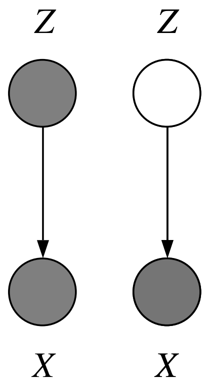
CSC412 – Midterm Test

Feb15, 2002
Time: 80 minutes
Total Points: 125
Worth: 25%

Name: Student Number:

1 Simple Directed Models (7 points)

Consider the two simplest possible directed graphical models, shown on the right.



In the table below, write the letter of the model that best fits the situation described by the row and column. Use each letter exactly once. Do not use the bottom left cell of the table marked by "XXX".

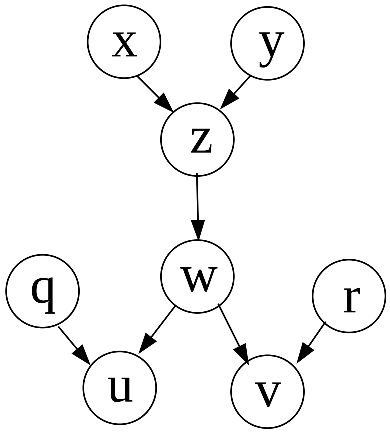
	x discrete observed	x continuous observed
z discrete, observed		
z discrete, unobserved		
z continuous, observed		
z continuous, unobserved	XXX	

A) Logistic/Softmax Regression
B) Naive Bayes Multinomial Classifier
C) Class-Conditional Gaussian Model
D) Mixtures of Gaussians Clustering
E) Linear Regression with Gaussian noise
F) Factor Analysis / Principal Component Analysis
G) Mixtures of Multinomials Clustering

2 Graphical Models (13 points)

Consider the two graphical models shown.

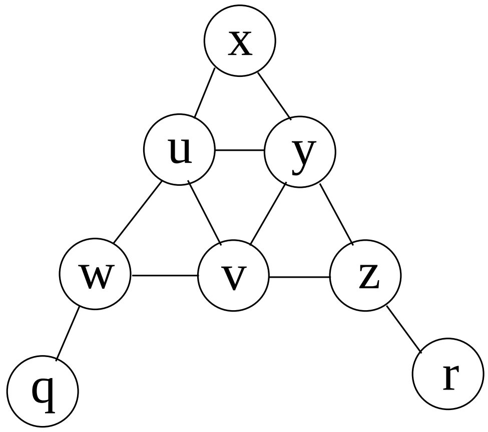
Model 1



- For each model, and each conditional independence statement below, write a "T" beside the statement if it is always true, or a "F" if the statement could be false.

Cond. Indep. Statement	Model 1 (T/F)	Model 2 (T/F)
$q \perp r$		
$x \perp y \mid u$		
$w \perp x, y \mid z$		
$w \perp x, y \mid u, v$		
$q \perp r \mid w$		
$q \perp r \mid u, v$		
$x \perp u \mid y$		
$x \perp u \mid y, v, w$		
$u \perp v \mid w$		

- Give the minimal Markov blanket of the node w in each model.



Model 2

- In Model2, list all the maximal cliques.

3 Short Answers (18 points)

Complete the statements in the space given.

- When speaking of "i.i.d. data", "i.i.d." stands for

  ____.

- In a directed tree, each node (except the root) has exactly ____________________________________.

- The key inequality used to lower bound the likelihood when deriving the EM algorithm is

  ________________________ inequality.

- Assume $\ell(\theta; X)$ is the likelihood function of a model for parameters θ and data X.

  $T(X)$ is a sufficient statistic for X if $\quad T(X^1) = T(X^2) \quad \Rightarrow$ ______________________________ $\forall \theta$.

- Random variables x and y are conditionally independent given z if $p(x, y|z) =$ __________________ $\forall y$.

  Random variables x and y are conditionally independent given z if $p(x|y, z) =$ __________________ $\forall y$.

- Consider a binary output y and some binary inputs $x_i, i = 1 \dots P$.
  A "Noisy-OR" model for y with failure probabilities α_i corresponding to each input x_i is:

  $p(y = 1|x_1 \dots x_P, \alpha_1 \dots \alpha_P) =$ ____________________________________

4 Maximum Likelihood: Poisson (12 points)

As a reminder, the *Poisson* distribution over a positive integer (count) random variable i is defined by the following mass function for a single positive parameter $\lambda > 0$.

$$p(i) = e^{-\lambda}\frac{\lambda^i}{i!} \quad i = 0, 1, \ldots, \infty$$

- Write the log likelihood of a dataset $i^1, \ldots, i^N$ in terms of λ.

- Find the maximum likelihood parameter λ in terms of $i^1, \ldots, i^N$.

5 Maximum Likelihood: Uniform (20 points)

The *uniform* distribution over a continuous scalar random variable x is defined by the following density, for two parameters a, b where $b > a$:

$$\begin{aligned} p(x) &= \frac{1}{b-a} && \text{if } a \le x \le b \\ p(x) &= 0 && \text{if } x < a \text{ or } x > b \end{aligned}$$

- Write the probability of a dataset $x^1, \ldots, x^N$ in terms of a and b.
  Be careful, and think about the various cases that are possible.

- Write down the maximum likelihood parameters a and b in terms of $x^1, \ldots, x^N$.

- What are the sufficient statistics of a dataset $x^1, \ldots, x^N$ for the uniform distribtion?

- Derive the mean and variance of x in terms of a and b.

6 Mixture of Gaussians Posteriors (25 points)

Consider a simple mixture of two one-dimensional Gaussians:

$$\begin{aligned} p(x) &= p(z=1)p(x|z=1) + p(z=2)p(x|z=2) \\ &= \alpha_1 \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + \alpha_2 \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} \end{aligned}$$

Where $\sigma_1, \sigma_2 > 0$; $0 < \alpha_1, \alpha_2 < 1$; $\alpha_1 + \alpha 2 = 1$.

- In general, there may be *more than one* value of x for which $p(z=1|x) = p(z=2|x)$ (i.e. the posteriors over the two components are equal).
  Write a quadratic formula $ax^2 + bx + c = 0$ that must be satisfied when $p(z=1|x) = p(z=2|x)$, expressing a, b, c in terms of $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \alpha_1, \alpha_2$.

- If $\sigma_1^2 = \sigma_2^2$, and $\mu_1 \neq \mu_2$, what is the single value of x such that $p(z = 1|x) = p(z = 2|x)$?

- Under what conditions are there infinitely many values of x such that $p(z = 1|x) = p(z = 2|x)$?

7 EM Algorithm for Unobserved Naive Bayes (30 points)

Consider the following "unobserved naive Bayes" model which has P *observed binary* variables $x_i \in \{0, 1\}$ ($i = 1 \ldots P$), and an *unobserved discrete* latent variable $z \in \{1, 2, \ldots, K\}$.

$$p(z = k) = a_k$$
$$p(x_i = 1|z = k) = b_{ik} \quad \forall i$$

- Calculate the posterior $p(z = k|x_1, x_2, \ldots, x_i, \ldots, x_P)$ of the latent variable given the binary observations.

- Calculate the marginal (incomplete) data log likelihood for some observed data x_i^n, $i = 1 \ldots P$, $n = 1 \ldots N$.

- Calculate the expected complete data log likelihood for the observed data x_i^n, $i = 1 \ldots P$, $n = 1 \ldots N$ under a distribution $p(z^n = k|x_1^n \ldots x_P^n) = q_k^n$.

- For a fixed q_k^n, and fixed observed data x_i^n, find the parameter settings a_k^* and b_{ik}^* which maximize the expected complete log likelihood.

- Assume we have some observed data x_i^n, $i = 1 \ldots P$, $n = 1 \ldots N$ and we want to fit this model using the EM algorithm. Using the results of the previous subquestions, write down the $E - step$ update for q_k^n and the $M - step$ updates for a_k and b_{ik}.
  Make sure the updates you write don't contain unspecified quantities.
  You should be able to turn your updates into code without any further derivations.