CS-GY 6923: Assignment 2

COLLABORATION POLICY: You may discuss general concepts relating to the homework questions with other students, but you must write up your solutions on your own, in your own words.

Hard Deadline: Assigment 2 is due on October 16.

1 Logistic Regression (15 points)

Please state your stochastic gradient descent algorithm for Logistic Regression.

2 Principal Component Analysis (PCA) (20 points)

Prove by induction that the linear projection onto an M-dimensional subspace that maximizes the variance of the projected data is defined by the M eigenvectors of the data covariance matrix S, given by

$$S = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)} - \bar{x})(x^{(i)} - \bar{x})^T$$

corresponding to the M largest eigenvalues. In class we proved for the case of M=1. Now suppose the result holds for some general value of M and show that it consequently holds for dimensionality $M+1$. To do this first set the derivative of the variance of the projected data with respect to a vector $u^{(M+1)}$ defining the new direction in data space equal to zero. This should be done subject to the constraints that $u^{(M+1)}$ be orthogonal to the existing vectors $u^{(1)}, ..., u^{(M)}$, and also that it be normalized to unit length. Use Lagrange multipliers to enforce these constraints. Then make use of the orthonormality properties of the vectors $u^{(1)}, ..., u^{(M)}$ to show that the new vector $u^{(M+1)}$ is an eigenvector of S. Finally, show that the variance is maximized if the eigenvector is chosen to be the one corresponding to eigenvalue λ_{M+1} where the eigenvalues have been ordered in decreasing value.

3 Characteristic Polynomial and Eigen-Decomposition(15 points)

1. (2 points) Find the characteristic polynomial of the matrix A=

$$\begin{bmatrix} 0 & 5 \\ 5 & 3 \end{bmatrix} \tag{1}$$

2. (2 points) Solve for the eigenvalues using the characteristic polynomial you calculated.
3. (3 points) Solve for the eigenvectors using the eigenvalues you calculated in the previous step.
4. (2 points) Validate your solution with MATLAB. (Please list the MATLAB command and returning results)
5. (3 points) Is the matrix A positive definite? Why?
6. (3 points) Now let us add to A a positive definite diagonal matrix

$$B = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} \tag{2}$$

Use MATLAB to calculate the eigenvalues of $A+B$. Is $A+B$ positive definite?

4 Covariance Matrix and PCA (20 points)

1. (3 points) Suppose you have three samples, $x^{(1)} = [3; 10; 1], x^{(2)} = [8; 2; 6], x^{(3)} = [5; 8; 0]$. Optionally you may form the sample matrix X as

$$X = \begin{bmatrix} 3 & 8 & 5 \\ 10 & 2 & 8 \\ 1 & 6 & 0 \end{bmatrix} \tag{3}$$

Please calculate the sample covariance matrix by hand.
2. (3 points) Is your sample covariance matrix positive semi-definite? Why? (Please do not use MATLAB for this question.)
3. (3 points) Use MATLAB to calculate all the principal components of your samples.
4. (2 points) Suppose you only have two samples $x^{(1)}$ and $x^{(2)}$, calculate your sample covariance matrix by hand again.
5. (4 points) Use MATLAB to calculate the principal components. What do you find? (Check your eigenvalues)
6. (5 points) Prove when the number of samples N is smaller than the dimension d of each sample, to reconstruct all your samples without loss, you only need to have at most N principal components.

5 EigenFace (30 points)

Download the YALE database following the link: `http://www.cad.zju.edu.cn/home/dengcai/Data/Yale/Yale_64x64.mat`

(See `http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html` for a description of the data, use MATLAB function 'reshape' for reshaping a vector into a matrix.)

1. (10 points) Use MATLAB to show the first 4 eigenfaces for that data set (as I did in class).

2. (10 points) Investigate the eigenvalues produced by the eigen-decomposition. How many principal components do you need to reconstruct the original images without any loss? (Hint: here our sample dimension d is much larger than the number of samples N.)

3. (10 points) How many principal components would you use in order to preserve at least 80% of the energy? That is, to determine the smallest d' such that $W \in R^{d\times d'}$, $W^TW = I_{d'}$, and

$$\sum_i \|W^T(x^{(i)} - \bar{x})\|_2^2 \geq 0.8 \times \sum_i \|(x^{(i)} - \bar{x})\|_2^2, \tag{4}$$

(Don't forget to subtract the mean.)