Lecture 13 Oct. 18 -2006

Data Visualization

STAT 442 / 890, CM 462

Lecture: Ali Ghodsi Scribes: Stefan Pintilie

1 Landmark MDS cont...

Previously we saw that we could approximate the symmetric matrix K by knowing only the A and B parts.

$$K = \begin{pmatrix} A & B \\ B^T & B^T A^{-1} B \end{pmatrix}$$

To get the values of the A and B parts we need to first consider the set of data points $X = (x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_n)$. We can select the first m of these points and call them R. The rest of the points can be called S. The set of points can be rewritten as $X = (R, S)$. The expressions for A and B become:

$$A = R^T R$$

$$B = R^T S$$

Since A is positive semi-definite we can also write A as:

$$A = U \Gamma U^T$$

Therefore,

$$R = \Gamma^{1/2} U^T$$

And,

$$S = R^{-T} B = \Gamma^{-1/2} U^T B$$

Recall that in MDS

$$X = \Lambda^{1/2} V^T$$

and

$$Y_d = \Lambda_d^{1/2} V_d^T$$

In the Nyström approximation the estimate for Y is:

$Y = \Lambda_d^{1/2} U_d^T$ if i is less than or equal to m.

$Y = \Lambda_d^{-1/2} U_d^T B$ if i is greater than m.

The low dimensional map Y can be approximated only by the eigendecomposition of A.

The approximation is exact if the rank of K is m or less. That is if $rank(K) \leq m$. Which implies that $||B^T A^{-1} B - C||^2 = 0$. Otherwise, the quality of the approximation depends on the value of $||B^T A^{-1} B - C||^2$.

Problem: So far we have been working with the kernel matrix K. In landmark MDS we do not begin with K. We begin with the distance matrix D.

$$D = \begin{pmatrix} E & F \\ F^T & G \end{pmatrix}$$

In order to turn the Nyström Approximation into Landmark MDS we need to be able to find A and B based on E and F. Recall from MDS that $K = \frac{-1}{2}HDH$ where $H = I - \frac{1}{n}ee^T$. This is equivalent to:

$$K_{ij} = d_{ij}^2 - \sum_i c_i d_{ij}^2 - \sum_j c_j d_{ij}^2 + \sum_{i,j} c_i c_j d_{ij}^2 = \frac{-1}{2}HDH$$

such that $\sum_i c_i = 1$ and $\sum_j c_j = 1$.

In order to find A and B from E and F set $c_i = \frac{1}{m}$ if $i \leq m$ and zero otherwise. Then A and B can be estimated as follows:

$$A_{ij} = d_{ij}^2 - \frac{1}{m}\sum_i d_{ij}^2 - \frac{1}{m}\sum_j d_{ij}^2 + \frac{1}{m^2}\sum_{i,j} d_{ij}^2 \tag{1}$$

$$B_{ij} = F_{ij}^2 - \frac{1}{m}\sum_i F_{ij}^2 - \frac{1}{m}\sum_j F_{ij}^2 \tag{2}$$

The last double summation term for B_{ij} is dropped, because it introduces an irrelevant shift of origin.