

# CSEP 573 Final Exam – March 12, 2016

**Name:**

This exam is take home and is due on **Sunday March 20th at 11:45 pm**. You can submit it in the online DropBox or to the course staff. This exam should not take significantly longer than 3 hours to complete if you have already carefully studied all of course material. Studying while taking the exam may take longer. :)

This exam is open book and open notes, but you must complete all of the work yourself with no help from others. Please feel free to post clarification questions to the course message board, but please do not discuss solutions.

If you show your work and *\*briefly\** describe your approach to the longer questions, we will happily give partially credit, where possible.

There are 8 pages in this exam.

Scores					
Q.1 (30)	Q.2 (30)	Q.3 (20)	Q.4 (20)	Q.5 (20)	Total (120)

## Question 1 – True/False – 30 points

Circle the correct answer for each True / False question. If you think a question is ambiguous, please add a very short explanation of the interpretation you are making, and we will do our best to grade accordingly.

1. True / False – Adding more edges to a Bayesian network can restrict the space of possible distributions it can represent. (3 pt)
2. True / False – For answering conditional queries in Bayesian networks, rejection sampling has generally been observed to provide worse estimates than likelihood weighting (when given the same number of samples). (3 pt)
3. True / False – Naive Bayes models always encode incorrect independence assumptions. (3 pt)
4. True / False – The Perceptron will always converge if the data is linearly separable. (3 pt)
5. True / False – Overfitting occurs when the test error is higher than the training error. (3 pt)
6. True / False – Inference by enumeration can produce incorrect results if the Bayes network is dense (has many edges). (3 pt)
7. True / False – The HMM forward inference algorithm takes time that is polynomial in the number of observations that have been received. (3 pt)
8. True / False – The choice of the variable ordering in variable elimination does not change the correctness of the algorithm (you will always get the correct answer for any ordering). (3 pt)
9. True / False – Naive Bayes, as presented in class, is an online learning algorithm. (3 pt)
10. True / False – The number of parameters in a Bayesian network grows exponentially with the highest out degree of a node in the network. (3 pt)

## Question 2 – Short Answer – 30 points

These short answer questions can be answered with a few sentences each. Please be brief, we will subtract points for very long responses (e.g. more than a sentence or two for each part of the question).

1. Short Answer – Briefly describe how you would decide which algorithm to use for answering queries to a Bayesian network. What is the key property of the network that, if known, would best help you make the appropriate decision. (5 pts)
  
2. Short Answer – In machine learning, explain generalization and over-fitting. Describe an experimental setup that correctly measures generalization. Assume that your algorithm has one hyperparameter that must be set. (5 pts)
  
3. Short Answer – Briefly describe a situation in which you would use Bayes rule, and why, from the examples we saw in class. (5 pts)
  
4. Short Answer – Briefly describe a sign of overfitting in Naive Bayes learning, and how it can be avoided. (5 pts)

5. Short Answer – Briefly describe when you would prefer to report precision and recall for a learned classifier, instead of accuracy. (5 its)

6. Short Answer – Briefly describe the difference between outcomes and events in joint probability models. (5 its)

### Question 3 – Hidden Markov Models: Tricky Coins – 20 points

Consider the following random process. A magician has two coins, each of which has an unknown type. They can either be fair coins (50/50 odds of heads vs tails), or trick coins that either (1) have heads on both sides or (2) have tails on both sides. A priori, each coin is equally likely to be any of the three possible types.

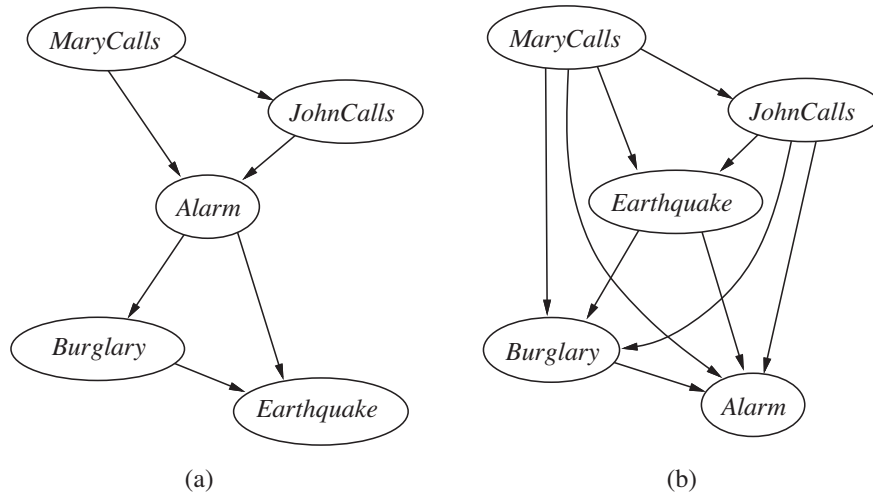
At every time step, the magician randomly picks a coin (without showing you which one was selected), flips it, and shows you the result. However, unfortunately, the magician only shows you the coin very briefly, and 10% of the time you make a mistake when you observe the true side of the coin (e.g. you see heads when it was actually tails).

1. Model this process as an HMM. Specify all of the necessary parameters. You do not have to write out all of the probability distributions explicitly, but be careful to specify what values they would have if you did the full enumeration. [15 pts]

2. Consider the Markov model that would result if you ran the process above and always observed heads. What is the stationary distribution of this model? [5pts]

## Question 4 – Bayesian Networks – 20 points

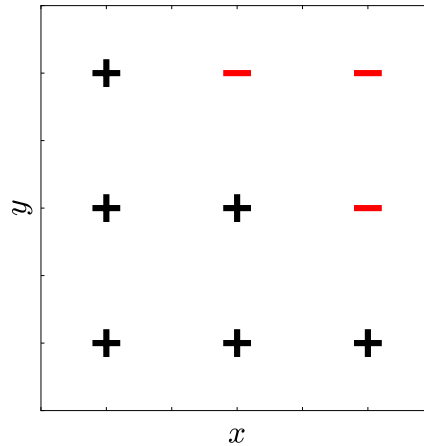
Consider the following two Bayesian networks, which are variations on the alarm network we discussed in class:



1. Based on the network structure alone, which network above makes the most independence (marginal or conditional) assumptions? [3 pts]
2. Draw a new Bayesian network with the same set of random variables that makes as many independence assumptions as possible. [5 pts]
3. Write down two conditional independence assumptions encoded by the structure of network (a). If there are not two, write as many as possible. [6 pts]
4. Write down two conditional independence assumptions encoded by the structure of network (b). If there are not two, write as many as possible. [6 pts]

## Question 5 – Perceptron – 20 points

Consider the following training set, where the  $x$  and  $y$  axis represent the values of two features and the examples are marked with  $+$  for the positive class and  $-$  for the negative class:



1. In the figure above, draw a decision boundary that the Perceptron could learn. [5 pts]
2. Briefly describe why you drew the line you did for the previous question. Are other separators possible? [5 pts]
3. The Perceptron is known to not converge in some situations. In the data above, circle one datapoint that, if you were to change its class, the Perceptron would no longer converge. [5 pts]
4. Now, given your new dataset from the last question, briefly describe a change that you could make which would, again, allow the Perceptron to converge. You cannot change the number of training examples or the labels they are assigned, but anything else is fair game. [5 pts]