CSL361 Programming assignment 3:

October 9, 2009

Latent semantic indexing:

1. Study the articles on *Latent Semantic Indexing* available at http://poorvi.cse.iitd.ernet.in/ suban/csl361/LSI.
2. Make a list of common CSE and Mathematics text books that you may have studied in your core and elective courses and a glossary of standard CSE terms to generate a *term-document* matrix similar to the examples in the articles.
3. Implement *LSI* in Matlab using both *SVD* and *QR* and demonstrate information retrieval. Verify whether *SVD* gives better clustering and noise reduction though the matrix approximation using *QR* "looks more similar" to the original matrix. Also try to generate two dimensional plots (different cross-sections) of the *terms* and *doc-*

uments with the queries to verify whether you indeed obtain meaningful clusterings. Be warned that you may have to tweak the weights a bit.

4. Generate a report explaining why the scheme works (if it works at all).

Principal component analysis:

1. Study
   (a) The PCA based face recognition paper by Turk and Pentland. A local copy is available at http://schnapps.cse.iitd.ernet.in/resources/papers/faces/face-recognition/mturk-CVPR91.pdf
   (b) The Face recognition Demo page at http://vismod.media.mit.edu/vismod/demos/facerec/
2. Show that in the projected space, the eigenfaces represent the *principal components*, i.e., the directions corresponding to maximum variances in the training data.
3. Obtain a database of "faces" of people in the class. Obtain about 10 images of each person with normalized lighting and camera settings. Please mount your camera on a tripod and make people stand so that the faces are registered in a common frame. Crop the faces interactively using any

stand software (such as http://www.gimp.org). Use 5 images of each person for building the PCA database, and the remaining for testing. You may borrow a tripod and a camera from the Vision laboratory.

4. Develop a piece of code (Matlab) for face recognition and experiment. In particular, experiment with various low-rank approximations (5,10,20,30,...) of the covariance matrix and find out where do you obtain the best results. Give reasons for what you observe. Display the average face and the eigenfaces. Also try to generate some interesting new faces by taking different linear combinations of the eigenfaces (e.g. Ankit+Prince/2).
5. Write a report elaborating your findings.