Due Date: October 22, 2015 in class.

Problem 1 (Mixture of Gaussians). Let $X \in \mathbb{R}^d$ be a random vector that is drawn from a mixture of Gaussians. More precisely, there are $k$ ($k \ll d$) Gaussian components, each with a center $\mu_i$ ($i \in \{1, 2, ..., k\}$). The random variable $X$ is sampled as

$$X \sim \mathcal{N}(\mu_i, \sigma^2 I) \text{ with probability } 1/k.$$ That is, first pick one of the $k$ Gaussians uniformly, and then sample $X$ from that Gaussian distribution. All the Gaussians have the same spherical covariance matrix $\sigma^2 I$ ($\sigma^2$ is known).

Given $n$ samples $X_1, X_2, ..., X_n$, let $A \in \mathbb{R}^{d \times n}$ be the matrix whose $i$-th column is equal to $X_i$. Let $C \in \mathbb{R}^{d \times n}$ be the (unknown) matrix whose $i$-th column is equal to the center for $X_i$.

(a) (5 points) How large is the spectral norm $\|A - C\|_2$? Your answer should be correct up to a constant factor with high probability. (Hint: By random matrix theory, a $d \times n$ matrix with independent standard Gaussian entries has spectral norm $\Theta(\sqrt{\min\{d,n\}})$).

(b) (10 points) Suppose the centers $\mu_i$’s are orthogonal to each other, and $\|\mu_i\|_2 = 1$ for all $i$. When $n \gg k \log k$ how large is the smallest nonzero singular value $\sigma_{\min}(C)$? Show your answer is correct (up to constant factor) with high probability.

(c) (5 points) Let $U$ be the column span of $C$, and $\hat{U}$ be the column span of the best rank-$k$ approximation of $A$. Show that under the assumption of (b), when $n \geq d \gg k \log k$ and $\sigma \ll 1/\sqrt{k}$, the distance between $U$ and $\hat{U}$ (measured in principal angle) is $O(\sigma \sqrt{k})$. (Hint: Use Wedin’s Theorem).

Theorem 1 (Wedin’s Theorem, Theorem 4.4, p. 262 in Stewart and Sun (1990)). Let $A, E \in \mathbb{R}^{m \times n}$ with $m \geq n$. Suppose $A$ has singular value decomposition

$$\begin{bmatrix} U_1^T & U_2^T & U_3^T \end{bmatrix} A \begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}.$$ Let $\tilde{A} := A + E$, with analogous singular value decomposition $(\tilde{U}_1, \tilde{U}_2, \tilde{U}_3, \tilde{V}_1, \tilde{V}_2, \tilde{\Sigma}_1, \tilde{\Sigma}_2)$. Let $\delta > 0$ be the minimum of $\min_{i,j} |\Sigma_1[i,i] - \Sigma_2[j,j]|$ and $\min_i \Sigma_1[i,i]$, if $\delta \geq 4\|E\|_2$ then the distance between $U$ and $\tilde{U}$ (measured in principal angle) is bounded by $O(\|E\|_2/\delta)$.
Problem 2 (Tensor Basics). Consider the following tensor

\[ T = \left( \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \right). \]

(a) (5 points) Write out the polynomial \( T(x, x, x) \) where \( x = (x_1, x_2) \in \mathbb{R}^2 \) as a sum of monomials.

(b) (5 points) Use Jenrich’s algorithm to decompose \( T \). In particular, let \( M_1 = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}, \)

\[ M_2 = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}, \]

do simultaneous diagonalization for \( M_1 \) and \( M_2 \). Finally write \( T \) as a sum of rank 1 tensors (use as few rank 1 tensors as possible).

Problem 3 (Noisy-Or Networks). Consider a probabilistic model for diseases and symptoms. There are \( n \) possible diseases and \( m \) symptoms. We use variables \( d_1, d_2, ..., d_n \in \{0, 1\} \) for diseases (\( d_i = 1 \) means the patient has the disease), and \( s_1, ..., s_m \in \{0, 1\} \) for symptoms (\( s_i = 1 \) means the patient has the symptom).

For each disease, there is a probability \( p_i (i \in \{1, 2, ..., n\}) \) that the patient has the disease, and all diseases are independent. The diseases and symptoms are connected by a weighted bipartite graph \( G = (D, S, E) \) (see Figure 1), on each edge the weight \( q_{i,j} \) represents the probability of a disease causing a symptom.

Each symptom may be caused by multiple diseases, and the probability of a symptom is

\[ \Pr[s_j = 0|d_1, d_2, ..., d_n] = \prod_{(i,j) \in E} (1 - d_i q_{i,j}). \]

This is called a “noisy-or” network, because if all the edge weights are 1, then \( s_j \) is just the logical or of all the diseases.

In this problem we consider a very simple network. There are only 4 diseases and 3 symptoms. Disease \( d_4 \) causes all 3 symptoms. Disease \( d_i \) for \( i = 1, 2, 3 \) causes only symptom \( s_i \).

(a) (10 points) Let \( T_{i,j,k}(i, j, k \in \{0, 1\}) \) be a \( 2 \times 2 \times 2 \) tensor, whose \( i, j, k \)-th entry is equal to \( \Pr[s_1 = i, s_2 = j, s_3 = k] \). Show that the tensor has rank (at most) 2. (Hint: Conditioned on \( d_4, s_1, s_2, s_3 \) are independent.)
(b) (10 points) Suppose all the conditional probabilities are in $(0, 1)$, and the probabilities of diseases are also in $(0, 1)$. Given a decomposition for tensor $T$ as $T = \lambda_1 u_1 \otimes v_1 \otimes w_1 + \lambda_2 u_2 \otimes v_2 \otimes w_2$, where $u_1, u_2, v_1, v_2, w_1, w_2 \in \mathbb{R}^2$ are unit vectors and $\lambda_1, \lambda_2$ are real numbers, describe an algorithm that can compute the conditional probabilities $q_{4j}$ for $j = 1, 2, 3$.

(Hint: The tensor decomposition is unique up to scaling and swapping the two components. The main difficulty is to find the correct scaling for the components $u, v, w$, and decide a correct ordering for the two components.)