**Linear Regression**

**(A Matlab Tutorial)**

Linear regression is finding out the best fit to the data features. It is used in prediction by estimating the regression function i.e., when the regression function is estimated, any unknown data can be compared with that function line to estimate its class (in classification) or behavior (in prediction).

Suppose we have a dataset related to cars and we want to estimate a regression line for its features: Weight and Milage. The weights of 93 cars in a dataset is shown below



**Fig. 1: Weight plot.**

Similarly, the milage of the same cars (in same order as in Fig. 1) is shown in Fig. 2.



**Fig. 2: MPG plot**

Let’s draw a **relationship** between these two features of the cars’ dataset. This simple plot one feature vs the other i.e., weight vs milage.



**Fig.3: Relationship between Weight and MPG of cars**

From Fig. 3, we see that there is an inverse relationship between these two features. So, we want to fit a line that approximate this behavior. One this is for sure that the slope of the line will be negative. Let’s write the equation of the line.

Here the slope is and y-intercept is . The slope can be written as

Here is the correlation coefficient which will tell you about the positive or negative slope. Also and are the standard deviation of and respectively.

Or

Now to find y-intercept we have to modify the equation of line (1) as the data is not the points but vectors, so the best idea is to rewrite the equation of line as

With and are mean values of and respectively. So

Put (5) in (1) and solve

The regression line is given as

