Introduction to Data Mining: Assignment #4
Summer 2014

Due: June 27th, 23:59:59 CST (UTC +8).

1.   Principal Component Analysis

Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Let us deepen our understanding of PCA (you should implement it in *pca.m*) by the following problems.

(a) Our actions to hack the CAPTCHA system last time have been detected, so they updated the system to add rotation in the CAPTCHA images:

Luckily, we have just learnt PCA. Your task is to implement *hack_pca.m* to recover the rotated CAPTCHA image using PCA.

Some rotated CAPTCHA samples are provided for you. It's OK for your algorithm to return grayscale image or image with size different form input image.

(b) Now let us apply PCA to a face image dataset in *ORL_data.mat*.

  (i) Run PCA on the dataset, visualize the learnt eigenvectors using *show_face.m*. You should see faces in the image, these faces are called Eigenface.

  (ii) Use PCA to do dimensionality reduction[1]. You should experiment with number of reduced dimensionality being 8, 16, 32, 64 and 128 respectively. For each test, run KNN on the dimensionality reduced dataset and report the testing error rate.

  (iii) Will dimensionality reduction cause loss of information? Let us see it visually. Do dimensionality reduction using PCA then recover the original image form the low dimensional feature vectors. Visualize the original and recovered image using *show_face.m*. Again, you should experiment with number of reduced dimensionality being 8, 16, 32, 64 and 128 respectively.

---

[1]You are strongly encouraged to try LDA on this part of problem. You should be impressed by its outstanding performance.

2.   Course Feedback

Please complete the following course feedback survey, your feedback will be highly helpful for us to improve this course in future offering.

(a) How do you think of the lectures? (Are they clear? Which parts do you like/dislike and why?)

(b) How do you think of the homework? (Is the amount of homework fair? Which assignments do you like/dislike and why? Do you have new ideas for homework problems?)

(c) Do you have any comments on the course materials? (Do you find any bugs in the slides? Are the slides clear enough? If not, which one? Do you need any supplementary readings on certain topics?)

(d) Do you have any suggestions on the course website?

(e) How do you think of the lecturer?

(f) How do you think of the TA?

(g) Any other comments or suggestions regarding the course? Or any other messages that you'd like to pass on to either Prof. Cai or the TAs?

Please submit your homework report in **pdf** format, with all your code in a zip archive.