Assignment 4

Using least squares:

At first, we could use lstsq with 1 parameter, a combination of parameters and then all parameters. The least standard deviation occurs in case of using all parameters and so it's better to consider all factors to be affecting the chance . We also see that standard deviations while using single parameters are minimum for CGPA which means CGPA has high chance of linearly affecting the probability (or even greater powers). Also, the linear coefficient is highest for CGPA, then GRE, TOEFL, LOR, University ranking, Research, SOP which implies a greater weight lying on `CGPA, GRE, TOEFL` if parameters vary linearly. University ranking has positive coefficient as its more difficult to get into university with rank 1 than into rank 5. Similarly, higher GRE score/ TOEFL score/ CGPA affects the chance of entering positively

Standard deviation in using least squares:

0.05950420877764954

Coefficients of factors in scaled factors:

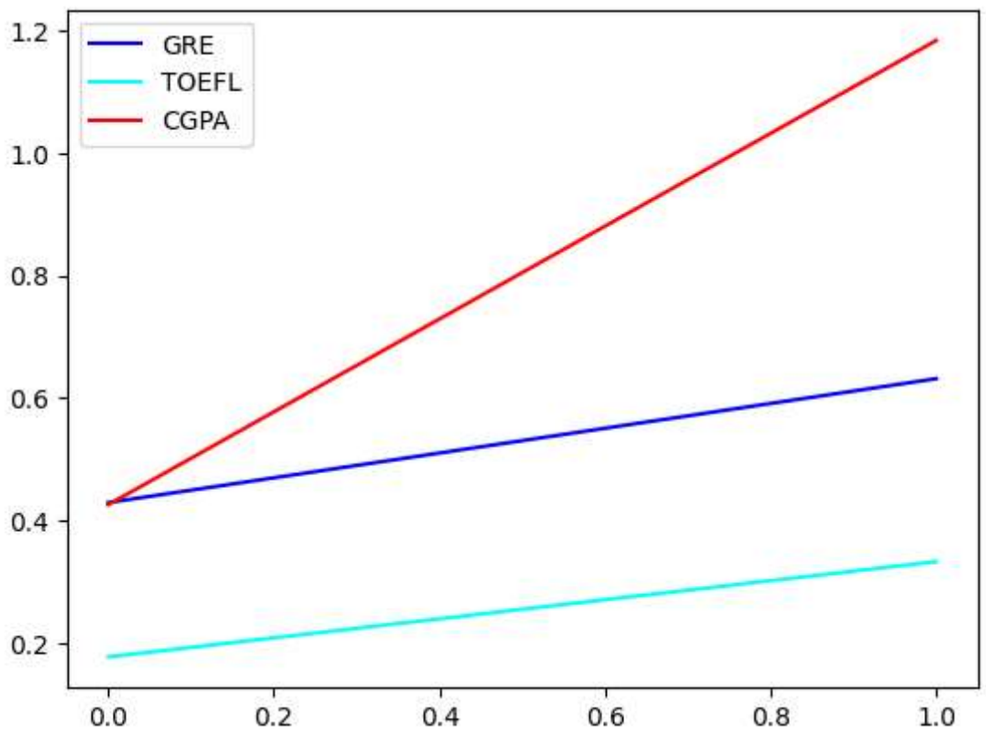
[0.6318922 0.33335669 0.02970684 0.00793069 0.08429371 1.18385053 0.02430748 -1.27572508]

Note: Increasing 1% of CGPA is more benefittable than increasing 1% of GRE then TOEFL if parameters affect linearly

Coefficients of factors in unscaled factors:

[0.00185851 0.00277797 0.00594137 0.00158614 0.01685874 0.11838505 0.02430748 -1.27572508]

It should be noted that increasing one unit of a factor, say CGPA give greater chance of admittance than increasing TOEFL then GRE. This makes sense as CGPA is on 10 while GRE is on 340 and TOEFL on 120.



This inference shows variation of CGPA, TOEFL & GRE from some minimum to maximum value and their corresponding chance of admission ignoring other factors

Using curvefit:

In curve fit, we could use any arbitrary function. To start with we could define each of the parameters raised to a certain power, as the function. Then we could alter powers to see if the standard deviation significantly drops. By this method, we would get varying powers. However using various functions like logarithms etc is also possible which could in fact give better results.

If we say change the powers of GRE, TOEFL and CGPA to 0.6, 0.4 and 2, we could get

Standard deviation:

0.05965940197351519

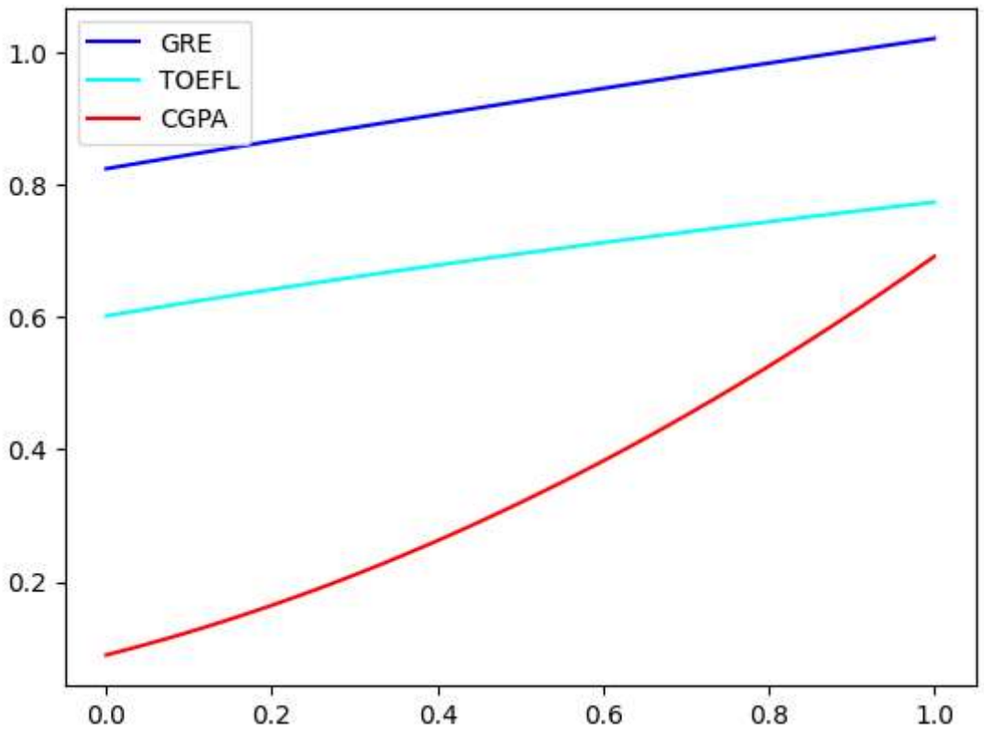
Coefficients

[1.01999345 0.77315859 0.02843898 0.00924342 0.08460998 0.69092552 0.02408125 -1.60101741]

It should be noted that the coefficients could be compared quantitatively only if all of them have same power and also all of them are normalised to a scale of 1

Thus to compare the parameters quantitatively we could plot a graph showing the variation of GRE scores, TOEFL scores and CGPA varying from (max-2*min) to the maximum value and their effect. Here minimum value is the minimum value obtained in given dataset. Assuming it takes equal amount of effort for a given person to increase their CGPA, GRE score/ TOEFL by some percentage from the minimum we have assumed and considering the person ignoring other criteria, this plot shows it is benefittable to increase GRE, then TOEFL than CGPA. However this is only inferred based on the model we have taken. It should also be noted that for a vast range of models especially linear models, CGPA criteria is more important than GRE

Thus the order varies as `GRE, TOEFL, CGPA` for this model



NOTE: This inference is assuming that minimum GRE score obtained is 238, TOEFL score obtained is 64 and minimum CGPA is 3.6