide it into the training and testing data sets and build the model on training data set and then apply that model in prediction on the testing data set which is nothing but a predictive analysis approach.

We have two **types of Linear Regression** :

1) **Simple linear regression** : Model with only 1 independent variable

2) **Multiple linear regression** : Model with more than 1 independent variables

In linear regression we have a cost function which is Residual sum of squared (RSS) and we need to minimize the same for which we can use differentiation(closed form method) or a gradient descent method.

Model formulation : $Y= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n$

Fitting a line : Minimize the RSS

Assessing the goodness of fit :

R squared = MSS/TSS = 1- RSS/TSS

Making predictions : For a new instance, predict the output

---

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

Ans : Assumptions of linear regression regarding residuals are as follows:

1) Error terms are **normally distributed** with **mean of Zero** (not the dependent and independent variables) : If they are not then the p-value obtained during the hypothesis test to determine the significance of the coefficients become unreliable.

2) Error terms are **independent** of each other

3) Error terms have constant variance(**homoscedasticity**) : The variance should not increase or decrease as the error value changes. The variance should not follow any pattern as the error terms change.

4) One more assumption which is not related to the error terms but related to linear regression is that the input variable and the target variable are **linearly dependent**.

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

Ans : **Coefficient of correlation** : is "R" value which is given in the summary table in the Regression output. It is the degree of relationship between two variables say x and y. It can take values in the range of -1 to 1 where 1 indicates that the two variables have perfect correlation. -1 means that the two variables are in perfect opposites. One goes up and other goes down, in perfect negative way. If they are not correlated then the correlation value can still be computed which would be 0.

Coefficient of determination : is the square of Coefficient of Correlation. That is nothing but the R-squared value = variability in Y(target) explained by model / the total variability in Y.

R squared = MSS/TSS = 1- RSS/TSS

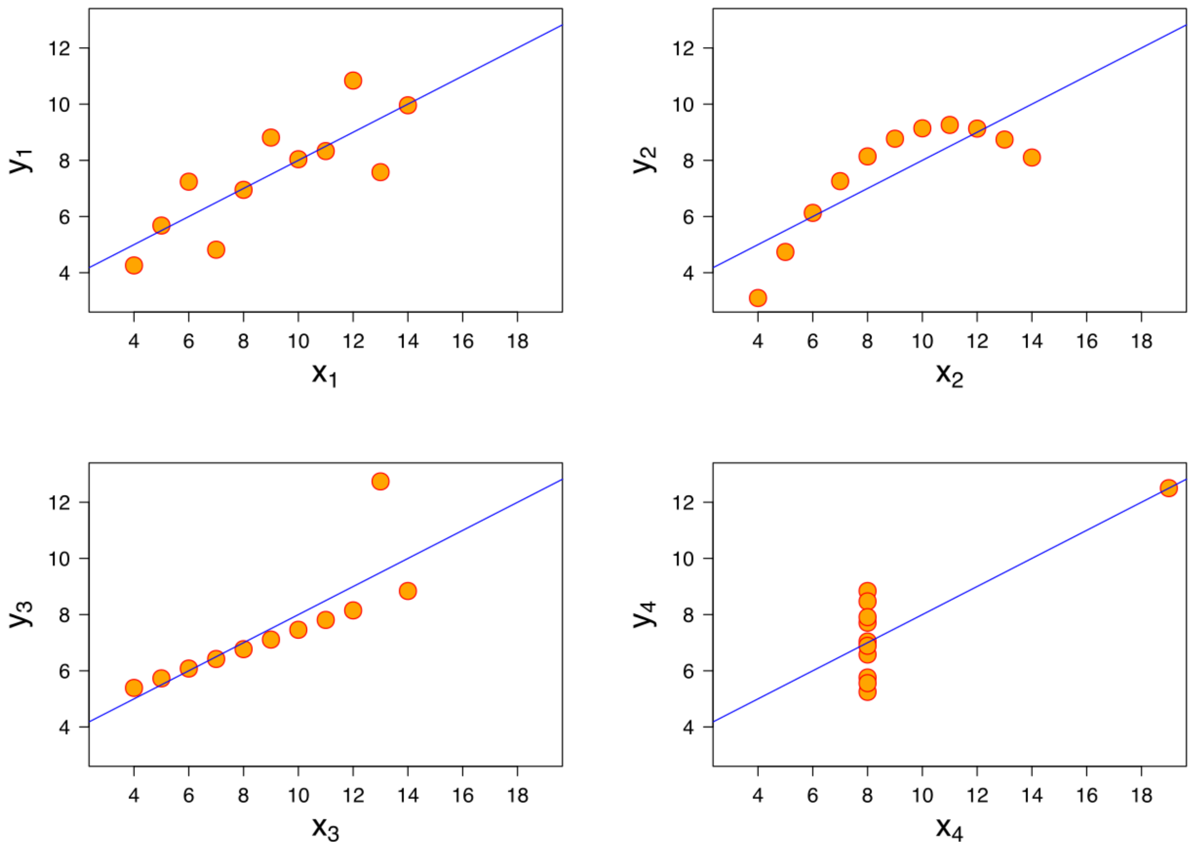
Higher the Coefficient of determination the better. It is always between 0 and 1. It can never be negative because it is a squared value.

----

4. Explain the Anscombe's quartet in detail.

Ans : 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 that are same sum, average and standard deviation. But things change completely**,** when they are graphed. Each graph tells a different story irrespective of their similar summary statistics.

	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



---

5. What is Pearson's R?

Ans : Pearson's R is a numerical summary of the strength of the linear association between the variables. **The Pearson's correlation** coefficient varies between -1 and +1 where:

r = 1 means the data is perfectly linear with a positive slope (i.e., both variables tend to change in the same direction)
r = -1 means the data is perfectly linear with a negative slope (i.e., both variables tend to change in different directions)
r = 0 means there is no linear association

---

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

Ans : **Scaling** : is a technique to standardize the independent features present in the data in a fixed range. It is performed during the data pre-processing to handle highly varying magnitudes or values or units. 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.

Why to perform scaling: If an algorithm is not using feature scaling method then it can consider the value 3000 meter to be greater than 5 km but that's actually not true and in this case, the algorithm will give wrong predictions. So, we use Feature Scaling to bring all values to same magnitudes and thus, tackle this issue.

Types of Scaling:

- **Normalized Scaling(Min-Max Scaling) :** This technique re-scales a feature or observation value with distribution value between 0 and 1.

$$X_{new} = \frac{X_i - \min(X)}{\max(x) - \min(X)}$$

- **Standardization :** It is a very effective technique which re-scales a feature value so that it has distribution with 0 mean value and variance equals to 1.

$$X_{new} = \frac{X_i - X_{mean}}{\text{Standard Deviation}}$$

---

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

Ans : An infinite VIF value indicates that the corresponding variable may be expressed exactly by a linear combination of other variables. In other words if they are having perfect correlation.

---

8. What is the Gauss-Markov theorem?

Ans : Gauss–Markov theorem states that the ordinary least squares (OLS) estimator has the lowest sampling variance within the class of best linear unbiased estimators(BLUE), if the errors in the linear regression model are uncorrelated, have equal variances and expectation value of zero. The errors do not need to be normal, nor do they need to be independent and identically distributed (only uncorrelated with mean zero and homoscedastic with finite variance).

---

9. Explain the gradient descent algorithm in detail.

Ans : It is an optimization algorithm which optimizes the cost function (which is RSS for linear regression) to reach the optimal solution. It is an iterative minimization method which reaches the minima step by step. It is basically used for updating the parameters of the learning model.

In gradient descent method we find local minima of the cost function by taking steps proportional to the negative of the gradient of the cost function at the current point.

We can use this method in logistic regression, linear regression and neural network as well.

We start with some value of θ^0 and calculate the new value say θ^1 with the help of the previous value of θ^0 as mentioned below:

$\theta^1 = \theta^0 - \eta * (\partial J/\partial \theta)$ at $\theta = \theta^0$

where ,

η = Learning rate which we generally use a very low value around 0.1, If we use larger values then there will be an oscillation and we might miss the local minima. So, to avoid that we use small learning rate.

J = Cost function

----

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

Ans : Quantile-Quantile (Q-Q) plot, is a graphical tool to help us assess if a set of data plausibly came from some theoretical distribution such as a Normal, exponential or Uniform distribution. Also, it helps to determine if two data sets come from populations with a common distribution.

Use and importance of a Q-Q plot in linear regression : This helps in a scenario of linear regression when we have training and test data set received separately and then we can confirm using Q-Q plot that both the data sets are from populations with same distributions.
