Multidimensional Scaling

Charlie Edelson

10/3/2017

Overview

- Classical Problem
  - Canonical Problem – Principle Coordinate Analysis
- Relationship to Principle Component Analysis
- General Problem
  - Metric vs Non-metric

Background and Motivation

Encountered over the summer

- Heavily used in genomics research
- Dimensional reduction technique

Classical Multidimensional Scaling

A Tale of Three Cities

Imagine we have three cities: A, B, and C

We wish to make a map of our cities, *but* we only know *distances* and not *locations*.

How do we place them on the map so that we preserve these distances?

Distance Matrix

We can introduce the distance matrix D, where, given a measure of distance d,

$$D = \{d_{ij}\} = \{d(\vec{r}_i, \vec{r}_j)\}. \tag{1}$$

For our city example, where $\vec{r}_i = (x_i, y_i)^T$, we'll use

$$\begin{aligned} d(\vec{r}_B, \vec{r}_A) &= \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2} \quad (2) \\ &= \|\vec{r}_B - \vec{r}_A\| \quad (3) \end{aligned}$$

Example - City Distances

For our cities $d_{ab} = d(\vec{r}_A, \vec{r}_B)$, so

$$D = \begin{bmatrix} d(\vec{r}_A, \vec{r}_A) & d(\vec{r}_A, \vec{r}_B) & d(\vec{r}_A, \vec{r}_C) \\ d(\vec{r}_B, \vec{r}_A) & d(\vec{r}_B, \vec{r}_B) & d(\vec{r}_B, \vec{r}_C) \\ d(\vec{r}_C, \vec{r}_A) & d(\vec{r}_C, \vec{r}_B) & d(\vec{r}_C, \vec{r}_C) \end{bmatrix}. \quad (4)$$

If $\vec{r}_A = (1,1)^T$, $\vec{r}_B = (1,2)^T$, $\vec{r}_C = (2,1)^T$, then

$$D = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & \sqrt{2} \\ 1 & \sqrt{2} & 0 \end{bmatrix}. \quad (5)$$

Rephrase the Problem

Instead of looking for the original vectors $\{\vec{r}_i\}_{i=1}^{n}$, we instead wish to find the set of vectors $\{\vec{z}_i\}_{i=1}^{n}$ s.t. they minimize

$$Stress_D = \sum_{i,j} \left(d_{ij} - d(\vec{z}_i, \vec{z}_j)\right)^2 . \tag{6}$$

In other words, find a set of vectors which have the same distances as the original vectors.

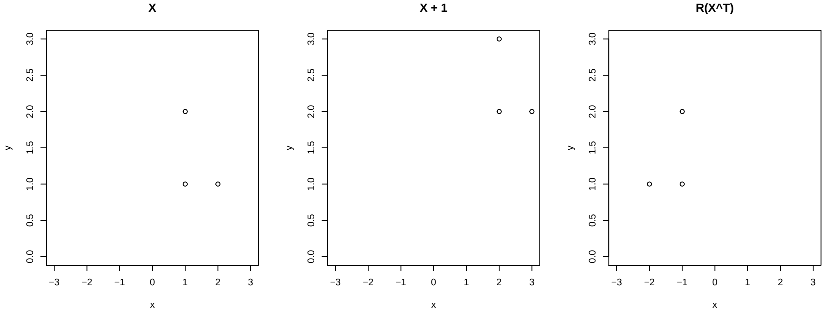
Difficulties Arise

Depending on the definition of $d(\vec{z}_i, \vec{z}_j)$, Eq.6 can difficult to solve analytically.

Additionally. . .

Unique Selections

...translations and rotations do not affect the distances!



Multiple solutions for a single distance matrix.

Altered Problem

Consider instead the distance matrix B where

$$B = \{b_{ij}\} = \{\langle \vec{x}_i - \bar{x}, \vec{x}_j - \bar{x} \rangle\}. \tag{7}$$

B is a matrix of mean centered inner products of $\vec{x}_i$.

Alternatively, for $X \in \mathbb{R}^{n \times p}$, where $X_{i.} = \vec{x}_i - \bar{x}$,

$$B = XX^T \tag{8}$$

Altered Problem Cont.

Lets review our goal – minimize

$$\sum_{i,j} \left(d_{ij} - d(\vec{z}_i, \vec{z}_j)\right)^2 . \tag{9}$$

For B

$$Stress = \sum_{i,j} \left(b_{ij} - d(\vec{z}_i, \vec{z}_j)\right)^2 \tag{10}$$

$$= \sum_{i,j} \left(\langle \vec{x}_i - \bar{x}, \vec{x}_j - \bar{x} \rangle - d(\vec{z}_i, \vec{z}_j)\right)^2 \tag{11}$$

Simple Minimum

It becomes evident that given

$$d(\vec{z}_i, \vec{z}_j) = \langle \vec{z}_i, \vec{z}_j \rangle \tag{12}$$

the minimum of

$$Strain_B = \sum_{i,j} (\langle \vec{x}_i - \bar{x}, \vec{x}_j - \bar{x} \rangle - d(\vec{z}_i, \vec{z}_j))^2 \tag{13}$$

occurs when $\{\vec{z}_i\}_{i=1}^n = \{\vec{x}_i - \bar{x}\}_{i=1}^n$.

Decompose Our Solution

For $B = XX^T$ we know finding X directly is a solution; since B is symmetric and semi-definite

$$B = E\Lambda E^{-1} \tag{14}$$

$$XX^T = E\Lambda^{1/2}\Lambda^{1/2}E^{-1}, \tag{15}$$

E_m is the matrix of eigenvectors of B and Λ is the diagonal matrix of eigenvalues.

Considering Eq. 15, we note $X = E\Lambda^{1/2}$.

Our Current Solution

We now know how to find a solution for B given $d(\vec{z}_i, \vec{z}_j) = \langle \vec{z}_i, \vec{z}_j \rangle$.

However, the original problem was $D = \{\|\vec{r}_i - \vec{r}_j\|\}$

We now ask if there exists a relationship between B and D.

Law of Cosines

Recall

$$\|\vec{r}_i - \vec{r}_j\|^2 = \|\vec{r}_i - \bar{r}\|^2 + \|\vec{r}_j - \bar{r}\|^2 - 2\langle \vec{r}_i - \bar{r}, \vec{r}_j - \bar{r} \rangle \tag{16}$$

Rearranging we find

$$\langle \vec{r}_i - \bar{r}, \vec{r}_j - \bar{r} \rangle = -\frac{\|\vec{r}_i - \vec{r}_j\|^2 - \|\vec{r}_i - \bar{r}\|^2 - \|\vec{r}_j - \bar{r}\|^2}{2}. \tag{17}$$

Perform this transform on D?

Centering Matrix

C_n is the centering matrix defined by

$$C_n = I_n - \frac{1}{n} 1_n 1_n^T \tag{18}$$

where 1_n is the column vectors of n 1s.

For a vector v

$$C_n v = I_n v - \frac{1}{n} 1_n v 1_n^T \tag{19}$$
$$= v - \bar{v} \tag{20}$$

Solution to Classical MDS

Now we can relate B to D by

$$B = -\frac{1}{2} C_n D^2 C_n. \tag{21}$$

So the $\vec{z}_i$ that minimize

$$Stress_B = \sum_{ij} (b_{ij} - \langle \vec{z}_i, \vec{z}_j \rangle)^2 \tag{22}$$

are given by finding $X = E\Lambda^{1/2}$.

Pinciple Coordiante Analysis and Pinciple Component Analysis

Principle Coordinate Analysis

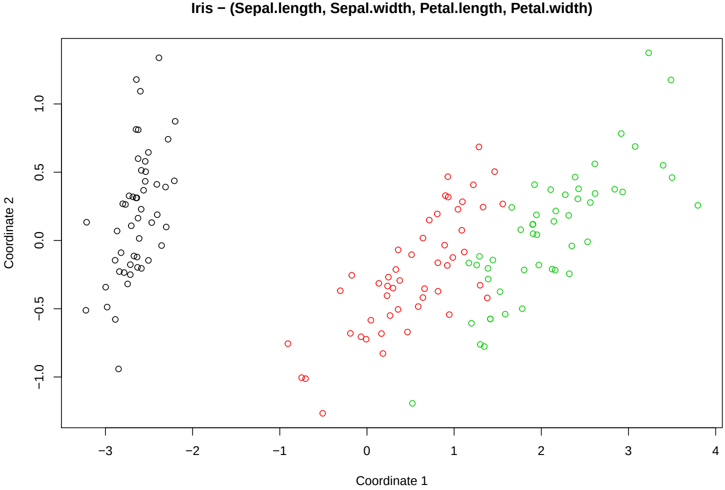
- Solution to Classical Problem
- Visualize high dimensional data along “principle coordinates"

Relationship to Principle Component Analysis

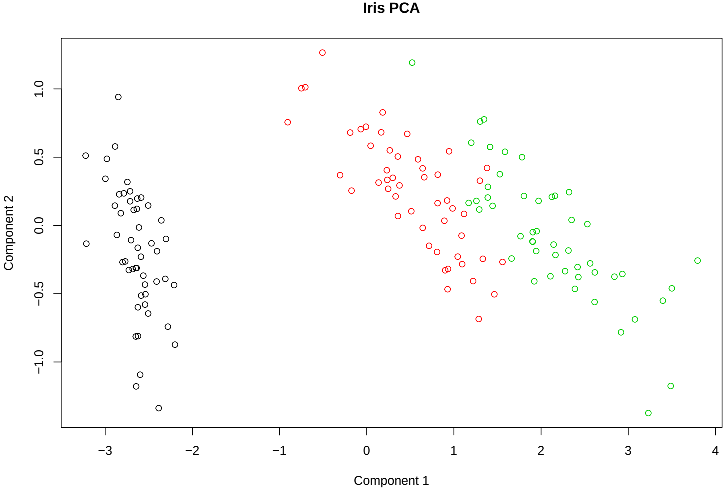
If $X \in \mathbb{R}^{n \times p}$, consider $\{E_m \Lambda_m : m < n\}$

- $XX^T \approx cov(X^T, X^T)$
- $E_m \Lambda_m$ are m largest terms
- Principle Component Analysis of X^T
- Maximizes dimensional variance along the first m components

Example – Iris Data PCoA



Example – Iris PCA



General Multidimensional Scaling

General Multidimensional Scaling

The general multidimensional scaling problem concerns itself with minimizing the $Stress_D$ for various different $d(\vec{x}_i, \vec{x}_j)$

General multidimensional scaling problems fall into two categories

- Metric
- Non-metric

Metric Multidimensional Scaling

Metric scaling has a stress function with an explicit dependence on the distance measure. In other words, it can be written as

$$Stress_D = \sum_{i,j}(d_{ij} - d(\vec{z}_i, \vec{z}_j))^2 \tag{23}$$

Non-Metric

Assumes distances may not be exact and that only order matters.

$$Stress_D = \sum_{i,j} (d_{ij} - f(d(\vec{z}_i, \vec{z}_j)))^2 \tag{24}$$

More general; can be fitted to many different dissimilarity measures.

Summary

- Multidimensional Scaling solves to problem of finding positions given distances
- Analytic solutions exist for classical multidimensional scaling
  - Equivalent to Principle Component Analysis
- A good tool for dimensional reduction
- Non-metric Multidimensional Scaling allows extremely general models