Assignment4Prediction

October 15, 2023

1 EE2703: Applied Programming Lab

1.1 Atreya Vedantam, EE22B004

1.2 Introduction

We are given a dataset and our task is to figure out how the data is interrelated. The first step in this analysis is called Exploratory Data Analysis. The idea is to get a visual sense of how the chance of admission into a top ranked university is dependent on each parameter. We can then deduce approximately which parameters are of great importance and which ones aren't as important. We can then conclude whether a linear fit makes sense of the given dataset.

1.3 Results

1.3.1 Can the chance of admission be predicted as a function of all the other parameters?

Yes, it can. A few plots using `partialDependencePlot` tells us that the data has relationships which we can exploit to analyse and predict.

1.3.2 What kind of relationship between the parameters do you think would give a good fit? Is there any reasonable way of doing this using just a linear fit (least squares) or by using some kind of nonlinear model?

A linear model seems to work the best. The coefficient of determination captures how good a fit is (it is the ratio of the squared error in the model and the variance in the target data), and it is 0.9065874141624964 in this case. This implies a strong linearity in the data.

1.3.3 Which of the parameters have the maximum impact on your chance of admission if your goal is to get admitted to a top ranked institution (rating 5)?

The answer is CGPA (closely followed by Research), followed by SOP. LOR and TOEFL scores have similar weightage towards increasing your chance of admission and GRE score seems to have comparatively little impact on your chance of admission. This inference is made simply by looking at the relative magnitudes of the weights.

1.3.4 Other Inferences

On observing the plots of the chances of admission into a top-rated university against the parameters, it becomes clear that not all of them are equally rewarding. For example, an increase in SOP from around 4 to 5 on an average does not help increase admission chances at all. Similarly, people

who have gotten an LOR of around 3 on an average are less likely to get admitted into a top-rated university than one with a 2. This might be because of inherent bias in the data (i.e. people who may focus on getting an LOR may not statistically focus on other more important aspects).

An increase in GRE score for all universities typically increases the chance of admission tremendously and also guarantees a degree of certainty in the sense that variance decreases for high GRE scores. This is not quite the case for only 5-rated universities where the variance remains around the same.

[6]:
```
from PIL import Image
import matplotlib.pyplot as plt
plt.grid(False)
plt.axis('off')
img = Image.open('a4_chancevsSOPFiveStar.png')
plt.imshow(img)
```

[6]:
```
<matplotlib.image.AxesImage at 0x7f7138226ac0>
```