Data Visualization
STAT 890 / 442, CM 462

Assignment 2
Fall 2006
Department of Statistics and Actuarial Science
University of Waterloo

Due: Monday October 30, at the start of class
Instructor: Ali Ghodsi
MS 6081G x37316, aghodsib@uwaterloo.ca

Policy on Lateness: Slightly late assignments (up to 24 hs after due date) are accepted with 10% penalty. No assignment are accepted after 24 hs after the due date.

1. The Matlab data file 2_3.mat contains 200 handwritten 2's and 3's images scanned from postal envelopes, like the ones shown below.

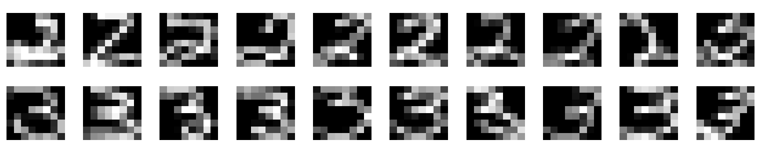

Figure 1:

These images are stored as a 64×400 matrix. Each column of the matrix is an 8×8 greyscale image (the pixel intensities are between 0 and 1). Figure 2 illustrates the two most significant dimensions found by PCA.

(a) Reproduce Figure 2. You only need to find the low-dimensional mapping Y_{PCA} by PCA, and then call the function *plotdigits* provided in the course web page.

(b) Each coordinate axis of the embedding correlates highly with one degree of freedom underlying the original data and articulates a major feature of the digits. Explain which features are captured by x and y axes.

(c) The '2's and '3's are clustered in Figure 2, but clusters overlap each other and are not linearly separable. Figure 2 illustrating the first two dimensions (the first dimension versus the second dimension). Plot all combinations of the first five dimensions (For example the first dimension versus the third one, or the second

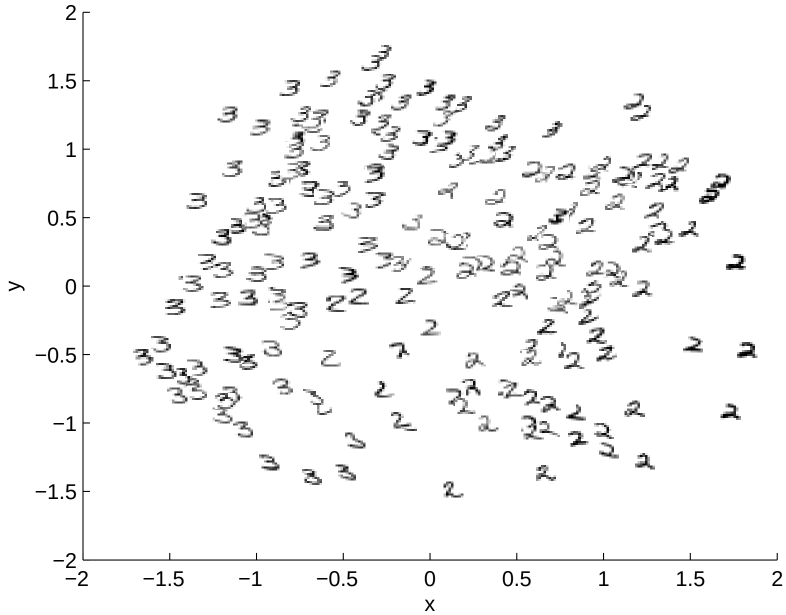

Figure 2: *A canonical dimensionality reduction problem from visual perception. The input consists of a sequence of 64-dimensional vectors, representing the brightness values of 8 pixel by 8 pixel images of digits 2 and 3. Applied to* $n = 400$ *raw images. A two-dimensional projection is shown, with the original input images.*

one versus the forth one) and find the dimension which separates 2s and 3s the best. All that needs to be handed in for this question is 10 plots for all 10 possible combinations. Identify the one which best separate 2s and 3s and don't forget to label axes.

2. Implement Kernel PCA (KPCA) with an appropriate kernel (and appropriate parameters) from your own choice and repeat part (c) of Question 1. That is find the five-dimensional mapping Y_{KPCA} by KPCA, and then call the function *plotdigits* for each combination of dimensions (For example the first dimension versus the third one, or the second one versus the forth one). If you chose the kernel and its parameter wisely you should observe improvement over PCA.

   (a) Hand in 10 plots for all 10 possible combinations.

   (b) Submit the code for your kernel PCA implementation.

   (c) As you probably observed, the low-dimensional map was affected by the choice of kernel function. Why the kernel that you have chosen works better than the others. Note that a kernel function implicitly maps the data points in a feature space and different kernels corresponds to different feature spaces.

Only for GradStudents
Due: Friday December 1, at the start of class

3. Suppose we approximately represent each point of the dataset as a linear combination of its k nearest neighbors. Let $(W_k)_{ij}$, be the weight of point x_j in the expansion of x_i minimizing the squred representation error.

   (a) Prove that $M_k(x_i, x_j) = ((I - W_k)^T (I - W_k))_{ij}$ is a positive semidefinite kernel on the domain $X = \{x_1, \ldots, x_n\}$.

   (b) Let λ be the largest eigenvalue of $(I - W_k)^T (I - W_k)$. Prove that the LLE kernel $K_k^{LLE}(x_i, x_j) = \left((\lambda - 1)I + W_k^T + W_k - W_k^T W_k \right)_{ij}$ is positive semidefinite on $\{x_1, \ldots, x_n\}$.

   (c) Prove that kernel PCA using the LLE kernel provides the LLE embedding coefficients for a $d-$dimensional embedding as the coefficient eigenvectors $\alpha^2, \ldots, \alpha^{d+1}$. Note that if the eigenvectors are normalized in $\mathcal{H}$, then dimension i will be scaled by $\lambda_i^{-1/2}$, $i = 1, \ldots, d$.

   (d) Prove that the pseudo-inverse of M is positive semidefinite.

   (e) Prove that performing kernel PCA on $M^\dagger$ (pseudo-inverse of M)is equivalent to LLE up to scaling factors.