EE 641 Final Exam

Fall 2013

Name: **Key** ______________________________

Instructions

- This exam contains 4 problems worth a total of 100 points.
- You may not use any notes, textbooks, or calculators.
- Answer questions precisely and completely. Credit will be subtracted for vague answers.

Good luck.

Problem 1. (25pt)
Show that if X is distributed as

$$p(x) = \frac{1}{z} \exp \left\{ - \sum_{\{s,r\} \in \mathcal{P}} b_{s,r} \rho(x_s - x_r) \right\} ,$$

where $\rho(\cdot)$ is a potential function and $\mathcal{P}$ is a set of pairwise cliques, then the conditional distribution of X_s given X_r for $r \neq s$ is given by

$$p(x_s | x_r \ r \neq s) = \frac{1}{z} \exp \left\{ - \sum_{r \in \partial s} b_{s,r} \rho(x_s - x_r) \right\} ,$$

and give an expression for z.

Solution:

First, we derive the form of the conditional distribution. We can factor the pair-wise Gibbs distribution into two portions: the portion that has a dependency on x_s, and the portion that does not. This results in the relation,

$$\begin{aligned} p(x) &= \frac{1}{z} \exp \left\{ - \sum_{\{s,r\} \in \mathcal{P}} b_{s,r} \rho(x_s - x_r) \right\} \\ &= \frac{1}{z} \exp \left\{ - \sum_{r \in \partial s} b_{s,r} \rho(x_s - x_r) \right\} f(x_{r \neq s}) , \end{aligned}$$

where $f(x_{r \neq s})$ is a function of the pixels x_r for $r \neq s$. Using the factored form, we may calculate the conditional distribution of X_s given the remaining pixels X_r for $r \neq s$ as

$$\begin{aligned} p_{x_s|x_{r \neq s}}(x_s | x_{r \neq s}) &= \frac{p(x)}{\int_{\Re} p(x_s, x_{r \neq s}) dx_s} = \frac{p(x_s, x_{r \neq s})}{\int_{\Re} p(x_s, x_{r \neq s}) dx_s} \\ &= \frac{\frac{1}{z} \exp \left\{ - \sum_{r \in \partial s} b_{s,r} \rho(x_s - x_r) \right\} f(x_{r \neq s})}{\int_{\Re} \frac{1}{z} \exp \left\{ - \sum_{r \in \partial s} b_{s,r} \rho(x_s - x_r) \right\} f(x_{r \neq s}) dx_s} \\ &= \frac{\exp \left\{ - \sum_{r \in \partial s} b_{s,r} \rho(x_s - x_r) \right\}}{\int_{\Re} \exp \left\{ - \sum_{r \in \partial s} b_{s,r} \rho(x_s - x_r) \right\} dx_s} . \end{aligned}$$

However, since this result is only a function of x_s and its neighbors, we have that

$$p_{x_s|x_{r \neq s}}(x_s | x_{r \neq s}) = p_{x_s|x_{\partial s}}(x_s | x_{\partial s}) ,$$

and the normalizing factor is given by

$$z = \int_{\Re} \exp \left\{ - \sum_{r \in \partial s} b_{s,r} \rho(x_s - x_r) \right\} dx_s .$$

Problem 2.(25pt)

Let $x \in \Re^N$ and $A \in \Re^{N \times N}$ be a full rank matrix.

a) Calculate a closed form expression for the ICD update when the MAP cost function has the form.

$$f(x;y) = ||y - Ax||^2 + ||x||_1$$

where $||x||_1$ is the L_1 norm of x, and A has rank N.

b) Prove that $f(x;y)$ is a strictly convex function of x.

c) Prove that this function has a unique global minimum which is also a local minimum.

Solution:

Part a)

Define $e = y - Ax$ to be the error vector, and $A_{*,i}$ to be i^{th} row of A. Also, let the updated value for the i^{th} pixel be $x_i + \alpha$.

Then the cost as a function of α is given by the following within a constant,

$$f(\alpha) = \theta_1 \alpha + \frac{1}{2}\theta_2 \alpha^2 + |\alpha + x_i| \ ,$$

where

$$\begin{aligned} \theta_1 &= -2e^t A_{*,i} \\ \theta_2 &= 2||A_{*,i}||^2 \ . \end{aligned}$$

So the ICD update is given by the solution to

$$x_i \leftarrow \arg\min_{\alpha} f(\alpha) \ .$$

This problem can be solved in two cases. The first case is when the absolute value term is positive, and the second is when it is negative.

First define,

$$z = \frac{-\theta_1}{\theta_2}$$

to be the solution ignoring the absolute value term.

case 1: If $\alpha + x_i > 0$, then

$$\alpha = z - \frac{1}{\theta_2}$$

case 2: If $\alpha + x_i < 0$, then

$$\alpha = z + \frac{1}{\theta_2}$$

case 3: If $\alpha + x_i = 0$, then $\alpha = -x_i$.

This solution can be expressed more compactly using soft thresholding in the form

$$x_i \leftarrow \left(|z + x_i| - \frac{1}{\theta_2} \right)_+ \text{sign}\,(z + x_i) \ \ ,$$

where the operation $(z)_+ = \max\{z, 0\}$ returns the maximum of the argument or zero. Expanding out the expression in terms of θ_1 and θ_2 yields

$$x_i \leftarrow \left(\left| \frac{\theta_1}{\theta_2} + x_i \right| - \frac{1}{\theta_2} \right)_+ \text{sign} \left(\frac{\theta_1}{\theta_2} + x_i \right) \ \ ,$$

and substituting in the expressions for θ_1 and θ_2 yields

$$x_i \leftarrow \left(\left| x_i - \frac{e^t A_{*,i}}{||A_{*,i}||^2} \right| - \frac{1}{2||A_{*,i}||^2} \right)_+ \text{sign} \left(x_i - \frac{e^t A_{*,i}}{||A_{*,i}||^2} \right) \ \ .$$

Part b) Since A is full rank, then $||y - Ax||^2$ must be strictly convex. Since $||x||_1$ is convex, the sum of the two functions must be strictly convex.

Part c)

In order to prove that the function has a unique global minimum, we need only show that the minimum exists within some compact set $||x|| \le d$ for some sufficiently large diameter d. So to do this, it is sufficient to show that for all $||x|| \ge d$, $f(x; y) > f(0; y)$.

In order to show this, we choose

$$d = 3\frac{||A^t y||}{\Sigma^2_{min}} \ \ ,$$

where Σ_{min} is the minimum singular value of A.

Then we have that for all $||x|| \ge d$

$$\begin{aligned}
f(x;y) &= ||y||^2 - 2x^t A^t y + x^t A^t A x \\
&\ge ||y||^2 - 2||x||\,||A^t y|| + \Sigma^2_{min}||x||^2 \\
&= \Sigma^2_{min}\left[\left(||x|| - \frac{||A^t y||}{\Sigma^2_{min}} \right)^2 - \left(\frac{||A^t y||}{\Sigma^2_{min}} \right)^2 \right] + ||y||^2 \\
&\ge \Sigma^2_{min}\left[3 \left(\frac{||A^t y||}{\Sigma^2_{min}} \right)^2 \right] + ||y||^2 \\
&\ge ||y||^2 \\
&= f(0;y)
\end{aligned}$$

Problem 3.(25pt)

Let X_n be N i.i.d. random variables with $P\{X_n = i\} = \pi_i$ for $i = 0, \cdots, M-1$. Also, assume that Y_n are conditionally independent given X_n and that the conditional distribution of Y_n given X_n is distributed as $N(\mu_{x_n}, \gamma_{x_n})$. Derive an EM algorithm for estimating the parameters $\{\pi_i, \mu_i, \gamma_i\}_{i=0}^{M-1}$.

Solution:

Given the assumptions of the problem, we know that $p(y, x|\theta)$ is an exponential distribution with natural sufficient statistics given by

$$\begin{aligned} N_m &= \sum_{n=1}^{N} \delta(x_n - m) && (1) \\ b_m &= \sum_{n=1}^{N} y_n \delta(x_n - m) && (2) \\ S_m &= \sum_{n=1}^{N} y_n y_n^t \delta(x_n - m) \ . && (3) \end{aligned}$$

The ML parameter estimate, $\hat{\theta}$, can then be computed from these sufficient statistics as

$$\begin{aligned} \hat{\pi}_m &= \frac{N_m}{N} && (4) \\ \hat{\mu}_m &= \frac{b_m}{N_m} && (5) \\ \hat{R}_m &= \frac{S_m}{N_m} - \frac{b_m b_m^t}{N_m^2} \ . && (6) \end{aligned}$$

In order to derive the EM update, we only need to replace the sufficient statistics in the ML estimate by their expected values. So to compute the EM update of the parameter θ, we first must compute the conditional expectation of the sufficient statistics. We can do this for the statistic N_m as follows.

$$\begin{aligned} \bar{N}_m &= \mathbb{E}\left[N_m | Y = y, \theta^{(k)}\right] \\ &= \mathbb{E}\left[\sum_{n=1}^{N} \delta(x_n - m) \middle| Y = y, \theta^{(k)}\right] \\ &= \sum_{n=1}^{N} \mathbb{E}\left[\delta(x_n - m) | Y = y, \theta^{(k)}\right] \\ &= \sum_{n=1}^{N} P\{X_n = m | Y = y, \theta^{(k)}\} \end{aligned}$$

Using a similar approach for all three sufficient statistics yields the E-step of

$$\begin{aligned}
\bar{N}_m &= \sum_{n=1}^{N} P\{X_n = m | Y = y, \theta^{(k)}\} \\
\bar{b}_m &= \sum_{n=1}^{N} y_n P\{X_n = m | Y = y, \theta^{(k)}\} \\
\bar{S}_m &= \sum_{n=1}^{N} y_n y_n^t P\{X_n = m | Y = y, \theta^{(k)}\} \ .
\end{aligned}$$

Then in order to calculate the M-step, we simply use these expected statistics in place of the conventional statistics in the equations for the ML estimator.

$$\begin{aligned}
\pi_m^{(k+1)} &= \frac{\bar{N}_m}{N} \\
\mu_m^{(k+1)} &= \frac{\bar{b}_m}{\bar{N}_m} \\
R_m^{(k+1)} &= \frac{\bar{S}_m}{\bar{N}_m} - \frac{\bar{b}_m \bar{b}_m^t}{\bar{N}_m^2} \ .
\end{aligned}$$

With each repetition of this process, we increase the likelihood of the observations.

Problem 4.(25pt)

Consider a birth-death process for which the state of the Markov chain, X_n, takes on values in the set $\{0, \cdots, M-1\}$, so that $P_{i,i+1} = \lambda$ is the birth rate for $i = 0, \cdots, M-2$; and $P_{i,i-1} = \mu$ is the death rate for $i = 1, \cdots, M-1$.

a) Calculate, $P_{i,j}$, the transition probabilities of the homogeneous Markov chain.

b) Calculate, π_i, the stationary distribution of the Markov chain.

Solution:

In a finite-state birth-death process the state of the Markov chain, X_n, takes on values in the set $\{0, 1, \cdots, M-1\}$. With each new time increment, the value of X_n is either incremented (i.e., birth occurs), decremented (i.e., death occurs), or the state remains unchanged. Mathematically, this can be expressed with the following transition probabilities.

$$P_{i,j} = \begin{cases} \lambda & \text{if } j = i+1 \text{ and } j < M-1 \\ \mu & \text{if } j = i-1 \text{ and } i > 0 \\ 1-\lambda-\mu & \text{if } j = i \text{ and } 0 < i < M-1 \\ 1-\lambda & \text{if } j = i \text{ and } i = 0 \\ 1-\mu & \text{if } j = i \text{ and } i = M-1 \\ 0 & \text{otherwise} \end{cases}$$

Here $0 < \lambda < 1$ is the probability of a birth and $0 < \mu < 1$ is the probability of a death, where we also assume that $\lambda + \mu < 1$.

In this case, the detailed balance equations have a simple solution. Let π_i be the steady-state probability that the Markov chain is in state i. Then the detailed balance equations require that

$$\pi_i \lambda = \pi_{i+1} \mu \ ,$$

for all $i \in \{0, \cdots, M-1\}$. This relations implies the recursion that $\pi_{i+1} = \frac{\lambda}{\mu} \pi_i$ must hold; so we know that the solution must have the form

$$\pi_i = \frac{1}{z} \left(\frac{\lambda}{\mu} \right)^i \ ,$$

where z is a normalizing constant. The value of z may then be calculated as

$$z = \sum_{i=0}^{M-1} \left(\frac{\lambda}{\mu} \right)^i = \frac{1 - \left(\frac{\lambda}{\mu}\right)^M}{1 - \frac{\lambda}{\mu}} \ .$$

where $z = M$ when $\frac{\lambda}{\mu} = 1$.

So the stationary distribution is given by

$$\pi_i = \left(\frac{\lambda}{\mu} \right)^i \frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^M} \ , \qquad (7)$$

when $\frac{\lambda}{\mu} \neq 1$, and it is given by $\pi_i = 1/M$ when $\frac{\lambda}{\mu} = 1$.

Since the distribution of equation (7) solves the detailed balance equations, it must also solve the full balance equations. In addition, we know three additional things. First, the number of states is finite. Second, all states communicate because $\left[P^M\right]_{i,j} > 0$, so the Markov chain is irreducible. Third, the Markov chain is aperiodic because $\left[P^k\right]_{i,j} > 0$ for $k = M, \cdots, 2M-1$. This implies that the Markov chain has a stationary distribution, and its stationary distribution is the unique solution to the full balance equations given by equation (7).

Moreover, since the distribution is also the solution to the detailed balance equations, the Markov chain must also be reversible.