Statistics 5525: Homework 4

For each homework assignment, turn in at the beginning of class on the indicated due date. Late assignments will only be accepted with special permission. Write each problem up *very* neatly (LaTeX is preferred). Show all of your work.

Problem 1

Read *Nonlinear Dimensionality Reduction by Locally Linear Embedding*, Sam T. Roweisand and Lawrence K. Saul, *Science* (2000).

Problem 2

Read *Probabilistic Principal Component Analysis*, Michael E. Tipping and Christopher M. Bishop (1999).

Problem 3

Obtain the "cereal" data set from the website. Let x_i be a vector of cereal measurements: *Calories, Protein, Fat, Sodium, Fiber, Carbo, Sugars, Shelf, Potass, Vitamins* for each of 22 cereals (i.e. $i = 1, \ldots, 22$).

Part a

Perform 2-D Classical MDS on the data set. That is, find lower dimensional coordinates ($z's$ in 2-D), such that the found $z's$ minimize the stress function:

$$\sqrt{\sum_{i<j}(||z_i - z_j||_2 - ||x_i - x_j||_2)^2}.$$

(Note: in Matlab, this is accomplished either via the cmdscale function, or the mdscale function, using Euclidean norms and the "strain" metric).

Part b

Perform a 2-D PCA projection of the data.

Part c

Verify that the *relative* distances between those found in the 2-D projections in parts a & b are the same. Make a conclusion.