Final Exam- To Be Done By Yourself
MSiA Program, Fall 2023
Optimization and Heuristics
Due: Wednesday, 12/6/2023, by 11:59 PM CT via Canvas

(If you can turn it in earlier—even days earlier, that would be great—and that will prevent last-minute glitches)

Instructions: This Word document is the final exam. Please complete the test in this document and submit it to Canvas when you are finished. Please rename the file to include your name or initials. Also, include your name within the document. The tests are due by Wednesday, but please feel free to turn them in early.

Scoring. The final exam is worth 20% of your grade. Questions 1-6 are worth 1 point each and should be short answers. Question 7 is worth 1.5 points, Questions 8, 10, and 11 are worth 2 points each, and Questions 9 and 12 are worth 3 points. Question 13, is an easy half point-- just provide an answer.

Please complete the test on your own and do not work with anyone else on it. This is not a group assignment. However, you may use the text and your notes. If you have any questions, please feel free to contact us.

Name: _____Samuel Swain (SMS5736) ______

1) Why is it important for a data scientist or a data science team to know about linear and integer programming? Or why isn't knowing about machine learning, deep learning, and reinforcement learning enough?

In data science, which revolves around leveraging data for problem-solving and business insights, a comprehensive understanding of various analytical methods is crucial. Beyond the realms of machine learning, deep learning, and reinforcement learning, linear and integer programming play a pivotal role in optimization, a key methodology for generating business value and insights. These programming techniques are instrumental in addressing specific data-driven challenges, particularly those involving constraints and resource allocation. Their inclusion in a data scientist's toolkit enables a more versatile approach to data analysis, ensuring that the most effective method is employed for each unique problem. This breadth of knowledge is essential in the ever-evolving landscape of data science, where the ability to adapt and utilize diverse techniques directly contributes to a team's effectiveness and innovation in data-driven decision making.

2) After this class, how can you more systematically spot optimization problems in the real world?

To systematically identify optimization problems in the real-world post-class, focus on recognizing three key elements: decision variables, an objective value to maximize or minimize, and the presence of constraints. This approach enables the detection of scenarios where optimization is applicable, helping to discern situations that call for strategic allocation of resources, efficiency improvements, or cost reduction.

3) **Assume that you have two manufacturing plants that need to make nine different products. Assume that annual demand for each product is 1 million units and that to achieve economies of scale you need to make at least 600,000 units at a single plant. You would also like to have each plant make the same number of units. What will happen to your model if you put in a constraint saying that plant capacity is 4.5 million units and at least 600,000 units of a product must be made at a plant?**

Adding the constraint of a plant capacity of 4.5 million units and a requirement to produce at least 600,000 units of a product at a single plant makes the model infeasible. This constraint results in a conflict between meeting minimum production requirements for each product and the plant's capacity limit. As a result, it becomes impossible to simultaneously satisfy both the minimum production requirement and the capacity constraint for all nine products (we could only get to eight). This means that the model cannot meet the total demand for all products while adhering to these constraints.

4) **Why should you be skeptical if a vendor (solving what you know to be solving a linear program with integer variables) presents a solution that claims to optimally solve *any* instance of a problem very fast?**

When a vendor claims to rapidly and optimally solve any instance of a linear program with integer variables, skepticism is advisable. Integer programming problems are typically NP-hard, meaning the complexity and time required to solve them can grow exponentially with problem size. Adding integer constraints to linear programs significantly increases their complexity. While heuristic methods can offer quick solutions, they do not always guarantee optimality.

5) **In this class, most models were deterministic LPs or IP's. For this problem, consider the JADE and other network design models or the Beer-Ale problem. In these cases, the underlying LP or IP model did not directly account for uncertainty. However, in each case, the real-world problem we were solving had uncertainty.**
    A. Why didn't the underlying models account for the uncertainty?
    B. How could these models be used to address the uncertainty without having to change the underlying LP or IP? (Hint: don't make these problems stochastic optimization problems)

A. In our class, the underlying models for JADE, other network design problems, and the Beer-Ale problem didn't account for uncertainty primarily to maintain simplicity and computational feasibility. Adding uncertainty into Linear

Programming (LP) or Integer Programming (IP) models increases their complexity and computational demand. We aimed to simplify these models to focus on their core aspects, making them more manageable and easier to analyze. This approach also helps when data on variability and probability distributions, crucial for modeling uncertainty, is not readily available or reliable. Moreover, deterministic models provide a clear baseline, offering insights under ideal conditions for easier comparison and understanding.

B. To address uncertainty without changing the underlying LP or IP models, we can employ several strategies. Scenario analysis is one, where we run the models under different assumptions to see how outcomes vary. Sensitivity analysis is also useful; by systematically varying parameters, we can assess the robustness of our solutions. Post-optimization analysis helps understand how these solutions might perform under uncertain conditions by simulating various scenarios. For supply chain models, incorporating safety stocks or redundancies can buffer against uncertainty. This method involves planning additional resources or capacities. Finally, a rolling horizon approach, where the model is regularly updated with new data, allows for adjustments in response to emerging trends or information, making our model more adaptable to changing circumstances.

6) **When solving an integer program, you often have to set an optimization gap (or 'branch and bound tolerance' in OpenSolver). What is this gap? (Explain it clearly). And, if you stop a program with a gap of 10%, does that mean you have not found a good solution?**

The concept of the optimization gap in integer programming, also known as the 'branch and bound tolerance', is crucial for understanding the feasibility and quality of solutions. The gap is the difference between the objective value of the best known feasible solution (the incumbent) and the best possible objective value as estimated by the solver. It's typically expressed as a percentage, representing the maximum allowable difference between the upper and lower objective bounds.

When the optimization gap becomes smaller than the specified gap parameter, the solver terminates, indicating that it considers the best-known solution sufficiently close to the optimal solution. This gap acts as a termination criterion and helps in controlling the trade-off between solution quality and computation time, allowing informed decisions in optimizing real-world problems.

Stopping a program with a 10% optimization gap doesn't inherently mean that a good solution has not been found. The acceptability of this gap largely depends on the specific requirements and context of the problem. In some cases, particularly where exact optimality is less critical, a solution within a 10% gap might be perfectly acceptable and significantly more efficient to compute. Thus, a 10% gap reflects a balance between solution quality and computational practicality, tailored to the specific needs and constraints of the problem.

7) **This question has two parts. Part A: In the Sports Team scheduling problem, we discussed using column generation. Why can you describe this approach as an "optimization-based heuristic?" Part B: Think about the solution to the revenue management homework problem—the problem with one price and a limited capacity. In class, I showed how to solve it with a simulation. How would you *best* describe my solution approach: was it closest to Optimization-based, Enumeration, Heuristic, or an Optimization-based Heuristic? Provide some explanation.**

   A. In the Sports Team scheduling problem, where column generation is used, describing this approach as an "optimization-based heuristic" is appropriate due to the way the method operates. Column generation is a technique typically used in solving large-scale linear programming problems. It works by starting with a smaller, more manageable subset of variables (or columns) and iteratively adding more variables based on a pricing problem that identifies which variables (or columns) could potentially improve the solution. This approach is 'optimization-based' because each iteration involves solving an optimization problem to identify the most promising columns to add. However, it's also a 'heuristic' in the sense that it doesn't initially consider all variables. Instead, it heuristically selects a subset of variables that are likely to lead to a good solution, which may not guarantee an optimal solution for the entire problem but often finds high-quality solutions more efficiently than considering all variables from the outset.

   B. The solution approach to the revenue management homework problem, which involved one price and limited capacity and was solved using simulation, can be best described as a heuristic method. This approach simulates various scenarios or outcomes based on the given data and constraints to find a satisfactory solution. The use of simulation here does not inherently involve optimizing an objective function in the traditional sense. Instead, it explores different possibilities or scenarios to see how they perform under the given constraints and conditions. The primary goal is not to find the optimal solution through a systematic optimization process or exhaustive enumeration but to approximate a good solution based on the outcomes of simulated scenarios. Therefore, it aligns more closely with the concept of a heuristic approach, focusing on practicality and efficiency in finding a sufficiently good solution rather than guaranteeing the optimality of the solution.

8) **ComplexCo makes a simple product but has some complicated purchasing rules. They purchase a rare earth element they call XT-10. They then process XT-10 and sell it. For our purposes, assume they need to buy 1 unit of raw XT-10 to make 1 unit of demand for the finished good. Assume that they have demand of 170 units for the finished good. They can buy raw XT-10 from one of three sources- A, B, and C. A, B, and C each have a limit of 100 units. Supplier A costs $1 (all $ figures are in millions) per unit, B costs $1.2 per unit, and C costs $1.5. Because of contracts, they have to buy at least twice of much from B as from A. And, they have to buy at least as much from C as from the other two**

suppliers combined. Formulate this as a linear program and solve it. What is the minimum cost solution? Provide some details on your formulation.

In solving ComplexCo's purchasing problem, the objective was to minimize the total cost of acquiring a rare earth element, XT-10, from three suppliers, A, B, and C, with costs of $1 million, $1.2 million, and $1.5 million per unit respectively. The decision variables represented the quantities to be purchased from each supplier, and the model was constrained by the need to meet a total demand of 170 units, with each supplier's limit set at 100 units. Additional contractual requirements stipulated that the amount purchased from B must be at least double that from A, and the quantity from C must be at least as much as the combined total from A and B. The optimal solution, achieved through linear programming, suggested purchasing 28 units from A, 57 from B, and 85 from C, thereby satisfying all constraints and achieving a minimum cost of approximately $223.9 million. This strategy not only adhered to the specified limits and contractual conditions but also represented the most cost-effective approach to fulfill the demand for XT-10.

9) **Feedco produces two types of cattle feed. They produce it based on the by-products (wheat and corn) rejected by a human food producer. There is a daily limit on the amount of wheat and corn every day—usually between 500 and 1000 pounds. They buy the wheat for $0.50 per pound and the corn for $0.45 per pound. They buy alfalfa as a filler every day for $0.40 with no limit. Both feeds consist totally of wheat, corn, and alfalfa. Feed 1 is their higher-valued product and sells for $1.50 per pound. Feed 2 sells for $1.20. Demand for each type of feed is unlimited. The high-value product must have a lot more wheat and corn than the low-value product, and the wheat is more valued than the corn. But both products have some of all three ingredients. Formulate a Linear Program to maximize Feedco's daily profit. [1] You'll need to develop reasonable constraints to make sure Feed 1 is better (has more wheat and corn) than Feed 2. Formulation should clearly show the decision variables, the objective function, and the constraints. The user should be able to adjust the model to create more or less differentiation between Feed 1 and Feed 2. Also, solve the LP (in Python or Excel) with some sample data and provide the optimal solution.**

Model Overview

The model is designed to optimize the production of two types of cattle feed by Feedco, considering resource availability, cost of ingredients, and selling prices. The goal is to maximize profit.

Decision Variables:

w1, c1, a1: Quantities of wheat, corn, and alfalfa in Feed 1.
w2, c2, a2: Quantities of wheat, corn, and alfalfa in Feed 2.
total_weight_feed1, total_weight_feed2: Total weight of Feed 1 and Feed 2.

Objective Function:

Maximize profit, calculated as the revenue from selling both feeds minus the costs of ingredients.

Constraints:

1. Ingredient Limits: Total wheat and corn used in both feeds cannot exceed their respective upper bounds.

2. Feed Composition:
   - Feed 1 must have at least a specified minimum percentage of wheat and corn.
   - Feed 2 must also meet a minimum percentage for wheat and corn, but less stringent than Feed 1.

3. Wheat to Corn Ratio:
   - In Feed 1, the quantity of wheat must be at least double that of corn.
   - In Feed 2, wheat must be at least 1.25 times the quantity of corn.

4. Minimum Ingredient Quantity: Both feeds must contain a minimum amount of wheat and corn.

5. Total Feed Composition: The sum of wheat, corn, and alfalfa in each feed type must not exceed the total weight allocated for each feed.

6. Non-Negativity: Ensures that all quantities are non-negative.

Scenario Analysis

Scenario 1: Upper Bounds for Wheat and Corn are 1000 Pounds

- Optimal Feed 1 Composition: Wheat: 722.22, Corn: 361.11, Alfalfa: 361.11.
- Optimal Feed 2 Composition: Wheat: 277.78, Corn: 222.22, Alfalfa: 1500.00.
- Total Profit: **$3059.72**.

Scenario 2: Upper Bounds for Wheat and Corn are 500 Pounds

- Optimal Feed 1 Composition: Wheat: 361.11, Corn: 180.56, Alfalfa: 180.56.
- Optimal Feed 2 Composition: Wheat: 138.89, Corn: 111.11, Alfalfa: 750.00.
- Total Profit: **$1529.86**.

Scenario 3: Wheat Availability is 1000 Pounds, Corn Availability is 500 Pounds

- Optimal Feed 1 Composition: Wheat: 750.0, Corn: 375.0, Alfalfa: 375.0.
- Optimal Feed 2 Composition: Wheat: 250.0, Corn: 125.0, Alfalfa: 1125.0.
- Total Profit: **$2725.0**.

Scenario 4: Wheat Availability is 500 Pounds, Corn Availability is 1000 Pounds

- Optimal Feed 1 Composition: Wheat: 291.67, Corn: 145.83, Alfalfa: 145.83.
- Optimal Feed 2 Composition: Wheat: 208.33, Corn: 166.67, Alfalfa: 1125.0.
- Total Profit: **$1776.04**.

Conclusion

The model adapts well to changes in resource availability. With more resources, Feedco can produce more of both feeds, leading to a higher profit. In the more constrained scenarios, the model efficiently allocates the limited resources to still maximize profit, albeit at a reduced scale. This flexibility and adaptability of the model make it a valuable tool for Feedco's production planning under varying conditions.

10) In the electricity unit commit model we covered in class, there was a logic constraint that helped ensure that the variables for turning on ("On"), turning off (Off), is on during the period ("Operating Now"), and was on in the last period ("Was Operating"). The constraint for every time period is (On – Off) – (Operating Now – Was Operating) = 0. In a situation where the plant was operating in the last period, and we want it to continue to operate, the Operating Now and Was Operating variables are both 1. To satisfy this constraint, the On and Off variables can be either both be 1 or both 0. Is there a flaw in this constraint? If so or not, please explain.

The constraint (On - Off) - (Operating Now - Was Operating) = 0 in the electricity unit commit model contains a logical flaw, especially evident when considering a power plant that was operating in the last period and continues to operate in the current period (both Operating Now and Was Operating are 1). In this case, the constraint simplifies to (On - Off) = 0, allowing both On and Off to be either 1 or 0. This presents a contradiction, as having both On and Off as 1 suggests the plant is being simultaneously turned on and off, which is not practically feasible. Ideally, for a continuously operating plant, we would expect On = 0 and Off = 0, which the current constraint fails to enforce. The model would benefit from a revised constraint that accurately reflects operational realities, ensuring that On and Off cannot both be 1 simultaneously and more accurately representing the physical state of the power plant.

11) Practice formulating Linear and Integer Programs. Formulate these for optimization:

A. Consider three binary variables: x, y, z. Formulate the two if-then constraints below:
If $x + y \geq 2$ then $z = 0$: $x + y \leq M(1 - z)$; $M \geq 2$
If $x + y \leq 1$ then $z = 1$: $x + y + M(z) \leq 1 + M$; $M \geq 1$

B. Formulate the following if-then constraints where $z_i, x_i \geq 0$ and $y_i \in \{0,1\}$
If $y_i = 1$, then $z_i \leq x_i$: $z_i - x_i \leq M(1 - y_i)$; $M \geq z_i - x_i + 1$
If $y_i = 0$, then $z_i \leq 0$: $z_i \leq My_i$; $M \geq z_i + 1$

C. Assume variables A, B, and C are continuous, non-negative decision variables. A sea port can load either 12 A's per week, or 44 B's, or 31 C's. What combinations of A, B and C can be loaded in 8 weeks?

Let $w_{A,i}, w_{B,i}$ and $w_{C,i}$ be binary decision variables for week i ($i = 1,2, \dots ,8$), indicating whether A,B, or C is loaded respectively.
In each week i, the port can choose to load A, B, or C exclusively.

- $w_{A,i} = 1$ if A is chosen in week i; 0 otherwise

- $w_{B,i} = 1$ if B is chosen in week i; 0 otherwise
- $w_{C,i} = 1$ if C is chosen in week i; 0 otherwise

We also have these capacity constraints:

- $a_i \leq 12 * w_{A,i}$
- $b_i \leq 44 * w_{B,i}$
- $c_i \leq 31 * w_{C,i}$

We can only load one of A, B, or C each week and have this constraint to deal with it:
$w_{A,i} + w_{B,i} + w_{C,i} \leq 1$ for each week i.

Finally, total quantity for 8 weeks:

- $a = \sum_{i=1}^{8} a_i$
- $b = \sum_{i=1}^{8} b_i$
- $c = \sum_{i=1}^{8} c_i$

12) You are advising a CEO who has just merged with another company. Let's call the two legacy companies Division 1 and Division 2. You have a pool of $200 million to invest in the merger. And you have identified 40 potential projects. (You can see the data in the spreadsheet posted with the final). Each project has an expected return to each of the divisions and an expected return to the company as a whole. This latter return is simply the sum of the returns to each division. You note that the two returns are not a 50/50 split—some of the investments will help one of the two divisions much more than the others. The CEO wants to make the right choice of investments to help the company. However, as with all mergers, the politics are tricky. The CEO needs both divisions to feel equal and work together as a team. Therefore, the CEO can't be perceived as favoring one division over the other. Run an optimization model for the CEO to help make a good decision on how to invest the $200 million. Explain your results.

In response to the CEO's request for an investment strategy post-merger, the optimization model's results indicate a well-balanced and profitable approach. The model, adeptly configured to maximize the total Net Present Value (NPV) within a $200 million budget, recommends investing in projects 8, 9, 11, 13, 14, 26, 30, 37, and 38. This selection is the outcome of an optimal solution found by the model, achieving a best objective of $321 million NPV, which signifies the highest possible return under the given constraints. Importantly, the model adheres to the critical balancing constraint, ensuring that the investment difference between Division 1 and Division 2 does not exceed $10 million. This consideration is vital for maintaining a harmonious relationship between the two divisions, addressing the CEO's concern about potential divisional favoritism in the wake of the merger. The chosen projects represent a strategic blend that maximizes overall returns while fostering equity and unity in the newly merged company. This solution not only promises financial efficacy but also aligns with the nuanced political dynamics of post-merger integration, offering a comprehensive roadmap for the CEO to navigate this complex scenario.

13) ***(Easy Half Point, if you answer it with an explanation):*** **(A). What are the most important lessons from this class** ***for you*****? What insights or ideas will be best for your future career? (B). What lectures, exercises, or readings were most impactful and memorable from this class?**

A. The most important lesson that I'm taking away from this class is that optimization exists and it can be very useful when approaching certain problems. Data scientists need to be able to recognize and solve a variety of problems and this class gives us a tool to do so. This is as stated in question 1 of this exam.

   I'm glad I took this course as it gives me the ability to understand problems in a different light and expand my field of thought.

B. The most important exercises were the ones I had to figure out by myself and struggle for hours on. One example of this was the cutting stock problem. During the initial few weeks of the course, I was still relatively new to optimization and wasn't quite used to formulating problems in the format of decision, constraints, and objective values. It forced me to take time practice thinking through creating the decision variables. Through that process, I got a lot better at it!

Thank you for the class. I had a great time and learned a lot. I'm glad I switched to this class from CS310! I hope you have a great rest of the quarter and a fantastic holiday! ☺

The End.

Thanks for your time this Quarter. We hope you learned a lot during course. And, we hope you find good ways to create value with optimization. Good luck with the start of your data science career. Keep in touch.

1 inspired by Wayne Winston's text book