Assignment 4: Graphical Models

Unsupervised Learning

Maneesh Sahani

Due: Thu Nov 30, 2006

1. Consider a multivariate Gaussian variable $\mathbf{x} = (x_1, \ldots, x_n)$ with given mean vector μ and covariance matrix Σ.

   2% Write out the probability density function for $\mathbf{x}$.

   13% Let $n = 4$, $\mu = (0, 1, 1, 0)$ and

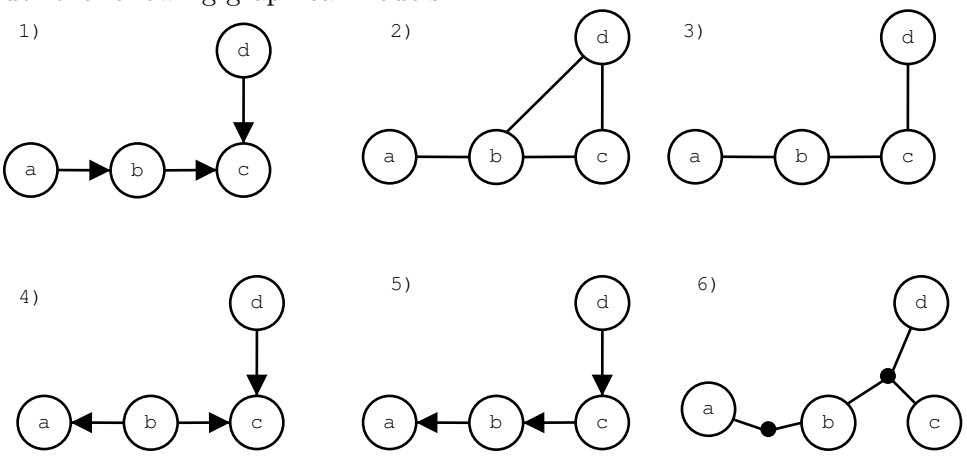
   $$\Sigma = \frac{1}{6} \begin{pmatrix} 7 & -2 & -2 & 1 \\ -2 & 7 & 1 & -2 \\ -2 & 1 & 7 & -2 \\ 1 & -2 & -2 & 7 \end{pmatrix},$$

   draw the corresponding undirected graph and define clique potentials consistent with the above Gaussian. [Hint: multiply out the terms that appear in the exponent.]

   2% Draw the corresponding factor graph.

   8% Describe and justify a *general* procedure for constructing the minimal factor graph for a multivariate Gaussian with known mean and covariance matrix.

2. Consider the following graphical models:



   15% For each graph, write down all the conditional independence relationships for variable c of the form $c \perp\!\!\!\perp x | y$, where x and y can be sets of other variables.

   10% Two graphs are equivalent if they have *all* the same marginal and conditional independence relationships between a, b, c, and d. Divide the above 6 graphs into the smallest number of non-overlapping sets of equivalent graphs.

3. You are the doctor on the Star Trek Enterprise and you are attempting to use Bayesian methods to help your diagnosis abilities. You would like to represent your knowledge about the following seven binary random variables describing the state of your patients on any given visit

```
M = has the disease microsoftus
L = has the disease linuxitis
A = has the disease applosis
V = is a vulcan (V=0 means ‘‘is a human’’)
H = has high temperature
P = likes pizza
B = has blue spots on face
```

You would like to build a directed graphical model which captures the following background knowledge:

```
Microsoftus is a rare disease.
Linuxitis and applosis are very rare diseases.
There are about four times as many humans as vulcans on the ship.
Vulcans have higher probability of getting microsoftus than humans.
Most vulcans like pizza, some humans like pizza.
Microsoftus usually causes high temperature and blue spots on the face.
Linuxitis always causes high temperature.
Applosis sometimes causes blue spots on the face.
```

5% Draw a directed graphical model representing the relationships between the above variables. If you need to make any additional assumptions to draw your graph, state clearly what they are.

10% For each variable in your graph, define a conditional probability table for that variable given the settings of its parents. Use the above background knowledge and convert those statements into probability tables which you think reasonably represent them. You will have to make up numbers for what terms like “rare”, “most”, and “usually” mean.

20% Using the graph that you’ve defined and the values of the conditional probability tables, compute the following probabilities. Show all your work:

```
P(patient likes pizza)
P(patient is a vulcan | patient has blue spots and high temperature)
P(patient has microsoftus | patient likes pizza and has blue spots)
```

Do these probabilities match your intuitions?

4. Following from the previous question... The Intergalactic Medical Journal publishes an article claiming that having the gene UNX triples the probabilities of having applosis and linuxitis. You are not sure you believe this article. You consult your medical records and you find that out of the 1000 passengers on the Enterprise, 70 were found to have only applosis, 40 were found to have only linuxitis, and 10 were found to have both applosis and linuxitis.

5% Ignoring this article and assuming applosis and linuxitis occur independently, how many patients would you expect on the Enterprise to have both applosis and linuxitis?

3% How would you augment the graph in the previous question to incorporate the UNX gene?

7% Are your medical records evidence in favour of or against the effect of the gene described in the journal article? Back up your claims, and state your assumptions. Is there any further information you would like to have? How would you reason about your degree of belief in the claims of this article given the data you have observed?

5. **(Bonus)** Consider the following two HMMs:

$$P_1(\mathbf{x}_{1:T}, \mathbf{s}_{1:T}) = P(\mathbf{s}_1)P(\mathbf{x}_1|\mathbf{s}_1)\prod_{t=2}^{T} P(\mathbf{x}_t|\mathbf{s}_t)P(\mathbf{s}_t|\mathbf{s}_{t-1})$$

and

$$P_2(\mathbf{x}_{1:T}, \mathbf{r}_{1:T}) = P(\mathbf{r}_1)P(\mathbf{x}_1|\mathbf{r}_1)\prod_{t=2}^{T} P(\mathbf{x}_t|\mathbf{r}_t)P(\mathbf{r}_t|\mathbf{r}_{t-1})$$

where $\mathbf{x}_t$ is the observation at time t, $\mathbf{s}_t$ and $\mathbf{r}_t$ are the hidden state variables for each HMM, respectively, and the notation $\mathbf{x}_{1:T}$, for example, denotes the sequence $\mathbf{x}_1 \ldots \mathbf{x}_T$, Now form a new model for the data by multiplying these two models and renormalizing:

$$P_3(\mathbf{x}_{1:T}, \mathbf{s}_{1:T}, \mathbf{r}_{1:T}) = \frac{1}{Z}P_1(\mathbf{x}_{1:T}, \mathbf{s}_{1:T})P_2(\mathbf{x}_{1:T}, \mathbf{r}_{1:T})$$

(a) **(Bonus +2%)** Draw a factor graph—with a node for each variable $\mathbf{x}_t$, $\mathbf{s}_t$, and $\mathbf{r}_t$—representing the conditional independence relationships in this new model, P_3.

(b) **(Bonus +8%)** Given a sequence $\mathbf{x}_{1:T}$, describe how you would compute $P(\mathbf{s}_t, \mathbf{r}_t|\mathbf{x}_{1:T})$. What is the time complexity of your algorithm?

(c) **(Bonus +5%)** If $\mathbf{s}_t$ and $\mathbf{r}_t$ are both discrete, taking on at most K states, is this model equivalent to an HMM with K^2 states? Why or why not?

(d) **(Bonus +5%)** Assume you want to learn the parameters of this model from data. Let's re-write the model more explicitly to make it clear:

$$P_3(\mathbf{x}_{1:T}, \mathbf{s}_{1:T}, \mathbf{r}_{1:T}|\theta, \phi) = \frac{1}{Z(\theta, \phi)}P_1(\mathbf{x}_{1:T}, \mathbf{s}_{1:T}|\theta)P_2(\mathbf{x}_{1:T}, \mathbf{r}_{1:T}|\phi)$$

where θ and ϕ are the usual transition, emission, and initial state HMM parameters for HMM 1 and 2, respectively, and Z is the normalization term, which depends on these parameters. What is the derivative of the log likelihood of P_3 with respect to the transition parameter, $\theta_{ij} \equiv P(\mathbf{s}_{t+1} = j|\mathbf{s}_t = i)$? Calculate the derivative of Z explicitly.

(e) **(Bonus +5%)** Assume there are L possible symbols: $\mathbf{x}_t \in \{1, \ldots, L\}$ and each HMM has K states. What is the maximum mutual information between $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ in this model, maximizing over all parameters?