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

Ans 1:

y = mx + c is the basis of linear regression. Where y is the predicted value given the value of dependent variable (known variable) x and bias term c found by optimisation algorithm.

The algorithm finds the value of m and c such that predicted value y hat is very close to actual value y.

Underlying algorithm used could be **gradient descent** or **least squared** sum to find tuneable parameter m and c such that difference between actual and predicted value is least possible.

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

Ans2:

Residuals are error term we get for each predicted value across the sample.

Residual / error = y hat – y

Distribution of these term should be normal for the linear regression to hold.

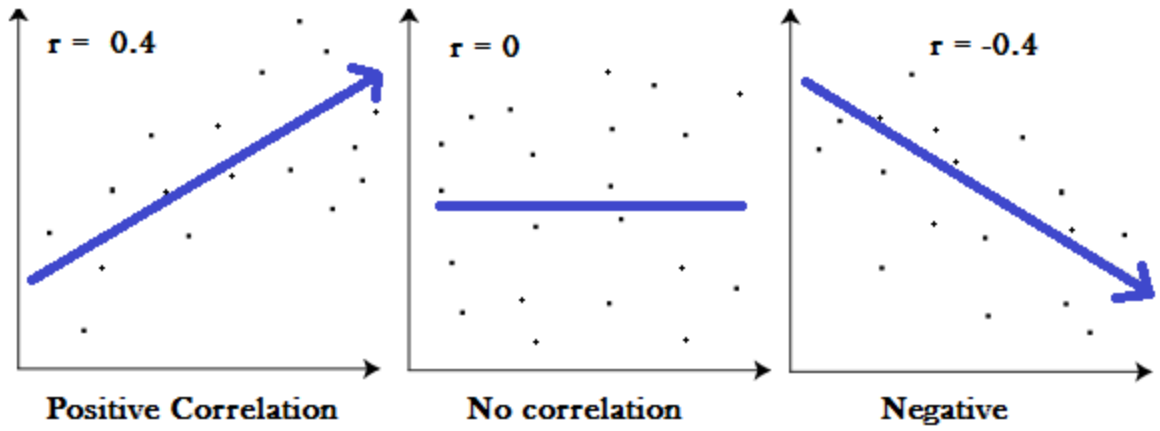
3. **What is the coefficient of correlation and the coefficient of determination?**

Ans3:

Correlation is numerical measure of how strongly two variable x and y are related.

It varies between -1 to 1 based on decrease or increase in y for increase or decrease in x.

Most used is Pearson Coefficient in linear regression.



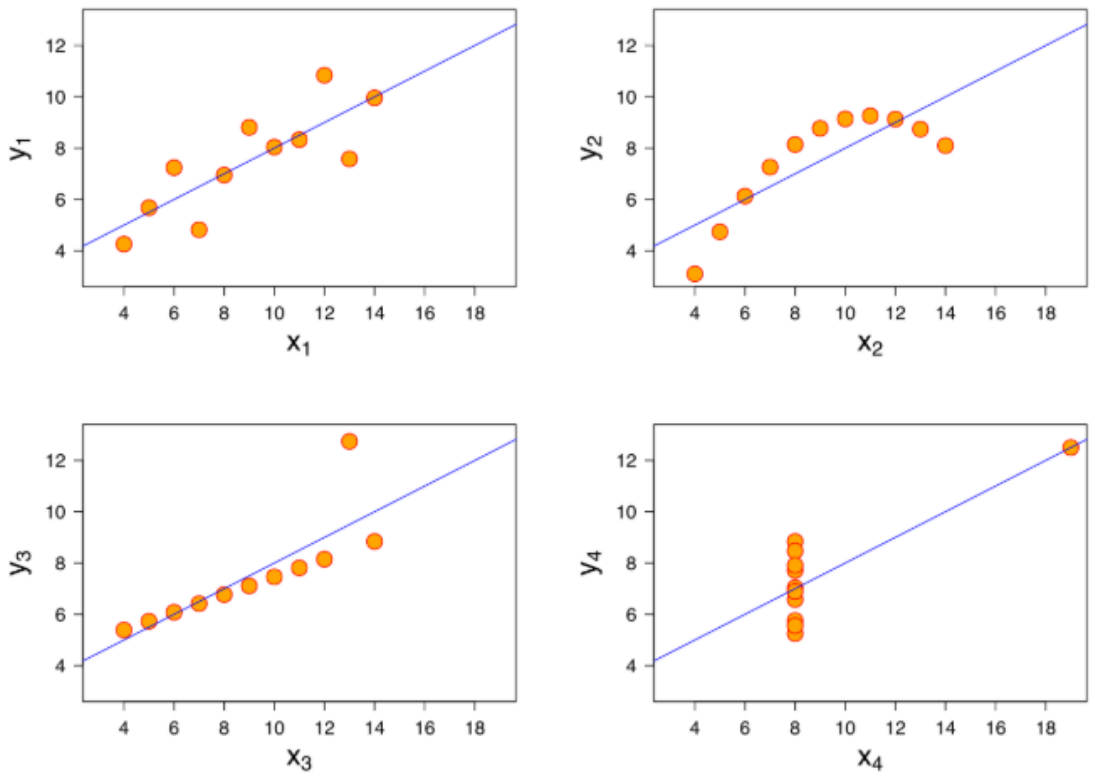
4: Explain the Anscombe's quartet in detail.

Ans 4:

Anscombe's quartet was developed by Francis Anscombe. **Four pair of datasets** were developed having **same mean, average and standard deviation** but when graphed show a very different relationship between x and y which linear regression fails to capture due to presence of outlier.

	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



5: What is Pearson's R?

Ans 5:

Pearson coefficient r which is numeric measure of how strongly y is related with x is given by r.

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

- n= number of the pairs of the variable x and y
- ∑xy = sum of products of the paired variable x and y
- ∑x = sum of the x scores
- ∑y= sum of the y scores
- ∑x2 = sum of the squared x scores
- ∑y2 = sum of the squared y scores

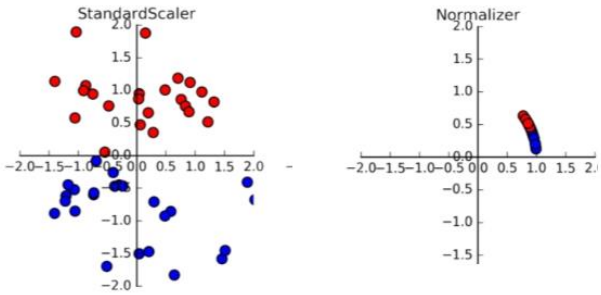
6: What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?

Ans 6:

Scaling is like change the values in centimetre to kilometres. In machine learning it is used to help the optimizer converge faster and get tuneable parameters quickly.

Standard Scaler: It changes the feature values such that it has 0 mean and 1 variance.

Normalizer: Projects the each datapoints on circle / sphere of radius 1.



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

Ans 7:

When there is perfect correlation between variables the VIF = infinite. A large value of VIF is indicator of correlation between variables and in layman's term features are duplicate of each other and does not contain any new information.

8: What is the Gauss-Markov theorem?

Ans 8:

The **Gauss Markov theorem tells if below five assumptions hold** true, the ordinary least squares estimate for regression coefficients gives you the *best linear unbiased estimate (BLUE)* possible.

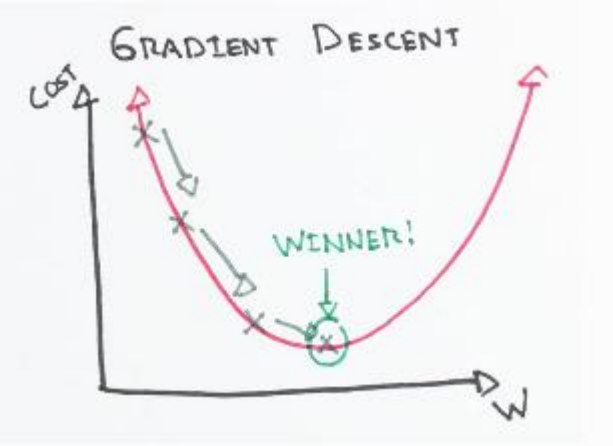
1. **Linearity**: the parameters we are estimating using the OLS method must be themselves linear.
2. **Random:** our data must have been randomly sampled from the population.
3. **Non-Collinearity:** the regressors being calculated aren't perfectly correlated with each other.
4. **Exogeneity**: the regressors aren't correlated with the error term.
5. **Homoscedasticity**: no matter what the values of our regressors might be, the error of the variance is constant.

9: Explain the gradient descent algorithm in detail.

Ans 9:

Gradient descent algorithm tries to minimise the below/cost function by getting the best possible values or m and b by moving in the direction of negative gradient and step equal to learning rate alpha. Mathematically we find m and b such that gradient is zero for this set (m and b) to get the local minimum.

$$f(m,b) = \frac{1}{N}\sum_{i=1}^{n}(y_i - (mx_i + b))^2$$



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

Ans 10:

This plot when linear confirms the assumption of liner regression that residuals should be normally distributed, or we need to pause and think if relationship could be described by linear regression.