CS440/ECE448 Fall 2015 Final Review

Be able to define the following terms and answer basic questions about them:

- **Probability**
  - Random variables
  - Axioms of probability
  - Joint, marginal, conditional probability distributions
  - Independence and conditional independence
  - Bayes rule

- **Bayesian inference**
  - Likelihood, prior, posterior
  - Maximum likelihood (ML), maximum a posteriori (MAP) inference
  - Naïve Bayes
  - Parameter learning

- **Bayesian networks**
  - Structure and parameters
  - Conditional independence assumptions
  - Calculating joint and conditional probabilities
  - Inference
  - Complexity of inference (worst-case, special classes of networks for which efficient inference is possible)
  - Parameter learning (very high level)
  - Hidden Markov models (definition, types of inference problems)

- **Markov decision processes**
  - Markov assumption, transition model, policy
  - Bellman equation
  - Value iteration, policy iteration

- **Reinforcement learning**
  - Model-based vs. model-free approaches
  - Exploration vs. exploitation
  - TD Q-learning

- **Machine learning**
  - Training, testing, generalization, overfitting
  - Supervised vs. unsupervised learning
  - Nearest neighbor classifiers
  - Perceptrons (incl. perceptron learning algorithm)
  - Support vector machines (incl. kernel support vector machines)
  - Neural networks (incl. training procedure)
  - Deep convolutional neural networks

Sample exam questions

1. Use the axioms of probability to prove that $P(\neg A) = 1 - P(A)$.

2. A couple has two children, and one of them is a boy. What is the probability of the other one being a boy?

3. Consider the following joint probability distribution:

   $P(A = \text{true}, B = \text{true}) = 0.12$
   $P(A = \text{true}, B = \text{false}) = 0.18$
   $P(A = \text{false}, B = \text{true}) = 0.28$
   $P(A = \text{false}, B = \text{false}) = 0.42$

   What are the marginal distributions of A and B? Are A and B independent and why?

4. A friend who works in a big city owns two cars, one small and one large. Three-quarters of the time he drives the small car to work, and one-quarter of the time he drives the large car. If he takes the small car, he usually has little trouble parking, and so is at work on time with probability 0.9. If he takes the large car, he is at work on time with probability 0.6. Given that he was on time on a particular morning, what is the probability that he drove the small car?

5. We have a bag of three biased coins, *a*, *b*, and *c*, with probabilities of coming up heads of 20%, 60%, and 80%, respectively. One coin is drawn randomly from the bag (with equal likelihood of drawing each of the three coins), and then the coin is flipped three times to generate the outcomes X_1, X_2, and X_3.

   a. Draw the Bayesian network corresponding to this setup and define the necessary conditional probability tables (CPTs).

   b. Calculate which coin was most likely to have been drawn from the bag if the observed flips come out heads twice and tails once.

6. Consider the data points in the table below representing a set of seven patients with up to three different symptoms. We want to use the Naive Bayes assumption to diagnose whether a person has the flu based on the symptoms.

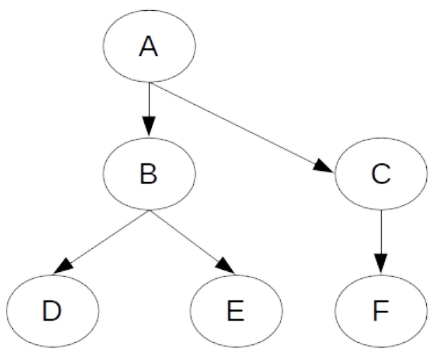
Sore Throat	Stomachache	Fever	Flu
No	No	No	No
No	No	Yes	Yes
No	Yes	No	No
Yes	No	No	No
Yes	No	Yes	Yes
Yes	Yes	No	Yes
Yes	Yes	Yes	No

a. Show the structure of the network and the conditional probability tables.

b. If a person has stomachache and fever, but no sore throat, what is the probability of him or her having the flu (according to your learned naive Bayes classifier)?

7. Consider a Naïve Bayes classifier with 100 feature dimensions. The label Y is binary with $P(Y=0) = P(Y=1) = 0.5$. All features are binary, and have the same conditional probabilities: $P(X_i=1|Y=0) = a$ and $P(X_i=1|Y=1) = b$ for $i=1, ..., 100$. Given an item X with alternating feature values ($X_1=1, X_2=0, X_3=1, ..., X_{100}=0$), compute $P(Y=1|X)$.

8. Consider the Bayesian network with the following structure and conditional probability tables (all variables are binary):



$P(A) = 0.8$

$P(B \mid A) = 0.5,\ P(B \mid \neg A) = 0.2$

$P(C \mid A) = 0.8,\ P(C \mid \neg A) = 0.6$

$P(D \mid B) = 0.5,\ P(D \mid \neg B) = 0.5$

$P(E \mid B) = 0.8,\ P(E \mid \neg B) = 0.8$

$P(F \mid C) = 0.2,\ P(F \mid \neg C) = 0.01$

a. Is this a polytree?

b. Are D and E independent? Are they conditionally independent given B?

c. How many numbers do you need to represent the full joint probability table?

d. If the variables were ternary instead of binary, how many values would you need to represent the full joint probability table and the conditional probability tables, respectively?

e. Write down the expression for the joint probability of all the variables in the network.

f. Find P(A = 0, B = 1, C = 1, D = 0).

g. Find P(B | A = 1, D = 0).

9. Two astronomers in different parts of the world make measurements M_1 and M_2 of the number of stars N in some small region of the sky, using their telescopes. Normally, there is a small probability e of error by up to one star in each direction. Each telescope can also (with a much smaller probability f) be badly out of focus (events F_1 and F_2), in which case the scientist will undercount by three or more stars (or if N is less than 3, fail to

detect any stars at all).

a. Draw a Bayesian network for this problem and show the conditional probability tables.

b. Write out the conditional distributions for $P(M_1 \mid N)$ for the case where $N \in \{1,2,3\}$ and $M_1 \in \{0,1,2,3,4\}$. Each entry in the conditional distribution table should be expressed as a function of the parameters *e* and/or *f*.

10. In micro-blackjack, you repeatedly draw a card (with replacement) that is equally likely to be a 2, 3, or 4. You can either *Draw* or *Stop* if the total score of the cards you have drawn is less than 6. Otherwise, you must *Stop*. When you *Stop*, your utility is equal to your total score (up to 5), or zero if you get a total of 6 or higher. When you *Draw*, you receive no utility. There is no discount ($\gamma = 1$).

    a. What are the states and the actions for this MDP?

    b. What is the transition function and the reward function for this MDP?

    c. Give the optimal policy for this MDP.

    d. What is the smallest number of rounds of value iteration for which this MDP will have its exact values (if value iteration will never converge exactly, state so).