CPSC 540 Assignment 4 (due March 15 at midnight)

The assignment instructions are the same as for the previous assignment.

1. Name(s):

2. Student ID(s):

1 Discrete Markov Chains

1.1 Sampling, Inference, and Decoding

The function *example_markovChain.jl* loads the initial state probabilities and transition probabilities for a Markov chain model,

$$p(x_1, x_2, \ldots, x_d) = p(x_1) \prod_{j=2}^{d} p(x_j \mid x_{j-1}),$$

corresponding to the "grad student Markov chain" from class.

1. Write a function, *sampleAncestral*, that uses ancestral sampling to sample a sequence x from this Markov chain of length d. Hand in this code and report the univariate marginal probabilities for time 50 using a Monte Carlo estimate based on 10000 samples. (You can use *sampleDiscrete* in *misc.jl* to sample from a discrete probability mass function).

2. Write a function, *marginalCK*, that uses the CK equations to compute the exact univariate marginals up to a given time d. Hand in this code, report all exact univariate marginals, and report how this differs from the marginals in the previous question.

3. Write a function, *viterbiDecode*, that uses the Viterbi decoding algorithm for Markov chains to find the optimal decoding up to a time d. Hand in this code and report the optimal decoding of the Markov chain up to time 50 and up to 100, and say how this differs from the sequence defined by the most likely states at each time up to time 100 (the states maximizing the marginal probabilities).

4. Report all the univariate conditional probabilities at time 50 if the student starts in grad school, $p(x_{50} = c \mid x_1 = 3)$ for all c.

Hint: for parts 2-3, you can use a 7 by d matrix M to represent the dynamic programming table, and for part 3 you can use another matrix B containing the argmax values that lead to each entry in the table. For the conditional question, you can answer it by changing the input to the existing code.

1.2 Conditioning Queries Requiring Inference

Next consider the following cases (which require implementing an extra rejection step or backward phase):

1. Report all the univariate conditional probabilities $p(x_5 = c \mid x_{10} = 6)$ ("where are you likely to be 5 years after graduation if you are in academia 10 years") obtained using a Monte Carlo estimate based on 10000 samples and rejection sampling. Also report the number of samples accepted among the 10000 samples.

2. Write a function, *sampleBackwards* that uses backwards sampling to sample sequences of length d given a particular value of the last value x_d. Hand in this code and report all the univariate conditional probabilities $p(x_5 = c \mid x_{10} = 6)$ obtained using a Monte Carlo estimate based on 10000 samples.

3. Write a function, *forwardBackwards* that is able compute all exact univariate conditionals $p(x_j \mid x_d = c)$ in $O(dk^2)$ for a length-d sentence. Hand in the code and report all the exact univariate conditionals $p(x_5 = c \mid x_{10} = 6)$.

2 Directed Acyclic Graphical Models

2.1 D-Separation

Consider a directed acyclic graphical (DAG) model with the following graph structure:
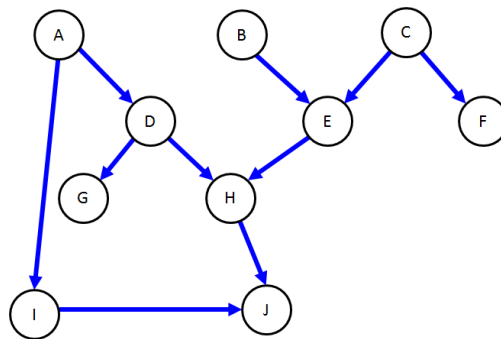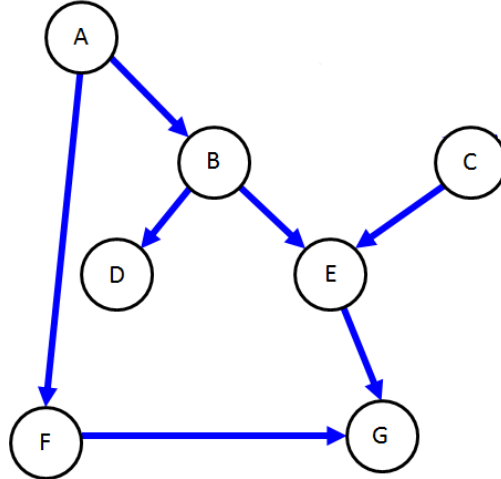

Assuming that the conditional independence properties are faithful to the graph, using d-separation briefly explain why the following are true or false:

1. $H \perp I$.

2. $H \perp I \mid A$.

3. $H \perp I \mid J$.

4. $H \perp I \mid A, J$.

5. $C \perp I$.

6. $C \perp I \mid J$.

7. $C \perp I \mid H$.

8. $C \perp I \mid A, H$.

9. $C \perp I \mid E, H, J$.

2.2 Exact Inference

Consider a directed acyclic graphical (DAG) model with the following graph structure:

Assume that all variables are binary and that we use the following parameterization of the network:

$$p(A = 1) = 0.7$$
$$p(B = 1 \mid A = 0) = 0.8$$
$$p(B = 1 \mid A = 1) = 1.0$$
$$p(C = 1) = 0.8$$
$$p(D = 1 \mid B = 0) = 0.8$$
$$p(D = 1 \mid B = 1) = 0.6$$
$$p(E = 1 \mid B = 0, C = 0) = 0.3$$
$$p(E = 1 \mid B = 0, C = 1) = 0.7$$
$$p(E = 1 \mid B = 1, C = 0) = 0.4$$
$$p(E = 1 \mid B = 1, C = 1) = 0.5$$
$$p(F = 1 \mid A = 0) = 0.5$$
$$p(F = 1 \mid A = 1) = 0.9$$
$$p(G = 1 \mid E = 0, F = 0) = 0.5$$
$$p(G = 1 \mid E = 0, F = 1) = 0$$
$$p(G = 1 \mid E = 1, F = 0) = 0.1$$
$$p(G = 1 \mid E = 1, F = 1) = 0.1$$

Compute the following quantities:

1. $p(A = 0)$.

2. $p(B = 1 \mid A = 0)$.

3. $p(B = 1)$.

4. $p(D = 1)$.

5. $p(B = 1 \mid D = 1)$.

6. $p(B = 1 \mid C = 1)$.

7. $p(B = 1 \mid A = 0, C = 1, F = 1)$.

Hints: some of the above quantities can be read from the table, some require using that probabilities sum to 1, some require the marginalization rule, some require Bayes rule, some require using [conditional] independence, and some will be simplified using calculations from previous sub-questions.

2.3 Inpainting

The function *example_fil.jl* loads a variant of the MNIST dataset. It contains all the training images but the test images are missing their bottom half. Running this function fits an independent Bernoulli model to the training set, and then shows the result of applying the density model to "fill in" four random test examples. It performs pretty badly because the independent model can't condition on the known top-half of the images.

1. Make a variant of the demo where you fit an inhomogeneous Markov chain to each image column. Hand in your code and an example of using this model to fill in 4 random test images.

2. Make a variant of the demo where you fit a directed acyclic graphical model to the data, using general discrete conditional probabilities and where the parents of pixel (i, j) are the other 8 pixels in the region $(i - 2 : i, j - 2 : j)$. Hand in your code and an example of using this model to fill in 4 random test images.

3. Make a variant of the demo where you fit a sigmoid belief network to the data, where the parents of pixel (i, j) are the other pixels in the region $(1 : i, 1 : j)$. Hand in your code and an example of using this model to fill in 4 random test images.

Hint: you may find it helpful to make an m by m array called *models* where element (i, j) contains the model for pixel (i, j). Included in *a4.zip* are *tabular.jl* and *logreg.jl* which implement solving a supervised learning problem using the tabular and sigmoid method (respectively). You may want to debug on a smaller version of the training set, since the runtime of fitting these models is non-trivial (solving 784 supervised learning problems takes time, although you could parallelize to make this go very quickly).

3 Very-Short Answer Questions

Give a short and concise 1-sentence answer to the below questions.

1. Describe a procedure to sample from the (continuous) exponential distribution.

2. Describe a procedure to sample from the (discrete) Poisson distribution.

3. What is an advantage of using Monte Carlo methods over the CK equations to estimate probabilities in Markov chains?

4. Why don't we normally consider the stationary distribution of inhomogeneous Markov chains?

5. Suppose we are using a hidden Markov model to track the location of a submarine using sonar measurements. What would x_j and z_j represent in this example?

6. Given a Markov model over x with discrete x_j, describe the parameterization of a hidden Markov models that defines the same distribution over x.

7. What is an advantage and a disavantage of using logistic regression to parameterize the CPDs in DAGs compared to a tabular representation?

8. When decoding a DAG, why does the order that we compute the messages matter?

9. What is the relationship between Ising models and UGMs?

10. What is the relevance of the Markov blanket in ICM?

11. Why do we have a "burn in" phase for Gibbs sampling?

4 Relevant Papers for Project

4.1 Finding Relevant Papers

To help you make progress on your project, for this part you should hand in a list of 10 academic papers related to your current project topic. Finding related work is often one of the first steps towards getting a new project started, and it gives you an idea of what has (and has not) been explored. Some strategies for finding related papers are:

1. Use Google: try the keywords you think are most relevant. Asking people in your lab (or related labs) for references is also often a good starting point.

2. Once you have found a few related papers, read their introduction section to find references that these papers think are worth mentioning.

3. Once you have found a few related papers, use Google Scholar to look through the list of references that are *citing* these papers (particularly for recent ones). You may have to do some sifting if there are a lot of citations. Reasonable criteria to sift through large reference lists include looking for the ones with the most citations or focusing on the more recent ones (then returning to Step 2 to find the more-relevant older references).

For this question you only need to provide a list, but in Assignment 5 you will have to do a survey of 10 papers. So it's worth trying to identify papers that are both relevant and important at this point. For some types of projects it will be easier to find papers than others. If you are having trouble, post on Piazza.

Although the papers do not need to all be machine learning papers, the course project does need to be related to machine learning in some way, so at least a subset of the papers should be machine learning papers. Here is a rough guide to some of the most reputable places to where you see machine learning works published:

- The International Conference on Machine Learning (ICML) and the conference on Advances in Neural Information Processing (NeurIPS) are the top places to publish machine learning work. The Journal of Machine Learning Research (JMLR) is the top journal, although in this field conference publications are usually viewed as more prestigious.

- Other good venues include AISTATS (emphasis on statistics), UAI (emphasis on graphical models), COLT (emphasis on theory), ICLR (emphasis on deep learning), ECML-PKDD (European version of ICML), CVPR and ICCV/ECCV (emphasis on computer vision), ACL and EMNLP (emphasis on language), KDD (emphasis on data mining), AAAI/IJCAI (emphasis on AI more broadly), JRSSB and Annals of Stats (emphasis on statistics more broadly), and Science and Nature (emphasis on science more broadly).

4.2 Paper Review

Among your list of 10 papers, choose one paper and write a review of this paper. It makes sense to choose a paper that is closely-related to your project or to choose one of the most important-looking papers. The review should have two parts:

1. A short summary of the contributions of the paper. Say what problem the paper is addressing, why this is an important problems, what is proposed, and how it is being evaluated.

2. A list of strengths and weaknesses of the paper, and whether the claims are appropriately evaluated. For ideas of what issues to discuss, see the JMLR guidelines for reviewers: `http://www.jmlr.org/reviewer-guide.html`

Note that you should include a non-empty list of strengths *and* weaknesses. Many students when doing their first reviews focus either purely on strengths or purely on weaknesses. It's important to recognize that all papers have weaknesses or limitations (even ones written by famous people or that are published in impressive places or that proved to be historically important) and all papers have strengths or at least a motivation (the authors must have thought it was worth writing for some reason).