CS725: Practice Problems 2

1. Understand the L_1 regularized logistic regression problem in `http://www.eecs.umich.edu/~honglak/aaai06_L1logreg.pdf` and understand the algorithm used to solve it. What are the gradient ascent steps for the regularized and for the non-regularized logistic regression objective? What are the Newton update rules for the regularized and for the non-regularized logistic regression objective?

2. Apply EM to mixtures of bernoullis (Sections 8 and 8.1), multivariate bernoullis with Nave Bayes assumption (Section 10.4) and univariate Gaussians with Nave Bayes assumption (Section 10.5) and mixture of multivariate Gaussians with common means (Section 11.1, page 64). All references to sections are to `http://www.cse.iitb.ac.in/~cs725/notes/classNotes/lecturenotes_cs725_aut11.pdf`.

3. Let $\mathbf{X}$ be a random vector and Γ its covariance matrix. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the n (normalized) eigenvectors of Γ. The n principal components of $\mathbf{X}$ are said to be $\mathbf{e}_1^T\mathbf{X}$, $\mathbf{e}_2^T\mathbf{X}$, ..., $\mathbf{e}_n^T\mathbf{X}$.

   The idea behind principal components is to find a rotation of the original coordinate system and to express $\mathbf{X}$ in that system so that each new coordinate expresses as much as possible of the variability in $\mathbf{X}$ as can be expressed by a linear combination of the n entries of $\mathbf{X}$.

   (a) Let $p(X_1) = \mathcal{N}(0,1)$ and $p(X_2) = \mathcal{N}(0,1)$ and $cov(X_1, X_2) = \theta$. Find all the principal components of the random vector $\mathbf{X} = [X_1, X_2]^T$.

   (b) Now, let $\mathbf{Y} = \mathcal{N}(\mathbf{0}, \Sigma)$ where $\Sigma = \lambda^2 I_{p\times p} + \alpha^2 ones(p,p)$ for any $\lambda, \alpha \in \Re$. Here, $I_{p\times p}$ is a $p \times p$ identity matrix while $ones(p,p)$ (following the scilab notation) is a $p \times p$ matrix of 1's. Find all the principal components of $\mathbf{Y}$.

4. Assume that a random vector $\mathbf{X} \in \Re^n$ comes from the commonly used exponential probability distribution of the form

$$\Pr(\mathbf{X} \mid \mathbf{w}) = \frac{1}{Z(\mathbf{w})} exp\left(\sum_{k=1}^{r} w_k \phi_k(\mathbf{X})\right) \tag{1}$$

where the forms of basis functions $\phi_k(.)$ are known and $\phi_k(\mathbf{x}_i)$ can be computed for any given data point $\mathbf{x}_i$. The parameters $\mathbf{w} = \{w_k\}$ are not known. $Z(\mathbf{w})$ is the normalization factor.

(a) Let us say you learn the maximum likelihood $\widehat{\mathbf{w}}_{ML}$ estimate using data set $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$. We can then derive the following equality;

$$\int_{\mathbf{x} \in \Re^n} \Pr(\mathbf{x} \mid \widehat{\mathbf{w}}_{ML}) \phi_k(\mathbf{x}) d\mathbf{x} = A \sum_i B(i) \qquad (2)$$

for any $1 \leq k \leq r$. Determine the values of A and $B(i)$ in terms of any of the known quantities in this problem. In the summation, over what range does i vary? You MUST show all the derivation steps.

(b) Now suppose you want to maximize the entropy of $\Pr(\mathbf{X} \mid \mathbf{w})$ subject to the constraint (2). Derive the form of $\Pr(\mathbf{X} \mid \mathbf{w})$ having maximum entropy, while satisfying constraint (2).

5. Prove that the k-means algorithm will converge in a finite number of steps. (As usual, the number of data points being clustered is n and k is fixed).

6. Solve the following problems from PRML (solutions on PRML book website): 1.1, 1.9, 1.10, 1.12, 1.22, 1.24, 1.29, 2.1, 2.32, 2.34, 2.36, 2.40, 2.56, 2.60, 3.4, 3.6, 4.2, 4.4, 4.9, 4.13, 6.1, 6.12, 7.1, 7.4, 7.8, 8.1, 8.8, 8.9, 8.12, 8.15, 8.18, 9.1, 9.3, 9.7, 9.8, 9.12, 9.15, 9.25, 9.26,

7. Solving the following harder problems from PRML (solutions on PRML book website) is recommended only for pratice: 1.15,, 1.20, 1.27, 1.35, 1.41, 2.17, 2.20, 3.5, 3.8, 3.10, 3.15, 4.19, 6.5, 6.7, 6.14, 8.2, 9.17, 14.5.