# 1 Data Analysis

I began by taking a look at the data. There doesn't seem to be any missing values or anomalies with the given dataset. Non-numeric features like LoR and SoP have been given numeric values.

To ensure that all variables are on a similar scale, I normalized the data by creating a function, normalizer.

To understand the relationship between different variables, I then plotted all the variables with the chance of getting in, separately. It was clearly seen that they all follow a linear relationship with the chance of admit i.e. they were directly proportional. So I used the least squares method.

The model that I am making is of the form:

$f(a, b, c, d, e, f, g, x_1, x_2, x_3, x_4, x_5, x_6, x_7) = a \cdot x_1 + b \cdot x_2 + c \cdot x_3 + d \cdot x_4 + e \cdot x_5 + f \cdot x_6 + g \cdot x_7 + constant$

# 2 Linear Regression

I created a matrix with seven plus one rows (for the constant term) and took the transpose of the matrix to use in lstsq function. I then extracted the coefficient values and calculated the predicted chance of admission. I then plotted the predicted chance of admission for all of the data and the given chance of admission on the same plot. (shown in Figure)

As seen in the figure, it is a pretty accurate prediction.

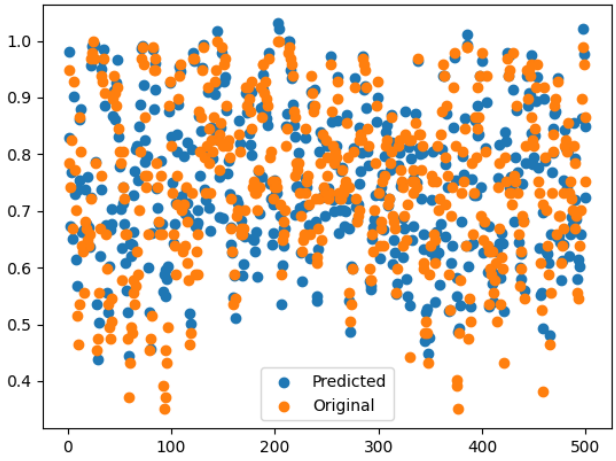


Figure 1: Predicted vs Original

## 3 Important Features

Now that we have the coefficients for the linear relationship between the factors and the chance of admission, we can compare them to see which is of higher importance (as they have already been normalized).

Here are the coefficients for each factor and the constant term:

- a = 0.6514352627870813 (GRE Score)
- b = 0.34366668759829977 (TOEFL Score)
- c = 0.030625608454519615 (University Rating)
- d = 0.008175966266838253 (SOP)
- e = 0.0869007337753541 (LOR)
- f = 1.210700752887382 (CGPA)
- g = 0.02505925627027436 (Research)
- constant = -1.3151804979350379 (Constant Term)

We can see that the most important factor is CGPA, followed by the GRE Score and then the TOEFL Score.

## 4 Top-Ranked Institutes

I also did the same analysis for top-ranked universities (rating 5) and got the same result, i.e., an aspiring student must direct their efforts towards CGPA, GRE Score, and then TOEFL Score.

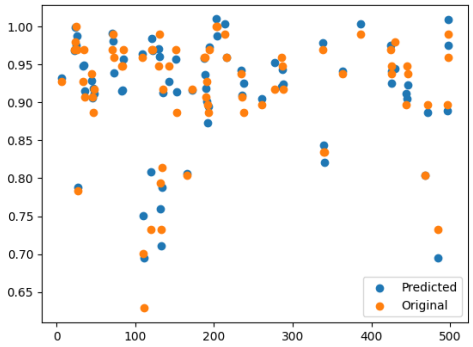


Figure 2: Universities rated 5