EE 641 Midterm Exam

Fall 2004

Key

Name: ______________________ Starting time: ______________________

Ending time: ______________________

Instructions

The following is a take home exam.

- This exam contains 4 problems and should be completed in 75 minutes.
- Answer questions precisely and completely. Credit will be subtracted off for vague answers.
- You may not use any notes, textbooks, or calculators; but you may use a single sided single page of hand copied notes.

Good luck.

Problem 1. (20pt)
Let $\{X_n\}_{n=0}^N$ be a Markov Chain with

$$P\{X_0 = k\} = \pi_k$$

for $1 \le k \le M$, and

$$P\{X_n = m | X_{n-1} = l\} = P_{l,m}$$

for $1 \le l, m \le M$. Let π denote the $1 \times M$ vector of initial probabilities, and let P denote the $M \times M$ transition matrix.

10 a) Derive an expression, $p(x|\theta)$, the likelihood for the entire sequence.

10 b) Derive the ML estimate of $\theta = (\pi, P_{l,m})$ given $\{X_n\}_{n=0}^N$.

a) $$p(x|\theta) = \pi_{x_0} \prod_{i=1}^{N} P_{x_{i-1} x_i}$$

b) $$\log p(x|\theta) = \sum_{i=1}^{N} \log P_{x_{i-1} x_i} + \log \pi_{x_0}$$

$$= \left[\sum_{k,m=1}^{M} h_{k,m} \log P_{km} \right] + \log \pi_{x_0}$$

where $h_{k,m} = \sum_{i=1}^{N} \delta(x_i - m)\,\delta(x_{i-1} - k)$

Notice that $\sum_{m=1}^{M} P_{km} - 1 = 0$

Constrained optimization for each k

$$\frac{h_{k,m}}{\hat{P}_{k,m}} - \lambda_k = 0$$

$$h_{k,m} = \lambda_k \hat{P}_{km}$$

$$\boxed{\hat{P}_{k,m} = \frac{h_{km}}{\sum_m h_{k,m}} \qquad \pi_k = \delta(k - x_0)}$$

Problem 2. (25pt)

Let $\{X_n\}_{n=0}^{N-1}$ be i.i.d. random variables with

$$P\{X_n = k\} = \pi_k$$

for $0 \le k \le M-1$.

Let $\{Y_n\}_{n=0}^{N-1}$ be random variables with conditional distribution

$$p(y|x,\theta) = \prod_{n=0}^{N-1} \frac{1}{\mu_{x_n}} \exp\{-y_n/\mu_{x_n}\} \mathbf{1}(y_n \ge 0)$$

where $\mathbf{1}(y_n \ge 0)$ is function which is 1 when $y_n \ge 0$ and zero otherwise, and $\theta = (\pi, \mu)$

6 a) Derive $p(y, x|\theta)$.

6 b) Derive $f(x_n|y_n, \theta) = p(x_n|y_n, \theta)$.

6 c) Derive the ML estimate of θ given both Y and X.

7 d) Derive the EM update equations for estimation of θ from Y.

a) $$p(Y, X|\theta) = \left[\prod_{n=0}^{N-1} \frac{1}{\mu_{x_n}} \exp\left\{-\frac{Y_n}{\mu_{x_n}}\right\} 1(Y_n \ge 0)\right] \prod_{n=0}^{N-1} \pi_{x_n}$$

b) $$f(x_n|Y_n, \theta) = \frac{\pi_{x_n} \frac{1}{\mu_{x_n}} \exp\left\{-\frac{Y_n}{\mu_{x_n}}\right\}}{\sum_k \pi_k \frac{1}{\mu_k} \exp\left\{-\frac{Y_n}{\mu_k}\right\}}$$

c) $$\hat{\theta} = \arg\max_\theta p(Y, X|\theta)$$

$$\hat{\mu}_k = \arg\max_{\mu_k} \sum_{n=0}^{N-1} \left(-\frac{Y_n}{\mu_k} - \log(\mu_k)\right) \delta(X_n - k)$$

$$= \arg\max_{\mu_k} \left\{-\frac{N_k S_k}{\mu_k} - N_k \log(\mu_k)\right\}$$

where $$S_k \triangleq \frac{1}{N_k} \sum_{n=0}^{N-1} Y_n \delta(X_n - k)$$

$$N_k \triangleq \sum_{n=0}^{N-1} \delta(X_n - k)$$

Name: ______________________

$$N_K \frac{S_K}{\hat{\mu}_K^2} - \frac{N_K}{\hat{\mu}_K} = 0$$

$$\hat{\mu}_K = S_K$$

$$\boxed{\hat{\mu}_K = \frac{1}{N_K} \sum_{n=0}^{N-1} Y_n \, \delta(x_n - K)}$$

$$\hat{\pi} = \arg\max_{\pi} \sum_{n=0}^{N-1} \log \pi_{x_n}$$

$$= \arg\max_{\pi} \sum_{K=1}^{M} N_K \log \pi_K \qquad \sum_{K=1}^{M} \pi_K = 1$$

$$\frac{N_K}{\hat{\pi}_K} - \lambda_K = 0$$

$$\boxed{\hat{\pi}_K = \frac{N_K}{N}}$$

d) $$\bar{N}_K' = \sum_{n=0}^{N-1} f(x_n \mid Y_n, \theta)$$

$$\hat{\pi}_K' = \frac{\bar{N}_K'}{N}$$

$$\hat{\mu}_K' = \frac{1}{\bar{N}_K'} \sum_{n=0}^{N-1} Y_n \, f(x_n \mid Y_n, \theta)$$

$$\theta' = (\hat{\pi}', \hat{\mu}')$$

Problem 3. (30pt)
Let X be an N pixel Gaussian MRF with density

$$p(x) = \frac{1}{(2\pi\sigma^2)^{1/2}}|B|^{N/2} \exp\left\{-\frac{1}{2\sigma^2}x^t Bx\right\}$$

Furthermore, let Y be an N pixel random field with pixels that are conditionally i.i.d. and Poisson given X, and

$$E[Y|X] = X$$

6 a) Calculate the maximum likelihood estimate of σ^2 given X and Y.
6 b) Calculate $p(y, x|\sigma)$, the joint probability density of Y and X given σ.
6 c) Calculate an expression for the MAP estimate of X given Y.
6 d) Derive an expression for the gradient descent update of the MAP cost functional.
6 e) Derive an expression for the ICD update of a signal pixel s.

a) $\hat{\sigma}^2 = \frac{1}{N} x^t B x$

b) $$p(y|x) = \left[\prod_{i=1}^{N} \frac{x_i^{y_i} e^{-x_i}}{y_i!}\right] p(x)$$

c) $\log p(y, x|\sigma) =$

$$= \sum_{i=1}^{N}\left(y_i \log x_i - x_i - \log y_i!\right)$$

$$- \frac{1}{2\sigma^2} x^t B x - \frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2)$$

$$+ \frac{N}{2}\log|B|$$

$$\hat{x} = \underset{x}{\arg\max} \underbrace{\left\{ \sum_{i=1}^{N}\left(y_i \log x_i - x_i - \log y_i!\right) - \frac{1}{2\sigma^2} x^t B x \right\}}_{c(x)}$$

Name: ______________________

d) $\nabla C(x) = \underbrace{\frac{y}{x}}_{} - 1 - \frac{1}{\sigma^2} x^* B$

row vector with elements

$\left[\frac{y_1}{x_1}, \ldots, \frac{y_N}{x_N}\right]$

$$x^{(k+1)} = x^{(k)} + \omega \left[\frac{y}{x} - 1 - \frac{1}{\sigma^2} x^* B \right]$$

e) $$x_i \leftarrow \arg\max_{x_i} \left\{ \sum_{i=1}^{N} (y_i \log x_i - x_i) - \frac{1}{2\sigma^2} x^* B x \right\}$$

$$= \arg\max_{x_i} \left\{ y_i \log x_i - x_i - \frac{1}{2\sigma^2} \sum_{\{i,j\} \in C} b_{ij'} (x_i - x_{j'})^2 - \frac{1}{2\sigma^2} \sum_i a_i x_i^2 \right\}$$

$$b_{ij'} = -\frac{1}{2} B_{ij'}$$

$$a_i = \sum_{j'} B_{ij'}$$

$$= \arg\max_{x_i} \left\{ y_i \log x_i - x_i - \frac{1}{2\sigma^2} \sum_{j \in \partial i} b_{ij'} (x_i - x_j)^2 - \frac{1}{2\sigma^2} a_i x_i^2 \right\}$$

Problem 4. (25pt)

Consider the cost functional

$$||y - Hx||^2 + x^t Bx$$

which takes on its unique minimum value at

$$\hat{x} = \arg \min_{x \in \mathbb{R}^N} \left\{ ||y - Hx||^2 + x^t Bx \right\}$$

6 a) Derive the gradient descent algorithm for this minimization problem, and show it has the form of the difference equation

$$x^{(k+1)} = (I - \omega P)\, x^{(k)} + \omega H^t y$$

where $P = H^t H + B$.

Further assume that $P = T^t \Lambda T$ were T is an orthonormal matrix, and Λ is a diagonal matrix with positive entrees $\lambda_1, \cdots, \lambda_N$.

6 b) Let $e^{(k)} = x^{(k)} - \hat{x}$, and derive a difference equation for $e^{(k)}$.

6 c) Determine the range of values of ω that will allow stable convergence of the gradient descent algorithm.

7 d) Explain what the trade-offs are in the selection of ω.

a) $$\nabla C(x) = -2(y - Hx)^t H + 2x^t B$$

$$x^{(k+1)} = x^{(k)} - \frac{\omega}{2} \nabla C(x)$$

$$= x^{(k)} + \omega\left[H^t(y - Hx) - Bx\right]$$

$$= x^{(k)} - \omega(H^t H x - Bx) + \omega H^t y$$

$$= (I - \omega\underbrace{(H^t H + B)}_{P})x + \omega H^t y$$

b) $$e^{(k+1)} = (I - \omega P)e^{(k)}$$

c) $$e^{(k+1)} = (I - \omega T^t \Lambda T)e^{(k)}$$

Name: ______________________

$$e^{(k+1)} = (T^*T - \omega T^*\Lambda T)e^{(k)}$$

$$= T^*(I - \omega\Lambda)Te^{(k)}$$

$$\underbrace{Te^{(k+1)}}_{\tilde{e}^{(k)}} = (I - \omega\Lambda)\underbrace{Te^{(k)}}_{\tilde{e}^{(k)}}$$

$$\tilde{e}^{(k)} = (I - \omega\Lambda)^k \tilde{e}^{(0)}$$

$$0 < \omega < \frac{2}{\max(\lambda_i)}$$

d) If ω is too large, then large eigenvalues are not convergent.

If ω is too small, then small eigenvalues converge slowly.