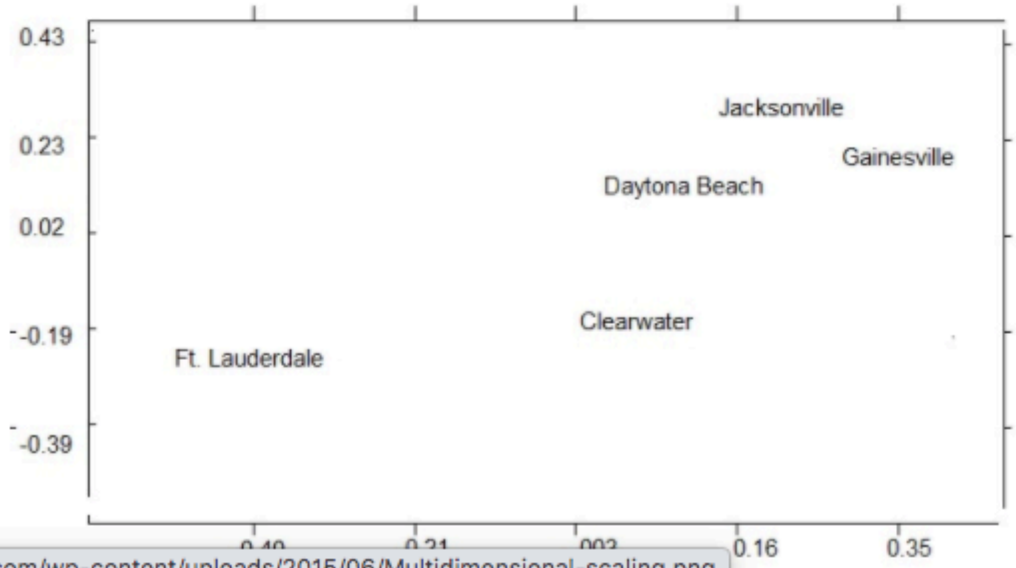
3.5 Multi-Dimensional Scaling (MDS)

3.5.1 Motivation

CITY	Clearwater	Daytona Beach	Ft. Lauderdale	Gainesville	Jacksonville
Clearwater	0	159	247	131	197
Daytona Beach	159	0	230	97	89
Ft. Lauderdale	247	230	0	309	317
Gainesville	131	97	309	0	68
Jacksonville	197	89	317	68	0

The scaling produces a graph like the one below.



Perception of Color in human vision

	445	465	472	490	504	537	555	584	600	610	628	651
445												
465	0.50											
472	0.56	0.19										
490	0.78	0.53	0.46									
504	0.91	0.83	0.75	0.39								
537	0.93	0.90	0.90	0.69	0.38							
555	0.93	0.92	0.91	0.74	0.55	0.27						
584	0.98	0.98	0.98	0.93	0.86	0.78	0.67					
600	0.96	0.99	0.99	0.98	0.92	0.86	0.81	0.42				
610	0.93	0.98	1.00	0.98	0.98	0.95	0.96	0.63	0.26			
628	0.89	0.99	0.99	0.99	0.98	0.98	0.97	0.73	0.50	0.24		
651	0.87	0.95	0.98	0.98	0.98	0.98	0.98	0.80	0.59	0.38	0.15	

MDS: a method to find a configuration of pts in d-dim space that preserves the pairwise dissimilarities.

Setting: Given n sample pts with pairwise distances, need to find "artificial features" so that the pts can be represented as $x^{(1)}, \dots, x^{(n)} \in \mathbb{R}^d$ with prescribed distances $\|x^{(i)} - x^{(j)}\|$.

Remark: As shifts of pts do not change pairwise distances, without loss of generality (WLOG), we assume $x^{(1)}, \dots, x^{(n)}$ are centered, i.e. $\mu \overset{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} x^{(i)} = 0$.

3.5.2 Theory.

Suppose we have found the features $x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \in \mathbb{R}^d$ so that the sample pts can be represented as $x^{(1)}, \ldots, x^{(n)}$ (and are centered). Form:

sample matrix: $$X = \begin{pmatrix} | & & | \\ x^{(1)} & \cdots & x^{(n)} \\ | & & | \end{pmatrix}$$

kernel matrix: $$K \overset{\text{def}}{=} X^T X = \begin{pmatrix} - & x^{(1)} & - \\ & \vdots & \\ - & x^{(n)} & - \end{pmatrix} \begin{pmatrix} | & & | \\ x^{(1)} & \cdots & x^{(n)} \\ | & & | \end{pmatrix}$$

<u>square distance matrix</u>: D with $D_{ij} \overset{\text{def}}{=} \| x^{(i)} - x^{(j)} \|^2$

Remark: (1) As $x^{(1)}, \ldots, x^{(n)}$ are centered,

$$X \cdot \mathbb{1} = \begin{pmatrix} | & & | \\ x^{(1)} & \cdots & x^{(n)} \\ | & & | \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} | & & | \\ x^{(1)} & + \cdots + & x^{(n)} \\ | & & | \end{pmatrix} = (n \cdot \mu) = 0$$

where $\mathbb{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$.

(2) K is sym, $K_{ij} = x^{(i)} \cdot x^{(j)}$,

and $K \mathbf{1} = X^T \underbrace{(X \mathbf{1})}_{=0} = 0$.

(3) D is sym, $D_{ii} = 0 \quad i = 1, \cdots, n$.

First, we investigate the relation between K and D.

Given K, we have:

$$D_{ij} \overset{\text{def of of } D}{=} \| x^{(i)} - x^{(j)} \|^2 = \left(x^{(i)} - x^{(j)}\right)^T \left(x^{(i)} - x^{(j)}\right)$$

$$\overset{\text{matrix multi.}}{=} x^{(i)T} x^{(i)} - x^{(i)T} x^{(j)} - x^{(j)T} x^{(i)} + x^{(j)T} x^{(j)}$$

$$= \underbrace{x^{(i)T} x^{(i)}}_{= x^{(i)} \cdot x^{(i)}} - 2 \underbrace{x^{(i)T} x^{(j)}}_{= x^{(i)} \cdot x^{(j)}} + \underbrace{x^{(j)T} x^{(j)}}_{= x^{(j)} \cdot x^{(j)}}$$

$$\overset{\text{def of } K}{=} K_{ii} - 2K_{ij} + K_{jj}.$$