Optimization
Project 2 – Integer Programming

Deliverables
One well-written pdf file and one Python code file (.py or .ipynb), submitted to Canvas. Your report should go into some detail about how you solved the problem, include some graphs that explain your results, and include relevant code chunks in the final output. Your report can be created by taking screenshots of the code/graph and assembling it in a word document, then export as a pdf file.

Problem Description
Marketing budgets now comprise 11 percent of total company budgets, based on a CMO survey sponsored by the Fuqua School of Business at Duke University, Deloitte LLP, and the American Marketing Association. However, the effectiveness of marketing varies significantly: on the one hand, P&G cut more than $100 million in digital marketing spending because their digital ads were largely ineffective; on the other hand, Netflix plans a 54% boost in ad spending because they got very positive feedback in international markets.

One potential reason for such variation is the way of making marketing budget allocations. Namely, how much to invest in each advertisement platform. As stated in the Handbook of Marketing Analytics:

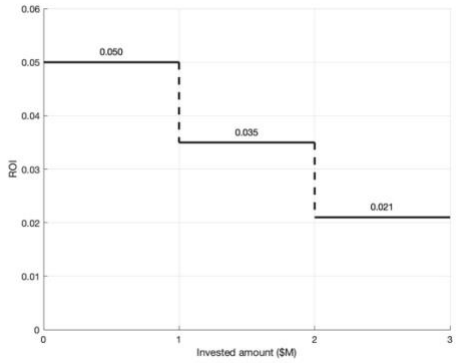
> *...budget decisions are often based on gut feelings or on the negotiation skills of individual managers. Consequently, politics and individual opinions tend to shape the decision process instead of fact-based discussions. Obviously, these rules and practices bear the risk of results far away from the optimal, profit-maximizing budget*.

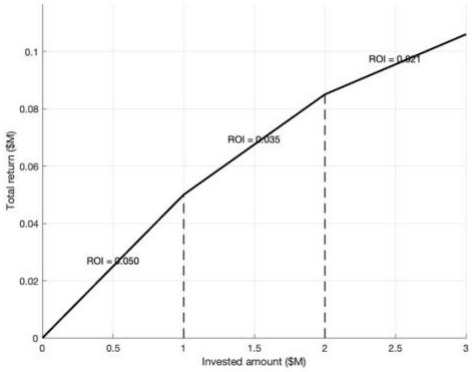
Indeed, the marketing strategy of Netflix seems to be steered by data.

In this project, we use integer programming to build a marketing budget allocation strategy.

Specifics

1) Assume that your company is deciding how to spend a marketing budget of $10M. You work in the marketing department as a data scientist and the chief marketing officer has asked you to write a report recommending how to spread this budget among several marketing mediums. Your department has employed an outside consulting firm to estimate the return on investment (ROI) of each marketing medium under consideration. The ROI for each marketing medium is modeled as a piecewise constant function with respect to the amount invested in the medium. As a result, the total return is a piecewise linear function with respect to the amount invested. For example, given the ROI in the left figure, there are three tiers: the first $1M invested yields a high return (5%); the next $1M (from $1M to $2M) yields a moderate return (3.5%); any investment beyond $2M up to $3M yields a lower return (2.1%). These ROI values are slopes of the piecewise linear function for total return vs. investment amount on the right.





The ROI data are in a CSV attached to this assignment ("roi_company1.csv").

2) On top of these ROIs, your boss has decided to constrain your budget as follows:
    a. The amount invested in print and TV should be no more than the amount spent on Facebook and Email. Surprisingly, email seems to be a great channel for reaching real people.
    b. The total amount used in social media (Facebook, LinkedIn, Instagram, Snapchat, and Twitter) should be at least twice of SEO and AdWords.
    c. For each platform, the amount invested should be no more than $3M.
3) For the given ROI data, the ROI is non-increasing with respect to the invested amount, so the total return is a concave function of the invested amount. To determine the optimal amount invested in different mediums, this marketing budget allocation problem can be formulated as a linear program when the return is concave with respect to ROI. You can let the amounts invested in each tier for each medium be continuous decision variables and then the total invested amount is the sum of them. The concavity of the return function guarantees that the invested amount in a tier is positive only when the invested amounts in previous tiers equal their upper bounds. Use gurobi to find the optimal budget allocation after formulating it as a linear program.
4) Your boss is happy to see the promising results presented by the marketing department. However, your boss is also very concerned because your boss recalls being somewhat disappointed after following such recommendations in the past. To be cautious about the decision, your team has decided to get another opinion about the ROI data and rerun the analysis. The second consulting firm returns the estimates of the ROI data in the CSV file "roi_company2.csv". After analyzing the second firm's ROI data, you find that their ROI's are not non-increasing with respect to the invested amount. As a result, the concavity of the total return is not retained, and it is possible that the tier with a higher ROI will be invested first, while its previous tiers will be skipped due to their lower ROI's if you still use the LP model.

    In this case, you must formulate the problem as a mixed integer program. Following is an example of using mixed integer program to express a piecewise linear function.

Suppose we would like to minimize function with respect to x over $[1,6]$:

$$f(x) = \begin{cases} 4-3x, & 1 \le x \le 3, \\ x-8, & 3 < x \le 6 \end{cases}$$

It is equivalent to the following mixed integer program:

$$\begin{aligned} \min\ & \lambda_0 f(1) + \lambda_1 f(3) + \lambda_2 f(6) \\ \text{s.t. } & z_1 + z_2 = 1 \\ & \lambda_0 + \lambda_1 + \lambda_2 = 1 \\ & \lambda_0 \le z_1 \\ & \lambda_1 \le z_1 + z_2 \\ & \lambda_2 \le z_2 \\ & z_1, z_2 \in \{0,1\},\ \lambda_0, \lambda_1, \lambda_2 \ge 0 \end{aligned}$$

z_1 and z_2 are binary decision variables that represent whether x belongs to $[1,3]$ and $(3,6]$, respectively. λ_0, λ_1 and λ_2 are continuous decision variables that represent the coefficients for the convex combination (a convex combination is a linear combination of points where all coefficients are non-negative and sum to 1) of the boundary values of the segments. Any $x \in [1,6]$ can be expressed as $x = 1 \cdot \lambda_0 + 3 \cdot \lambda_1 + 6 \cdot \lambda_2$, $\sum_{i=0}^{3} \lambda_i = 1$, which is a convex combination of boundary values of the segments. Similarly, $f(x)$ can be expressed as a convex combination of the function values at these boundary points: $f(x) = \lambda_0 f(1) + \lambda_1 f(3) + \lambda_2 f(6)$. The first constraint restricts that x can only belong to one segment. The second constraint ensures that we have a convex combination. The third, fourth and fifth constraints link λ and z variables.

Inspired by this, to solve the marketing allocation problem when the total return is nonconcave, you need a binary decision variable for each medium and each tier to determine whether the invested amount falls in that tier. You also need continuous decision variables to determine the coefficients for convex combination. You will add constraints like the above ones in the example. The three constraints above given by your boss are still valid, so think about how you express them using these decision variables. Formulate the problem as a mixed integer program and then solve it using gurobi.

5) Are the allocations the same? Assuming the first ROI data is correct, if you were to use the second allocation (the allocation that assumed the second ROI data was correct) how much lower would the objective be relative to the optimal objective (the one that uses the first ROI data and the first allocation)? Assuming the second ROI data is correct, if you used the first allocation how much lower would the objective be relative to the optimal objective? Do you think the third constraint above, based on your boss' experience, is useful?
6) After conducting preliminary marketing research, your company find that there is a minimum amount for each medium such that any investment below this amount would be insufficient to

generate meaningful impact. In this case, for each medium, you will either not invest or invest at least that minimum amount. These minimum amounts are given in the attached csv file "min_amount.csv". Embed this rule in the mixed integer program model and solve the problem again with the second firm's ROI data. What is the optimal allocation?

7) Your boss has gained permission to reinvest half of the return. For example, if the marketing obtains a 4% return in January, the budget of February will be \$10M + \$10M × 4% × 50% = \$10.2M. The monthly ROI for next year is given in the file "roi_monthly.csv". The three constraints given by your boss and the constraint for the minimum amounts are still in place for each month. You do not need to implement the minimum amount constraint in Part 6). What is the optimal allocation for each month?
8) A *stable budget* is defined as a monthly allocation such that for each platform the monthly change in spend is no more than \$1M. Is the allocation you found stable? If it isn't, you do not need to solve a new optimization model. Describe how you might model this?
9) Write an pdf file that does all of this and summarizes your analysis with graphs, text, and code chunks. The first few lines of the python file, in the first code chunk at the beginning, should include pd.read_csv call that reads the csv file that has the ROI data. Be sure to include a noticeable comment that lets us know where you read the csv files. The template csv files fit the correct format. To be graded, we will change this pd.read_csv call to load a new csv file and re-run your code to see if you get the right answer on new data. Failure to run will automatically reduce your grade by 10 percentage points! Be sure that all your analysis in the python code file is generalized, so when we load the new csv files the output will be for the new data, instead of the template data. That means you should not hard code any numbers. Instead you should reference variable names for your output.
10) Grading will be based 66% on whether you get the right answer or not when we re-run your analysis with new data. If you don't get the right answer or your python code file doesn't run, we will go through your code and give partial credit accordingly. The easier it is to read your code the easier it is for us to understand what you're doing, so use a lot of comments in your code! The remaining 34% of your grade is based on the quality of your analysis and presentation of results in the pdf file. Write this as if you were actually going to submit it to your boss.