Homework 3 - Due Mon. Feb. 20th
AS.110.446, EN.550.416

Dr. Mauro Maggioni **Office**: 405 Krieger Hall
Web page: www.math.jhu.edu/˜ mauro **E-mail**: myfirstname.maggioni@jhu.edu

Homework Policies

As in the first homework set.

Exercises

Exercise 1 (20 pts). [Inspired by Ex. 5.3 in Trefethen's book] Consider the matrix

$$A = \begin{bmatrix} 3 & -8 \\ -4 & 2 \end{bmatrix}$$

- Compute, on paper, a real SVD of A in the form $A = U\Sigma V^T$. The SVD is not unique, so find the one with the minimal number of minus signs in U and V.
- List the singular values, left singular vectors, right singular vectors.
- Draw an accurate picture of the unit ball in $\mathbb{R}^2$ and its image under A, together with the singular vectors, with their coordinates marked.
- What are the $1-$, $2-$, $\infty-$ and Frobenius norms of A?
- What is the rank of A? How can it be read from the SVD?
- Find A^{-1} not directly, but via the SVD
- Verify that $|\det A| = \sigma_1\sigma_2$.
- What is the area of the ellipsoid onto which A maps the unit ball of $\mathbb{R}^2$? [Hint: use the geometric interpretation of determinant of a linear map/think about the map as a change of variable, and about its Jacobian]
- What is the best rank-1 approximation to A?

Exercise 2 (10 pts). [Ex. 3.16 from Heath's book] Let A be a $m \times n$ matrix. Under what conditions on A is the matrix A^TA:

- Symmetric?
- Nonsingular?
- Positive definite? (Recall that a matrix is positive definite if $\langle Av, v\rangle > 0$ for all $v \neq 0$.)

In each case fully justify your answer with an argument as rigorous as you can, and/or by providing counterexamples as needed.

Exercise 3 (10 pts). Which of the following matrices are orthogonal? $A_1 := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, $A_2 := \begin{bmatrix} 4 & 0 \\ 0 & \frac{1}{4} \end{bmatrix}$, $A_3 := \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$, $A_5 := A_1A_3$, $A_6 := A_3A_2$, $A_7 := A_2A_3A_2^{-1}$.

Exercise 4 (60 pts). Extend your work in Part II of Exercise 4 of the previous homework as follows:

- For all 10 classes, compute the Principal Components of the points (one point corresponds to an image, viewed as a vector in $\mathbb{R}^{28^2}$), obtaining the left principal vectors $U_i \in \mathbb{R}^{28^2 \times 28^2}$ and the mean $m_i \in \mathbb{R}^{28^2}$ for each class $i = 0, \ldots, 9$.
- For each $k = 1, \ldots, 28^2$, define a map $f_K : \mathbb{R}^{28^2} \to \{0, \ldots, 9\}$ that projects a point x to the closest affine subspace $m_i + \langle U_i(:, 1:k) \rangle$, i.e.

$$f_K(x) = \arg\min_{i=0,\ldots,9} ||x - \underbrace{(U_i(:, 1:K)U(:, 1:K)^T(x - m_i) + m_i)}_{\mathrm{Proj}_{m_i + \langle U_i(:,1:K) \rangle} x}|| \, .$$

  Explain geometrically what this map does and explain the formula just given. f_K is called a classifier since it maps points to their (predicted) class. It is defined for any point, both in the training data you used to construct f_K, and for any other point in $\mathbb{R}^{28^2}$.

- If $f_K(x)$ is the correct class of the point x, we say that x is classified correctly, otherwise we say that x is classified incorrectly. The confusion matrix C_K is the 10×10 matrix whose entry $(C_K)_{ii'}$ is equal to the number of points x of class i such that $f_K(x) = i'$, i.e. the set of points with true label i that are classified by f_K as being in class i'.
  - What is the ideal confusion matrix C_K, for a classifier that classifies correctly most or all the time?
  - For various values of K, compute the confusion matrix C_K, and comment on any trends you may see in the confusion matrix. Can you choose a value of K that works best? How do you define best?
  - Display some of the points the are classified correctly and some of those that are classified incorrectly.
- The confusion matrices above are not a good way to look at the performance of f_K: they are constructed by evaluating f_K on the training data, that was also used to construct f_K itself. This may not be telling us much about the performance on new data, that was not used to construct f_K. So: load the MNIST test data set (from the MNIST webpage), and test f_K on test data, and construct the corresponding confusion matrices C_K, for various values of K. Can you choose a value of K that works best?

Note: You are not required to try all possible values of K, in order to find the one that works best; think of and argue for possible strategies for only checking some values of K. Thinking about this is more important than obtaining the very best value of K, which would indeed require trying all possible values.