# Assignment 4

**Pallakonda Sahithi - EE22B080**

## Predicting the function for Chance of Admit

Import the necessary libraries such as `numpy` which is used to estimate the Feature importance values.

**The python code `ee22b080_Assignment4.py` performs the following tasks**:

Considering the function for chance of admit is a linear model which depends on the features gre scores, toefl scores, rating, sop, lor, cgpa.

- First, it extracts the data points from the file `Admission_Predict_Ver1.1.csv` and stores the fraction $\frac{Value\ of\ the\ feature}{Maximum\ value\ of\ the\ feature}$ into different lists `gre_score[]`, `toefl_score[]`, `rating[]`, `sop[]`, `lor[]`, `cgpa[]`, `research[]`,`chance[]` with the appropriate data type.

Let **N** denote the number of sets of data points.

**Constructing the matrix M**:

- **M** is a matrix consists of the values of the features `gre_score, toefl_score, rating, sop, lor, cgpa` in different columns. Here, it is assumed that the data points, which are having the research is equal to 1 are considered for estimating the equation for chance of admit. Along with this, a new list **b**[ ] also created, an element in b[] represents the given chance of admit correspondingly. Let **U** denote the number of sets of data points which are having research equal to 1.

- Order of matrix M will become $U \times 7$.

- The first column of **M** are occupied by **gre scores**, second column with **toefl scores**, so on up to the **cgpa**, the last column is occupied by **ones**.

**Estimation of Feature Importances**:

Since it is assumed that the function for chance of admit as a linear model, so the typical equation for chance of admit is given as,

–> Chance of admit $= a_1(f_1) + a_2(f_2) + a_3(f_3) + a_4(f_4) + a_5(f_5) + a_6(f_6) + a_7$

where, $a_i$ represents the value of feature importance corresponding to $f_i$.

Using the function **`np.linalg.lstsq()`** in `numpy`, we can estimate $a_i$s based on the matrix equation $Ma = b$.

**Estimated Equation for Chance of Admit(y)**:

y $= 0.652 \times (\frac{gre\ score}{360}) + 0.345 \times (\frac{toefl\ score}{120}) + 0.061 \times (\frac{rating}{5}) + 0.017 \times (\frac{sop}{5}) + 0.066 \times (\frac{lor}{5}) + 1.147 \times (\frac{cgpa}{10}) - 1.229$

**Errors**:

- Square Root of mean square error is 0.06144.

- Average absolute error is 0.04262.

## Conclusions

- As we can see that error is not so large by computing with linear model. Therefore, linear regression model estimates the expected values upto 96% accuracy.

- Average ratio of estimated chance of admit to given chance of admit is 0.98571, this value tells us that the probability of being admitted and the probability of being admitted using the expected function are almost the same.

- The value of $a_i$ is maximum for **CGPA**, therefore, **CGPA** affects the chance of admit by more amount as compared to other features. So you should focus more on CGPA to get admitted in top rated university.