ISYE 6501 - Homework 11

November 1, 2019

Question 15.1

We can think of an optimization problem where a first year MSA student needs to allocate her weekly schedule across four different classes, ISyE 6501 (x), CSE 6040 (y), CSE 6242 (z), ISyE 6333 (w). Her objective is to maximize her GPA, therefore the four decision variables $\{x, y, z, w\}$ represent the study time allocated to each class. For simplicity, we can think of the outcome variable GPA as the raw sum of scores in the four subjects: a value of 400 implies that she achieved a perfect score in each class and therefore obtained 100 out of the 100 available points in each class. The coefficients $\{a, b, c, d\}$ represent the returns on each hour of study for each subject. Obtaining reliable estimates for $\{a, b, c, d\}$ is far from an easy task, and there are reasons to believe that these parameters are not constant across students but rather functions of their individual characteristics. However, survey data from prior cohorts of MSA students can be used to perform a regression analysis and obtain reasonable estimates, under the rather strong but nonetheless necessary assumption that $\{a, b, c, d\}$ are the same for all students.
Next, we need to define the constraints of the optimization problem. Our student has 70 hours/week available to allocate, which means that the sum of $\{x, y, z, w\}$ has to be less or equal than 70. Also, she knows that her weakest skill is coding and therefore she wants to allocate at least 35% of the available time to CSE 6242 (z).By contrast, she is pretty confident with her stats background and wants to allocate less time to ISyE 6501 than she will to CSE 6040, and less time to ISyE 6501 than she will to ISyE 6333. Lastly, we should not forget the nonnegativity constraints and we should therefore require that each class is being allocated a positive amount of hours. We can write out the maximization problem as follows:

$$
\begin{aligned}
\underset{x,y,z,w}{\text{maximize}} \quad & GPA = a \cdot x + b \cdot y + c \cdot z + d \cdot w \\
\text{subject to} \quad & x + y + z + w \leq 70, \\
& z \geq 0.35(x + y + z + w), \\
& x \leq y, \\
& x \leq w, \\
& x \geq 0, \\
& y \geq 0, \\
& z \geq 0, \\
& w \geq 0
\end{aligned}
\tag{1}
$$

Question 15.2

The .py file with the full code is attached. The problem requires to minimize the costs of a diet subject to some constraints. The decision variables are:

- A continuous decision variable representing the quantity chosen for each food type;
- A binary helper variable taking the value of 1 for foods that have been included in the diet, O for foods that did not.

The constraints are the following:

- For each nutrient, there are a min and a max daily intake that the diet has to abide by;
- The minimum serving size cannot be any less than 10% of a serving;
- Either celery or broccoli might be chosen but not both;
- The diet needs to entail at least 3 different sources of proteins.

The PuLP's solver returned the following cost-minimizing diet:

```
Chosen_Poached_Eggs = 1.0
Chosen_Scrambled_Eggs = 1.0
Chosen_Tofu = 1.0
Quantity_Celery,_Raw = 52.298911
Quantity_Frozen_Broccoli = 0.24876754
Quantity_Lettuce,Iceberg,Raw = 64.546946
Quantity_Oranges = 2.3250062
Quantity_Poached_Eggs = 0.1
Quantity_Popcorn,Air_Popped = 13.852241
Quantity_Scrambled_Eggs = 0.1
Quantity_Tofu = 0.1
The total cost of the diet is: $4.38
```

For the second dataset, we resorted to a little bit more hard-coding than we wanted, mainly because of the rather sketchy tabulation of the dataset. We could have used Excel workarounds; however, hard-coding comes somehow with the "benefit" of making a problem's statement – more so, its constraints – even clearer to the audience. One further problem with the dataset was the amount of missing data. We already showed in the imputation assignment how missing data can be handled by way of a variety of techniques. For the sake of this assignment, we used a quick-and-dirty solution that excluded the columns with at least 30-percent of missing data, and filled the remaining missing cases with 0s. This is far from the optimal approach, but our focus for this assignment is on optimization and, especially, on how to set up meaningful constraints. This is even more important with this second dataset, which returns an optimal solution representing a diet which a human being would be hardly ever down to eat.

```
Quantity_Egg,_white,_dried = 0.24
Quantity_KRAFT,_CRYSTAL_LIGHT_Sugar_Free_Low_Calorie_Soft_Drink_Mix_Lemo = 0.44130035
Quantity_Kanpyo,_(dried_gourd_strips) = 1.8856248
Quantity_Leavening_agents,_baking_powder,_double_acting,_sodium_aluminum = 0.057445037
Quantity_Leavening_agents,_baking_soda = 0.01925441
Quantity_Oil,_bearded_seal,_(oogruk_oil)_(Alaska_Native) = 0.59681698
Quantity_Seeds,_chia_seeds,_dried = 0.070942666
Quantity_Vegetable_oil,_palm_kernel = 1.1129138
Quantity_Vital_wheat_gluten = 0.5052402
```

```
Quantity_Water, _bottled, _non_carbonated, _CALISTOGA = 9999.5573
The total cholesterol of the diet is: 0.0
```

Clearly, a whole set of constraints need be engineered to make this problem meaningful. Much like the previous problem, it seems reasonable that different patients will have preferences for combinations of foods. By this token, we might require that only one among one or more lists of foods is chosen. Also, we might want to enforce *at least*-type of constraints for designated categories of food, such as proteins or carbohydrates. Differently than before, we might want to specify *no more than*-type of constraints for specific types of foods such as oils, beverages, or powders, which are very low on cholesterol but are hardly appetizing by themselves. A reasonable approach in this sense would be that of requiring that, say, oils are consumed with some complementary ingredients to cook much more fulfilling dishes. For instance, we might require that at least 10 times as much flour is used if a given oil is used, so that some tasty cake can be prepared. This constraint would have the mathematical form $x_{flour} \geq 0.1 y_{oil}$ where x_i represents the quantity of a food i and y_i is an indicator variable taking on the value 1 if the food is chosen and 0 elsewhere. Other reasonable constraints would be to require that at least a given amount c of a food is consumed, if the food is chosen ($x_i = c y_i$); that at least a given amount w of foods are in the diet ($\sum_i y_i \geq w$); or that some foods are removed from the diet ($x_i = 0$).