Solutions to Sample Final Exam A

Learning Objective:

- Create Python code to automate a given task.
- Formulate linear optimization models to inform a business decision.

Instructions:

The final exam tests your mastery of skills taught in Weeks 7-12, which culminates in creating linear optimization models to inform a given business decision. There are three questions, worth a total of 30 points. The exam is 100 minutes, and is open-notes but closed-computer. You can bring paper notes or books of any kind, but no computers, tablets, or cell phones are allowed. Do not share your solutions with a student who has not yet completed an exam and do not look at other people's solutions. **Any violation of academic integrity will result in a zero grade for the exam for everyone involved.**

As long as you fulfill all the specifications described in the problem description, it doesn't matter how you model the problem or how efficient is your code. As with the midterm, partial credits will be given for any fragments of correct solutions, or for English descriptions that would lead to a solution.

Note: All three questions in this particular exam are motivated by the taxi and ride-hailing industry.

Q1. Airport Transportation Service (Concrete Formulation; 11 points)

Pedro owns a family business providing rides to and from the airport. His primary vehicle is a large van, which he operates himself. He owns a second vehicle, a small sedan, which is operated by his wife. Ride requests that are not fulfilled by him or his wife are passed on to his network of business partners. He has many business partners and you can assume that he can find available business partners to pick up every customer if needed.

Pedro would like to compute an optimal plan for fulfilling ride requests so as to maximize the total profit earned by him and his wife. (His business does not profit from requests passed on to his business partners.) The following table summarizes the ride requests his business received for a given morning.

Customer	Van Needed	Pickup Time	Pickup Location	Dropoff Location	Profit (dollars)
Anne	Yes	09:00	Suburb	Airport	60
Beth	No	10:00	Downtown	Airport	30
Cui	No	10:30	Airport	Suburb	50
Dwane	No	11:00	Airport	Downtown	35
Edna	No	11:30	Downtown	Airport	30
Frank	No	12:00	Suburb	Airport	55

Below are the travel times (in minutes) between each pair of locations. You can assume that these estimates are conservative, so that regardless of traffic conditions, one can always travel between any spot in the Airport and any spot in Downtown within 30 minutes, and between any spot in the Airport and any spot in the Surburb within 45 minutes, and so on.

Origin / Destination	Airport	Downtown	Suburb
Airport	0	30	45
Downtown	30	0	30

Origin / Destination	Airport	Downtown	Suburb
Suburb	45	30	0

Each customer is only requesting one car, so if Pedro picks up a customer, then his wife can be simultaneously fulfilling another request. Moreover, a ride needs to be completed before the same driver can pickup another customer. A plan for fulfilling ride requests also needs to satisfy the following:

- A request with a "Yes" in the "Van Needed" column can be assigned to Pedro, but it cannot be assigned to his wife.
- Each assigned driver needs to be able to arrive on time to each pickup location, based on the travel times above. For example, the same driver cannot pickup both Anne and Beth, because if the driver departs from the Suburb at 09:00, he/she would drop off Anne at the Airport at 09:45. If the driver immediately proceed to pick up Beth, he/she would arrive at Downtown at 10:15, which is later than the 10:00 pickup time requested by Beth. On the other hand, the same driver can pick up both Beth and Cui, because if the driver pickups Beth at Downtown at 10:00, he/she would drop her off at the airport at exactly 10:30, which is the perfect time to pickup Cui. You should examine the tables above to infer all such scheduling constraints. When doing so, you just have to make sure that after dropping off the previous customer, the driver can arrive on time at the next pickup location assigned to him or her, according to the driving time estimates. Moreover, you should not plan for extra margins: the driving times can be thought of as already incorporating everything needed for a trip, such as finding the passenger within the given location, loading or unloading luggage, and collecting payments.

A final consideration is as follows: that morning, his wife has other things she would like to attend to, so there is a large opportunity cost on her time. Pedro estimates that having his wife drive at all that morning is equivalent to incurring a fixed cost of 90 dollars, whereas if he does not assign any rides to her, he can save this cost. **This cost needs to be accounted for in the total profit being maximized.**

Formulate a linear optimization formulation for the above problem. You only need to write a concrete formulation, so you don't need any data variables but can simply plug in the numbers above. However, you cannot just solve the problem by hand and write down the optimal solution. In particular, if the "Profit (dollars)" column changes to other numbers, then one should be able to simply update the corresponding numbers in your formulation, and it should still be correct.

Sample Solution

Decision Variables:

- $X_A, X_B, X_C, X_D, X_E, X_F$: whether Pedro assigns the requests by Anne, Beth, Cui, Dwane, Edna, or Frank to himself. (Binary)
- $Y_A, Y_B, Y_C, Y_D, Y_E, Y_F$: whether Pedro assigns the requests by Anne, Beth, Cui, Dwane, Edna, or Frank to his wife. (Binary)
- Z: whether Pedro assigns any customers to his wife. (Binary)

Objective:

Maximize: $60(X_A+Y_A)+30(X_B+Y_B)+50(X_C+Y_C)+35(X_D+Y_D)+30(X_E+Y_E)+55(X_F+Y_F)-90Z$

Constraints:

$$X_A + Y_A \leq 1$$
$$X_B + Y_B \leq 1$$
$$X_C + Y_C \leq 1$$
$$X_D + Y_D \leq 1$$
$$X_E + Y_E \leq 1$$
$$X_F + Y_F \leq 1$$
$$Y_A = 0$$
$$X_A + X_B \leq 1$$
$$X_C + X_D \leq 1$$
$$X_C + X_E \leq 1$$
$$X_E + X_F \leq 1$$
$$Y_A + Y_B \leq 1$$
$$Y_C + Y_D \leq 1$$
$$Y_C + Y_E \leq 1$$
$$Y_E + Y_F \leq 1$$
$$Y_A, Y_B, Y_C, Y_D, Y_E, Y_F \leq Z$$

Q2. Pooling in Ride-Hailing (Abstract Formulation; 11 points)

This problem asks you to write a linear optimization formulation to solve a simplified version of the pooling problem faced by ride-hailing companies such as Uber and Lyft. These companies provide discounts to riders who are willing to let their trips be "pooled" with another trip, so that a driver would pick up both riders before dropping off each of them at their respective destinations. Through pooling, the travel time of each rider would increase, but driver capacity is more efficiently utilized. This practice is only profitable if the company finds good matches between trips so as to maximize the total benefit of pooling.

For simplicity, assume that each trip can be pooled with at most one other trip. The ride-hailing company has estimated the benefit of pooling for each pair of trips, which accounts for the potential cost savings from pooling and the potential inconveniences incurred for the pooled customers. As an illustration, suppose there are six trips, labelled A through F. The following table shows what the benefit values may look like.

Benefit of Pooling	A	B	C	D	E	F
A	0	6	4	3	1	1
B	6	0	5	5	2	3
C	4	5	0	1	4	3
D	3	5	1	0	2	1
E	1	2	4	2	0	4
F	1	3	3	1	4	0

Note that the table is symmetric, which means that it remains the same if it is transposed. In other words, pooling A-B is the same as pooling B-A; their order does not matter. As an illustration of how to use the table, suppose that trips B and C are pooled together, and trips E and F are pooled together, while trip A and trip D are not pooled; then the total benefit is $5 + 4 = 9$. As another example, suppose that trips A and B are pooled together, and none of the other trips are pooled; then the total benefit is 6.

The ride-hailing company would like to pool trips so as to maximize the total benefit of pooled trips subject to the following constraints:

- Each trip can be pooled with at most one other trip. (It is also possible for a trip to be not pooled, as in the examples above.)

- For a pair of trips, if the benefit of pooling is strictly less than a threshold t, then the trips cannot be pooled. For example, if $t = 3$, then the above table implies that A-D is a valid pooling, but B-E is not.
- The total number of pooled pairs is at most k. In the above example, if $t = 0$ and $k = 3$, then the optimal solution is to pool A-C, B-D, and E-F, which yields a total benefit of $4+5+4 = 13$. However, if $t = 0$ and $k = 1$, then the optimal solution is to only pool A-B, which yields a benefit of 6.

Your abstract formulation needs to be able to correctly handle arbitrarily many trips, arbitrary benefits of pooling, and arbitrary values of t and k. You may assume that the table is always symmetric as above, with zeros on the diagonal and non-negative entries everywhere else. Moreover, you may assume that t is always a non-negative number, and k is always a positive integer.

Sample Solution 1

Data:

- S: set of trips.
- v_{ij}: the benefit of pooling trip $i \in S$ with trip $j \in S$.
- t: the threshold on the benefit of pooling below which pooling for the given pair is not allowed.
- k: the maximum number of pooled pairs allowed.

Decision Variables:

- x_{ij}: whether to pool trip $i \in S$ with trip $j \in S$. (Binary)

Objective:

$$\text{Maximize: } 0.5 \sum_{i \in S, j \in S} v_{ij} x_{ij}$$

Constraints:

$$\begin{aligned}
\sum_{j \in S} x_{ij} &\leq 1 && \text{for each trip } i \in S. \\
x_{ii} &= 0 && \text{for each trip } i \in S. \\
x_{ij} &= x_{ji} && \text{for each pair } i \in S, j \in S. \\
x_{ij} &= 0 && \text{for each pair } i \in S, j \in S \text{ such that } v_{ij} < t. \\
0.5 \sum_{i \in S, j \in S} x_{ij} &\leq k
\end{aligned}$$

Sample Solution 2

Data:

- S: set of trips.
- P: set of pairs of distinct trips. Within this set, each pair would appear at most once: For example, if (A, B) already appears, then (B, A) would not appear.
- P_i: subset of P in which trip $i \in S$ is part of the pair.
- v_p: the benefit of pooling pair $p \in P$.
- t: the threshold on the benefit of pooling below which pooling for the given pair is not allowed.
- k: the maximum number of pooled pairs allowed.

Decision Variables:

- x_p: whether to pool the given pair of trips $p \in P$. (Binary)

Objective:

$$\text{Maximize: } \sum_{p \in P} v_p x_p$$

Constraints:

$$\begin{aligned}
\sum_{p \in P_i} x_p &\leq 1 && \text{for each trip } i \in S. \\
x_p &= 0 && \text{for each pair } p \in P \text{ such that } v_p < t. \\
\sum_{p \in P} x_p &\leq k
\end{aligned}$$

Q3. Strategic Driving (Gurobi Coding; 8 points)

A driver in a ride-hailing platform may increase earnings by strategizing about where to go next, rather than simply following the App. For example, at certain times in the day, waiting at Downtown for ride requests may lead to short trips that don't pay well. An alternative strategy might be to move to the Airport in hope of getting assigned to a long trip from there. Of course, whether this strategy is actually better depends on the expected earnings of each trip, the demand patterns at each location, and the amount of time left before the driver intends to go home.

The following linear optimization formulation solves a simplified version of the above optimization problem for a single driver. The formulation assumes that the time needed to travel between each pair of locations is one time period.

Data:

- L: the set of locations.
- s: the driver's initial location at time 0.
- n: the number of time periods to plan. The driver needs to go home at time n.
- $T = \{0, 1, 2, \cdots, n\}$: the set of time periods.
- $r_{i,j}$: the expected earnings of a trip from location $i \in L$ to $j \in L$. Note that it is possible that $i = j$, as this can correspond to short trips within a local region.
- $p_{i,j}$: the probability that if a driver turns on the App at location $i \in L$ in a given time period, he/she will receive a ride request from location i to location j in the same time period.

Decision Variables:

- $x_{i,t}$: the earnings-to-go if the driver is at location $i \in L$ at time $t \in T$. Earnings-to-go is defined as the total expected earning from time t until time n, assuming optimal behavior by the driver.

Objective and Constraints:

$$\begin{aligned}
\text{Minimize:} \quad & x_{s,0} \\
\text{s.t.:} \quad & \\
& x_{i,t} \geq x_{j,t+1} && \text{for each } i \in L,\ j \in L, \text{ and } t \in \{0, 1, \cdots, n-1\}. \\
& x_{i,t} \geq \sum_{j \in L} p_{i,j}(r_{i,j} + x_{j,t+1}) && \text{for each } i \in L \text{ and } t \in \{0, 1, \cdots, n-1\}. \\
& x_{i,t} \geq 0 && \text{for each } i \in L \text{ and } t \in T.
\end{aligned}$$

An explanation of the above formulation is outside the scope of this exam and is not necessary for solving this problem. You should simply trust that the formulation works.

Write a function called "plan_route" with three input arguments:

- **inputFile**: the filename to the input file. The format is explained below.
- **s**: a string representing the parameter s in the formulation above.
- **n**: a positive integer representing the parameter n in the formulation above.

The input file is an Excel spreadsheet with two sheets. The first sheet, called "Earnings", encodes the earnings $r_{i,j}$. It has the following format:

	A	B	C	D	E
1		Downtown	Airport	Suburb A	Suburb B
2	Downtown	10	60	30	25
3	Airport	70	0	80	90
4	Suburb A	35	70	0	30
5	Suburb B	30	80	30	0
6					

The second sheet, called "Transition Probabilities", encodes the probabilities $p_{i,j}$. It has the following format:

	A	B	C	D	E
1		Downtown	Airport	Suburb A	Suburb B
2	Downtown	0.75	0.05	0.1	0.1
3	Airport	0.3	0.2	0.2	0.3
4	Suburb A	0.3	0.2	0.4	0.1
5	Suburb B	0.4	0.2	0.1	0.3
6					

In both tables, the row corresponds to the first index, which is the origin, and the column corresponds to the second index, which is the destination. These tables are not necessarily symmetric. For example, in the above, $r_{Downtown,Airport} = 60$ while $r_{Airport,Downtown} = 70$. In other words, a ride from Downtown to Airport earns 60, whereas a trip from the Airport to Downtown earns 70. Your code needs to be able to handle arbitrary input data of the same format.

Your function should return two objects:

- **total_earnings**: a float corresponding to the optimal objective value of the above formulation, **rounded to two decimal places.**
- **earnings_table**: a DataFrame where the rows are the locations $i \in S$, and the columns are $t \in \{1, 2, \cdots, n\}$. Each entry contains the corresponding optimal value of $x_{i,t}$, **rounded to two decimal places.**

See the sample run for an illustration. (The sample run uses the data shown above.)

[3]:
```python
# Write all of your final code here
import pandas as pd
from gurobipy import Model, GRB
def plan_route(inputFile,s,n):
    r=pd.read_excel(inputFile,sheet_name='Earnings',index_col=0)
    p=pd.read_excel(inputFile,sheet_name='Transition Probabilities',index_col=0)
    L=r.index
    T=range(n+1)
    mod=Model()
    x=mod.addVars(L,T)
    mod.setObjective(x[s,0])
    for i in L:
        for j in L:
            for t in range(n):
                mod.addConstr(x[i,t]>=x[j,t+1])
    for i in L:
        for t in range (n):
            mod.addConstr(x[i,t]>=sum(p.loc[i,j]*(r.loc[i,j]+x[j,t+1]) for j in L))
    mod.setParam('OutputFlag', False)
    mod.optimize()
    total_earnings=round(mod.objval,2)
    earnings_table=pd.DataFrame(index=L,columns=range(1,n+1))
    for i in L:
        for t in range(1,n+1):
            earnings_table.loc[i,t]=round(x[i,t].x,2)
    return total_earnings,earnings_table
```

[4]:
```python
# Sample run
total_earnings,earnings_table=plan_route('sample-final-A-input.xlsx','Suburb A',6)
print('Optimal earnings starting at Suburb A:',total_earnings)
earnings_table
```

```
Optimal earnings starting at Suburb A: 206.22
```

	1	2	3	4	5	6
Downtown	169.86	134.48	96.4	64.0	16.0	0.0
Airport	206.22	169.86	134.48	96.4	64.0	0.0
Suburb A	169.86	134.48	97.98	64.0	27.5	0.0
Suburb B	173.22	136.7	101.48	64.0	31.0	0.0