1. Explain the linear regression algorithm in detail.

Linear regression literally means best fitting line. Although linear is not equal to line but shows the properties of linear relationship. Linear regression is zeroth algorithm for machine learning since it is base for logestic regression and all deep learning. There are hundred ways of performing linear regression. Simple but powerful algorithm

Linear regression is good for;

Finding relationship insights between two variables

How two variables are correlated?

How every variable effect other?

How accurate we can predict the variables?

How strong is the relationship is?

Linear regression uses linear systems

$$Y = \beta_0 + \beta_1 X$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

To find out the best suitable set of coefficients for number of given variables.

$$e_i = y_i - \hat{y}_i$$

We check the distance between the true and predicted one and nudge the coefficient values

$$RSE = \sqrt{\frac{1}{n-2} RSS} = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

We check performance of a model by RSE, Formula given above. We have gradient descent algorithm to adjust and fine tune with iterations. We use partial derivatives to calculate the coefficients.

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

Normality assumption: errors are normally distributed

Zero mean assumption: the error terms are normally distributed and they are peaked at zero and their addition is always a zero

Constant variance assumption: All error terms show same variance.

Independent error assumption: Residual terms are not related and independent of each other

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

   coefficient of correlation: Is a statistical measure to tell how two variables are related. How change in one variable will linked to other variable. A positive coefficient means a positive relation. If one variable increases other ones also increases.

   coefficient of determination: Simply R^2. It is interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variable.

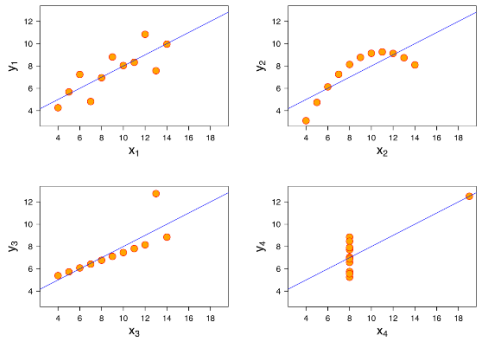
   Value range from 0-1 and 0 means model can't predict anything with variables and 1 mean its predicting without errors and both 0 and 1 are dangerous. Under fitting and overfitting. More is good but not all the way up.

$$R^2 = \{ (1 / N) * \Sigma [(x_i - x) * (y_i - y)] / (\sigma_x * \sigma_y) \}^2$$

4. Explain the Anscombe's quartet in detail.

   Linear regression has few rules before we insert data into algorithm. Linear regression is sensitive to outliers and linear regression can only explain linear relationship between the variables. It can't explain a exponential or polynomial relationship.

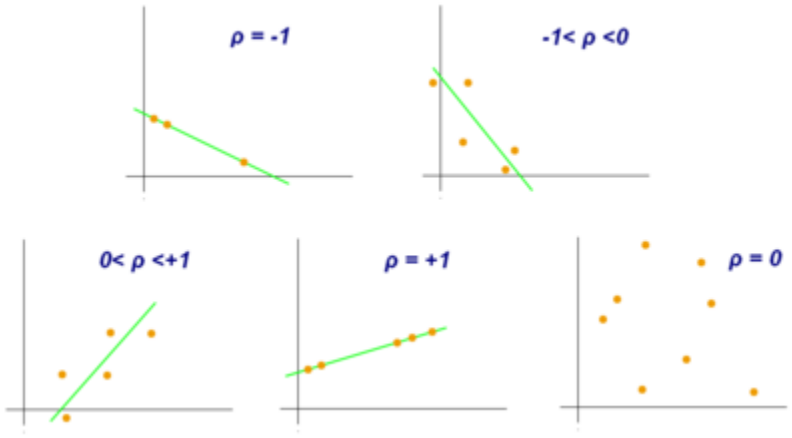
   Ansombe quartet will explain with example.



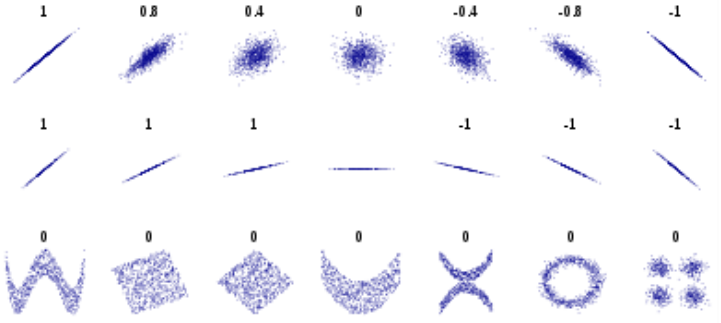
   As you can see how easy we can fool algorithm.

5. What is Pearson's R?

   As we have correlation between two variables but it doesn't always mean that they have linear relationship. Pearson R tells you linear correlation between variables. How one can predict other with linear relationship.



Linear relationship and p is linear correlation.



Pearson correlation will tell you what is happening with data.

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

   Data have variables and they have different magnitudes and distributions.

   For example: no of bedrooms and area while predicting house price.

   Every bedroom increases many feet of area. The Eucledian distance between the both of variables. So we scale it into a fixed range to bring them into same scale. It doesn't increase any accuracy but help algorithm converge faster.

   We have few types of scaling like, Standard scaling, min max scaling.

   Normalized scaling or normalization. Converts distribution into normal (bell shaped curve). Simple guassian distribution.

   Standard scaling or standardization or z score normalization.

   Standardizing, which means subtracting the mean and dividing by the standard deviation,
   You might have observed that sometimes the value of VIF is infinite. Why does this happen?

7. What is the Gauss-Markov theorem?

The Gauss Markov theorem says that, under certain conditions, the ordinary least squares estimator of the coefficients of a linear regression model is the best linear unbiased estimator, that is, the estimator that has the smallest variance among those that are unbiased and linear in the observed output variables. So we can get best OLS with less variance

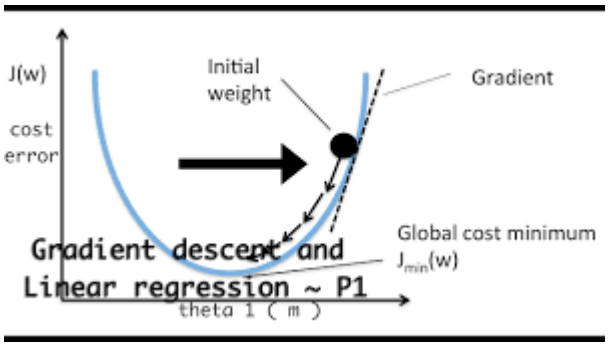
8. Explain the gradient descent algorithm in detail.

   Gradient descent algorithm is iterative optimization algorithm. Deals with first order derivatives. Used to find minimum of a function. It is used to decrease error in learning algorithms. This is most prominently used in almost all machine learning algorithm. Backpropagation uses it to nudge weights in neurons.

   For example we will consider linear regression.

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} \left(\frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right)^2 \right)$$

   Above given function decreases cost in iterations. It takes partial derivatives and nudges each coefficient independently.

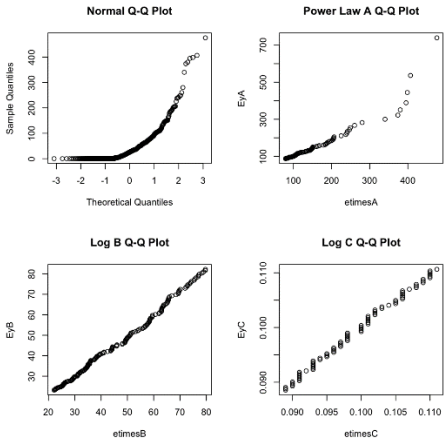


   As I said earlier it's an iterative optimization algorithm.

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

   Q-Q plot is probability distribution plot against two variables. Just like distribution plot but against two variables. Its most commonly used in statistics intervals are chosen. point on the plot links to one of the quantiles of the another distribution plotted against the same quantile of the first distribution.

   Below is image showing plot with different distribution plots.


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-3
-2
-1
0
1
2
3
0
100
200
300
400

Power Law A Q-Q Plot
EyA
etimesA
100
200
300
400
100
300
500
700

Log B Q-Q Plot
EyB
etimesB
20
30
40
50
60
70
80
30
40
50
60
70
80

Log C Q-Q Plot
EyC
etimesC
0.090
0.095
0.100
0.105
0.110
0.090
0.100
0.110