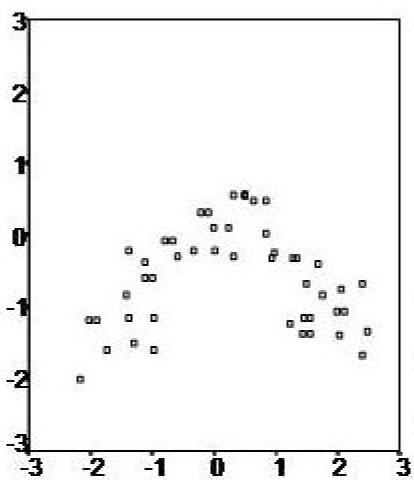
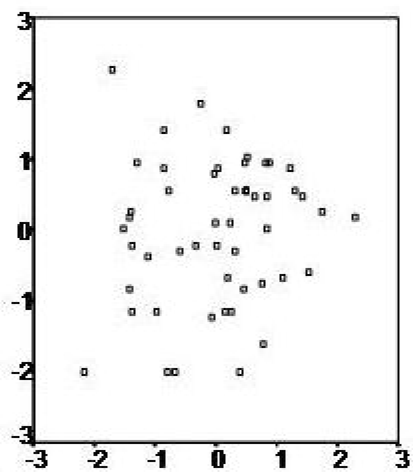
Question-1:
List down at least three main assumptions of linear regression and explain them in your own words. To explain an assumption, take an example or a specific use case to show why the assumption makes sense.

While there are a few assumptions made in Linear Regression, here are three such assumptions and their description. I would like to take the

- Linear regression needs the relationship between the independent and dependent variables to be linear





  - **Example:** The plots don't represent a linear relation as the first one shows a parabolic relation and the second one shows a scattered relation. If we try to build a model that would fit a straight line through all these points, the model will be completely overfit while training or it will be underfit. Hence it is assumed that the attributes that are used to build the model have to be linearly dependent on each other.

- Linear regression assumes that there is little or no multicollinearity in the data
  - Multicollinearity occurs when the independent variables are too highly correlated with each other. There are ways to identify this correlativity.
    - Correlation matrix – when computing the matrix on all independent variables the correlation coefficients need to be smaller than 1.
    - Tolerance – the tolerance measures the influence of one independent variable on all other independent variables. Tolerance is defined as $T = 1 – R^2$ for these first step regression analysis. With $T < 0.1$ there might be multicollinearity in the data and with $T < 0.01$ there certainly is.

        - Variance Inflation Factor (VIF) – the variance inflation factor of the linear regression is defined as VIF = 1/T. Higher the value, higher the multicollinearity.
    - **Example:** Variables like car height, car width and car length with the volume of the car which is a derived metric by doing a multiplication on all the three values. These values would be highly collinear because of the nature of the object it refers to.
- Linear regression analysis requires that there is little or no autocorrelation in the data. Autocorrelation occurs when the residuals are not independent from each other.
    - **Example:** In cross sectional data, if the change in the income of a person A affects the savings of person B (any person other than person A), then autocorrelation is present.

Question-2:

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).

$J(x) = x^2 + x + 1$ $\qquad \eta = 0.1$

$x_0 = 1$

Differentiating, $J(x)$

$\cancel{J'(x) = 2x + 1}$ $\quad J'(x) = 2x + 1$

Gradient Descent

Iteration 1 :

$$x_1 = x_0 - \eta \left.\frac{\partial J}{\partial x}\right|_{x = x_0}$$

$$= x_0 - \eta (2x_0 + 1) \Rightarrow \text{Substituting } x_0$$

$$= 1 - 0.1(2 \times 1 + 1)$$

$$= 1 - 0.3$$

$$x_1 = 0.7$$

Iteration 2 :

$$x_2 = x_1 - \eta \left.\frac{\partial J}{\partial x}\right|_{x = x_1}$$

$$= 0.7 - 0.1(2x_1 + 1)$$

$$= 0.7 - 0.1(2(0.7) + 1) = 0.7 - 0.24$$

$$x_2 = 0.46$$

Closed form soln: $\quad J'(x) = 0$

$$\therefore 2x + 1 = 0$$

$$\Rightarrow x = \frac{1}{2} = 0.5$$

Closed form soln $= 0.5$

$0.5 \approx 0.46 \longrightarrow$ Over iterations the value converges at the minima

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.

$$J(x,y) = x^2 + 2xy + y^2 \qquad \eta = 0.1$$

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

<u>closed form</u> :

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

$$\partial J/\partial x = 2x + 2y = 0 \implies x = -y$$

$$\partial J/\partial y = 2x + 2y = 0 \implies y = -x$$

1st <u>Iteration</u>

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} - \eta \left(\begin{bmatrix} \partial J/\partial x \\ \partial J/\partial y \end{bmatrix}\right)$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0.1 \begin{pmatrix} 2x_0 + 2y_0 \\ 2x_0 + 2y_0 \end{pmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0.1 \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.6 \end{bmatrix}$$

2nd iteration: $\cancel{\partial}$ $\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - \eta \begin{bmatrix} \partial J/\partial x \\ \partial J/\partial y \end{bmatrix}$

$$= \begin{bmatrix} 0.6 \\ 0.6 \end{bmatrix} - 0.1 \begin{bmatrix} 2x_1 + 2y_1 \\ 2x_1 + 2y_1 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.6 \end{bmatrix} - 0.1 \begin{bmatrix} 2.4 \\ 2.4 \end{bmatrix}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0.36 \\ 0.36 \end{bmatrix}$$

we can notice that both the values for x & y are the same on the two iterations.