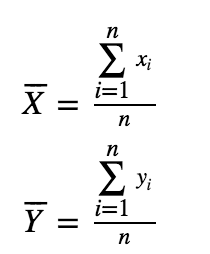
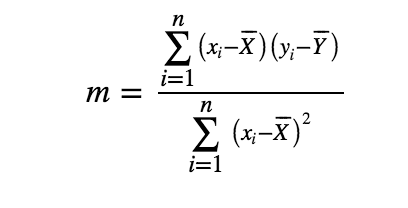
Fit a line to given data points. Linear regression, Root Mean Square & Prediction

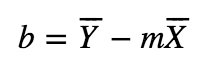
A line of best fit is a straight line that is the best approximation of the given set of data. A line of best fit can be roughly determined using an eyeball method by drawing a straight line on a scatter plot so that the number of points above the line and below the line is about equal (and the line passes through as many points as possible).

A more accurate way of finding the line of best fit is the least square method.

Use the following steps to find the equation of line of best fit for a set of ordered pairs (x1,y1),(x2,y2),...(xn,yn)(x1,y1),(x2,y2),...(xn,yn).

Step 1: Calculate the mean of the x-values and the mean of the y-values.  


Step 2: The following formula gives the slope of the line of best fit.  


Step 3: Compute the y-intercept of the line by using the formula.  


Step 4: Use the slope m and the y-intercept b to form the equation of the line.

Year House Prices  
1995 53807  
1996 55217  
1997 55209  
1998 55415  
1999 63100  
2000 63206  
2001 63761  
2002 65766

You can use a graph and pen & paper to solve this, if time allows do in python.

Draw a scatter plot for above data points on x-y graph, where x is year and y the house prices. You can convert the year to continuous points - 1 for 1995, 2 for 1996 and so on. You can start your coordinates from (0,0)  
Calculate the mean for x and y. Then find the slope - m and y-intercept - b as given in the line equation, as per the formula.  
Draw the line in a graph crossing those points.

Come up with a best-fit line equation using this.  
Do the prediction for the year 2010 and year 2017 using the line equation.