

EE381V: Large Scale Learning — Spring 2013

ASSIGNMENT TWO

Caramanis/Sanghavi

Due: Tuesday, Feb. 19, 2013.

Computational Problems.

1. Mixture model problem

This problem shows that projecting onto proper subspace helps greatly in learning mixture models compared to projecting onto random subspace.

- (a) Create a mixture model with the following specifications.
 $X \in R^d$ is a random variable whose distribution is a mixture of 4 gaussian random variables with means at e_1, e_2, e_3 and e_4 (e_i are standard basis) of magnitude 1 and variance 0.1. All 4 random variables are chosen uniformly at random. Draw $\lfloor 4 * d \log(d) \rfloor$ samples of X for $d=10, 100$ and 500 .
- (b) Cluster the samples based on nearest neighbour. One way to do it is to classify any two points as from the same cluster if the distance between them is less than a threshold r (Look for function 'pdist' in matlab). Count the fraction of pairs in error. The error will depend on the threshold chosen r . One good value for threshold is the diameter of a cluster, which is $\sqrt{2 * \sigma^2 * d}$.
- (c) Now project all the points into a low dimensional subspace spanning the means of the four gaussian random variables. Take the space spanned by $\{e_1, \dots, e_4\}$ for this case. Compute the same error for different values of d .
- (d) This time project all points into a randomly generated subspace of dimension 10 (One way is to use the range space of a random matrix). Compute error in this case for different values of d .
- (e) Give a plot comparing errors in the above three scenarios.

2. Planted Model

Recall the planted model from class. Let $n = 200$, and $k = 5$. Form the matrix P as in class, that has p on the five equal blocks on the diagonal, and $q = 1 - p$ everywhere else. Let A be the random matrix, as in class, where each entry A_{ij} is a Bernoulli random variable with probability P_{ij} . Note that you will have to construct A as a symmetric matrix, so generate elements above the diagonal, and then just replicate them below.

- (a) For $p = 0.8, 0.7, 0.6$ and 0.55 , generate the eigenvalues of P and of (the random matrix) A and plot them. How many eigenvalues of P are non-zero?
- (b) Now run the spectral clustering algorithm on A (rather than the Laplacian). The points that you get will be 5 dimensional. Pick a random projection onto two dimensions, and project all of the resulting five dimensional points onto these two randomly chosen dimensions. Plot the results for the different values of p .

(c) Cluster according to k -Means.

Linear Algebra.

The following problems focus on important concepts and ideas from linear algebra that are critical in this class. Please read up about singular value decomposition, and other topics as necessary. Problems marked with a * are optional.

1. Range and Nullspace of Matrices: Recall from class the definition of the null space and the range of a linear transformation, $T : V \rightarrow W$:

$$\begin{aligned}\text{null}(T) &= \{\mathbf{v} \in V : T\mathbf{v} = 0\} \\ \text{range}(T) &= \{T\mathbf{v} \in W : \mathbf{v} \in V\}\end{aligned}$$

- Suppose A is a 10-by-10 matrix of rank 5, and B is also a 10-by-10 matrix of rank 5. What is the **smallest** and **largest** the rank the matrix $C = AB$ could be?
 - Now suppose A is a 10-by-15 matrix of rank 7, and B is a 15-by-11 matrix of rank 8. What is the **largest** that the rank of matrix $C = AB$ can be?
2. Riesz Representation Theorem: Consider the standard basis for $\mathbb{R}^n$: $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, etc. Recall that the inner-product of two vectors $\mathbf{w}_1 = (\alpha_1, \dots, \alpha_n)$, $\mathbf{w}_2 = (\beta_1, \dots, \beta_n) \in \mathbb{R}^n$, is given by:

$$\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = \sum_{i=1}^n \alpha_i \beta_i.$$

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a linear map. Show that there exists a vector $\mathbf{x} \in \mathbb{R}^n$, such that

$$f(\mathbf{w}) = \langle \mathbf{x}, \mathbf{w} \rangle,$$

for any $\mathbf{w} \in \mathbb{R}^n$.

Remark: It turns out that this result is true in much more generality. For example, consider the vector space of square-integrable functions (something we will see much more later in the course). Let F denote a linear map from square integrable functions to $\mathbb{R}$. Then, as a consequence similar to the finite dimensional exercise here, there exists a square integrable function, g , such that:

$$F(f) = \int fg.$$

3. Recall from class that the spectral theorem for symmetric $n \times n$ real matrices, says, among other things, that if A is a symmetric (real) $n \times n$ matrix, then it has an basis of orthonormal eigenvectors, $\{v_1, \dots, v_n\}$. Use A and $\{v_i\}$ to construct a matrix T , such that the matrix $T^\top AT$ is diagonal.
4. Let A be a rank n matrix (of any dimensions, not necessarily square). Let its singular value decomposition be given by

$$A = U\Sigma V^*,$$

where Σ is the matrix of singular values given in *descending order*: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$. Let Σ_k denote the matrix with $\sigma_{k+1}, \dots, \sigma_n$ set to zero, and let $\hat{A} = U\Sigma_k V^*$. Show that $\hat{A}$ solves the optimization problem:

$$\min_{\hat{A} : \text{rk}(\hat{A}) \leq k} \|A - \hat{A}\|_F.$$

Hint: you can use the fact that the optimal solution should satisfy $(A - \hat{A}) \perp \hat{A}$, where orthogonality is defined with respect to the natural matrix inner product compatible with the Frobenius norm:

$$\langle M, N \rangle = \sum_{i,j} M_{ij} N_{ij} = \text{Trace}(M^* N).$$

If you choose to use this hint, please do show that $\hat{A}$ and A satisfy the orthogonality, as claimed.

5. * Prove the above hint, namely, that the optimal solution must satisfy the orthogonality condition (in the previous problem, you are assuming that this is true).
6. Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Recall that by the spectral theorem, A will have real eigenvalues. Therefore we can order the eigenvalues of A : $\lambda_1(A) \geq \lambda_2(A) \geq \dots \geq \lambda_n(A)$. Show that:

$$\begin{aligned} \lambda_1(A) &= \max_{y \neq 0} \frac{\langle y, Ay \rangle}{\langle y, y \rangle}, \\ &\vdots \\ \lambda_k(A) &= \max_{V: \dim V = k} \min_{0 \neq y \in V} \frac{\langle y, Ay \rangle}{\langle y, y \rangle}. \end{aligned}$$

7. * The spectral radius of a matrix A is defined as:

$$\rho(A) = \max\{|\lambda| : \lambda \text{ an e-value of } A\}.$$

Note that the spectral radius is invariant under similarity transformations, and thus we can speak of the spectral radius of a linear operator.

(a) Show that $\rho(A) \leq \sigma_1(A)$.

(b) Show that

$$\rho(A) = \inf_{\{S: \det S \neq 0\}} \sigma_1(S^{-1}AS).$$

8. A linear operator N is called *nilpotent* if for some integer k , $N^k = 0$. Show that if N is nilpotent, then $(I + N)$ has a square root. (Hint: consider the Taylor expansion of $\sqrt{1+x}$, and use that as a starting point).
9. Find an example of a matrix that is diagonalizable, but not unitarily so. That is, produce an example of a $n \times n$ matrix A (n up to you) for which there is some invertible matrix T that satisfies $T^{-1}AT = D$ for some diagonal matrix, but the columns of T cannot be taken to be orthonormal. Hint: related to one of the problems above.
10. Suppose $A \in \mathbb{C}^{m \times n}$ has full column rank n . Show that:

$$\min_{\Delta \in \mathbb{C}^{m \times n}} \{ \|\Delta\|_2 \mid \text{rank}(A + \Delta) < n \} = \sigma_n(A),$$

where $\sigma_n(A)$ denotes the smallest singular value of A .