Linear Regression

Linear Regression is a **Supervised** Learning Algorithm.
Linear Regression is used for Predictive analysis on Samples

Basic Idea :

The Algorithm fits the best possible line in the Multi-Dimensional plot such that Sum of Squares of Distance between Y(Target Variable) and the Line is Minimum.

Hypothesis :

Hypothesis(H) is a Function which predicts the Target Variable (Y).In the Case of Linear Regression H is a Linear Function of Features/Inputs(X).

H (Hypothesis) = $\sum_{i=0}^{N}(\theta_i)(X_i)$ (Where, X_0 = 1)

θ — Weight/Parameter/Regression Coefficients
N — Number of Features/Inputs
[**Note :** H has N + 1 terms where θ_0 is Y-Intercept]

Cost Function :

$$J = (\tfrac{1}{2M})\sum_{i=0}^{M}(H_i - Y_i)^2$$

M — Number of Samples given in the Dataset
J is the Cost function Which is the average of errors between H and Y in the prediction.

Gradient Descent :

Gradient Descent Algorithm is used in order to reduce the Value of J.
θ is initialised to some value.

$$\frac{\partial J}{\partial \theta} = \frac{\partial}{\partial \theta}((\tfrac{1}{2M})\sum_{i=0}^{M}(H_i - Y_i)^2)$$

$$\frac{\partial J}{\partial \theta} = (\tfrac{2}{2M})(\sum_{i=0}^{M}(H_i - Y_i)(\frac{\partial}{\partial \theta}(\theta X_i - Y_i)))$$

$$\frac{\partial J}{\partial \theta} = (\tfrac{1}{M})(\sum_{i=0}^{M}(H_i - Y_i)(X_i)$$ (As Y is Constant)

$$\theta_j = \theta_j - \alpha\frac{\partial}{\partial \theta_j}(J)$$

$$\theta_j = \theta_j - (\tfrac{\alpha}{M})\sum_{i=0}^{M}(H_i - Y_i)(X_i)$$

Where J tends form 0 to N.
Here the θ(Weight) is updated such that Value of J becomes minimum.

The basic idea behind the Gradient Descent algorithm is to reach the Global minimum of the Cost Function(J).

Questions :
1) Is the Cost function (Mean Square Error) Convex or Non-Convex for Linear Regression?
   The Cost function here is Convex.
2) Why is (1/2) multiplied to Cost Function?
   (1/2) is multiplied so that the further calculation becomes easier as on taking derivative (2) gets multiplied.
3) Can the term "Epoch" be used in the above explanation?If so where and how can it be used?
   Epoch term can be used to represent the number of times a model undergoes training in a dataset. The θ is varied by passing N number of samples. Which means the Model undergoes N Epochs.

4) Does change Threshold value impact the Accuracy of the model?

Yes,The change in Threshold value impacts the Accuracy of the model.

5) What is assumed before creating this model?

Linear Relationship and No/Little Co-Relativity between Independent variables.