Key: IPBR Problem Set 1

due June 21st, 2017

My answers and comments are in red. Remember that for many of these questions, there are multiple possible solutions. Especially when it comes to designing algorithms and writing code, my answers are just one way to solve the problem, not the only way.

Question 1

Evaluate (indicate whether they are true or false) the following logical expressions:

a. not 12 < 10 **true**

b. (6 > 5 or 1 < 0) or 7 < 8 **true**

c. 1 > 2 or (not 6 < 5) **true**

d. (1 = 1 or (not 12 < 11)) and (2 = 1 and 4 < 7) **false**

Evaluate the value of the logical variable *ans* in the following expressions:

e. myList = [1 6 8 2]
   ans = length of myList < 6
   ans = 4 < 6
   ans = true

f. A = true
   ans = not A
   ans = not true
   ans = false

g. B = false
   myArray = [6 5 4
   3 2 1]
   ans = myArray(1,2) > 4 or B
   ans = 5 > 4 or false
   ans = true

Write a logical expression that:

h. Always evaluates as true
   6 > 5

i. Evaluates as true if A is greater than 3 or if B is less than 7, false otherwise
   A > 3 or B < 7

Question 2

Consider the following array A, and answer the questions below:

A = [4.07 5.48 6.12
4.52 5.36 5.98
5.01 5.76 6.18
5.02 6.37 6.76]

The columns of this array represent the prices for different fruits: column 1 is apples, column 2 is oranges, and column 3 is pears. The rows of this array represent years from 1981 (first row) to 1984 (last row).

a. What are the dimensions (what is the size) of A?
4x3 because A has 4 rows and 3 columns

b. What element of the array is found at indices A(1,1)? What does this number represent? (I.e. the price of ___ in the year ___.)
A(1,1) = 4.07, the price of apples in the year 1981

c. What element of the array is found at indices A(3,2)? What does this number represent?
A(3,2) = 5.76, the price of oranges in the year 1983

d. If I wanted to know the price of pears in the year 1981, what indices would I use to find this information in the array? (i.e. A(_,_))
A(1,3)

e. Write an algorithm to find the difference between the price of oranges in 1981 and 1983. Use array indexing to find the data that is needed for the algorithm.
pDiff = A(3,2) – A(1,2)

f. Suppose I wanted to add some data on kiwi prices to this array. Kiwi prices in the years 1981-1984 were $2.12, $2.27, $2.19, and $2.46. How can we modify array A to include this new data? What does the new version of this array look like?
We can add this data to a fourth column in the array:
A = [4.07 5.48 6.12 2.12
4.52 5.36 5.98 2.27
5.01 5.76 6.18 2.19
5.02 6.37 6.76 2.46]

g. Suppose that there is a second variable B = 10. If I enter the command `A(3,1) = B`, what would happen to the values of A and B? What happens if I instead enter the command `B = A(3,1)`?

For the command A(3,1) = B:
A = [4.07 5.48 6.12 2.12
4.52 5.36 5.98 2.27
10 5.76 6.18 2.19
5.02 6.37 6.76 2.46]
B = 10

For the command B = A(3,1)
A = [4.07 5.48 6.12 2.12
4.52 5.36 5.98 2.27
5.01 5.76 6.18 2.19
5.02 6.37 6.76 2.46]
B = 5.01

Question 3

Suppose you are interested in mouse behavior, and you want to know what odors scare mice. When mice are afraid, they stop moving and freeze. To test which odors are fear inducing, you record a movie of the mouse walking around its cage, and randomly release one of ten odors. You record the mouse for 2 hours, and expose the mouse to a randomly selected odor once every 2 minutes (for 30 seconds).

Somebody else's program automatically records the results in two variables. The first variable contains the coordinates of the mouse, recorded every second. The second variable records which odor is being released (stored as the numbers 1-10, with a number 0 if no odor is being released), also recorded every second. Your goal is to create a bar graph that plots the average speed of the mouse when each odor is exposed.

a. What are the sizes of your two starting variables?
   Coordinate variable: 7200x2
   This is the number of recordings (each second for 2 hours, 60*60*2) by the number of coordinates recorded (x and y)
   Odor variable: 7200x1 *(can also be written as 7200)*

b. What do the different dimensions of your two variables represent?
   Coordinate variable: rows = time, columns = coordinate
   Odor variable: rows = time

c. What is the size of your final variable going to be?
   We want to store the mean velocity for each odor, and the control; our final variable will be size 11x1.

d. What do the different dimensions of your final variable represent?
   Rows: odor.

e. Describe an algorithm that takes you from the first two variables to the final desired variable.

   First, we will calculate the mouse's speed at each time point:

   1. **Using the coordinate variable, calculate the difference of each value in column 1 from the value above it. This tells you how far in x the mouse moved between two frames. Do the same for column 2, which tells you the difference in y traversed in every frame.**
   2. **Square every element in your array.**
   3. **Then, for every row, sum the values in its two columns. The resulting array is a 7200x1 array that tells you the speed (distance/s) of the mouse at each time point. Call this the SpeedArray.**

Then, we will match the speed at each time point to the correct odor

4. **For every odorant i (in the range 0-10), find the rows in your odor variable that recorded odor i. For those same rows in SpeedArray, get the velocity values in those rows. These are the instantaneous velocity values recorded while odor i was being exposed. Take the average of these velocities, and store at in the ith row of your final variable**
5. **You can then use the values in the final variable to generate a bar graph, where each bar corresponds to a particular odor and the heights of the bars are the mean velocity values recorded in the final variable.**

Question 4

Suppose you are interested in the stress response of yeast cells to different concentrations of sodium. You want to know how the expression of 1000 different genes changes across 5 different sodium concentrations. To do this, you perform single cell RNA-seq from 100 cells in each condition to get the expression levels of each of the 1000 different genes. Suppose you want to know for each condition which genes have expression levels 10x greater than in control conditions (in answering the questions below, assume that there are 2,5,50 and 55 genes that satisfy this condition for the 4 experimental concentrations).

1. How might you store this data in MATLAB? How many variables would you use? What are the sizes of each of the dimensions of your variables? What do these dimensions represent?
I would use a single variable: a 3-dimensional matrix. The matrix would have dimensions 1000x5x100, which represent genes, salt concentrations, and cells, respectively. For the columns (salt concentrations), the experimental conditions are in columns 1-4, and the control condition is in column 5.

2. What is(are) the size(s) of your final variable(s) going to be?
I will make four final variables: a 1x2 list, a 1x5 list, a 1x50 list, and a 1x55 list, representing the identified genes for each of the four experimental conditions.

3. What might be a simple algorithm to take you from your initial variables to your final variable?
    1. **For the initial 3D matrix, average across the third dimension to generate an average expression level for each gene. The resulting matrix will have dimensions 1000x5**
    2. **Now, we can calculate the fold expression change for each gene. For each experimental condition column (columns 1-4), divide the expression level of each gene by the value in the control condition (column 5). Store the results in a new matrix, foldChange.**
    3. **For each column in foldChange, go through the column and find the values in the column that are greater than 10. For those values, store the corresponding gene (either record the row number, or use the row number to identify the gene name) in a list. Make one list for each column.**