Admission Predict

M.Saketh Sai Ram - EE22B022

Import the necessary libraries such as `numpy` to solve the Matrix Equation. The Python script `ee22b022_Assignment4.py` contains the code that would estimate a Linear Function for the Chance of Admit in terms of the parameters, GRE Score, TOEFL Score, University Rating, SOP, LOR, CGPA, and Research.

1 Description of the Code

1.1 Extraction of Data Points

The Data points were appended to the lists `x1, x2, x3, x4, x5`. `xi` for `i = 1,2,3,4,5`; contains the data points with the value of $(Given\ Value)^i$

1.1.1 Structure of the List

Let **N** denote the number of sets of Data points. The structure of the list `xi` for $i = 1, 2, 3, 4, 5$; is given as,

$$\begin{bmatrix} (GRE_1)^i & (TOEFL_1)^i & (Rating_1)^i & (SOP_1)^i & (LOR_1)^i & (CGPA_1)^i & (Research_1)^i & 1 \\ (GRE_2)^i & (TOEFL_2)^i & (Rating_2)^i & (SOP_2)^i & (LOR_2)^i & (CGPA_2)^i & (Research_2)^i & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (GRE_N)^i & (TOEFL_N)^i & (Rating_N)^i & (SOP_N)^i & (LOR_N)^i & (CGPA_N)^i & (Research_N)^i & 1 \end{bmatrix}$$

1.2 Typical Model of the Function

The assumed typical model of the function to predict the Chance of Admit is given as,

$$Chance\ of\ Admit = a_1(P_1)^n + a_2(P_2)^n + a_3(P_3)^n + a_4(P_4)^n + a_5(P_5)^n + a_6(P_6)^n + a_7(P_7)^n + a_8 \quad (1)$$

where P_1 corresponds to the GRE Score, P_2 corresponds to the TOEFL Score, similarly for all the rest of the Parameters. $a_1,\ a_2$ ... a_7 are the coefficients corresponding to the $P_1,\ P_2$ ... P_7 in the Equation(1).

1.3 Calculation of the Coefficients

The Parameter's Coefficients can be computed using the function `numpy.linlag.lstsq()`, it takes the Input Data(xi) and the Output Data(Given Chance of Admit) as its arguments and would solve for the $a_1,\ a_2$ ... a_7 for n = 1, 2, 3, 4, 5.

All these coefficients don't matter, we need only the Error(say Root Mean Square Error) to predict the proper fit for Chance of Admit.

1.3.1 RMS Values for different orders of function

The Root Mean Square Errors for n = 1, 2, 3, 4, 5 are given as,

- Order(1) : 0.05950420877764951
- Order(2) : 0.0598106002775882
- Order(3) : 0.06055045591489407
- Order(4) : 0.06156499467200284
- Order(5) : 0.07434225941433129

From the above analysis, we can conclude that Linear Regression is the best fit since it has less error.

1.3.2 Estimation of Coefficients and the function for Chance of Admit

This can be done by using the function `linalg.lstsq()`. Therefore, the coefficients of the parameters for the Linear Function are given below,

- GRE Score: 0.00185851
- TOEFL Score: 0.00277797
- University Rating: 0.00594137
- SOP: 0.00158614
- LOR: 0.01685874
- CGPA: 0.11838505
- Research: 0.02430748
- Constant term: -1.27572508

Hence the approximate equation for Chance of Admit (C) would become,

$$C = 0.0019(P_1) + 0.0028(P_2) + 0.0059(P_3) + 0.0016(P_4) + 0.0169(P_5) + 0.1184(P_6) + 0.0243(P_7) - 1.2757$$

2 Conclusions

- The Percentage Error in predicting the Chance of Admit by using the Linear Regression Model is around 6%, the error will increase as the order increases.
- If we plot a graph between the Given Chance of Admit and the Estimated Chance of Admit, the graph will be a straight line with a slope equal to 1. In this case, the rough estimate of slope is 0.9998971962833128, which is almost nearer to 1.
- Parameter Importance is defined as the product of the coefficient of a parameter in the Linear function by its maximum value. This term tells about the parameter that creates more impact on the Chance of Admit. Parameter Importance values are,

  1. GRE Score: 0.669
  2. TOEFL Score: 0.333
  3. University Rating: 0.029
  4. SOP: 0.008
  5. LOR: 0.084
  6. **CGPA: 1.183**
  7. Research: 0.024

  Based on the above observations, we can say that **CGPA** will affect more, **SOP** will affect less on Chance of Admit. So, you should focus more on CGPA to get admitted to a top-rated university.