UC Berkeley
Department of Electrical Engineering and Computer Science
Department of Statistics

EECS 281A / STAT 241A Statistical Learning Theory

Problem Set 5

Fall 2011

Issued: Thursday, November 10, 2011 **Due:** Wednesday, November 30, 2011

Problem 5.1

Convexity of loglikelihood in mixture weights: Given k fixed distributions $p_1(x), p_2(x), \ldots, p_k(x)$, consider the problem of fitting the mixture weights $\theta = (\pi_1, \pi_2, \ldots, \pi_k)$ (where $\sum_i \pi_i = 1$, $\pi_i \geq 0$) to the mixture distribution

$$p(x|\theta) = \pi_1 p_1(x) + \pi_2 p_2(x) + \cdots + \pi_k p_k(x).$$

Prove that, given n i.i.d. samples of X, the loglikelihood $\ell(\theta|\mathcal{D})$ is concave in θ.

Problem 5.2

HMM with mixture model emissions.

A common modification of the HMM involves using mixture models for the emission probabilities $p(y_t|q_t)$. For concreteness, let's assume that the y_t are real-valued vectors, and thus our model involves a mixture of Gaussians for each value of the state.

(a) Draw the graphical model for this modified HMM, identifying clearly the additional latent variables that are needed.

(b) Write the expected complete log likelihood for the model and identify the expectations that you need to compute in the E step.

(c) Outline an algorithm for computing the E step, relating it to the standard alpha and beta recursions.

(d) Write down the equations that implement the M step.

Problem 5.3

Factor analysis and principal component analysis

We have supplied you with two 2-dimensional data sets that illustrate subspace methods: `pca1.dat` and `pca2.dat`. To generate the data, we first chose a line through the origin and chose random samples from a univariate standard Gaussian distribution along that line. We then "corrupted" these data in two different ways: (1) by using an additive two-dimensional Gaussian with equal covariances in the y_1 and y_2 directions (`pca1.dat`), and (2) by using an additive two-dimensional Gaussian with greater covariance in the y_2 than in the y_1 direction (`pca2.dat`). You are to compare the factor analysis and the principal component fits to these two data sets and comment on what changes and what stays the same.

(a) Write a Matlab or R implementation of PCA: For each data set you should compute the sample covariance matrix, determine the principal eigenvector, and project the data onto the corresponding subspace.

(b) Write a Matlab or R implementation of factor analysis using the EM algorithm discussed in class. Once you've determined the parameters, for each data point you can compute the posterior probability $p(x|y)$; this is the factor analysis equivalent of projecting onto the principal subspace.

(c) Compute the fits for both data sets and plot the resulting projections. What changes and what stays the same?

Problem 5.4

Gibbs sampling and mean field: Consider the Ising model with binary variables $X_s \in \{-1, 1\}$, and a factorization of the form $p(x;\theta) \propto \exp\big\{\sum_{s\in V} \theta_s x_s + \sum_{(s,t)\in E} \theta_{st} x_s x_t\big\}$ To make the problem symmetric, assume a 2-D grid with toroidal (donut-like) boundary conditions, as illustrated in Figure 1.

(a) Derive the Gibbs sampling updates for this model. Implement the algorithm for $\theta_{st} = 0.2$ for all edges, and $\theta_s = 0.2 + (-1)^s$ for all $s \in \{1, \ldots, 49\}$ (using the node ordering in Figure 1). Run a burn-in period of 1000 iterations (where one iteration amounts to updating each node once). For each of 1000 subsequent iterations, collect a sample vector, and use the 1000 samples to form Monte Carlo estimates $\widehat{\mu}_s$ of the moments $\mathbb{E}[X_s]$ at each node. Output a 7×7 matrix of the estimated moments. Repeating this same experiment a few times will provide an idea of the variability in your estimate. Hand in print-outs of your code, as well as your results.

(b) Derive the naive mean field updates (based on a fully factorized approximation), and implement them for the same model. Compute the average ℓ_1 distance $\frac{1}{49}\sum_{i=1}^{49} |\tau_s - \widehat{\mu}_s|$ between the mean field estimated moments τ_s, and the Gibbs estimates $\widehat{\mu}_s$. Hand in print-outs of your code, as well as your results.

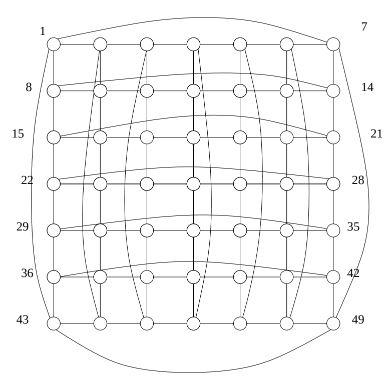

Figure 1: A two-dimensional grid graph with toroidal boundary conditions.