IPBR Problem Set 1

due June 21st, 2017

This problem set is chance to practice basic algorithm design, variable assignment, and array indexing. It is not necessary to use MATLAB for this problem set, although you can if you would like. Remember that problem sets are not graded- this is just a chance for you to practice what you learned in class, and you do not need to complete every section. For feedback, submit this problem set by e-mail to squyres@g.harvard.edu, or in hard copy at the start of lecture on Wednesday.

Question 1

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

a. not 12 < 10

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

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

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

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

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

f. A = true
ans = not A

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

Write a logical expression that:

h. Always evaluates as true

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

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?

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 ___.)

c. What element of the array is found at indices A(3,2)? What does this number represent?

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(_,_))

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.

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?

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)`?

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?

b. What do the different dimensions of your two variables represent?

c. What is the size of your final variable going to be?

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

e. Describe an algorithm that takes you from the first two variables to the final desired 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 1 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?
2. What is(are) the size(s) of your final variable(s) going to be?
3. What might be a simple algorithm to take you from your inital variables to your final one