EECS 349 Problem Set 4
Due 11:59PM Thursday, June 2

Submit a single PDF file.

1. **(2.5 points)**
   In the first homework, you were asked to try to find two algorithms that gave very different performance on a given data set. In this homework question, you'll instead try to create a data set that gives very different performance for two given algorithms. Specifically, can you create a data set of 1000 examples with at most 1000 attributes such that Weka gives very different 10-fold cross-validation performance (in absolute value) for nearest neighbor (IBk) vs. decision trees (J48)? You only need to create one data set, for which one of the two algorithms wins. You can choose which of IBk or J48 you want to be the winner.

   How large is “very different”? Here is what to aim for in this question and in Question 2:
   - 10%+ difference in accuracy: half credit
   - 20%+ difference in accuracy: 75% credit
   - 30%+ difference in accuracy: full credit

   (however, 40%+ differences are definitely possible for both questions)

   Important: you should use the default settings in Weka for each algorithm -- you can find IBk under "lazy" in the "classifiers" section. Report exactly how you generated the data set (including the code if applicable) and a table giving the 10-fold CV accuracy of each algorithm on your data. You do not have to include an explanation why your data set works the way it does, or include the data itself. Just describe the data clearly and include the results.

2. **(2.5 points)**
   Repeat Question 1, except using multi-layer perceptrons (found under "functions" in Weka) vs. Naive Bayes classifiers (found under "bayes" in Weka).

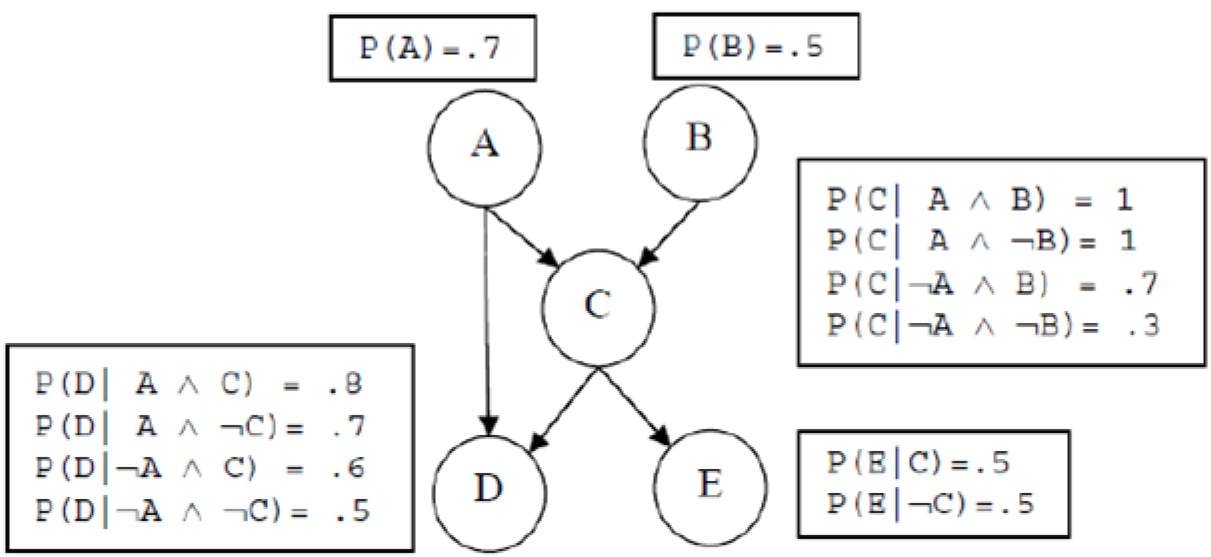
   Yes, you *can* use the same data set in Question 1 as in Question 2, if you like.

3. **(2 points)**
   For each of the following approaches, give an example of an optimization problem which is more easily solved by that approach than by any of the other approaches. For each problem, explain why it's so well-suited for the approach you chose (0.33 points x 3 = 1.0 points) and why it’s not so well-suited for the other approaches (0.33 points x 3 = 1.0 points). Note, for this question, you can choose in each case whether the function to be optimized is known in mathematical form to the algorithm.
   a. Genetic Algorithms
   b. Gradient Descent
   c. Hill Climbing

4. **(2.5 points)**
Consider six binary variables related to scoring well on an exam: G (got a good night's sleep), S (studied a lot), I (find the material interesting), E (exam is easy), A (got an A on exam), R (recommend class to friends at end of quarter).

   a. (1 point) Draw a Bayes Net representing this situation. There are multiple different reasonable networks for this domain. You should try to exploit conditional independencies and end up with relatively few edges.
   b. (0.25 points) What is the minimum number of probabilities you would need in order to specify all the conditional probability tables for your Bayes Net?
   c. (0.25 points) How many independent parameters are needed to specify the full joint distribution over the six variables, assuming no independencies?
   d. (0.5 points) Using your network, is E (whether the exam is easy) independent of G (whether you got a good night's sleep) when we are not given the value of any other variables? In a sentence, justify why or why not.
   e. (0.5 points) Using your network, assume we know A = true (you got an A on the exam). Is E conditionally independent of G given A? In a sentence, say why or why not.

5. **(0.5 points)**
Consider the Bayes Net pictured below.



What is P(C) for the Bayes Net? Show your work (Hint: this should be about two lines of math).