REGRESSIONS IN MACHINE LEARNING

Neel Nakrani

CSE, BITS Edu Campus

WHAT IS REGRESSION: -

DEFINITION:-In statistical modelling, regression analysis is a set of statistical processes for estimating the relationships among variables.

=>In Statistical Analysis Regression is study of how independent variable affects the dependent variable to estimate or predict mean or average values of former in terms of the known or fixed values of latter. Here regression does not imply that the causes of dependent variable is independent variable it just finds the relationship between independent variable and dependent variable.

2)**TYPES OF REGRESSION: -**
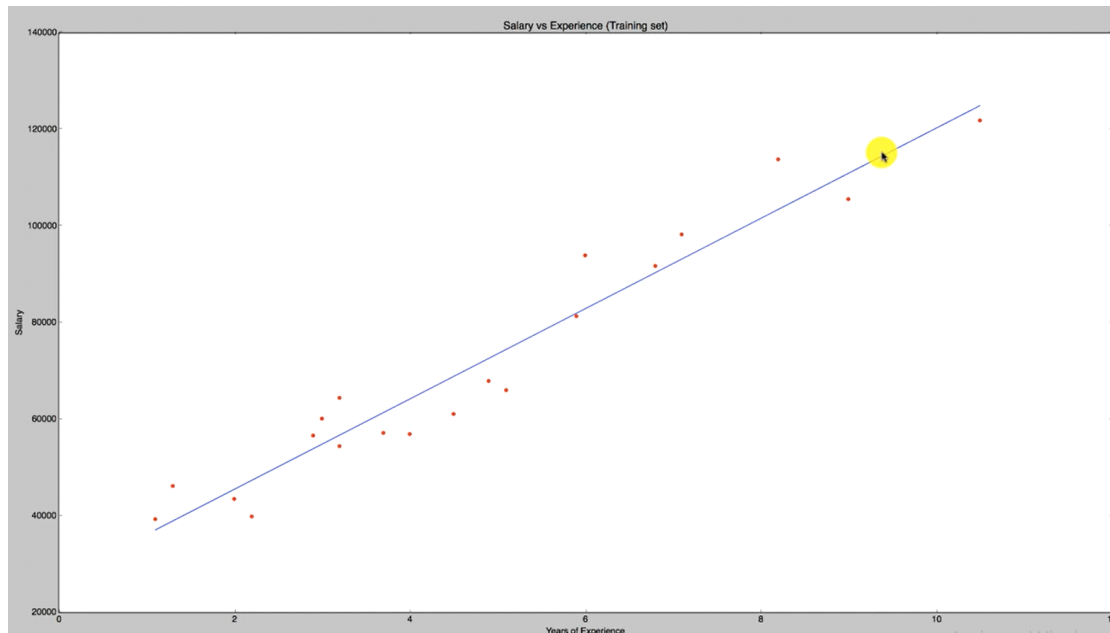
a) **Simple Linear Regression: -**

=>A simple Linear Regression is basically this formula
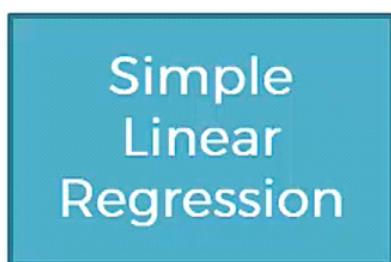
$$y = b_0 + b_1 x_1$$

where y is dependent variable, x_1 is independent variable and b_0 is constant. In simple linear regression there is only one independent variable which affects the value of dependent variable.

=>In below chart salary vs years of experience it is example of simple linear regression where red dots are real data and blue line is repressor line predicted by simple python program.

b) **Multiple Linear Regression: -**

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

Where X_1, X_2, X_3 . . . are independent variable and Y is dependent variable.

=>As seen in formula multiple linear regression is same as simple linear regression only that there are more variable so graph of multiple linear regression is same as simple linear regression.
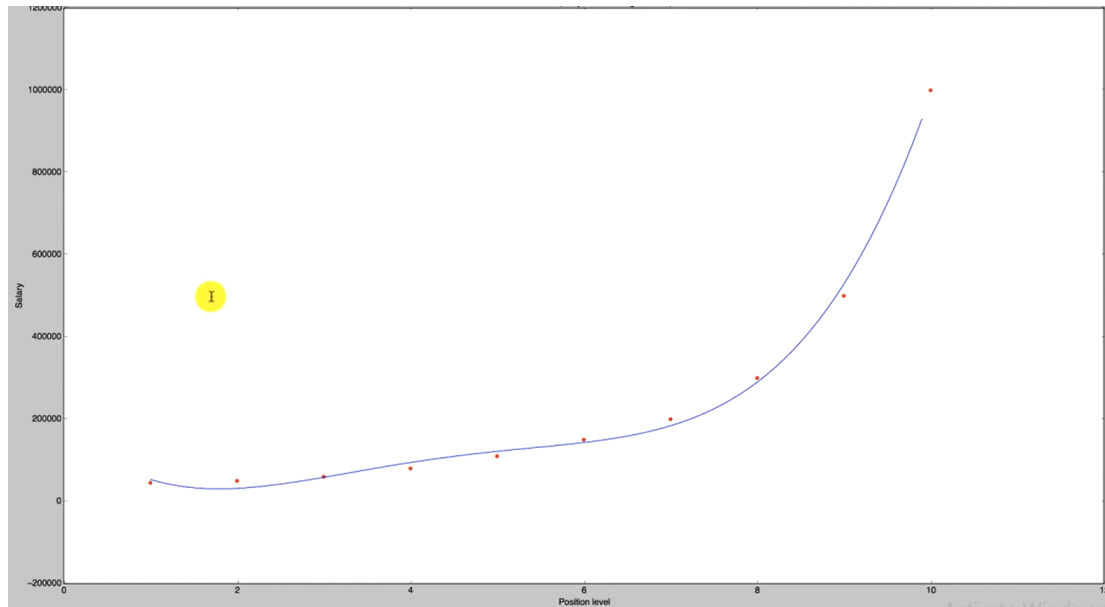
c)**Polynomial regression: -**

Polynomial Linear Regression

$$y = b_0 + b_1 x_1 + b_2 x_1^2 + \dots + b_n x_1^n$$

Here, there is only one independent variable but in different powers.

=>Below graph is about position level vs salary where red point are real data and blue curve is predicted regression by python program.


d)**Support Vector Regression:**

=>Support Vector Machines support linear regression and nonlinear regression that we can refer as SVR.

=>Instead of trying to fit the largest possible street between two classes while limiting margin violations, SVR tries to fit as many instances as possible on the street while limiting margin violations. Here width of street is controlled by hyper parameter Epsilon.

=>SVR performs linear regression on higher dimensional space.
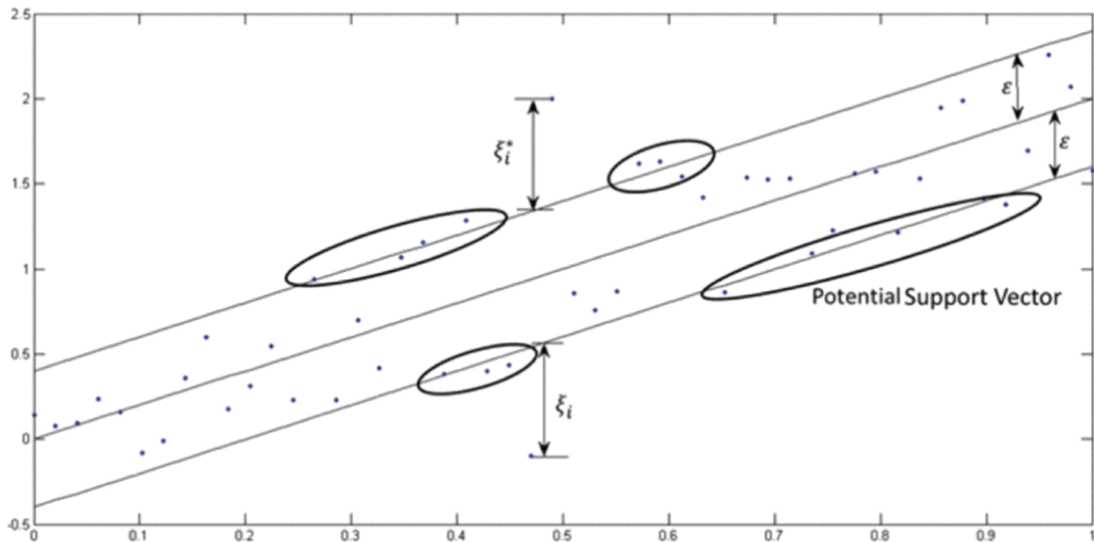
=>We can think of SVR as if each data point in the training represents its own dimension. When we evaluate the kernel between a test point and a point in the train set the resulting value gives the coordinate of our test point in that dimension.

=>The vector we get when we evaluate the test point for all points in the train set is called k vector which is simply a representation of test point in higher dimension.
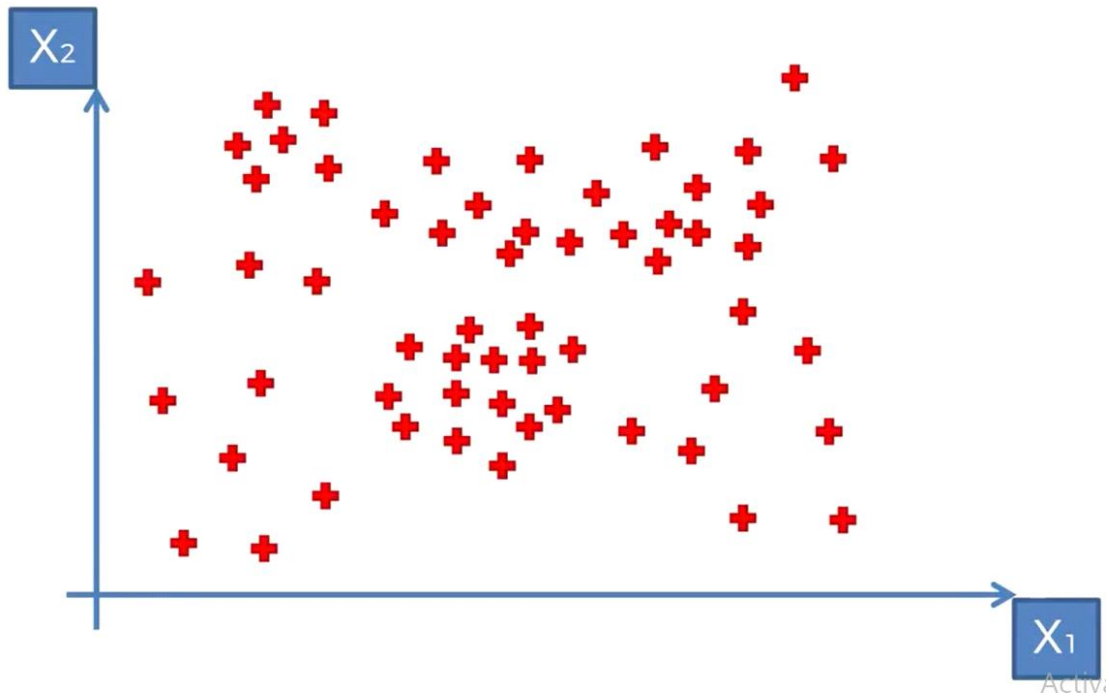
=>Once we have this vector in higher dimension we use it to perform linear regression.

=>Below is graph of SVR for position level vs salary. Here blue line is SVR predicted by program. Here red dot in right top corner is considered as outlier to increase accuracy.
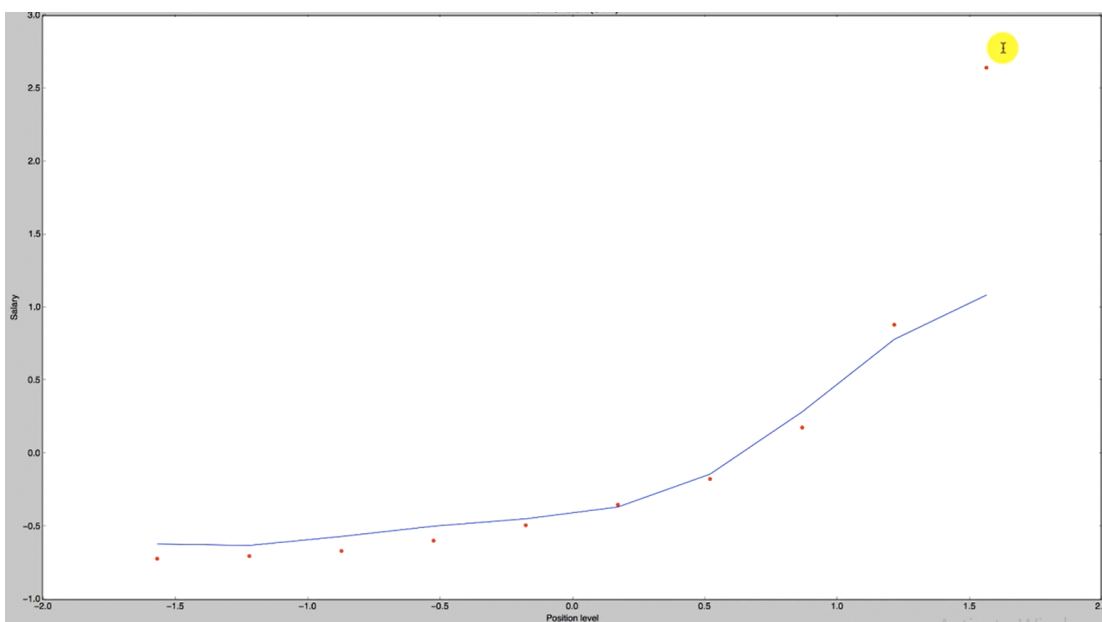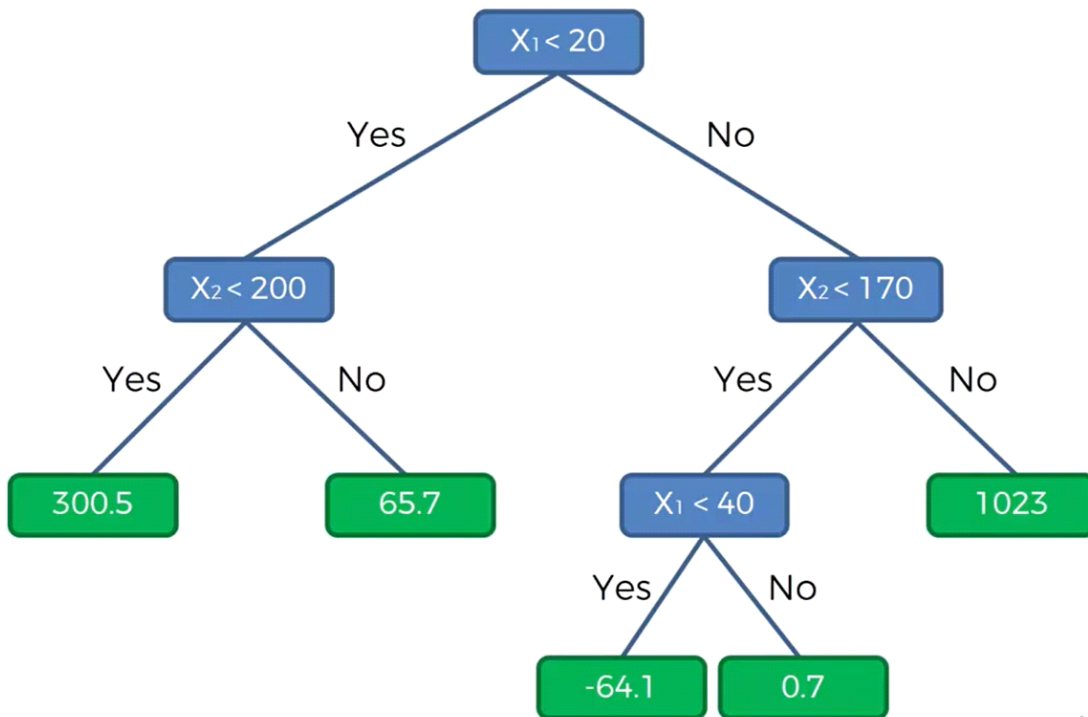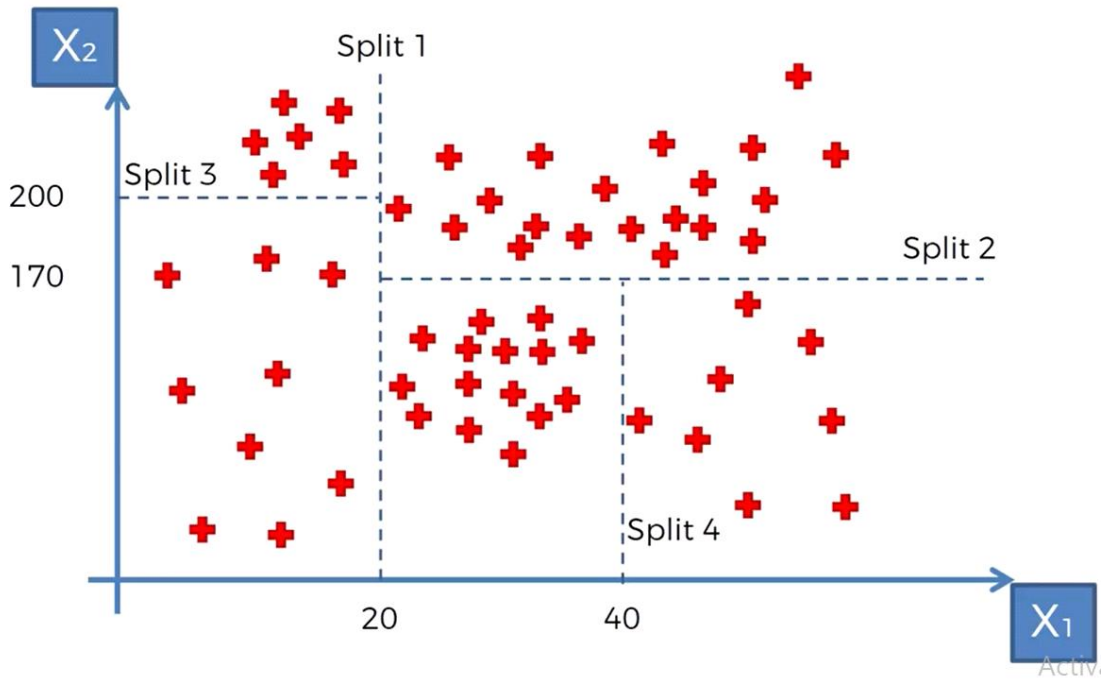
e) **Decision Tree Regression:**

=>Consider below graph.

=>Here x_1, x_2 both are independent variable and dependent variable(y) is on 3rd dimension.

=>Decision tree will divide this graph into some groups so that when we calculate average of y for that particular group it will increases the probability of predicted value to be true.as shown in below.

=>Here it is decision tree to our given graph where in blue boxes there are conditions for split and in green boxes there are average of particulargroup.

=>Below is graph of decision tree regression created by python program.
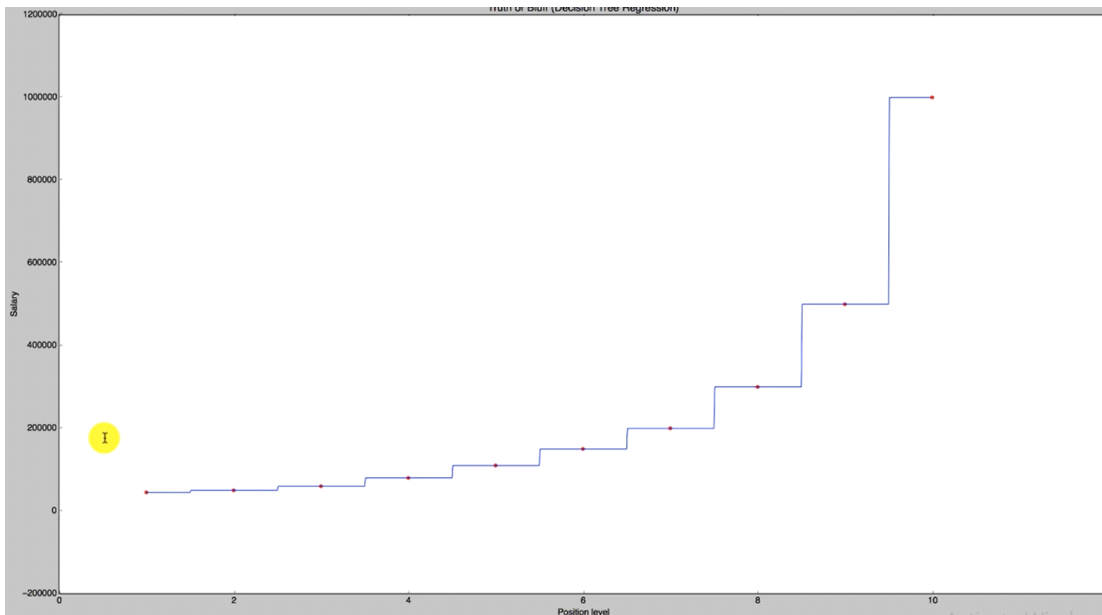
f) Random Forest regression:

=>In many applications decision tree regression is not so much accurate so to get more accurate predictions we use random forest regression where average of group of decision tree (i.e. 100) is taken.

=>Below is graph of random forest regression which looks same as decision tree regression. Here red dot in right most corner is considered as outlier to increase accuracy of prediction.