Artificial Intelligence
Fall 2014
Homework 3

Show all your work for full credit!

1. Suppose a person goes for a heart check-up and the doctor tries to ascertain whether that person has a heart condition based on three characteristics: a patient's *gender*, *blood pressure*, and *electrocardiograph (ECG) reading.* The characteristics can take on the following values:

   - Gender: Male or Female
   - Blood pressure: Low or High
   - ECG reading: Normal or Abnormal

   We are given the following information:

   $$P(\text{having a heart condition}) = 0.1$$

   $$P(\text{gender is male} \,|\, \text{a patient does not have a heart condition}) = 0.8$$
   $$P(\text{gender is male} \,|\, \text{a patient has a heart condition}) = 0.6$$

   $$P(\text{high blood pressure} \,|\, \text{a patient does not have a heart condition}) = 0.6$$
   $$P(\text{high blood pressure} \,|\, \text{a patient has a heart condition}) = 0.7$$

   $$P(\text{abnormal ECG} \,|\, \text{a patient does not have a heart condition}) = 0.5$$
   $$P(\text{abnormal ECG} \,|\, \text{a patient has has a heart condition}) = 0.8$$

   You may assume the patient's gender, blood pressure, and ECG reading are all conditionally independent of each other given the presence or absence of a heart condition.

   (a) A female patient has low blood pressure and an abnormal ECG. What are the maximum likelihood and MAP hypotheses regarding whether or not she has a heart condition?

   (b) What is the posterior probability of this patient having a heart condition?

   (c) A male patient has low blood pressure and a normal ECG. What are the maximum likelihood and MAP hypotheses regarding whether or not he has a heart condition?

2. You are training your Naive Bayes spam classifier with the following data set:
   - You have a training set of 100 emails.
   - 58 of the 100 emails are spam.
   - 47 of the 58 emails that are spam have the word "buy"
   - 41 of the 58 emails that are spam have the word "win"
   - 42 of the 100 emails aren't spam
   - 3 of the 42 emails that aren't spam have the word "buy"
   - 5 of those 42 emails that aren't spam have the word "win"

   (a) Show the prior probabilities calculated for each of the hypotheses (spam and not-spam), as well as the probabilities for each of the features given each of the hypotheses.

   Remember that the probabilities for each feature given the hypothesis should be smoothed, but the priors for the hypotheses should not be smoothed. See the class webpage for notes if that doesn't ring a bell.

   (b) A new email arrives which includes both the words "buy" and "win." Tell if the new email should be classified as spam or not-spam (show the calculations for the MAP hypothesis).

   Hint: the posterior probability of spam for the newly-arrived email should be about 0.984.

3. *Note: This problem involves multiplying vectors and matrices. You are welcome to use software to do the multiplications itself, however, you should show all your work. Another way of thinking about this is that you should understand the math well enough to do it by hand if I asked you to. That is, you should know how to multiply 2-by-2 matrices and vectors by hand. (Hint, hint.)*

   Professor Somnus is investigating whether the students in his class are getting enough sleep. He collects some data and deduces that the probability that a student will get enough sleep on a given night only depends on whether they got enough sleep the night before. If a student gets enough sleep the previous night, the probability they will get enough sleep tonight is 0.8. If they didn't get enough sleep the previous night, the probability they will get enough sleep tonight is only 0.3.

   (a) Formulate this problem using a Markov chain.

   (1) Draw the Markov chain diagram showing the probability of transitioning between "enough sleep" and "not enough sleep." (This is the not the Bayes net diagram; this is the diagram that shows up at the top of the page at en.wikipedia.org/wiki/Markov_chain.)

   (2) Write down the transition matrix T.

   (b) Assuming you got enough sleep on night 0, what's the probability you get enough sleep on night 3?

(1) Show the initial state vector, v_0.

(2) Calculate and show the probability distribution for nights 1, 2, and 3 by multiplying by T.

(c) In the far, far, future, what is the probability of getting enough sleep on some night?

4. Continuing with the situation in Problem 3, Professor Somnus has no way of directly observing whether or not his students are getting enough sleep. All he can observe is whether they are sleeping in his class or not. He knows that if a student gets enough sleep on some night, the next day there is a 0.1 probability that they will fall asleep in class. If they don't get enough sleep the night before, there's a 0.3 probability of falling asleep in class.

   For this question, assume that night x is followed by day x.

   (a) Professor Somnus observes a student falling asleep in class on day 1, staying awake on day 2, but falling asleep on day 3. Calculate the probability, using the forward algorithm, that the student got enough sleep on night 3, given that sequence of observations. (Remember, night 3 happens right before day 3). The professor assumes there's an equal prior probability of enough sleep/not enough sleep on night zero.

   (1) Write out the appropriate matrices O_1, O_2, and O_3.

   (2) Write out the initial vector f_0.

   (3) Use the forward algorithm to calculate the probability distributions for enough sleep/not enough sleep on nights 1, 2, and 3. Show your work.

   (b) Professor Somnus remembers that in Hidden Markov Models, the observations can affect our beliefs about the past, as well as the future. Use the backward algorithm to re-calculate Professor Somnus's belief of getting enough sleep on night 1, given the the sequence of observations. Show your work.