GEORGIA INSTITUTE OF TECHNOLOGY
SCHOOL of ELECTRICAL and COMPUTER ENGINEERING

ECE 6250 Fall 2018
Problem Set #10

Assigned: 14-Nov-18
Due Date: 26-Nov-18

As stated in the syllabus, unauthorized use of previous semester course materials is strictly prohibited in this course.

All assignments will be due at the beginning of class on the due date.

PROBLEM 10.1:
Using your class notes, prepare a 1-2 paragraph summary of what we talked about in class in the last week. I do not want just a bulleted list of topics, I want you to use complete sentences and establish context (Why is what we have learned relevant? How does it connect with other things you have learned here or in other classes?). The more insight you give, the better.

PROBLEM 10.2:
Let $\boldsymbol{A}$ be an $N \times N$ symmetric matrix. Find the best approximation of $\boldsymbol{A}$ that is a scalar multiple of the identity. That is, solve

$$\min_{\gamma \in \mathbb{R}} \ \|\boldsymbol{A} - \gamma \mathbf{I}\|_F^2.$$

PROBLEM 10.3:
Suppose we make a noisy observation of $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, with

$$\boldsymbol{A} = \begin{bmatrix} 2 & 4 & -1 \\ 1 & -2 & 1 \\ 4 & 0 & 1 \\ 5 & 6 & -1 \\ 8 & -4 & 2 \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -2 \\ 5 \end{bmatrix}$$

(a) Find the total-least squares solution to the above linear inverse problem. (Use MATLAB.)

(b) What is the residual error $\|\boldsymbol{\Delta}\|_F^2$? What are the $\boldsymbol{\Delta A}$ and $\boldsymbol{\Delta y}$ corresponding to your solution?

Last updated 16:32, November 16, 2018

PROBLEM 10.4:

The variable `faces` in `faces.mat` contains $D = 2414$ examples of images of human faces. Each face is a 32×32 pixel image, stored as a vector of length 1024. You have been provided with a function `plotFaces.m` which will plot the first several images (columns) in a matrix that you provide it. For example, the command `plotFaces(faces,4,4)` will plot a 4×4 table with the first 16 images in the matrix `faces`.

In this problem we will show that this data is relatively tightly clustered around a comparatively low-dimensional subspace of $\mathbb{R}^{1024}$.

(a) Based on the discussion of PCA in class, write code which can compute the optimal affine approximation to the data in `faces`. Specifically, let $\boldsymbol{x}_n$ denote the n^{th} column in `faces`. We are looking for the optimal approximation of the form $\boldsymbol{x}_n \approx \boldsymbol{\mu} + \boldsymbol{Q}\boldsymbol{\theta}_n$. Plot the resulting approximation for the first 16 faces when $r = 16$ (using `plotFaces`).

(b) Plot each of the basis vectors for the subspace identified in the previous part when $r = 16$ (again using `plotFaces`).

(c) How large does r need to be to ensure that the relative error between the original dataset and our approximation is less than 5%? How does this relate to the singular values of the original matrix?

(d) Comment on the qualitative differences between the original dataset and the approximation produced by PCA and how this changes as we vary r.

PROBLEM 10.5:

Write a MATLAB function that uses the power iteration procedure described in the class notes to find the largest eigenvalue of a matrix $\boldsymbol{A}$ and its associated eigenvector. Keep track of how many iterations it takes until the estimated eigenvalue does not change more than `eps` (i.e., machine precision) between iterations for the following matrices:

- A random symmetric 10×10 matrix.
- A random symmetric 100×100 matrix.
- A random symmetric 1000×1000 matrix.

You can form a random symmetric matrix using the `randn` command to generate a random $\boldsymbol{B}$ and then forming $\boldsymbol{A} = \boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}}$. Repeat the above for several different random matrices and comment on the following:

(a) Does the number of iterations seem to increase with the size of the matrix? Why should or should not that be the case? (You may wish to compare the number of iterations in each trial to $|\lambda_1|/|\lambda_2|$, where λ_1 is the eigenvalue with largest magnitude, and λ_2 is the eigenvalue with second largest magnitude.)

(b) How do the eigenvalues found using your code compare with the largest (magnitude) eigenvalues found using the MATLAB `eig` command?

PROBLEM 10.6:
Suppose that you wish to solve for $\boldsymbol{x}$ given that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ where

$$\boldsymbol{A} = \begin{bmatrix} 10000 & 10001 \\ 10001 & 10002 \\ 10002 & 10003 \\ 10003 & 10004 \\ 10004 & 10005 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 20001 \\ 20003 \\ 20005 \\ 20007 \\ 20009 \end{bmatrix}.$$

Note, the exact solution is $\boldsymbol{x} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$.

(a) Determine the condition number of $\boldsymbol{A}^T\boldsymbol{A}$.

(b) Compute the least-squares solution using the formula $\hat{\boldsymbol{x}} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}A^T\boldsymbol{b}$ explicitly.

(c) Compute the solution using the QR decomposition. You should show the matrix equations that you used to arrive at your solution along with any MATLAB code or commands (you can use the MATLAB `qr( )` command).

(d) Compute the solution using the Cholesky factorization. You should show the matrix equations that you used to arrive at your solution along with any MATLAB code or commands (you can use the MATLAB `chol( )` command).

(e) Compute the solution using the MATLAB backslash command "\". View the online documentation of the backslash command (also known as `mldivide()` and determine which method it is using to solve for $\hat{\boldsymbol{x}}$.

(f) Compare the answers and comment.

Include your code with your assignment submission.

Last updated 16:32, November 16, 2018

PROBLEM 10.7 (OPTIONAL):

(a) Let $\mathbf{\Sigma}$ be an $N \times N$ diagonal matrix. What is the diagonal matrix that has rank at most r which is closest to $\mathbf{\Sigma}$? That is, solve

$$\min_{\mathbf{\Sigma}' \in \mathbb{R}^{N\times N}} \|\mathbf{\Sigma} - \mathbf{\Sigma}'\|_F^2 \quad \text{subject to} \quad \mathbf{\Sigma}' \text{ is diagonal, } \operatorname{rank}(\mathbf{\Sigma}') \le r.$$

(b) Show that the Frobenius norm is invariant under rotations. That is, if $\boldsymbol{X} \in \mathbb{R}^{M\times N}$ is an arbitrary matrix and $\boldsymbol{W} \in \mathbb{R}^{M\times M}$ and $\boldsymbol{Z} \in \mathbb{R}^{N\times N}$ are orthonormal matrices ($\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} = \boldsymbol{W}\boldsymbol{W}^{\mathrm{T}} = \mathbf{I}$ and similarly for $\boldsymbol{Z}$), then

$$\|\boldsymbol{X}\|_F = \|\boldsymbol{W}\boldsymbol{X}\|_F = \|\boldsymbol{X}\boldsymbol{Z}\|_F = \|\boldsymbol{W}\boldsymbol{X}\boldsymbol{Z}\|_F.$$

(c) Let $\mathbf{\Sigma}$ and $\mathbf{\Sigma}'$ be $N \times N$ diagonal matrices, and let W and Z be $N \times N$ orthonormal matrices. Argue that

$$\|\mathbf{\Sigma} - \mathbf{\Sigma}'\|_F^2 \le \|\mathbf{\Sigma} - \boldsymbol{W}\mathbf{\Sigma}'\|_F^2,$$

and

$$\|\mathbf{\Sigma} - \mathbf{\Sigma}'\|_F^2 \le \|\mathbf{\Sigma} - \mathbf{\Sigma}'\boldsymbol{Z}\|_F^2.$$

(d) Let $\mathbf{\Sigma}$ be an $N \times N$ diagonal matrix. For convenience, we will assume that the entries along the diagonal are in decreasing order of magnitude, so that $|\mathbf{\Sigma}_{1,1}| \ge |\mathbf{\Sigma}_{2,2}| \ge \cdots \ge |\mathbf{\Sigma}_{n,n}$. Show that the solution $\hat{\boldsymbol{X}}$ to

$$\min_{\boldsymbol{X} \in \mathbb{R}^{N\times N}} \|\mathbf{\Sigma} - \boldsymbol{X}\|_2^2$$

is also diagonal. Proving this will requires several steps. First, note that using the SVD, the program above becomes

$$\min_{\substack{\boldsymbol{U},\boldsymbol{V},\{\alpha_k\} \\ \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\mathbf{I},\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=\mathbf{I}}} \left\| \mathbf{\Sigma} - \sum_{k=1}^{r} \alpha_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}} \right\|_F^2.$$

(a) Fix $\boldsymbol{U}, \boldsymbol{V}$. Find the α_k that minimize

$$\min_{\{\alpha_k\}} \left\| \mathbf{\Sigma} - \sum_{k=1}^{r} \alpha_k \boldsymbol{u}_k \boldsymbol{v}_k^{T} \right\|_F^2.$$

Hint: the $N \times N$ matrices $\boldsymbol{u}_k \boldsymbol{v}_k^T$ are orthonormal under the trace inner product. So this is just finding the best approximation in a subspace for which you have an orthonormal basis ...

(b) Use your answer to previous part to show that the bases that yield the smallest approximation error can be found using

$$\max_{\substack{\boldsymbol{U},\boldsymbol{V} \in \mathbb{R}^{N\times r} \\ \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\mathbf{I},\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=\mathbf{I}}} \sum_{k=1}^{r} |\langle \mathbf{\Sigma}\boldsymbol{u}_k, \boldsymbol{v}_k \rangle|^2$$

(c) Argue that if the $\boldsymbol{v}_k = \boldsymbol{u}_k$, then

$$\sum_{k=1}^{r} |\langle \boldsymbol{\Sigma}\boldsymbol{u}_k, \boldsymbol{v}_k\rangle|^2 = \sum_{k=1}^{r} \|\boldsymbol{\Sigma}\boldsymbol{u}_k\|_2^2 = \|\boldsymbol{\Sigma}\boldsymbol{U}\|_F^2.$$

(d) Argue that

$$\max_{\substack{\boldsymbol{U},\boldsymbol{V}\in\mathbb{R}^{N\times r}\\ \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\mathbf{I},\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=\mathbf{I}}} \|\boldsymbol{\Sigma}\boldsymbol{U}\|_F^2$$

is maximized by $\boldsymbol{u}_k$ with

$$u_k[i] = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases}.$$

A key step in this is to argue that if $\boldsymbol{U}$ is an $N \times r$ matrix whose columns are orthonormal and if $\boldsymbol{u}^j \in \mathbb{R}^r$ is the j^{th} **row** of $\boldsymbol{U}$, then $\|\boldsymbol{u}^j\|_2^2 \leq 1$.

(e) Use the above to show that if $\boldsymbol{A}$ is a $M \times N$ matrix with SVD

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} = \sum_{k=1}^{p} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T,$$

then the solution to

$$\min_{\boldsymbol{X}\in\mathbb{R}^{M\times N}} \|\boldsymbol{A} - \boldsymbol{X}\|_F^2 \quad \text{subject to} \quad \operatorname{rank}(\boldsymbol{X}) \leq r$$

is

$$\widehat{\boldsymbol{A}} = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T.$$