

CS 8750

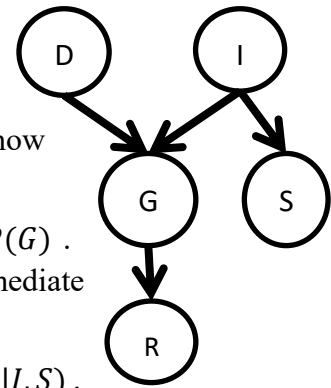
HW #3: Probabilistic inference and machine learning (10 points)

Spring 2017

(Due 2/28, Tuesday, midnight)

Part I (8 points)

1. (6 points) Given the following Bayesian network of Boolean random variables. Let $P(I) = 0.1$, $P(D) = 0.2$, $P(S|I) = 0.3$, $P(S|\neg I) = 0.4$, $P(G|I, D) = 0.1$, $P(G|I, \neg D) = 0.2$, $P(G|\neg I, D) = 0.3$, $P(G|\neg I, \neg D) = 0.4$, $P(R|G) = 0.1$, $P(R|\neg G) = 0.2$.



- 1) Use the Variable Elimination algorithm to compute $P(R)$. Eliminate variables in the order of D, S, I, and G. Show intermediate steps.
 - 2) Use the Variable Elimination algorithm to compute $P(G)$. Eliminate variables in the order of D, S, I, and R. Show intermediate steps.
 - 3) Use the Variable Elimination algorithm to compute $P(G|I, S)$. Eliminate variables in the order of D and R. Show intermediate steps.
 - 4) Use the Direct Sampling algorithm to generate one sample. During sampling, choose the more likely value for each variable.
 - 5) Given evidence I and S, use the Likelihood Weighting algorithm to generate one sample and its corresponding weight. During sampling, choose the more likely value for each variable.
 - 6) Apply Gibbs Sampling algorithm. Starting from initial $(\neg D, \neg I, \neg G, \neg S, \neg R)$, resample variables in the order of D, I, G, S, and R, to generate five samples. During sampling, choose the more likely value for each variable.
2. (2 points) Given the following training dataset for junk email detector.

SPAM:

CLICK HERE TO WIN MONEY

GAMBLING CLICK HERE

HAM:

GO TO LAS VAGAS CLICK HERE

LAS VAGAS IS GAMBLING CITY

GAMBLING COSTS MONEY

- a. Using the Bag of Words method, what is the size of vocabulary that contains all of the words in the messages?
- b. What is the ML (maximum likelihood) solution for learning $P(SPAM)$?
- c. Using Naïve Bayes model and ML learning, what is the ML solution for $P("CLICK"|SPAM)$?
- d. Using Naïve Bayes model and ML learning, for message $M=$ “GAMBLING IS HERE”, What is $P(SPAM|M)$?

Part II (2 points)

Find and read a paper related to learning probabilistic models, such as Bayesian networks, from real world data and write a 1-page review. Specifically, the requirements are as follows:

1. Find a technical paper published in the recent 5 years on your topic, related to learning probabilistic models, such as Bayesian networks, from real world data.
2. In your submission, explain briefly why you select this paper, and then write a 1-page review of it (should contain a paragraph to describe the real-world data used by the paper) by following the guideline in [How to read a paper \(slides\)](#).

Later, everyone will be given 2 minutes in class to present these results. You don't need to prepare slides, just to tell the class what you did and learned.