1   Notations and Definitions

1. $U_j^{[i]}$ = university j of class i

2. $U_o^{[i]}$ = benchmark university for class i

3. $A_j^{[i]}$ = amount invested in university j for class i

4. $T_j^{[i]}$ = Tuition in university j for class i

5. $n_j^{[i]}$ = number of students receving our grant in university j for class i

6. $IP_j^{[i]}$ = poverty rate of students entering university j for class i

7. $OP_j^{[i]}$ = poverty rate of students exiting university j for class i

8. $\phi_j^{[i]}$ = number of students in university j for class i

9. $g_j^{[i]}$ = maximum amout of grants to be given to university j for class i

10. <u>Group</u>, refers to the Carnegie Classification of the university

11. <u>Class</u>, refers to which benchmark within the group, the university is closest to.

2   Benchmark Setting

First we omitted any below medium size universities or communal universities such as racial or religious schools.

Then we partition all universities by number of years required for graduation.

Lastly, we rank the universities by the amount of their poverty elevation standards:
$R_j^{[i]} = \frac{IP_j^{[i]} - OP_j^{[i]}}{IP_j^{[i]}}$.

Next we choose the top, highest ranking universities per 'years of graduation'-group. We set these chosen universities as our benchmarks for each class in each group. I.e.

$$o^{[i]} = \operatorname*{argmax}_{j} R_j^{[i]}$$

Where o, is to represent the benchmark index of the class i.

3   Identifying matches

For each group, we identify which universities are closest to which benchmark within the group. We apply the kNN algorithm on all the variables listed below.

<center>LISTOFVARIABLES</center>

(We only omitted informative variables from the original dataset).
For this purpose and obtain the following classification criteria:
Assuming $j \in [1, n]$, let $y_j^{[i]}$ be the j-th variable for the i-th class benchmark.
Let x_j be the j-th variable of a particular university.
A university u, is considered member of the k-th class if the variables x_i of that university are 'close' to a particular variable of a given class in euclidean-space.
That is, if u belongs to the k-th class then

$$k = \underset{i}{\operatorname{argmin}} \sqrt{\sum_{j=1}^{n} (x_j - y_j^{[i]})^2}$$

4   Determine Investment

4.1   Introduction

4.2   Number of student-Helped

Let $A_j^{[i]}$ be the amount we invest in university j of the class i, and $T_j^{[i]}$ be the tuition of low-income students.
Then we can compute the number of students who we help as

$$n_j^{[i]} = \lfloor \frac{A_j^{[i]}}{T_j^{[i]}} \rfloor$$

$$n^{[i]} = \sum_{j=1}^{\infty} n_j^{[i]}$$

$$n = \sum_{i=1}^{\infty} n^{[i]}$$

Please note that the number of universities is <u>finite</u>, we use infinity here in order to avoid the introduction of more variables.

4.3   Adjusted Poverty

Next we computed the adjusted poverty rate. We make a crucial assumption here about the correlation between grant-money and retention rates. We assume that every student who receive the grant will stay in the program and university (this assumption can be adjusted by a future study on the subject matter).

Let $\phi_j^{[i]}$ be the total number of students in university j for class i.
The adjusted rate of existing poor students based on our assumptions is:

$$OP_j^{[i]*} = \frac{OP_j^{[i]} \phi_j^{[i]} - n_j^{[i]}}{\phi_j^{[i]}}$$

We can then provide a rank to this university, in the similar fashion we computed the rank for the benchmark universities:

$$R_j^{[i]*} = \frac{IP_j^{[i]} - OP_j^{[i]*}}{IP_j^{[i]}}$$

We use this to compute the increase to the social well-being of students going to this university.

4.4   Return on Investment

The return on investment for our model will be the change in the level of alleviation of poverty provided by our investment.

We do this as follows:
$$R_j^{[i]}(1 + r_j^{[i]}) = R_j^{[i]*}$$
where $r_j^{[i]}$ is the return on investment for a particular school.

We also note that this can be rewritten as $r_j^{[i]} = \frac{OP_j^{[i]} - OP_j^{[i]*}}{IP_j^{[i]} - OP_j^{[i]}}$. This will later be used for simplification purpose of the maximization problem.

We then average these rates, to get the approximated return:

$$\bar{r} = \frac{\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} r_j^{[i]}}{\#(\{\forall i, j : U_j^{[i]}\})}$$

For simplification, the total number universities, or $\#(\{\forall i, j : U_j^{[i]}\})$ will be written as Φ, so

$$\bar{r} = \frac{\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} r_j^{[i]}}{\Phi}$$

4.5 Maximization of return

Now that we have a measure of return on investment, we wish to maximize this measure based on the physical restriction that are given to us.
Since we only have a limit of \$100M, and would not want to overshoot our investment by giving any particular university more than what the top-tier schools are given. We put the following restriction forth.

Let $g_j^{[i]} = \frac{G_o^{[i]}}{\phi_o^{[i]}} - \frac{G_j^{[i]}}{\phi_j^{[i]}}$, be the restriction on the per-student investment.
We do this, in order to make sure that our investment has a reasonable bound that does not exceed the higher-tiered school.

So now our problem can be formulated as follows:

$$\max \bar{r}$$
$$\text{subj} \sum \sum A_j^{[i]} = 100,000,000$$
$$\forall i, j \, 0 \leq A_j^{[i]} \leq g_j^{[i]} \phi_j^{[i]}$$

4

4.6 Simplifying the complexity

In order to simplify the complexity of the previous problem we outlined before. We reformulated the problem as follows:

First recall, results:

(1) $r_j^{[i]} = \frac{OP_j^{[i]} - OP_j^{[i]*}}{IP_j^{[i]} - OP_j^{[i]}}$

(2) $OP_j^{[i]*} = \frac{OP_j^{[i]} \phi_j^{[i]} - n_j^{[i]}}{\phi_j^{[i]}}$

, this can be further simplified:

$$r_j^{[i]} = \frac{1}{\phi_j^{[i]}(IP_j^{[i]} - OP_j^{[i]})} n_j^{[i]} = \frac{1}{\phi_j^{[i]}(IP_j^{[i]} - OP_j^{[i]})} \lfloor \frac{A_j^{[i]}}{T_j^{[i]}} \rfloor$$

$$\approx \frac{1}{\phi_j^{[i]}(IP_j^{[i]} - OP_j^{[i]})T_j^{[i]}} A_j^{[i]} = \Delta_j^{[i]} A_j^{[i]}$$

We note that $\Delta_j^{[i]}$ is the information given by our data, whereas $A_j^{[i]}$ is our investment, which is a variable that can be controlled by us. Hence the mean return can be written as:

$$\bar{r} = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \frac{1}{\Phi} r_j^{[i]} = \frac{1}{\Phi}\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \Delta_j^{[i]} A_j^{[i]}$$

So our new maximization problem can be formulated as follows:

$$\max \frac{1}{\Phi}\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \Delta_j^{[i]} A_j^{[i]}$$

$$\text{subj} \sum\sum A_j^{[i]} = 100,000,000$$

$$\forall i,j \, 0 \le A_j^{[i]} \le g_j^{[i]} \phi_j^{[i]}$$