# Linear Regression- Assignment

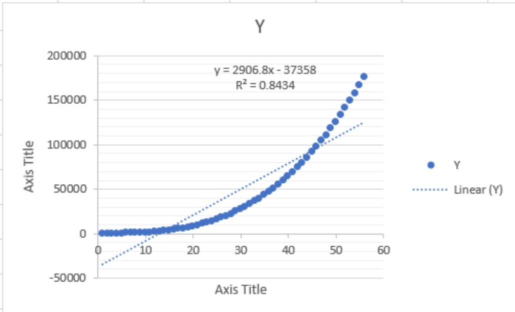
**Question-1. List down at least three main assumptions of linear regression and explain them in your own words. To explain an assumption, take an example or a specific use case to show why the assumption makes sense.**

The following are the assumptions of Linear Regression:

1. **Linearity**:
   The dependent variable should have linear relationship with the predictor variable, i.e. it should be proportionate to predictor variable such that if we increase the predictor by 1 unit, the dependent variable will either increase or decrease by a proportional value.

   **Example**: Consider if the relationship between the predictor and dependent variable is non linear which is given by relationship Y=X^3, In Linear Regression we fit a Line model for prediction, while the actual trend is a curve.

| X | Y | Y_Pred |
|---|---|---|
| 1 | 1 | -34451.2 |
| 2 | 8 | -31544.4 |
| 3 | 27 | -28637.6 |
| 4 | 64 | -25730.8 |
| 5 | 125 | -22824 |
| 6 | 216 | -19917.2 |
| 7 | 343 | -17010.4 |
| 8 | 512 | -14103.6 |
| 9 | 729 | -11196.8 |
| 10 | 1000 | -8290 |

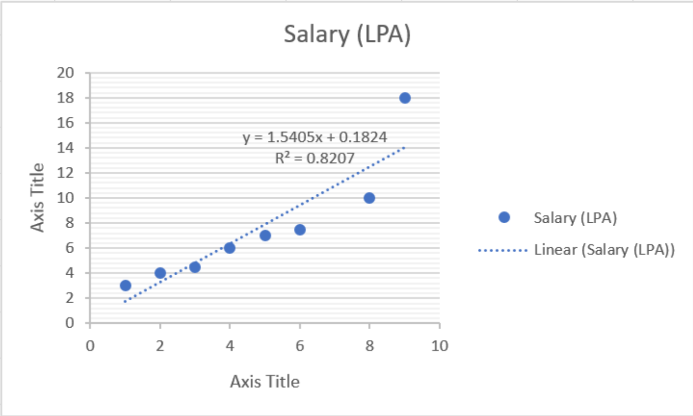



   Notice the Y_Pred is completely different from actual Y, which clearly shows that Linear Regression cannot be used to model this scenario.

2. **Outliers**:
   There should not be outliers in the training data, as the best fit line will get distorted because of the outliers, as it tries to minimize the distance between all the points including the outlier, which is having highest distance from the mean of the other data points.

   **Example**: Dataset for linear regression with 1 predictor variable no. of years of experience and Salary is taken with outlier and best fit line is calculated.

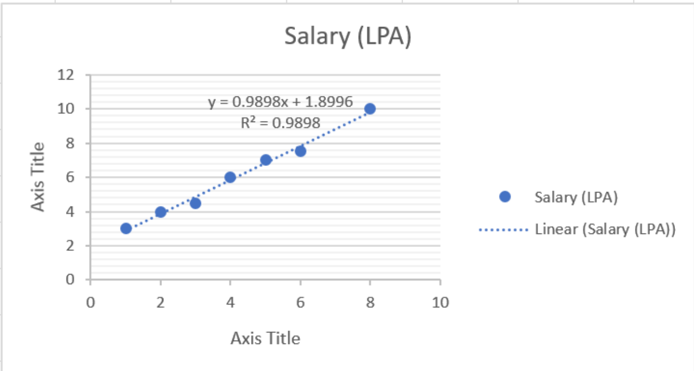
| No. of Years Experience | Salary (LPA) |
|---|---|
| 8 | 10 |
| 1 | 3 |
| 9 | 18 |
| 4 | 6 |
| 5 | 7 |
| 2 | 4 |
| 3 | 4.5 |
| 6 | 7.5 |




Notice the R Squared is 0.82. It would be better if we remove the outlier which is at 9 years of experience, the Salary is comparatively high 18 LPA, and is a outlier in the dataset.

Remove the outlier and fit the line:

| No. of Years Experience | Salary (LPA) |
|---|---|
| 8 | 10 |
| 1 | 3 |
| 4 | 6 |
| 5 | 7 |
| 2 | 4 |
| 3 | 4.5 |
| 6 | 7.5 |




Notice the model is performing better with R Squared of 0.98.

3. **Multicollinearity**:
   Multicollinearity is where there is correlation between different predictor variables and one or more variable can be derived by combining one or more predictor variables.
   It means there is redundancy in the predictor variables used. When we eliminate the Variables, we will be looking at top 15 or top 20, and this redundant variables will be given more weightage and some other important variable which must have been considered in the model will be eliminated, resulting in less performing model.

   **Example**:
   In the Linear Regression Assignment where we predict the price of car from other variables, Length and Width are positively correlated among themselves.

   Similarly, Wheel_Base, Length, Width, Weight, Engine_Size and Bore_Ratio form a cluster among themselves which needs to be carefully considered while including variables for the model, so that other variables will get their actual significance in the model.

**Question-2: Explain the gradient descent algorithm in the following two parts:**

1. **Illustrate at least two iterations of the algorithm using the univariate function J(x)=x2+x+1. Assume a learning rate η=0.1 and an initial guess x0=1 and demonstrate that the iterations converge towards the minima. Also, report the minima (which you can compute using the closed form solution).**

   Please refer image below where the problem is solved in paper:

   **Steps**:

   1. We start with initial value for x and a learning rate.
   2. Find the minima by differentiating the equation with respect to x and then equating to 0.
   3. With the above equation, find the value of x which is the Minima = -0.5.
   4. Then find the Minima using Gradient Descent method, and at step 13, we get -0.546 where we are close to the minima found by Closed form method.

$J(x) = x^2 + x + 1$

$\eta = 0.1$

$x_0 = 1$

Closed Form Solution to Compute minima :

$J'(x) = 2x + x + 1 = 0$

$\Downarrow$ Differentiation $\Downarrow$ Find Minima

$2x + 1 = 0$

$$\boxed{x_{opt} = -0.5}$$

Gradient Descent :

$$x_{new} = x_{old} - \eta \frac{\partial J}{\partial x} \Big|_{x = x_{old}}$$

$$= x_{old} - \eta J'(x)$$

$$= x_{old} - \eta (2x_{old} + 1)$$

$$x_{new} = 1 - 0.1(2(1) + 1)$$

$$\boxed{x_{new} = 0.7} \rightarrow \text{step } 1$$

$$x_{new} = 0.7 - 0.1(2(0.7)+1)$$

$$\boxed{x_{new} = 0.46} \Rightarrow \text{Step 2}$$

$$x_{new} = 0.46 - 0.1(2(0.46)+1)$$

$$\boxed{x_{new} = 0.268} \Rightarrow \text{Step 3}$$

$$x_{new} = 0.268 - 0.1(2(0.268)+1)$$

$$\boxed{x_{new} = -0.78} \Rightarrow \text{Step 4}$$

$$x_{new} = -0.78 - 0.1(2(-0.78)+1)$$

$$\boxed{x_{new} = -0.729} \Rightarrow \text{Step 5}$$

$$x_{new} = -0.729 - 0.1(2(-0.729)+1)$$

$$\boxed{x_{new} = -0.6832} \Rightarrow \text{Step 6}$$

$$x_{new} = -0.68 - 0.1(2(-0.68)+1)$$

$$\boxed{x_{new} = -0.64} \Rightarrow \text{Step 7}$$

$$\vdots$$

$$\boxed{x_{new} = -0.546} \Rightarrow \text{Step 13}$$

2. **Illustrate at least two iterations of the algorithm using the bivariate function of two independent variables J(x,y)=x2+2xy+y2. Assume a learning rate η=0.1 and an initial guess (x0,y0)=(1,1) . Report the minima and show that the solution converges towards it.**

   Please refer image below where the problem is solved in paper:

   **Steps**:

   1. We start with initial value for x and y, and a learning rate.
   2. Find the minima by differentiating the equation with respect to x and with respect to y and equating both to 0.
   3. With the above equations, find the values of x and y which is the Minima = (0,0).
   4. Then find the Minima using Gradient Descent method, and at step 1, we get – (0,0), where we get the minima found by Closed form method.

$$J(x,y) = x^2 + 2xy + y^2$$

$$\eta = 0.1$$

$$(x_0, y_0) = (1, 1)$$

Minima using Closed Form Solution:

$$\nabla J(x,y) = 0$$

$$\nabla J = \begin{bmatrix} \frac{\partial J}{\partial x} \\ \frac{\partial J}{\partial y} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{array}{l} \Rightarrow 2(x+y) = 0 \\ \Rightarrow 2(x+y) = 0 \end{array}$$

$$\boxed{x_{opt} = 0, \quad y_{opt} = 0}$$

Gradient Descent:

$$\begin{bmatrix} x^{new} \\ y^{new} \end{bmatrix} = \begin{bmatrix} x^{old} \\ y^{old} \end{bmatrix} - \eta \begin{bmatrix} \frac{\partial J}{\partial x} \\ \frac{\partial J}{\partial y} \end{bmatrix}_{x_{old},\, y_{old}}$$

$$\frac{\partial J}{\partial x} = 2(x+y) \qquad \frac{\partial J}{\partial y} = 2(x+y)$$

$$\begin{bmatrix} x^{new} \\ y^{new} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.1 \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x^{new} \\ y^{new} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \longrightarrow \text{Step } 1.$$