Multivariate Analysis (Slides 11)

- In today's class we consider multidimensional scaling.

- First we examine the methodology before applying it to some examples.

Multidimensional Scaling

- Multidimensional Scaling (MDS) is used in problems of the following form:

  - For a set of dissimilarities between every pair of n items, find a representation of the items in $\mathbb{R}^d$ ($d < n$) such that the inter-point distances match the original dissimilarities as closely as possible.

- Hence MDS seeks to produce a lower dimensional representation of the data such that distances between points i and j in the representation, δ_{ij}, are close to the dissimilarities between these points, d_{ij}, for all i, j.

- MDS thus essentially produces a 'map' of the observations.

- To reproduce the dissimilarities between n observations as closely as possible, up to $(n-1)$ dimensions may be required. The objective of MDS is to produce an 'optimal' configuration of the observations in a smaller number of dimensions.

Distances Between European Cities

- Road distances (in km) between 21 European cities:

	Athens	Barcelona	Brussels	Calais	Cherbourg	...
Barcelona	3313					
Brussels	2963	1318				
Calais	3175	1326	204			
Cherbourg	3339	1294	583	460		
⋮	⋮	⋮	⋮	⋮	⋮	⋮

- The objective is to use multidimensional scaling to take this set of dissimilarities and return a set of two dimensional points such that the distances between the points are approximately equal to the dissimilarities.

Multidimensional Scaling methods

- Different approaches to the idea of 'matching' $\{d_{ij}\}$ and $\{\delta_{ij}\}$ give rise to two different techniques, both of which aim to find a representative set of points such that:

$$\delta_{ij} = f(d_{ij})$$

  – **Metric MDS** refers to when f is a continuous and monotonic function, *e.g.*, the identity function or a function converting dissimilarities into a distance like form.

  – **Non-metric MDS** only makes use of the rank order of the dissimilarities. As such, the transformation f need only obey the monotonicity constraint:

$$d_{ij} < d_{kl} \Rightarrow f(d_{ij}) \leq f(d_{kl})$$

  Such an f need only preserve rank order.

Multidimensional Scaling methods

- We consider two possible Metric MDS methods:

  1. Classical metric scaling

  2. Metric least squares scaling.

- We also consider 'Kruskal's non-metric scaling'.

- It is not always true that different MDS techniques will produce the same results for the same data set, so we will consider **Procrustes analysis**, which is a technique for matching one configuration with another and for producing a measure of the match.

Classical Metric Scaling

- Classical metric scaling produces a configuration of points from the dissimilarities via an algebraic reconstruction method.

- Assume that the (continuous monotonic) function of dissimilarities is the identity function. The objective is to find the coordinate location of each observation, *i.e.*,

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1d} \\ x_{21} & x_{22} & \ldots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nd} \end{pmatrix}$$

- Here d is the dimension of the configuration (generally we want $d = 2$ or 3).

- Since such a configuration of coordinates is invariant to rotations, reflections, and translations, we fix the centroid to be at the origin.

6

Classical metric scaling

- For ease of interpretation assume observation i is located at the origin and that there are two further observations j and k. The distances between them are given by d_{ij}, d_{ik}, and d_{jk} (these are available from the dissimilarity matrix).

$$d_{jk}^2 = d_{ij}^2 + d_{ik}^2 - 2d_{ij}d_{ik}\cos(\theta_{jik})$$

$$\Rightarrow -1/2(d_{jk}^2 - d_{ij}^2 - d_{ik}^2) = d_{ij}d_{ik}\cos(\theta_{jik})$$

- Here θ_{jik} is the angle between d_{ij} and d_{ik}.

- Since $||\mathbf{x}_j|| = d_{ij}$ and $||\mathbf{x}_k|| = d_{ik}$ by definition, and because:

$$\mathbf{x}_j^T \mathbf{x}_k = ||\mathbf{x}_j||||\mathbf{x}_k||\cos(\theta_{jik})$$

Then

$$-1/2(d_{jk}^2 - d_{ij}^2 - d_{ik}^2) = \mathbf{x}_j^T \mathbf{x}_k$$

- Note that the vectors $\mathbf{x}_j$ are what we are seeking.

How do we obtain the $\mathbf{x}_j$s?

- Let $\mathbf{B}$ denote a matrix with elements $b_{jk} = -1/2(d_{jk}^2 - d_{ij}^2 - d_{ik}^2)$.

- From the previous slide we note that:

$$\mathbf{B} = \mathbf{X}\mathbf{X}^T$$

- As $\mathbf{B}$ is symmetric it can be decomposed into:

$$\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$$

  with $\mathbf{V}$ a matrix of eigenvectors and $\mathbf{\Lambda}$ a diagonal matrix of eigenvalues.

- Proof omitted but this is what was used in PCA.

- Hence:

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

Dimensionality

- An important issue with any MDS technique is the number of dimensions required to represent the configuration of points.

- We could express the (squared euclidean) distance between two points $\mathbf{x}_i$ and $\mathbf{x}_j$ in an $(n-1)$ dimensional space as $\delta_{ij} = \sum_{k=1}^{n-1} \lambda_k (v_{ik} - v_{jk})^2$.

- Hence any 'small' eigenvalues contribute little to the squared distance, and so the d eigenvectors associated with the d 'large' eigenvalues can be used to form the space representing the configuration of points.

- Note this is $n-1$ and not n because a dissimilarity matrix is not of full rank (we could always calculate the final row or column from the remainder of the matrix).

Dimensionality

- The general aim of MDS is to represent the observations graphically with $d = 2$ or $d = 3$ dimensions.

- To choose the appropriate d we could consider something like $\sum_{k=1}^{d} \lambda_k / \sum_{k=1}^{n-1} \lambda_k$, which is a measure of the proportion of variation explained by using only d dimensions.

Pseudocode:

1. Obtain the dissimilarities $\{d_{ij}\}$.

2. Form $\mathbf{B}$, each element of which is given by $b_{jk} = -1/2(d_{jk}^2 - d_{ij}^2 - d_{ik}^2)$, with i representing the centroid/origin of all the observations.

3. Create matrix $\mathbf{\Lambda}$ from the eigenvalues $\lambda_1, \ldots, \lambda_{n-1}$ and the matrix $\mathbf{V}$ from the associated eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ of $\mathbf{B}$.

4. Choose an appropriate number of dimensions d using a suitable measure.

5. The coordinates of the n required points that are used to represent the n observations in d-dimensional space are given by

$$x_{ij} = \lambda_j^{1/2} v_{ij}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, d$.

# Multidimensional Scaling Results

- Recall the data concerning chromosome dissimilarity between different animal species.

- A subset of this is:

	Man	Monkey	Horse	Pig	Pigeon	Tuna	Mould	Fungus
Man	0	1	17	13	16	31	63	66
Monkey	1	0	16	12	15	32	62	65
Horse	17	16	0	5	16	27	64	68
Pig	13	12	5	0	13	25	64	67
Pigeon	16	15	16	13	0	27	59	66
Tuna	31	32	27	25	27	0	72	69
Mould	63	62	64	64	59	72	0	61
Fungus	66	65	68	67	66	69	61	0

Multidimensional Scaling Results

- The results of a multidimensional scaling in this case is:
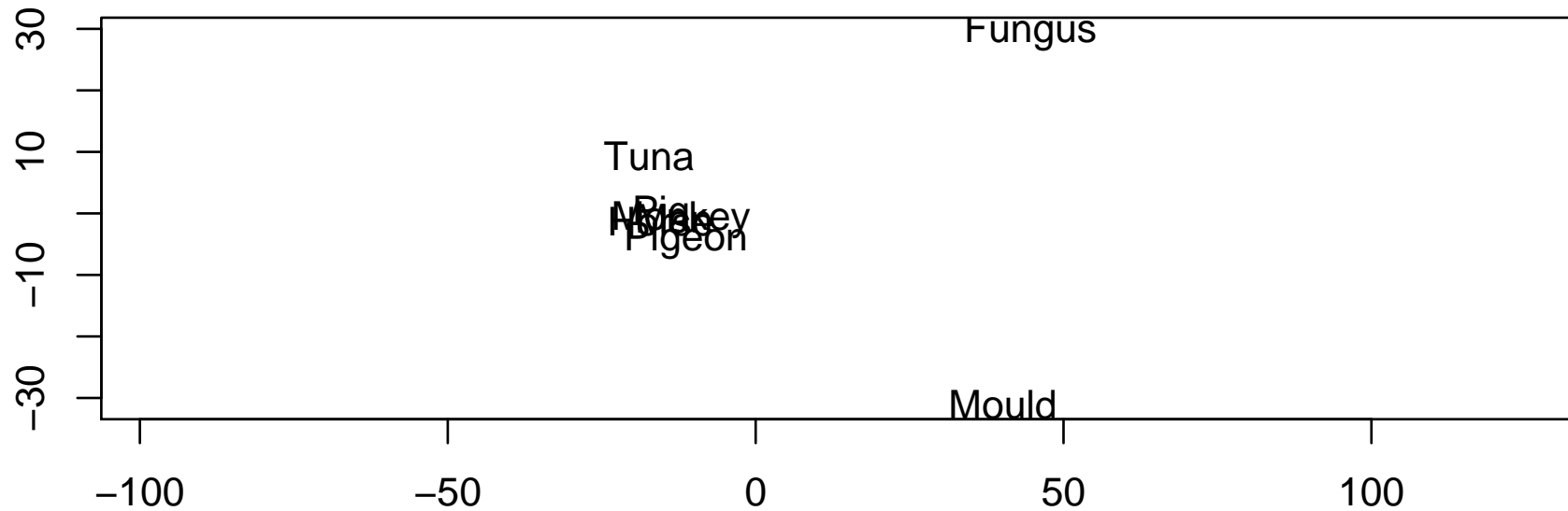
Invariance: Rotation

- If we rotate the points in the plot around some point, then the results are equally interpretable:
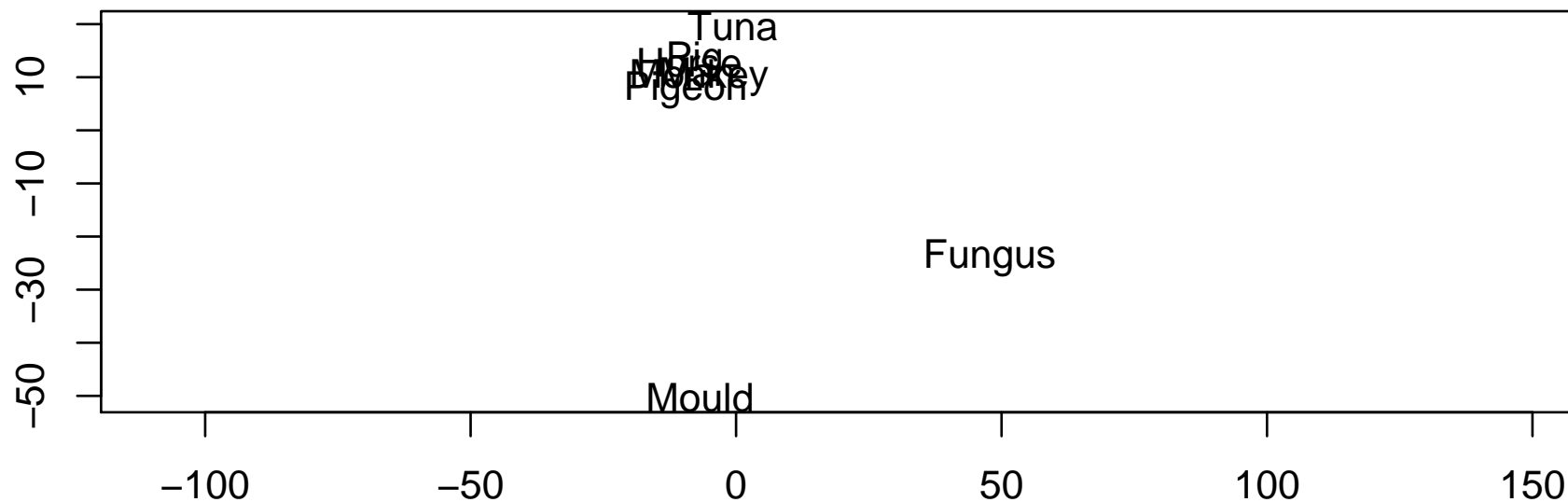
- The same is true when reflecting in a line.

How Well Did We Do?

- We can construct a distance matrix for the points in the MDS plot.

- The hope is that it should not be very different from the original dissimilarity matrix.

```
          Man Monkey  Horse     Pig Pigeon   Tuna Mould
Monkey  1.128
Horse   2.198  3.273
Pig     2.238  3.360  1.128
Pigeon  4.011  3.422  5.105  5.839
Tuna   10.933 11.575 10.748  9.647 14.925
Mould  61.338 60.273 63.021 63.535 57.990 70.203
Fungus 65.289 64.405 67.401 66.869 65.373 65.102  60.6
```

Stress

- The **stress** of a MDS is defined to be $\sum_{i=2}^{n} \sum_{j<i} (\delta_{ij} - d_{ij})^2$.

- Again δ_{ij} is the distance between i and j in the plot and d_{ij} is the distance between i and j in the dissimilarity matrix.

- For our data $\delta_{ij} - d_{ij}$ was found to be:

```
            Man   Monkey    Horse      Pig  Pigeon     Tuna Mould
Monkey    0.128
Horse   -14.802 -12.727
Pig     -10.762  -8.640  -3.872
Pigeon  -11.989 -11.578 -10.895  -7.161
Tuna    -20.067 -20.425 -16.252 -15.353 -12.075
Mould    -1.662  -1.727  -0.979  -0.465  -1.010 -1.797
Fungus   -0.711  -0.595  -0.599  -0.131  -0.627 -3.898  -0.4
```

Sammon Stress

- The Sammon measure of stress takes into account the size of the distances being approximated.

$$\text{Stress}_{\text{Sammon}}(\mathbf{d}, \delta) = \sum_{i=1}^{n} \sum_{j \neq i} d_{ij}^{-1}(d_{ij} - \delta_{ij})^2 / \sum_{i=1}^{n} \sum_{j \neq i} d_{ij}$$

- Now small dissimilarities have more weight in the loss function than large ones, providing motivation for them to be reproduced more accurately.

```
         Man Monkey  Horse    Pig Pigeon    Tuna Mould
Monkey −0.010
Horse   0.143 −0.230
Pig     0.818  0.392  0.258
Pigeon  1.403  1.413  2.673  0.926
Tuna    4.088  4.683  2.740  1.815 −8.082
Mould   6.697  5.825 −4.811 −1.563  2.600 −10.602
Fungus  7.732  5.890 −6.594 −2.868 −4.942   0.235 0.323
```

Metric least squares scaling.

- Metric least squares scaling finds a configuration by matching $\{d_{ij}\}$ to $\{\delta_{ij}\}$ by minimizing a loss function S, which possibly involves a continuous monotonic transform of the dissimilarities.

- An iterative numerical approach (method of steepest descent) is taken to minimize Sammon's loss function.

- Classical MDS is the situation in which a Euclidean distance model is chosen and the objective is to minimize the Stress value.

- Alternative metric MDS options specifically are designed to incorporate a different distance method, or to minimize an altered loss function such as Sammon stress.

Non-metric Multidimensional Scaling

- In non-metric MDS only the rank order of the dissimilarities must be preserved by the transformation f of dissimilarities.

$$\text{Stress}^2_{\text{Kruskal}} = \sum_{i=1}^{n} \sum_{j \neq i} [f(d_{ij}) - \delta_{ij}]^2 \Big/ \sum_{i=1}^{n} \sum_{j \neq i} \delta_{ij}^2$$

- Kruskal's approach chooses a d-dimensional configuration to minimize the 'stress'. However, the input distances are allowed to be subject to a non-monotonic transformation.

- Similar to Sammon's method an iterative algorithm is used to minimize the stress (again a steepest descent method is used).

- As for the dimension d, Kruskal proposed experimenting with several values of d and plotting the resulting stress against d and choosing the d value where a 'statistical elbow' appears in the plot (the stress should always decrease as dimensionality increases).
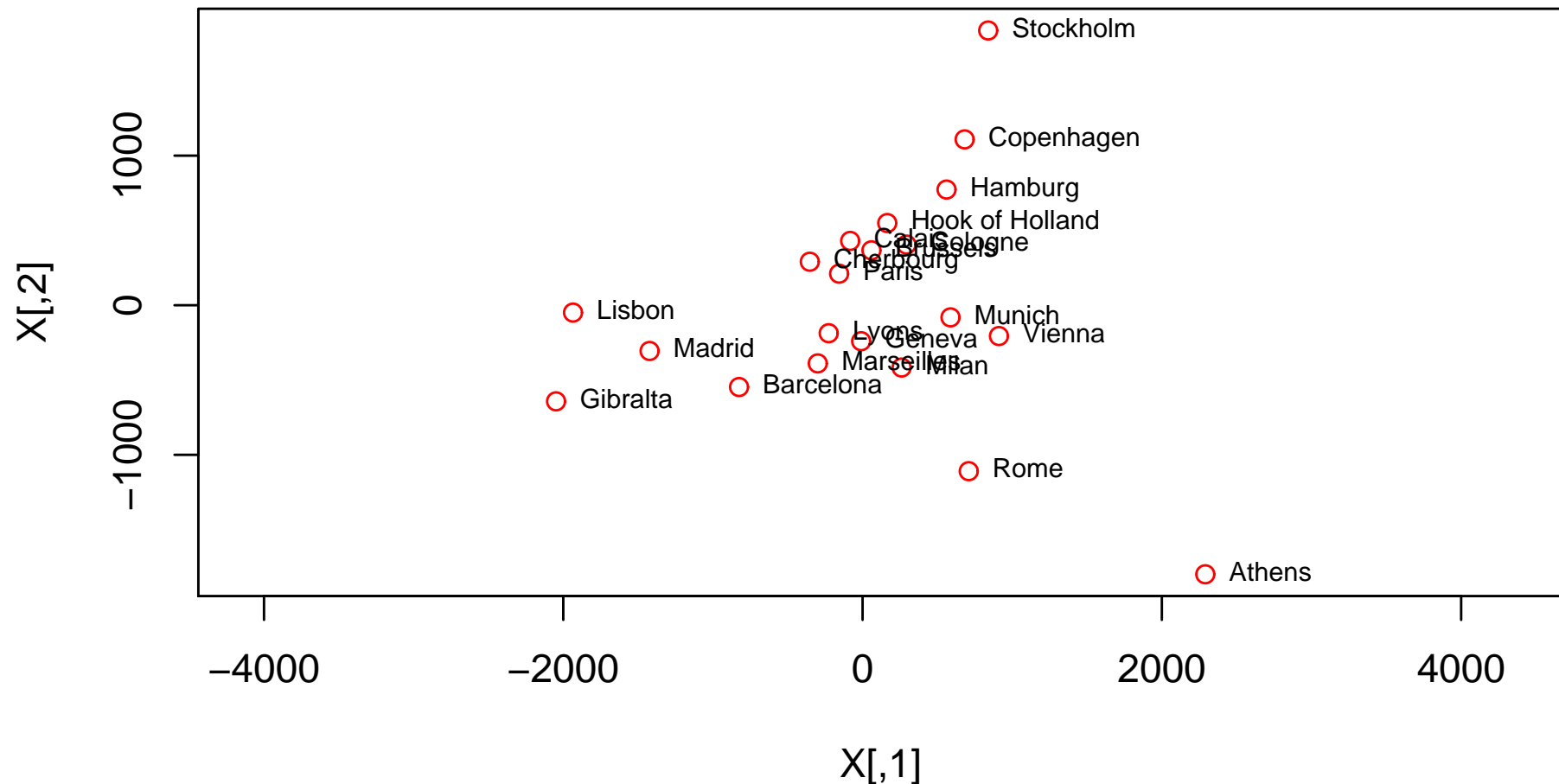
Iris Data

- A multidimensional scaling of the iris data (Euclidean distance) follows:
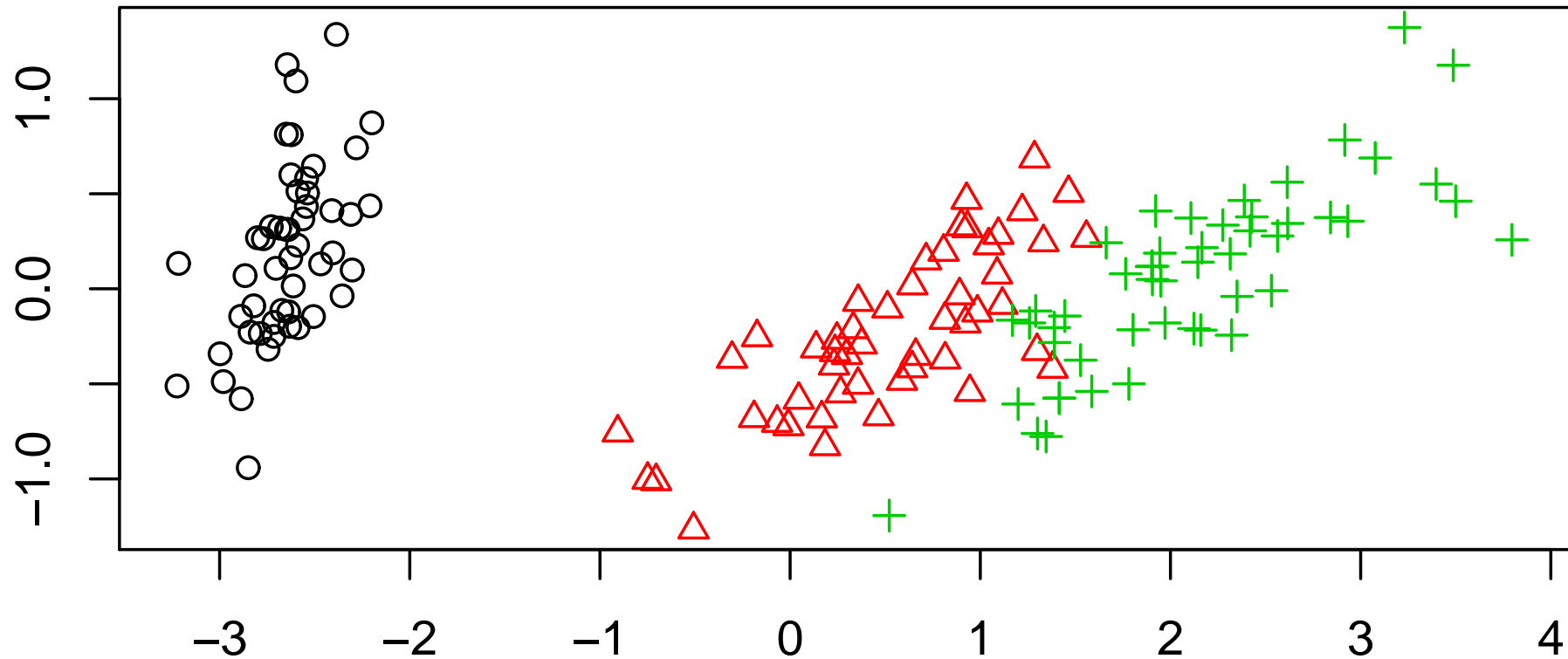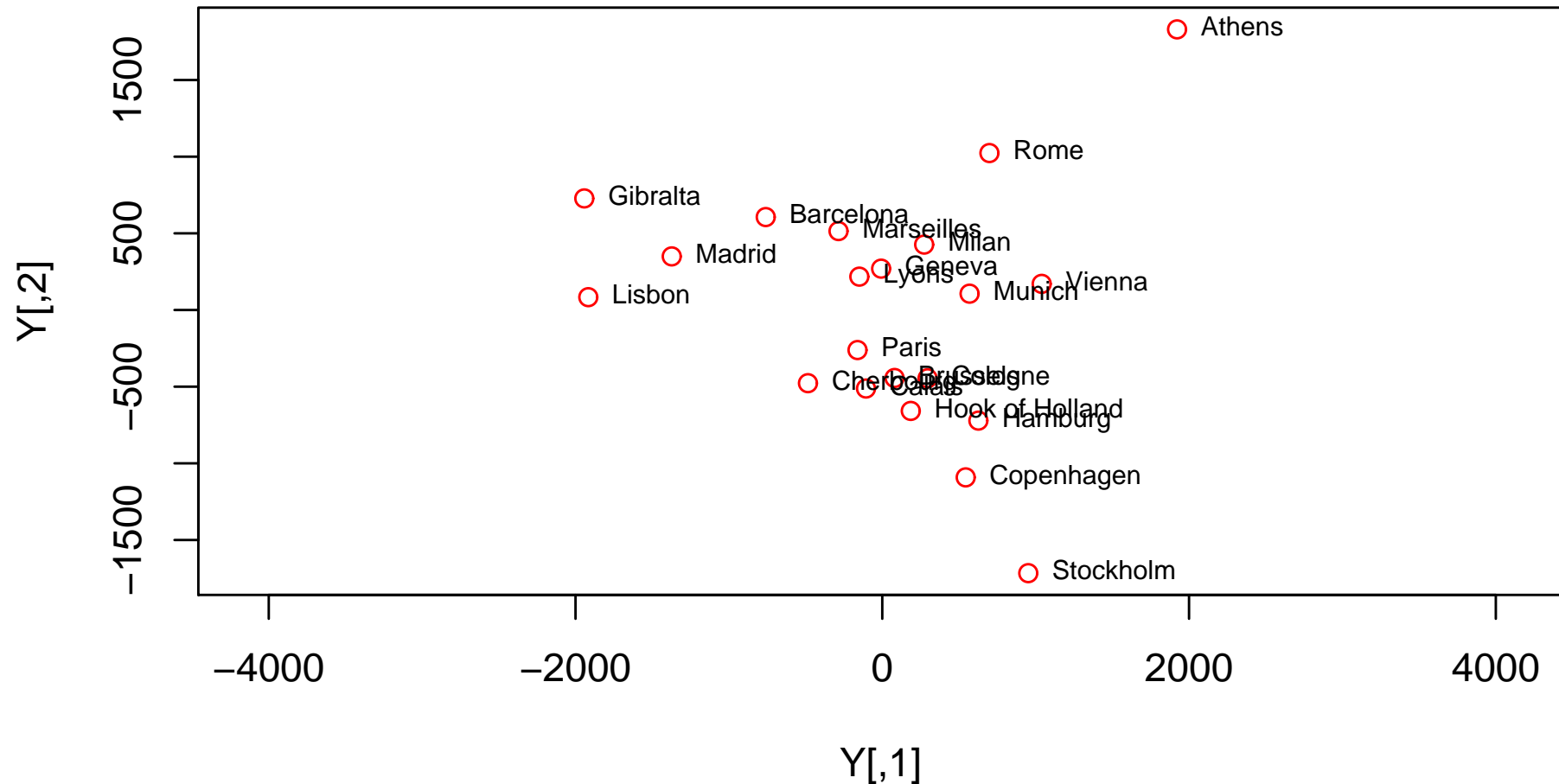
Road Distances: MDS Results

- The results of a multidimensional scaling of the road distance data are shown (the ○ shows the estimated location of the points):

Road Distances: MDS (Sammon) Results

- Using Sammon stress:

- (Flip the plot over and it looks more familiar)

Procrustes Analysis

- Procrustes analysis matches one MDS configuration with another by dilation, rotation, reflection and translation.

- Say two MDS methods have been applied to a set of n points, resulting in coordinate matrices $\mathbf{X}$ and $\mathbf{Y}$, respectively.

- There is a one-to-one mapping from the ith point in $\mathbf{X}$ to the ith point in $\mathbf{Y}$.

- The sum of squared distances between corresponding points in the two configurations is:

$$R^2 = \sum_{i=1}^{n} \sum_{j=1}^{d} (y_{ij} - x_{ij})^2$$

- To match the configurations, one of them is kept constant (the **reference** configuration), whilst the other is transformed.

Procrustes Analysis

- Assume that $\mathbf{Y}$ is the reference configuration, whilst $\mathbf{X}$ is to be transformed to achieve the best match with $\mathbf{Y}$.

- The new coordinates of the points in the $\mathbf{X}$ space will be:

$$\mathbf{x}'_i = \rho \mathbf{A}^T \mathbf{x}_i + \mathbf{b}$$

  Where

$$
\begin{aligned}
\rho &= \quad \text{a dilation matrix} \\
\mathbf{A} &= \quad \text{orthogonal matrix causing a rotation and/or reflection} \\
b &= \quad \text{a translation factor.}
\end{aligned}
$$

- The new sum of squared distances between points is therefore:

$$R^2 = \sum_{i=1}^{n} (\mathbf{y}_i - \rho \mathbf{A}^T \mathbf{x}_i - \mathbf{b})^T (\mathbf{y}_i - \rho \mathbf{A}^T \mathbf{x}_i - \mathbf{b}). \tag{1}$$

- By minimizing this we can estimate the optimal ρ, $\mathbf{A}$ and $\mathbf{b}$.

Procrustes Analysis: Goodness of Fit

- The measure of the 'match' between the two configurations is the minimized value of R^2. This is known as the **Procrustes sum of squares**.

- We can also examine plots of pointwise residuals between the final transformed configuration and the 'reference' configuration.

Pseudocode:

1. Translate the configurations to the origin by subtracting the average vector for each configuration from its coordinate values.

2. Compute the optimal values for $\rho, \mathbf{A}$ and $\mathbf{b}$ and apply it to the non-reference configuration.

3. Calculate the Procrustes sum of squares (smaller values indicate likeness).

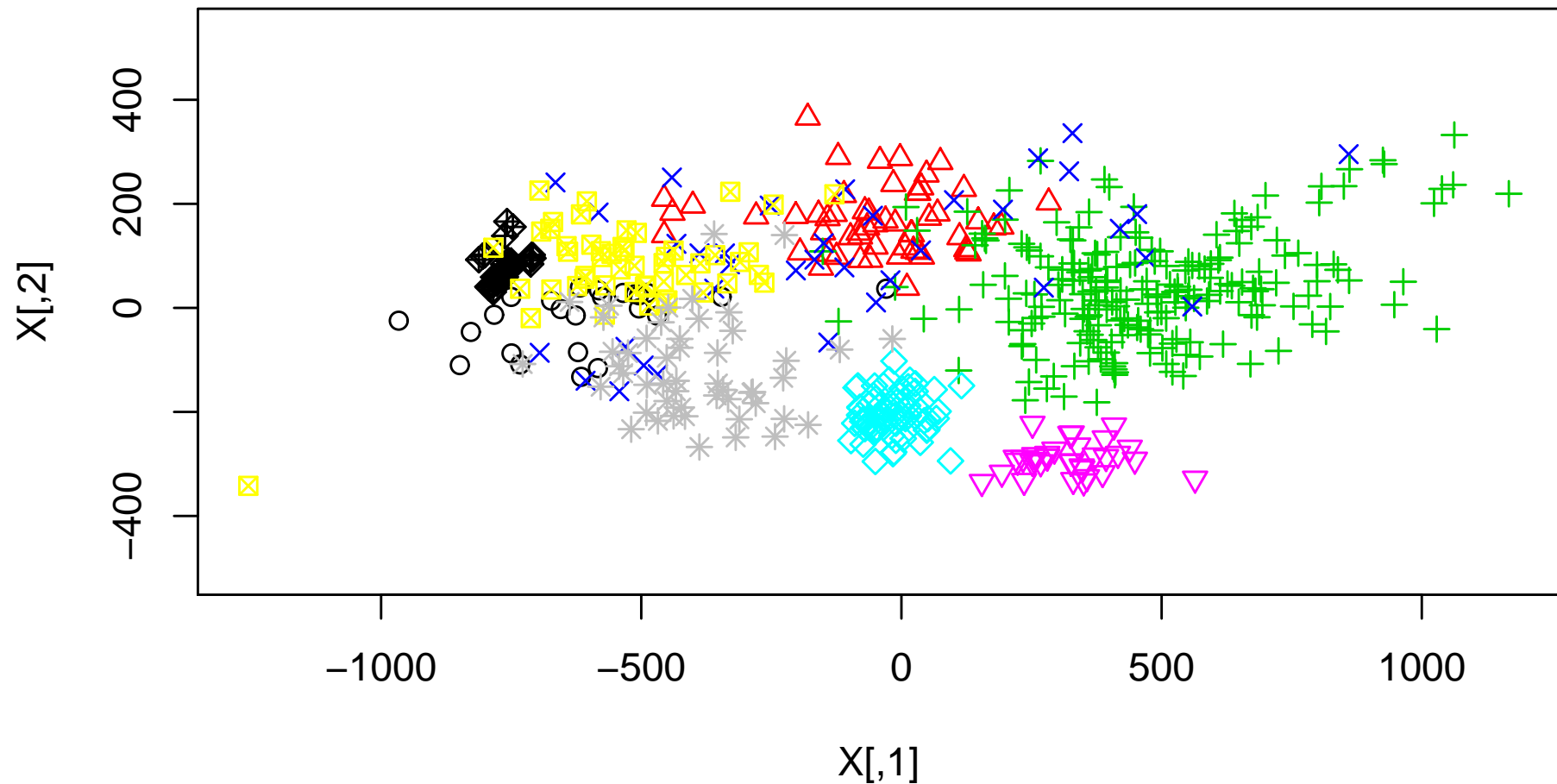# Procrustes Analysis: European cities

- Procrustes methods can be used to match the second data set to the first.

Procrustes errors
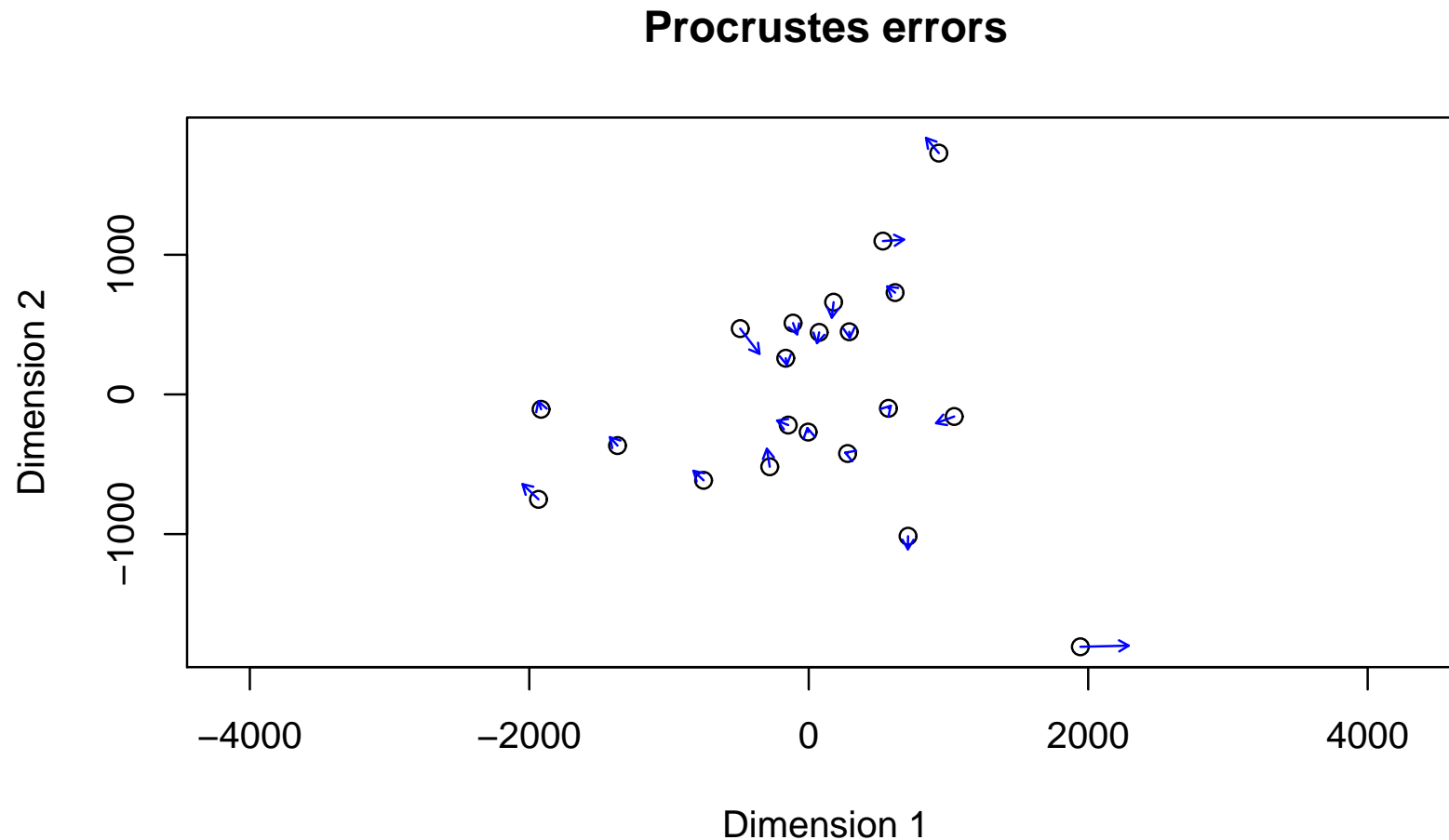
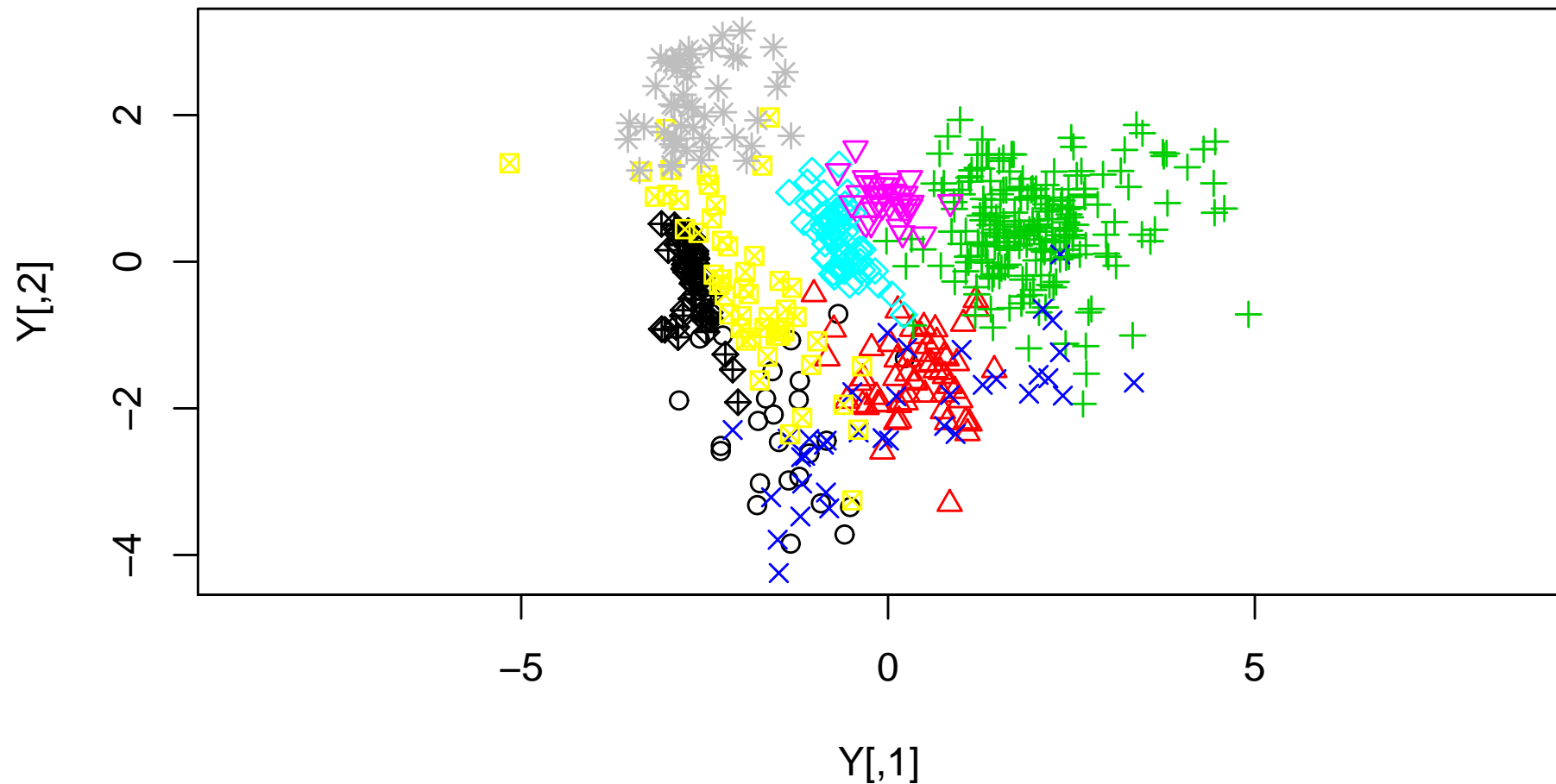- The small arrows show that the points agree quite well after matching.

MDS Results: Olive Oil

- A multidimensional scaling of the Olive oil data with Euclidean distance is shown (regions plotted in different colours and symbols).
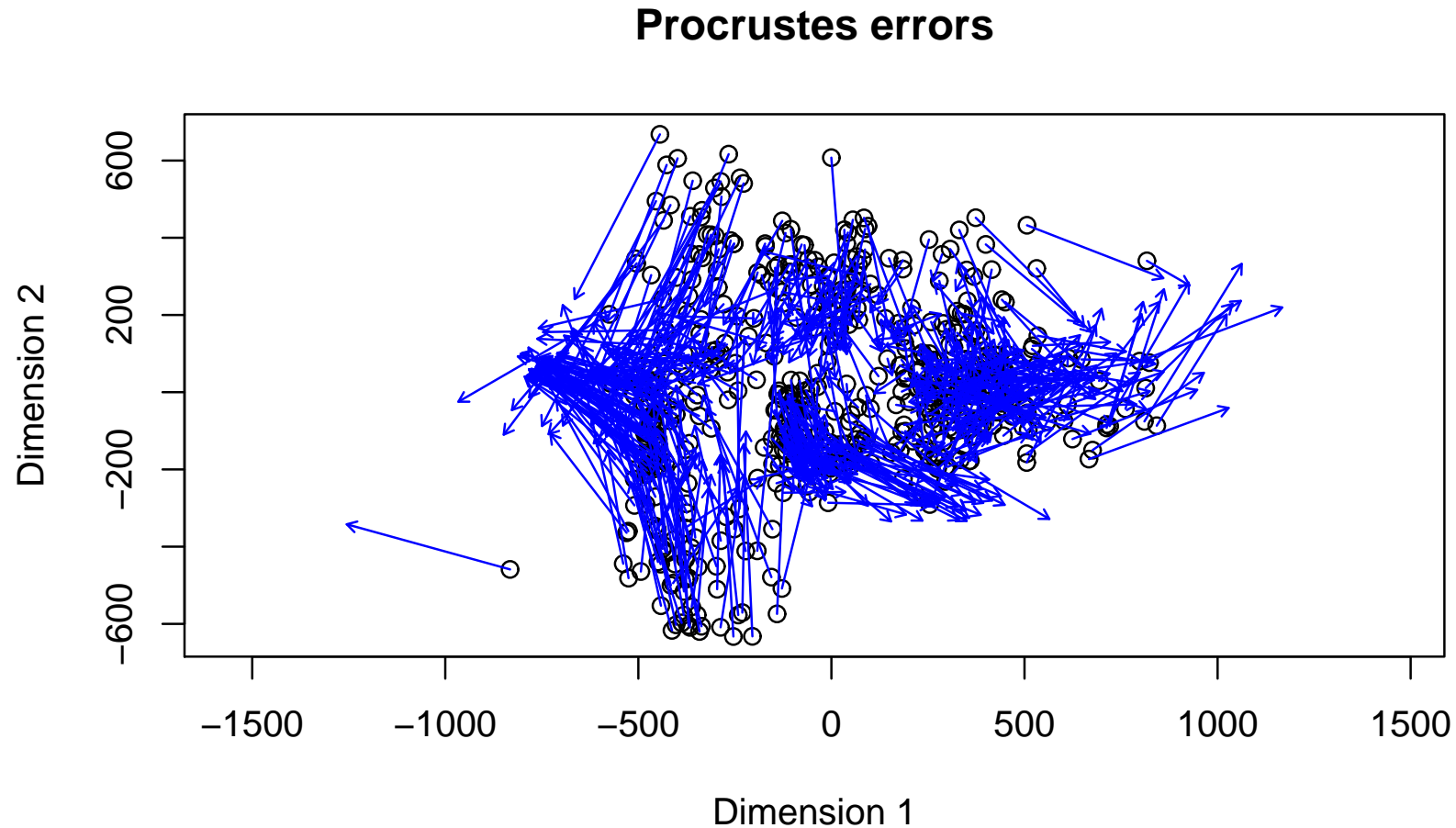
MDS Results: Standardized Data

- If we standardize the data before computing the distance, then we get the
  following:

Procrustes Analysis

Procrustes errors

- Note the effect that standardization has had.

Connections with PCA

- PCA is performed by eigen-decomposition of the data covariance matrix to provide new variables that are formed from linear combinations of the original variables. The new variables are uncorrelated and account for maximum variance in the original variables.

- Classical MDS performs eigen-decomposition of the data dissimilarity matrix to find a low-dimensional configuration of the entities such that distances are preserved as closely as possible in a least-squares sense.

- When Euclidean distance is used within Classical MDS, the resulting low-dimensional co-ordinates are the same as the principal co-ordinates that would be obtained from PCA.

Miscellaneous

- There are many alternative methods of multidimensional scaling.

- The book Multidimensional Scaling by Cox and Cox (2001) gives an exhaustive review of the subject.

- Procrustes analysis is an area of statistics in its own right.

- Procrustes methods are used to study the shapes of objects.

- Landmarks are chosen on the boundary of the object, and Procrustes methods are used to compare the shapes of different objects by aligning the landmarks so that they best match.