M532 **Spring 2018**
Problem Set 5. Due Wednesday, April 25

Part A: Theory and Foundations

Problem 1

Show that the number of zero eigenvalues indicates the number of connected components in Laplacian eigenmaps problem.

Problem 2

Show that the eigenvalues in the Local Linear Embedding problem are all non-negative.

Problem 3

In this problem we explore the details of the LLE derivation were the dimension reduced data matrix is given by $Y = [y_1, \ldots, y_P] \in \mathbb{R}^{k \times P}$. You will derive, in a sequence of steps, the LLE eigenvector equation

$$MY^T = Y^T \Lambda$$

where $M = (I - W)^T (I - W)$. For a single point x the weight vector w^* is the solution to the minimization problem

$$E(w*) = \text{minimize}_w\, E(w)$$

subject to $\sum_j w_j = 1$, where

$$E(w) = \|x - \sum_{j \in N} w_j x_j\|^2.$$

a) Show that

$$E(w) = \sum_{jk} w_j w_k C_{jk}$$

where $C_{jk} = (x - x_j)^T (x - x_k)$.

b) Using the result in a) show that show that

$$w_j^* = \frac{\sum_k C_{jk}^{-1}}{\sum_{lm} C_{lm}^{-1}}$$

OR, alternatively, show that

$$Cw_j^* = e$$

where e is a vector of ones; note that you have to rescale this solution such that the weights sum to one. This latter formula is the one you will use in the computation of LLE.

c) Show that

$$M_{ij} = \delta_{ij} - W_{ij} - Wji + \sum_k W_{ki} W_{kj}$$

d) Now show that

$$\sum_i \|y_i - \sum_j W_{ij} y_j\|^2 = \sum_{i,j} M_{ij} y_i^T y_j$$

e) Using this result show

$$\sum_i \|y_i - \sum_j W_{ij} y_j\|^2 = trace(YMY^T)$$

f) Lastly, show that the eigenvector problem comes from the optimization problem

$$\text{minimize}\, trace(YMY^T)$$

subject to $YY^T = I$.

Part B: Computing

Problem 1

Write a code to implement the LLE algorithm and apply it to Kohonen's animal data set. Compare with your SOM and Laplacian Eigenmaps data reduction results.

Problem 2

Write a code to implement the SVM algorithm using the quadratic programming subroutine in Matlab. Test your code using both separable and non-separable data sets in the plane. Plot the solution including the data (identifying the classes clearly) and the three SVM hyperplanes.

Problem 3

Write a code to implement the Sparse SVM algorithm using the linear programming subroutine in Matlab. Test your code using both separable and non-separable data sets in the plane. Plot the solution including the same data as in problem 2 (again, identifying the classes clearly) and the three SSVM hyperplanes.

Problem 4

Using the two class hyperspectral data matrices X, Y (on canvas in the file HSIdata), build a classifier using the first 50% of each data set. Provide a confusion matrix for your predictions on the test data set consisting of the remaining 50% of the data.