Homework Due 2014-11-21 by 23:59 EDT

1 General Instructions

Please read these instructions carefully for each assignment, though they are generally do not vary between the assignments.

1. You need to follow carefully the specific instructions for the assignment as written below.
2. *If you have questions concerning this homework email the course assistant responsible for this assignment:* Heng Lyu, mailto:hl1463@nyu.edu. The course assistant is ready to help you understand the assignment but cannot modify the assignment in any way unless a change is posted on NYU Classes.

   If you still need to follow up and communicate with the course instructor please email Zvi Kedem *in the way specified in the course description on NYU Classes.*

   Of course, Zvi is available to help you understand the course material as you need it, so you should set up an appointment to see him if you need help. Meeting in person is more effective for this than communicating by email.
3. Submit your homework in electronic form by uploading it to NYU Classes by the due date and time. Use only permitted software and format. E.g., if you are asked to run on a Matlab or a clone, that is what you must do.
4. If you submit a scanned, handwritten assignment, it has to be written neatly, that is, it should be neatly divided into lines just as a typeset document, etc.
5. Show *all* your applicable work (other than for reading assignments, if any).
6. If you want to refer to a specific line in this document, refer to the small numbers in the left margin.
7. Be sure that you submitted what you were supposed to submit and wanted to submit. You are responsible for that. The best thing to do is to download what you have submitted and to look at it and make sure of that.
8. You are unable to submit your solutions by the deadline, write Zvi Kedem as soon as possible.
9. **Be sure to follow the academic integrity rules listed in the course description posted on NYU Classes**. The department and the GSAS treat academic integrity very seriously and I am required to report all possible violations.

2 Assignments

Note: No claim of originality is made with respect to any of the assignments below. They may have been taken from various published materials.

For non-programming assignments, you may handwrite them (*neatly*) and scan them. But in any case they must be in PDF (not Word, or similar). Put all the answers in one document. Clearly indicate the beginning of each problem/subproblem, identifying it using the numbers given the problems below. Do not use a computer for computations, unless specified as part of the assignment.

For Matlab problems/subproblems, submit a main script "main.m" and all your functions. For instance you are asked to write two functions: foo and bar, then you need to submit three files for the programming part: "main.m", "foo.m", "bar.m". You may define private helper functions in each function file. Put your name, NetID, and N-number as comment at the beginning of "main.m". Clearly indicated the beginning of each problem/subproblem, identifying it using the numbers given the problems below. Clearly define input and output of each problem/subproblem. Your "main.m" must be executable and reflect the output of each problem/subproblem after execution. You don't need to submit the running log this time.

For Octave users, please submit function script files and a running log.

In questions below

- C_0 is a fair coin; it comes up heads 50% of the time
- C_1 is an unfair coin; it comes up heads 80% of the time.
- C_2 is an unfair coin; it has heads on both sides so it comes up heads 100% of the time.
- D_0 is a fair die, each number comes up with the probability 1/6.
- D_1 is a die in which various numbers comes with various probabilities, as specified below:

$$\Pr[i] = \begin{cases} 0.25 & \text{for } i = 1 \\ 0.25 & \text{for } i = 2 \\ 0.20 & \text{for } i = 3 \\ 0.15 & \text{for } i = 4 \\ 0.10 & \text{for } i = 5 \\ 0.05 & \text{for } i = 6 \end{cases}$$

1. Make sure that you read and understand the material in the "What we have covered so far" file, posted after the class of Monday, 2014-11-03. The non-MATLAB questions are minimal just for you to make sure you understand the material.

2. (a) Using a suitable drawing and additional text, explain why

$$\Pr[A \mid B] + \Pr\left[\bar{A} \mid B\right] = 1$$

   for any event B such that $\Pr[B] \neq 0$.

   (b) Let A_1, A_2, A_3 be events such that

i.

$$\Pr[A_i, A_j] = 0 \quad \text{for} \quad i \neq j, \quad \text{and}$$

ii.

$$\sum_{i=1}^{3} \Pr[A_i] = 1.$$

Explain why

$$\sum_{i=1}^{3} \Pr[A_i \mid B] = 1$$

for any event B such that $\Pr[B] \neq 0$.

3. Alice tosses C_0, then rolls D_0. If the result is ≤ 2, she tosses C_0, otherwise, she tosses C_1.

   What is the probability that she gets exactly one H in the two tosses?

4. On the table there are 8 C_0 coins and 3 C_1 coins.

   Alice picks a coin randomly without looking at it and tosses it some n times. She reports the results to Bob

   (a) $n = 1$ and I got one H. What is the probability that the coin is C_0?

   (b) $n = 2$ and I got two H. What is the probability that the coin is C_0?

   (c) $n = 2$ and I got one H and one T but I am not telling you the order, so it was either HT or TH. What is the probability that the coin is C_0?

   Please use Bayes' theorem.

5. There are three urns, each containing two balls

   - Urn 1 has two black balls
   - Urn 2 has two white balls
   - Urn 3 has one white ball and black ball

   Alice picks an urn randomly and out of the urn a ball randomly. The ball is black. What is the probability that the other ball in that urn is black too. Please use Bayes' theorem.

6. Alice tosses D_1. Bob asks her yes/no questions to find out what the result was. Bob uses the specific "search tree algorithm" described in `Probability` and shown in Fig. 18 there.

   (a) Compute with the precision of two decimal digits of accuracy after the decimal point what is the expected number of questions Bob asks.

   (b) Compute with the precision of two decimal digits of accuracy after the decimal point what is the entropy.

7. There are two parts: (a) and (b). You may submit part (b) with the next assignment instead of the current one, if you prefer.

(a) Write a Matlab function P = NBProbability(S, D, N) that does the following:
S: m-by-n boolean matrix of symptom history where S(i, j) = 1 indicates that in the i'th patient record symptom j was observed.
D: m-by-1 boolean vector where D(i) = 1 indicates that the i'th patient was diagnosed with a flu (assume these records are correct).
N: 1-by-n boolean vector where N(j) = 1 indicates that the symptom j is observed in a new case. Assume each symptom is conditionally independent with each other.
P: compute the Naive Bayes probability that the new patient has a flu, i.e. eliminate the evidence term in the Bayes theorem since it is invariant whether the patient has a flu or not and thus it is useless in classification problems.

For instance, S = logical([1,0,0,1;1,1,0,0;1,0,1,1;0,1,0,1;0,0,0,0;0,1,1,1;0,1,1,0;1,1,0,1]); D=logical([0;1;1;0;1;0;1]); N=logical([1,0,0,0]);, then P = 0.0090.

(b) Further assume that each symptom has multiple states instead of two, plus there are k diseases to be diagnosed. S, D, N are thus of type uint8 instead of boolean. Write a Matlab function [P, I] = NBProbability2(S, D, N) that returns the probability of each disease I where P and I are k-by-1 vectors. Normalize P such that the sum of all probabilities equals 1.