UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering

ECE 544NA Pattern Recognition
Fall 2016

Homework in lieu of EXAM 2

Due Thursday, November 10, 2016 by 7:00pm

Feel free to work with other students, and/or to seek help from Raymond and Prof. Mark. Homework should be submitted **in your own handwriting**, to Professor Mark's mailbox in Beckman Institute by **7:00pm, Thursday November 10**. Note that the doors of the Beckman Institute are locked every night at roughly 7:00pm, so if you arrive at the last minute, you may not be able to submit on time.

Problem 1 (EM and GMM)

DLR77 = (Dempster, Laird and Rubin, 1977). JB97 = (Jeff Bilmes, 1997). This problem will use the notation of DLR77. Thus, in this problem, all vectors are row vectors rather than column vectors. The vector y is observed, z is hidden, and $x = [y, z]$. Equations in DLR77 will be referred to by numbers as (DLRn.nn).

(a) The fundamental problem of maximum likelihood parameter estimation is to find a set of parameters, ϕ, that will maximize the log likelihood of the data, the quantity $L(\phi)$ defined in Eq. (DLR2.4). When some of the variables are hidden, maximum likelihood parameters can be found by maximizing $Q(\phi'|\phi)$ from Eq. (DLR2.17) instead.

Let's substitute the pair of variables (y, z) everywhere that the paper uses x. Notice that with this substitution, Eq. (DLR2.5) becomes

$$k(z|y,\phi) = k(y,z|y,\phi) = f(x|\phi)/g(y|\phi)$$

Expand Eq. (DLR2.17) as

$$Q(\phi'|\phi) = \int k(z|y,\phi) \log f(z,y|\phi')dz$$

Define the cross-entropy between ϕ and ϕ' to be

$$H(\phi'|\phi) = \int k(z|y,\phi) \log k(z|y,\phi')dz \tag{1}$$

Use Eq. (DLR1.1), Eq. (DLR2.4), Eq. (DLR2.5), Eq. (DLR2.17), and Eq. (1) to find $Q(\phi'|\phi)$ in terms of $L(\phi')$ and $H(\phi'|\phi)$.

(b) Recall that the goal of maximum likelihood estimation is to find parameters that maximize $L(\phi)$. Suppose we start with one set of parameters, $\phi^{(p)}$, and we want to find another set, $\phi^{(p+1)}$, such that $L(\phi^{(p+1)}) > L(\phi^{(p)})$.

Use Eq. (DLR3.3), and your result from part (a), to show that

$$L(\phi^{(p+1)}) - L(\phi^{(p)}) \geq Q(\phi^{(p+1)}|\phi^{(p)}) - Q(\phi^{(p)}|\phi^{(p)}) \tag{2}$$

Given Eq. (2), how should we choose $\phi^{(p+1)}$, and why?

(c) Use Eq. (DLR2.1) to find $Q(\phi^{(p+1)}|\phi^{(p)}) - Q(\phi^{(p)}|\phi^{(p)})$ for distributions in the exponential family. Your answer should include terms related to $\log a(\phi^{(p)})$, $\log a(\phi^{(p+1)})$, $\phi^{(p+1)}$, $\phi^{(p)}$, and the term $t^{(p)T}$ defined in Eq. (DLR2.2).

(d) Use Eq. (DLR2.11) and your answer to the previous section to prove that $Q(\phi|\phi^{(p)})$ is maximized by the value of ϕ specified in Eq. (DLR2.3).

(e) Use Eqs. (DLR1.1), (DLR2.1), (DLR2.5), and (DLR2.7) to derive (DLR2.8). Hint: you may find this easier if you substitute (y, z) in place of x everywhere that x occurs. Thus, for example, the region of integration $\mathcal{X}(y_0)$ can be defined as $\mathcal{X}(y_0) = \{y, z : y = y_0\}$

(f) Consider the case of a GMM. Define y to be a set of N i.i.d. observation vectors, each of which is M-dimensional:

$$y = \{\vec{y}_1, \ldots, \vec{y}_N\}, \quad \vec{y}_n \in \Re^M$$

Let z be the matching set of hidden variables, each of which is just an integer $1 \le z_n \le K$, thus

$$z = \{z_1, \ldots, z_N\}, \quad z_n \in \{1, \ldots, K\}$$

Finally, define their joint pdf to be

$$f(y, z|\phi) = \prod_{n=1}^{N} c_{z_n} \frac{1}{\det(2\pi\Sigma_{z_n})} \exp\left(-\vec{r}_{z_n}\left((\vec{y}_n - \vec{\mu}_{z_n}) \otimes (\vec{y}_n - \vec{\mu}_{z_n})\right)^T\right) \tag{3}$$

where the following special notation has been defined. First, $\vec{r}_z$ is a $1 \times M^2$ vector, created by listing all of the elements of the $M \times M$ square matrix Σ_z^{-1} into one long row vector (for example, in a pseudo-Matlab notation, we could write `r=Sigmainverse(:)'`). Second, $\vec{y}_n \otimes \vec{y}_n$ is the tensor product vector, defined as follows. If $\vec{y}_n = [y_{n1}, \ldots, y_{nM}]$, then

$$\vec{y}_n \otimes \vec{y}_n = [y_{n1}\vec{y}_n, y_{n2}\vec{y}_n, \ldots, y_{nM}\vec{y}_n]$$

Finally, you may find it useful to define the unit indicator function as

$$[\![p]\!] = \begin{cases} 1 & p \text{ is true} \\ 0 & p \text{ is false} \end{cases} \tag{4}$$

Show that the GMM PDF (Eq. (3)) is a member of the exponential family. Find the vector of sufficient statistics, $t(y, z)$, and the vector of parameters, ϕ, such that the only interaction between them is their inner product, as shown in Eq. (DLR2.1). Remember that $t(y, z)$ must contain only simple functions of the data (observed vectors $\vec{y}_n$ and hidden variables z_n), while ϕ must contain only simple functions of the trainable parameters (c_z, $\vec{\mu}_z$, $\vec{r}_z$, and/or Σ_z^{-1}).

(g) Use your result from Sec. (f) to compute the observation-conditional expectation $t^{(p)}$, defined in (Eq. (DLR2.2)). Your answer should contain only the observation vectors $\vec{y}_n$, and the gamma function defined as follows:

$$\gamma_k^{(p)}(n) = k(z_n = k|y_n, \phi^{(p)}) = \frac{c_k \mathcal{N}(\vec{y}_n|\vec{\mu}_k, \Sigma_k)}{\sum_{\ell=1}^{K} c_\ell \mathcal{N}(\vec{y}_n|\vec{\mu}_\ell, \Sigma_\ell)}$$

(h) Use your result from Sec. (f) to compute the unconditional expectation $E(t(x)|\phi^{(p+1)})$. Notice that this expectation does *not* depend on $\vec{y}_n$.

(i) Show that if you set the results of the previous two sections equal to one another, as in Eq. (DLR2.3), then you wind up with re-estimation equations that are identical to those of page 7 in JB97.

Problem 2 (Restricted Boltzmann Machines)

Smolensky, 1986 (S86) defines the knowledge vector

$$\vec{a} = [a_1, \ldots, a_\alpha, \ldots]^T, \quad a_\alpha \in \{0, 1\}$$

and the feature vector

$$\vec{r} = [r_1, \ldots, r_i, \ldots]^T, \quad r_i \in \{0, 1\}$$

The harmony between them, $H(\vec{r}, \vec{a})$, is given by his Eq. (1) (which we will call (S1)). Define $\vec{e} = [1, 1, \ldots]^T$, and define W as on p. 232 of S86 to be

$$W_{i\alpha} = (\vec{k}_\alpha)_i \frac{\sigma_\alpha}{\|\vec{k}_\alpha\|_1}$$

Then in terms of W, $\vec{e}$, and Smolensky's backoff coefficient κ, it's possible to re-write Eq. (S1) as

$$H(\vec{r}, \vec{a}) = (\vec{r}^T W - \kappa \vec{e}^T)\vec{a} \tag{5}$$

(a) Eq. (S2) (equation (2) in Smolensky, 1986) defines $p(\vec{r}, \vec{a}) \propto e^{H/T}$ for some arbitrary constant T. Show that this pdf fits the exponential family of distributions defined in (DLR2.1). Create a table in which the DLR77 notation is in the left column, and the corresponding S86 notation is in the right column. Put the following entries in the left column: y, z, ϕ, $t(x)$, $a(\phi)$, and $b(x)$. Specify, in the right column, terms using Smolensky's notation that correspond to each of these DLR77 terms, in order to demonstrate that Eq. (S2) is a pdf in the exponential family.

(b) Use Eq. (S1) and (S2) to derive the conditional expectations $E(r_i|\vec{a})$ and $E(a_\alpha|\vec{r})$.

Problem 3 (Parzen Windows)

Suppose that we have a training dataset $\mathcal{D} = \{x_1, \ldots, x_N\}$ whose samples were generated i.i.d. from an unknown pdf $f(x)$. We estimate $f(x)$ using the following estimator:

$$f_N(x) = \frac{1}{hN} \sum_{n=1}^{N} K\left(\frac{x - x_n}{h}\right)$$

where

$$K(x) = \begin{cases} 1 & |x| \leq 0.5 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $E[f_N(x)]$ in terms of $f(x)$.

(b) Find $E[f_N^2(x)]$ in terms of $f(x)$.

(c) Choose a value of h, as a function of N, such that $f_N(x)$ is a consistent and asymptotically unbiased estimator.