Math 4630/5630

Homework 1 Solution

Problem 1. Facility Location problem.

- Define the decision variables in the following way.

  For i = NY, LA, Ch, At, let

  $$W_i = \begin{cases} 1 & if\ \ a\ warehouse\ is\ opened\ in\ city\ i \\ 0 & otherwise \end{cases}$$

  For i=NY,LA,Ch,At, and j=1,2,3, let

  X_{ij} = number of units sent from warehouse i to region j .

- Then the total cost is

  $400W_{NY} + 500W_{LA} + 300W_{Ch} + 150W_{At}$ (fixed costs)
  $+ 20X_{NY,1} + 40X_{NY,2} + 50X_{NY,3}$ (shipping costs)
  $+ 48X_{LA,1} + 15X_{LA,2} + 26X_{LA,3}$
  $+ 26X_{Ch,1} + 35X_{Ch,2} + 18X_{Ch,3}$
  $+ 24X_{At,1} + 50X_{At,2} + 35X_{At,3}$

We need the following constraints.

- We can't send units from a warehouse unless it is open:

  $X_{NY,1} + X_{NY,2} + X_{NY,3} \le 100\ W_{NY}$
  $X_{LA,1} + X_{LA,2} + X_{LA,3} \le 100\ W_{LA}$
  $X_{Ch,1} + X_{Ch,2} + X_{Ch,3} \le 100\ W_{Ch}$
  $X_{At,1} + X_{At,2} + X_{At,3} \le 100\ W_{At}$

  Note that, by choosing the big M number equal to 100, we take care of warehouse capacity constraints too.

- The demands of the three regions should be satisfied:

  $X_{NY,1} + X_{LA,1} + X_{Ch,1} + X_{At,1} \ge 80$ (region 1)
  $X_{NY,2} + X_{LA,2} + X_{Ch,2} + X_{At,2} \ge 70$ (region 2)
  $X_{NY,3} + X_{LA,3} + X_{Ch,3} + X_{At,3} \ge 40$ (region 3)

- The three restrictions:
  1. If the New York warehouse is opened, then the Los Angeles warehouse must be opened:
     $$W_{LA} \ge W_{NY}$$
  2. At most three warehouses can be opened:
     $$W_{NY} + W_{LA} + W_{Ch} + W_{At} \le 3$$
  3. Either the Atlanta or the Los Angeles warehouse must be opened:
     $$W_{LA} + W_{At} \ge 1$$

Summarizing, we have the following IP model:

Minimize $400W_{NY} + 500W_{LA} + 300W_{Ch} + 150W_{At}$
$+ 20X_{NY,1} + 40X_{NY,2} + 50X_{NY,3}$
$+ 48X_{LA,1} + 15X_{LA,2} + 26X_{LA,3}$
$+ 26X_{Ch,1} + 35X_{Ch,2} + 18X_{Ch,3}$
$+ 24X_{At,1} + 50X_{At,2} + 35X_{At,3}$

subject to

Constraints (i) relating W_i's and X_i's, (ii) giving capacity limits.

$X_{NY,1} + X_{NY,2} + X_{NY,3} \le 100\, W_{NY}$

$X_{LA,1} + X_{LA,2} + X_{LA,3} \le 100\, W_{LA}$

$X_{Ch,1} + X_{Ch,2} + X_{Ch,3} \le 100\, W_{Ch}$

$X_{At,1} + X_{At,2} + X_{At,3} \le 100\, W_{At}$

Demand constraints:

$X_{NY,1} + X_{LA,1} + X_{Ch,1} + X_{At,1} \ge 80$ (region 1)

$X_{NY,2} + X_{LA,2} + X_{Ch,2} + X_{At,2} \ge 80$ (region 2)

$X_{NY,3} + X_{LA,3} + X_{Ch,3} + X_{At,3} \ge 80$ (region 3)

The extra restrictions:

$W_{LA} \ge W_{NY}$

$W_{NY} + W_{LA} + W_{Ch} + W_{At} \le 3$

$W_{LA} + W_{At} \ge 1$

Set constraints:

W_i binary for i=NY,LA,Ch,At

X_j integer for j=1,2,3.

Problem 2. Blending problem.

(a) Let x_1 , x_2 and x_3 denote the number of kilograms of corn, tankage and alfalfa respectively. Then we have the following IP:

Minimize $84x_1 + 72x_2 + 60x_3$ (cost in cents)

subject to

$90x_1 + 20x_2 + 40x_3 \ge 200$ (carbohydrates)

$30x_1 + 80x_2 + 60x_3 \ge 180$ (protein)

$10x_1 + 20x_2 + 60x_3 \ge 150$ (vitamins)

$x_1 , x_2 , x_3 \ge 0$ integer

Note that it would also be reasonable to have continuous (not necessarily integer) variables for this problem.

(b) For i=1,2,3, let

$$y_i = \begin{cases} 1 & \text{if feed type } i \text{ is in the mix} \\ 0 & \text{otherwise} \end{cases}$$

Then we need to add the following constraints to part (a).

$y_1+y_2+y_3 \le 2$

$x_i \le My_i$ for i=1,2,3 and large positive M

y_i binary for i=1,2,3

(c) Change the constraints of part (a) to the following.

$90x_1 + 20x_2 + 40x_3 \ge 200 - 200y_1$

$30x_1 + 80x_2 + 60x_3 \ge 180 - 180y_2$

$10x_1 + 20x_2 + 60x_3 \ge 150 - 150y_3$

$y_1+y_2+y_3 = 1$
y_i binary for i=1,2,3
$x_1, x_2, x_3 \geq 0$ integer

Problem 3. Generalized knapsack problem.
Note that b[i,k] is not a linear function on k, and that complicates the problem. To overcome this difficulty, introduce the following decision variables. For each item i (i=1,2,3) and for k copies (k=0,1,...,L(i)):

$$y_{i,k} = \begin{cases} 1 & \text{if } k \text{ copies of item } i \text{ are taken} \\ 0 & \text{otherwise} \end{cases}$$

Then the total benefit is $\sum_{i=1}^{3} \sum_{k=1}^{L(i)} b[i,k] y_{i,k}$.

The knapsack capacity constraint is: $\sum_{i=1}^{3} \sum_{k=1}^{L(i)} k \cdot w_i \cdot y_{i,k} \leq W$. (W=20)

And we need the following constraints that force $y_{i,k}$ to its meaning:

$\sum_{k=0}^{L(i)} y_{i,k} = 1$, for each i=1,2,3.

Summarizing, we have the following IP model:

Maximize $\sum_{i=1}^{3} \sum_{k=1}^{L(i)} b[i,k] y_{i,k}$

subject to $\sum_{k=0}^{L(i)} y_{i,k} = 1$, for each i=1,2,3

$\sum_{i=1}^{3} \sum_{k=1}^{L(i)} k \cdot w_i \cdot y_{i,k} \leq W$

$y_{i,k}$ binary for i=1,2,3 and k=0,1,...,L(i)

Problem 4. Portfolio selection.

Define the following decision variables. For j=A,B,C,D, let
X_j=number of shares invested in alternative j.

Then the total investment can't exceed $200,000:

$100X_A + 50X_B + 80X_C + 40X_D \leq 200{,}000$

The constraint providing that the annual rate of return must be at least 9% is:

$0.12*100X_A + 0.08*50X_B + 0.06*80X_C + 0.1*40X_D \geq 0.09*200{,}000$

No one stock can count for more than 50% of the total dollar investment. The corresponding constraints are:

$100X_A \leq 0.5*(100X_A + 50X_B + 80X_C + 40X_D)$
$50X_B \leq 0.5*(100X_A + 50X_B + 80X_C + 40X_D)$
$80X_C \leq 0.5*(100X_A + 50X_B + 80X_C + 40X_D)$
$40X_D \leq 0.5*(100X_A + 50X_B + 80X_C + 40X_D)$

a) The total risk is $0.1*100X_A + 0.07*50X_B + 0.05*80X_C + 0.08*40X_D$.
Thus, the IP model for this case is:

Minimize $0.1*100X_A + 0.07*50X_B + 0.05*80X_C + 0.08*40X_D$
Subject to $100X_A + 50X_B + 80X_C + 40X_D \leq 200{,}000$
$0.12*100X_A + 0.08*50X_B + 0.06*80X_C + 0.1*40X_D \geq 0.09*200{,}000$
$100X_A \leq 0.5*(100X_A + 50X_B + 80X_C + 40X_D)$
$50X_B \leq 0.5*(100X_A + 50X_B + 80X_C + 40X_D)$
$80X_C \leq 0.5*(100X_A + 50X_B + 80X_C + 40X_D)$
$40X_D \leq 0.5*(100X_A + 50X_B + 80X_C + 40X_D)$
$X_A, X_B, X_C, X_D \geq 0$ integer

b) The projected return on investment is
$0.12*100X_A + 0.08*50X_B + 0.06*80X_C + 0.1*40X_D$.
Thus, the IP model for this case is:

Maximize $0.12*100X_A + 0.08*50X_B + 0.06*80X_C + 0.1*40X_D$
Subject to $100X_A + 50X_B + 80X_C + 40X_D \leq 200{,}000$
$0.12*100X_A + 0.08*50X_B + 0.06*80X_C + 0.1*40X_D \geq 0.09*200{,}000$
$100X_A \leq 0.5*(100X_A + 50X_B + 80X_C + 40X_D)$
$50X_B \leq 0.5*(100X_A + 50X_B + 80X_C + 40X_D)$
$80X_C \leq 0.5*(100X_A + 50X_B + 80X_C + 40X_D)$
$40X_D \leq 0.5*(100X_A + 50X_B + 80X_C + 40X_D)$
$X_A, X_B, X_C, X_D \geq 0$ integer

Problem 5. Linear regression.

The difficulty here lies in the fact that he optimization problem as it is stated in the problem set is not linear: the absolute value or the maximum functions are not linear. So we need to reformulate these somehow using simple tricks that make the problems linear.

a) Note that our goal is to find values for a_1 and a_0 which minimize $\sum_{i=1}^{n} |y_i - (a_1 x_i + a_0)|$.

Thus, a_1 and a_0 are variables, and x_i's and y_i's are given data. However, the above function is not linear. To make it linear, we need to introduce new variables. For i=1,…,n, let $z_i = |y_i - (a_1 x_i + a_0)|$. Then the new model is:

Minimize $\sum_{i=1}^{n} z_i$

subject to $z_i = |y_i - (a_1 x_i + a_0)|$, for each i=1,…,n

However, now we have non-linear functions in the constraints.
Suppose for each i=1,…n, we substitute $z_i = |y_i - (a_1 x_i + a_0)|$ by a pair of related constraints:

$z_i \geq y_i - (a_1 x_i + a_0)$ (1)

and $z_i \geq -y_i + (a_1 x_i + a_0)$ (2)

Note that (1) and (2) provide that $z_i \geq |y_i - (a_1 x_i + a_0)|$. But since our model is trying to minimize z_i's, in the optimal solution the value of each z_i will be taken all the way down to $|y_i - (a_1 x_i + a_0)|$. Summarizing, the linear program is:

Minimize $\sum_{i=1}^{n} z_i$

subject to $z_i \geq y_i - (a_1 x_i + a_0)$, for each i=1,…,n (1)

$z_i \geq -y_i + (a_1 x_i + a_0)$, for each i=1,…,n (2)

b) We want to $\min_a \max_i |y_i - (a_1 x_i + a_0)|$. a_1 and a_0 are variables, and x_i's and y_i's are given data. But the maximum of absolute values is not a linear function. To make it linear, we need to introduce a new variable. Let $z = \max_i |y_i - (a_1 x_i + a_0)|$. Then the new model is:

Minimize z

subject to $z = \max_i |y_i - (a_1 x_i + a_0)|$

Now we have a non-linear function in the constraint. However, the following equivalent formulation takes care of that problem.

Minimize z

subject to $z \geq y_i - (a_1 x_i + a_0)$, for each i=1,…,n (1)

$z \geq -y_i + (a_1 x_i + a_0)$, for each i=1,…,n (2)

Note that (1) and (2) provide that $z \geq \max_i |y_i - (a_1 x_i + a_0)|$. But since our model is trying to minimize z, in the optimal solution the value of each z will be taken all the way down to $\max_i |y_i - (a_1 x_i + a_0)|$.

c) Just replace $(a_1 x_i + a_0)$ above with $(a_k x_i^k + a_{k-1} x^{k-1}{}_i + \ldots + a_1 x_i + a_0)$.