ASSIGNMENT 4 : Data Analysis

I and my classmate Ilin (EE22B094) discussed ideas together in modelling this analysis.

Scatter plots: With the data given, I made arrays of each parameter given in the csv file.

- I tried plotting each parameter against the chance of admission, with scatter plot.
- I observed that GRE, TOEFL, CGPA had nearly linear relationship i.e that with increase in these scores, increased the chance of admission.
- And for the other parameters, I could barely observe any relation. They were parallel to y - axis.
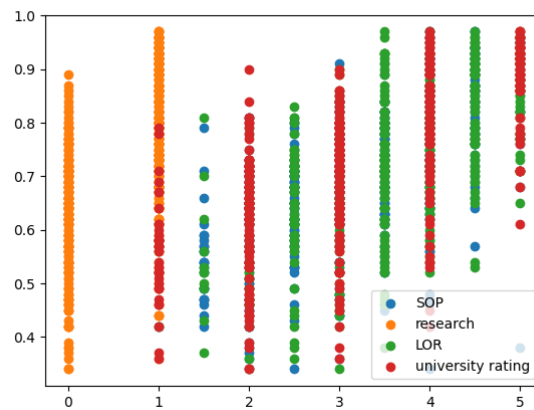

Figure 1: Scatter plot with research, SOP, LOR, university rating

Least squares method:

- I defined a straight line function `stline` assuming that GRE, TOEFL, CGPA have a linear relation.
- I used least squares method to estimate the slope and interecept. With these values, I found yest and plotted it.
- The graphs plotted were not that accurate, there were of points that were not covered, because there was huge scattering in the values.
- I had to conclude that there was no proper linear relation betwwen these parameters and chance of admission, but they had some non - linear relation.
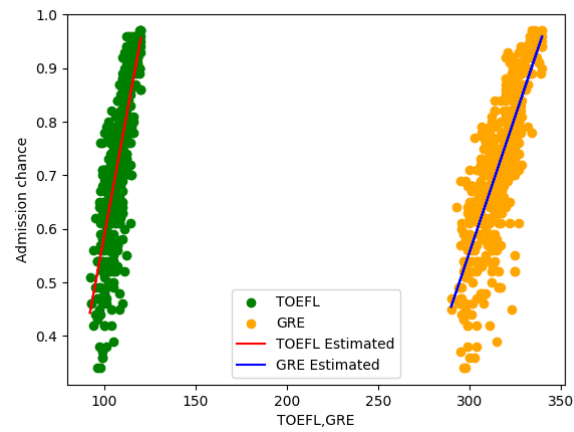
Curve Fit method:

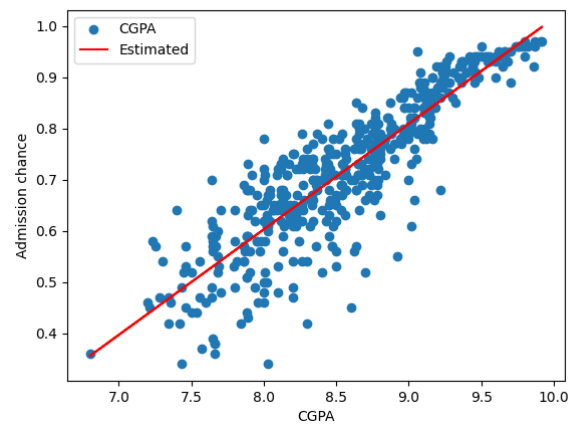Figure 2: TOEFL, GRE plot using least square method

Figure 3: CGPA plot using least square method

- I defined a polynomial function with all 7 parameters as independent variables - **CGPA ^ t1 + SOP ^ t2 + LOR ^ t3 + GRE ^ t4 + TOEFL ^ t5 + university_ranking ^ t6 + research ^ t7** . I thought this could give a better relation.
- But when I find out the t1, t2,...t7, I got negative values, which did't make sense, so I had to find another way.
- I defined a different fucntion with all 7 parameters as independent variables - **p1 * CGPA ^ t1 * SOP ^ t2 * LOR ^ t3 * GRE ^ t4 * TOEFL ^ t5 * university_ranking ^ t6 * research ^ t7** .
- With estimated initial values, I got values of p1, t1, t2,...t7. The values obtained were good enough with the observations made from scatter plots.
- We can conclude that CGPA(highest), GRE, TOEFL have more impact on chance of admission, than SOP, LOR, university_ranking and research. Research has least impact, and CGPA has greatest impact according to this model.

For getting into Top Ranked University:

- I separated the parameters for univeristy_rating 5, and applied the same function.
- I observed that for CGPA, TOEFL had more impact.
- So, it can be concluded that, to get into a top ranked university :
    - Focus more on CGPA and TOEFL score.
    - One also has to get better score in GRE.
    - SOP also helps a bit in getting into top - ranked university.
    - Doing a research, having LOR doesn't contribute much to get into the college.