Data Visualization
STAT 890 / 442, CM 462

Assignment 1
Fall 2006
Department of Statistics and Actuarial Science
University of Waterloo

Due: Friday October 13, 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. **Image denoising** In PCA, a set of observations X can be reconstructed by $U_d U_d^T x$, where U_d is a matrix consisting of the top d eigenvectors of the covariance matrix of X. If $x \in R^D$ is a noisy observation generated from a low-dimensional structure f whose dimensionality is $d < D$, then intuitively the first d eigenvectors should contain most of the information about f and the remaining eigenvectors should just contain noise. Therefore, PCA can be used as a data denoising technique. Suppose the observed data are noisy. If one finds the correct intrinsic dimensionality of the data d and reconstructs the data by using only the first d significant eigenvectors, one should expect to be able to filter out the noise while still capturing the important information in the data. As an experiment, we use the data set "*noisy.mat*" available at the course web site. The data set consists of 1965 20-pixel-by-28-pixel grey-scale images distorted by adding Gaussian noises to each pixel with s=25.

   a) Apply PCA to the noisy data. Suppose the intrinsic dimensionality of the data is 10. Compute reconstructed images using the top $d = 10$ eigenvectors and plot five original and reconstructed images (select five images randomly).
   If original images are stored in matrix X (it is 560 by 1965 matrix) and reconstructed images are in matrix $\hat{X}$, you can type in *colormap gray* and then $imagesc(reshape(X(:, 10), 20, 28)'$
   $imagesc(reshape(\hat{X}(:, 10), 20, 28)'$ to plot the 10^{th} original image and its reconstruction.

   b) Repeat part *a* with $d = 2$ and $d = 30$.

c) The eigenvectors serve as basis images for the reconstruction. Plot the first 30 eigenvectors. Compare the first 10 and the last 20 basis images (the first 10 and the last 20 eigenvectors) and comment on your observation.