The Technion (and The University of Texas at Austin)
Department of Electrical and Computer Engineering

Large Scale Learning — Fall 2013

Assignment 1

Constantine Caramanis Not Due

There are numerous problems marked as *Exercise* given during the class, that are meant to fill in some missing details. These are not replicated here. The point of the exercises below is more of the same: to provide practice and review, and also to fill in details left out in class.

Some linear algebra review. Most of the below are standard.

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

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

   - 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. Let A be an $n \times m$ real matrix, and B a $k \times m$ real matrix. Suppose that for every $x \in \mathbb{R}^m$, $Ax = 0$ only if $Bx = 0$, that is,
$$Ax = 0 \Rightarrow Bx = 0.$$
Show that there exists a $k \times n$ real matrix C such that $CA = B$.

3. Recall the definition of rank, and show the following.
   - For A an $m \times n$ matrix, $\text{rank}A \le \min\{m, n\}$.
   - For A an $m \times k$ matrix and B a $k \times n$ matrix,
   $$\text{rank}(A) + \text{rank}(B) - k \le \text{rank}(AB) \le \min\{\text{rank}(A), \text{rank}(B)\}.$$
   - For A and B $m \times n$ matrices,
   $$\text{rank}(A + B) \le \text{rank}(A) + \text{rank}(B).$$
   - For A an $m \times k$ matrix, B a $k \times p$ matrix, and C a $p \times n$ matrix, then
   $$\text{rank}(AB) + \text{rank}(BC) \le \text{rank}(B) + \text{rank}(ABC)$$

4. Consider a mapping $T : V \to V$. If the vector space V is finite dimensional, show that then if $\text{null}T = \{0\}$, then T is surjective. Conversely, if T is surjective, then $\text{null}T = \{0\}$, and $T\boldsymbol{v} = 0$ implies $\boldsymbol{v} = 0$.

5. Show that both directions of the above can fail if V is not finite dimensional. That is, give an example of an infinite dimensional vector space, V, and a linear operator $T : V \to V$, such that T is surjective, but null$T \neq \{0\}$. Then, give an example of an infinite dimensional vector space, V, and a linear operator $T : V \to V$, such that null$T = \{0\}$, but T is not surjective.

6. 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, $\{v_1, \ldots, v_n\}$. Use A and $\{v_i\}$ to construct a matrix T, such that the matrix $T^\top AT$ is diagonal.

7. 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*: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$. Let Σ_k denote the matrix with $\sigma_{k+1}, \ldots, \sigma_n$ set to zero, and let $\hat{A} = U\Sigma_k V^*$. Show that $\hat{A}$ solves the optimization problem:
$$\min_{\hat{A} : \mathrm{rk}(\hat{A}) \leq k} \|A - \hat{A}\|_F.$$
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:
$$\langle M, N \rangle = \sum_{i,j} M_{ij} N_{ij} = \mathrm{Trace}(M^* N).$$
If you choose to use this hint, please do show that $\hat{A}$ and A satisfy the orthogonality, as claimed.

8. 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).

9. 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: $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$. Show that:
$$\begin{aligned} \lambda_1(A) &= \max_{y \neq 0} \frac{\langle y, Ay \rangle}{\langle y, y \rangle}, \\ &\vdots \\ \lambda_k(A) &= \max_{V : \dim V = k} \min_{0 \neq y \in V} \frac{\langle y, Ay \rangle}{\langle y, y \rangle}. \end{aligned}$$

10. 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
    $$\rho(A) = \inf_{\{S : \det S \neq 0\}} \sigma_1(S^{-1}AS).$$

11. 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).

12. 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.

13. Suppose $A \in \mathbb{C}^{m\times n}$ has full column rank n. Show that:

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

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

Other Problems

1. **Planted Model**

We will discuss the planted model in class in a lecture or two. In this model, we explore how much noise spectral methods can tolerate while still finding the right clusterings. It is used as a model for community detection. The basic model is as follows: Recall that if we have a $n \times n$ similarity matrix with k blocks that have all 1's inside the blocks, and then 0's outside the blocks, that spectral clustering performs a perfect clustering of the blocks. Instead, the planted model puts a 1 in each element of the blocks with probability p, and puts a 1 in each entry *outside* the blocks with some probability q. For this to make sense, we need $p > q$. Therefore, there are missing 1's (or edges) inside clusters with probability $(1 - p)$, and there are 1's outside, or equivalently edges between clusters, with probability q.

How small can $(p - q)$ be, before we fail to detect the clusters?

Let $n = 200$, and $k = 5$. Form the matrix P as described above, that has p on the five equal 40×40 blocks on the diagonal, and $q = 1-p$ everywhere else. Let A be the resulting (random) similarity matrix 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 matrix A. Permute the rows and columns (same permutation!) to get an appreciation of how non-trivial it is to find the blocks.

(b) Compute the eigenvalues of the deterministic matrix P and of (the random matrix) A and plot them. How many eigenvalues of P are non-zero?

(c) 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.

(d) Cluster according to k-Means.

2. **Gershgorin Circle Theorem**

Let A be some $n \times n$ matrix. If for some invertible matrix, X, $X^{-1}AX = D$, a diagonal matrix, then we know the eigenvalues of A are the elements of D. If $X^{-1}AX$ is *approximately diagonal*, i.e.,

$$X^{-1}AX = D + F,$$

where F has zeros on the diagonal, then how close are the eigenvalues of A to the elements of D? You answer that here.

(a) Show that for any square matrix, M, the induced infinity norm satisfies:

$$\|M\|_\infty = \sup_{\|x\|_\infty = 1} \|Mx\|_\infty = \max_i \sum_j |M_{ij}|.$$

(b) For $\lambda \in \lambda(A)$, and assuming that $\lambda \neq d_i$ for any i, show that $(D - \lambda I) + F$ is singular.

(c) Again assuming $\lambda \in \lambda(A)$, and assuming that $\lambda \neq d_i$ for any i, show that

$$1 \leq \|(D - \lambda I)^{-1} F\|_\infty = \sum_{j=1}^{n} \frac{|f_{kj}|}{|d_k - \lambda|}.$$

(d) Conclude that

$$\lambda(A) \subseteq \bigcup_{i=1}^{n} \mathcal{D}_i,$$

where

$$\mathcal{D}_i = \{z \in \mathbb{C} \, : \, |z - d_i| \leq \sum_{j=1}^{n} |f_{ij}|\}.$$