CS504 - Assignment 3

(dimensionality reduction with pca)

REWARD: 150 points

DUE: Mon Mar 20 at 5pm PST

TEST DATA: testDataA3.tar

the task

Consider data coming in as a matrix of R samples of C dimensional data that is represented as a R x C matrix. First read in a number k that is the number of eigenvectors you want to keep. Then read in the matrix in standard form of the number of rows, the number of columns, then the data. Then translate the data into Z-score by column, that is, convert the numbers in each column to the Z-score for in column. ◀ NEW

Write a C/C++ program called pca that uses principle component analysis to pick the top k eigenvalues and encodes the data in k dimensional space instead of C dimensions. Essentially you will be compressing the data by translating the data to the k most significant eigenvectors. Eigenvalues and eigenvectors can be computed in the matix library. Here is a help sheet on PCA.

Your output is

- Print out the eigenvectors as a matrix with eigenvectors in rows.
- Print out the eigenvalues as a row matrix.
- A matrix in the same format which is a reconstruction of the data in only k dimensions so it is an R x k matrix with the first line being the number of rows and columns.
- Then translate the compressed data back to R x C matrix being sure to undo the Z-score and print that. ◀ NEW
- Finally, print out the component matrix.

submission

Homework will be submitted as an **uncompressed** tar file to the homework submission page linked from the class web page. You can submit as many times as you like. **The LAST file you submit BEFORE the deadline will be the one graded.** For all submissions you will receive email at your uidaho.edu mail address giving you some automated feedback on the unpacking and compiling and running of code and possibly some other things that can be autotested.

Have fun.