CSC412/2506 – Example Midterm Test

2006
Time: 60 minutes
Aid sheet and calculators permitted.

Name: Student Number:

1 Directed Models

- What does the statement "x is conditionally independent of y given z"
  imply about the joint distribution $p(x, y, z)$?

- Consider a distribution on three binary variables a, b, c which factors according to $p(a, b, c) = p(a)p(b|a)p(c|a)$.
  Draw the corresponding graphical model.
  If $p(a = 1) = 1/8, p(b = 1|a = 1) = 1/50, p(b = 1|a = 0) = 1/70, p(c = 1|a = 1) = 1/3, p(c = 1|a = 0) = 1/3$,
  What is $p(b = 1)$?

  Is b independent of c (for the specific parameters above)? Why or why not?

2 Maximum Likelihood: Uniform

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

$$p(x) = \frac{1}{b-a} \quad \text{if } a \leq x \leq b$$
$$p(x) = 0 \quad \text{if } x < a \text{ or } x > b$$

- 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.

- Derive the maximum likelihood parameters a and b in terms of $x^1, \ldots, x^N$. Justify your answer.

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

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

3 Maximum Likelihood: Multivariate Gaussian

As you know, the *Gaussian* distribution over a vector random variable $\mathbf{x}$ is defined by the following density function for a mean vector $\mathbf{m}$ and a symmetric positive definite matrix $\mathbf{V}$. (Read $|\mathbf{M}|$ as the determinant of the matrix $\mathbf{M}$.)

$$p(\mathbf{x}) = (2\pi|\mathbf{V}|)^{-d/2} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^\top \mathbf{V}^{-1}(\mathbf{x}-\mathbf{m})\right]$$

- Under this distribution, what is the expected value of $x_i x_j$?

- Write the log likelihood of a dataset $\mathbf{x}^1, \ldots, \mathbf{x}^N$ in terms of $\mathbf{m}$, $\log|\mathbf{V}^{-1}|$ and $\mathbf{V}^{-1}$.

- What are the sufficient statistics of a dataset $\mathbf{x}^1, \ldots, \mathbf{x}^N$ for the Gaussian distribtion? Justify your answer by indicating how the likelihood depends only on these sufficient statistics.
  Hint: Remember that $\mathbf{x}^\top \mathbf{M}\mathbf{x} = \text{Trace}[\mathbf{M}\mathbf{x}\mathbf{x}^\top]$ and that Trace is a linear operation.

- Find the maximum likelihood mean parameter $\mathbf{m}^*$ in terms of $\mathbf{x}^1, \ldots, \mathbf{x}^N$. Show your work.

- [Harder] Find the maximum likelihood parameter $\mathbf{V}^*$ in terms of $\mathbf{x}^1, \ldots, \mathbf{x}^N$ and $\mathbf{m}^*$. Show your work.
  Hint: take the derivative of the log likelihood with respect to $\mathbf{V}^{-1}$, leaving an expression in $\mathbf{V}$.

4 Numeric Distributions

This question tests your thinking about distributions as multidimensional tables, a concept which is crucial when you actually implement code for probabilistic models.

- Consider three multidimensional arrays (tables) `pA,pB,pC`, each representing a joint probability distribution over discrete random variables. Each dimension of the table represents one discrete random variable, and the size of the dimension is equal to the number of possible values the variable can take on. For example `pA(3,4,5,1)` is the probability under distribution A that $\mathbf{x}_1$ takes on its third value, $\mathbf{x}_2$ takes on its fourth value, $\mathbf{x}_3$ takes on its fifth value and $\mathbf{x}_4$ takes on its first value.
- Using the operations "select slice i along dimension j", "sum along the kth dimension" and "divide each element in the table by a scalar", describe, in order, the numerical operations that would be necessary to calculate the following marginal and conditional distributions and say what the size of the output array (table) would be if the original tables have size 5x5x5x5.
  1. $pA(\mathbf{x}_3, \mathbf{x}_4 | \mathbf{x}_1$ takes its third value)
  2. $pB(\mathbf{x}_4 | \mathbf{x}_2$ takes its first value)
  3. $pC(\mathbf{x}_1)$
- How would you decide if each of the following statements was true, given a multidimensional array representing a joint distribution?
  1. $\mathbf{x}_1$ is conditionally independent of $\mathbf{x}_2$ given $\mathbf{x}_3$
  2. $\mathbf{x}_1$ is conditionally independent of $\mathbf{x}_2$ given $\mathbf{x}_3$ and $\mathbf{x}_4$
  3. $\mathbf{x}_1$ is marginally independent of $\mathbf{x}_2$

5 Continuous Factors for Discrete Observations

Consider the two-node directed graphical model, shown on the right.
The node z represents a continuous (scalar or vector) but unobseved factor, and the node x represents a discrete (categorical) observed variable (or a vector of discrete observed variables).

- Write the factorization of the joint distribution $p(x, z)$ is implied by the model?.

- Suggest specific forms for each of the terms in the factorization you gave above.

- What is the expression for the marginal density over the observed variable(s) x ?

- If you had to chose between calling this model a "classification" model or a "clustering" model, which would be more appropriate?

6 Training Logistic Regression

- Consider a logistic regression model with binary input features $\mathbf{f}_i$ and a class label c, trained to maximize average conditional likelihood $p(c|\mathbf{f})$ with no regularization. As a reminder, the model is given below:

$$p(c|\mathbf{f}) = \frac{\exp\{\mathbf{w}_c^\top \mathbf{f}\}}{\sum_k \exp\{\mathbf{w}_k^\top \mathbf{f}\}}$$

- Derive the gradient $\frac{\partial \ell}{\partial \mathbf{w}_c}$

- Describe what condition is satisfied when this gradient is zero.

- What happens to the weight w_{ci} for feature i and class c if, using binary features, on the finite sample available:

  A certain component $\mathbf{f}_i$ is always 1 for data in class c?

  A certain component $\mathbf{f}_i$ is always 0 for data in class c?

- What can happen if, for all training examples n in class c, $\mathbf{f}_i^n = \mathbf{f}_j^n$, that is features i and j always agree with each other (when the class label is c)?

7 Short Answers

Complete the statements in the space given.

- 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 $p(x|y, z) =$ _____ $\forall y$.

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

- Maximum likelihood structure learning in fully observed tree models involves solving a _____

  _____ problem, for example using _____

  algorithm.

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

  $p(y = 1|x_1 \ldots x_P, \alpha_1 \ldots \alpha_P) =$ _____

- In factor analysis the covariance of the posterior distribution over the latent variable given an observation is

  _____ of the observation.