

Linear Regression Report
Graduation Admitance on a University

Luis Enrique Meza

Introduction:

Using a DataSet from the UCLA university, I developed a linear regression model that helps stimate how probable is for you to be admitted in the university, based on different aspects such as TOEFL Score, etc.

Data:

The data is for 7 parameters that end up on a Y value, being this one, the probability to be admitted on the university.

Also, I did a Coefficient Correlation calculation, in order to see how the parameters were linearly related to the result (which is the probability to be admitted).

I obtained that only 4 of 7 out of them, were symbolic. The other 3, which were SOP, LOR and CGPA, were just noise to my model, so in order to train, I decided to take them off.

	Chance of Admittance
<i>SOP</i>	0,024817219
<i>LOR</i>	0,079989628
<i>CGPA</i>	0,17054571
<i>Research</i>	0,291675821
<i>TOEFL Score</i>	0,379870683
<i>GRE Score</i>	0,41030869
<i>University Rating</i>	0,443283513

Normalization:

Before starting both the training and the testing, I needed to normalize the values in order for the model to work, because the numbers between parameters are a little bit apart from each other.

In order to do this process, I needed to calculate the Average from the Data, and the Standard Deviation as well:

	Averages:	Std_Dev:
X1	316,766917	11,4592706
X2	107,385965	6,05804449
X3	3,08521303	1,14424797
X4	17,6290727	16,2951038

X5	16,2180451	16,2085986
X6	699,882206	323,982681
X7	0,54636591	0,4984706
Y	66,5137845	22,6991286

Training:

In order to train, I used the Gradient Descend Model, and the Mean Square Error just to see the behavior of the model itself. Also, to lower the error and compare how my model was working.

Testing:

I split the samples into 80% of them being to train, and the other 20% was for testing the data:

The results that were given had around 20% error, meaning that it vary with more or less 80% accuracy.

Results:



Our results were the one expected. The error went down as I needed to do, and according to the graph of MSE (Mean Square Error). And also, the difference between the real data and the data from the tests was not that big.

Conclusion:

I concluded that my model worked pretty well taking into consideration that the parameters had a lot of noise and had a correlation coefficient to relation to the percentage of being admitted.