MLL

OBJECTIVE: MULTIDIMENSIONAL SCALING

Goal is given the following form:
- For a set of data points, every n items, find a rep of items in $\mathbb{R}^d$ ($d \ll n$). Such that inter-point distance matrix of original data is preserved
- MDS seeks to produce a low dimensional rep of the data such that distance between points i and j in the rep δ_{ij} are close to the data between d_{ij} for all i, j
- Effectively produces a map of data
- To reproduce data from n dim. is closely as possible up to n-1 dimensions may be required
- Aim of MDS is to produce a 'optimal' config of obj in a space $\neq$ dimen

Scaling Methods

- 2 approaches to "measuring" $\{d_{ij}\}$ and $\{\delta_{ij}\} \Rightarrow$ 2 techniques, both which aim to find set of points such that $\delta_{ij} \approx f(d_{ij})$

Metric MDS - refers to when f is continuous and monotonic function, eg the identity function or a first order polynomial with a positive slope

Non Metric MDS - only makes use of the rank order of the d_{ij} s.t. such, the function f need only obey the monotonicity constraint.
$d_{ij} < d_{kl} \Rightarrow f(d_{ij}) \leq f(d_{kl})$
Sum on f need only preserve rank order.

- Classical metric Scaling, Metric least squares scaling, Weighted generalised Scaling
- Procrustes Analysis technique for matching one configuration with another and for producing a measure of the match

Classical Metric Scaling

- Produces a config of points from the dissimilarity matrix (distance matrix)
- Allows for the (continuous monotonic) function of dissimilarity function. Objective is to find the coordinate matrix of the objects i.e.:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix}$$

- Have d the dimension of the configuration (generally want d=2, or 3)
- Since Sum a config of coordinates is invariant to rotations, reflections and translations we fix the centroid to be at the origin.

- For ease of interpretation assume axis 1 is located at origin and the further axis J and k. Distance: d_{ij}, d_{ik}, d_{jk}.
$-1/2 (d^2_{jk} - d^2_{ij} - d^2_{ik}) = x_{ij} x_{ik}$
- Leading value x_{11}.

After Smit ⇒ $X = VA^{1/2}$
V-matrix of e vectors
A-diagnd matrix of e-values

Dimensionality
- Could explain (Squared euclidean) distance between two points x_i and 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" e-value contribute little to squared distance and so the d e-vectors associated with the d 'large' e-values can be used to form the space representing the configuration of points.
- Aim is $d=2$ or $d=3$
- To choose appropriate d we could consider $\sum_{k=1}^{d} \lambda_k / \sum_{k=1}^{n-1} \lambda_k$ which is a measure of the proportion of variation explained by using only d dimens.

Pseudo Code.
1. Obtain dissimilarity $\{d_{ij}\}$
2. Form B, each element of which is given by $b_{jk} = -\frac{1}{2}(d_{jk}^2 - d_{ij}^2 - d_{ik}^2)$ with i representing the centroid/origin of all observations
3. Create matrix Λ from e-values $\lambda_1, \ldots \lambda_{n-1}$ and matrix V from the associated e-vectors $v_1 \ldots v_{n-1}$ of B
4. Choose an appropriate # dimensions d using suitable measure
5. The coordinates of the n required points that are used to represent the n objects in d-dimensional space are given by:
$$x_{ij} = \lambda_j^{1/2} v_{ij}$$
for $i = 1 \ldots n$ and $j = 1 \ldots d$

- Can construct a distance matrix for the points in the MDS plot.
- Hope that it should not be very different from the original diss matrix.

STRESS
- stress of a MDS defined to be $\sum_{i=2}^{n} \sum_{j<i} (\delta_{ij} - d_{ij})^2$
δ_{ij} is distance between i and j in plot and d_{ij} is dissimilarity matrix

Sammon Stress
- Sammon measure of stress takes into account the size of distances being approximated
$$S_1 = (d, \delta) = \sum_{i=1}^{n} \sum_{j \neq i} d_{ij}^{-1} (d_{ij} - \delta_{ij})^2 / \sum_{i=1}^{n} \sum_{j \neq i} d_{ij}^2$$
- Now small dissimilarities have more weight in the loss function (than large ones) providing motivation for them to be reproduced more accurately

08/07/17. MULTIDIMENSIONAL SCALING

Metric (or Classical) Scaling
- Finds a config by matching $\{d_{ij}\}$ to $\{\delta_{ij}\}$ by minimizing a loss function, which possibly involves a continuous monotonic transformation of the dissimilarities.
- An iterative approach (numerical) (method of steepest descent) is used to minimize Sammon's loss function
- Classical MDS is the situation in which a Euclidean dist model is chosen and obtained to minimize the stress value
- Alternate 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 MDS
- Rank order of dissimilarities must be preserved by the transformation f of dist

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

- Kruskal's approach should a d-dimensional config to minimize the stress. However, the input dissimilarities are allowed to be subject to a monotonic transformation
- Similar to Sammon's method on iterative algorithm is used to minimize the stress (steepest descent method)
- D. Kruskal proposed experimenting with several values of d and plotting resulting stress against d and choosing d value where a 'statistical elbow' appears in plot (stress should always decrease & diminishing increases)

Procrustes analysis
- Matches are MDS configurations with another by dilations, rotations, reflections and translations
- Says 2 MDS method has been applied to a set of n points, resulting in coordinate matrices X and Y respectively
- There is a one to one mapping from i^{th} point in X to i^{th} point in 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 configurations one of them is kept constant (like reference config) while the other is transformed
- Assume Y is ref config, while X is to be translated to better match with Y
- The new coordinates of points in X space will be $x_i' = \rho A' x_i + b$
  - ρ = a dilation matrix
  - A = orthogonal matrix causing a rotation and/or reflection
  - b = a translation factor

MLA

14/04/15 MULTI DIMENSIONAL SCALING

- MDS used in problems of 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
- Here 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 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. Objective of MDS is to produce an "optimal" configuration of the observations in a smaller number of dimensions

- Example, road distances in europe
⇒ Objectives of MDS is to take the set of dissimilarities and return a set of two dimensional points, such that the distances between these points are approximately equal to the dissimilarities

MDS Methods
- Different approaches to idea of matching d_{ij} and δ_{ij}, 2 different techniques 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

Metric Method - Classical metric scaling and metric least squares scaling.
⇒ Kruskal non metric scaling

- Not always true that different MDS techniques will produce the same results from dataset, so we will consider PROCRUSTES analysis which is a technique for matching one configuration with another and for producing a measure of match

Classical Metric Scaling.

- Produces a configuration of points from the dissimilarities via an algebraic reconstruction model
- Assume that the (continuous monotone) function of dissimilarity is the identity function. The objective is to find the coordinate location of each observation, i.e

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ \vdots & & & \\ x_{n1} & x_{n2} & & x_{nd} \end{pmatrix}$$

- Here d is the dimension of the configuration (generally 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.

- Assume observation i is located at the origin and that there are two further observations J and K. Distances between them are given by d_{ij}, d_{ik}, d_{jk}

$$d^2_{jk} = d^2_{ij} + d^2_{ik} - 2 d_{ij} d_{ik} \cos(\theta_{jik})$$
$$\Rightarrow -\tfrac{1}{2}(d^2_{jk} - d^2_{ij} - d^2_{ik}) = d_{ij} d_{ik} \cos(\theta_{jik})$$

θ_{jik} is angle between d_{ij} and d_{ik}

Since $\|x_j\| = d_{ij}$ and $\|x_k\| = d_{ik}$ by definition hence:

$$x_j^T x_k = \|x_j\| \|x_k\| \cos(\theta_{jik})$$

Then:

$$-\tfrac{1}{2}(d^2_{ij} - d^2_{ij} - d^2_{ik}) = x_j^T x_k$$

- Vectors x_j are what we are seeking

MLA

15. MULTI DIMENSIONAL SCALING.

How do we obtain the X_I's?

- Let B denote a matrix with elements $b_{JK} = -\frac{1}{2}(d^2_{JK} - d^2_{iJ} - d^2_{iK})$
- We know that $B = XX^T$
- As B is Symetric it can be decomposed into $B = V\Lambda V^T$
  with V a matrix of eigenvectors and Λ a diagonal matrix of eigen values
- Hence $X = V\Lambda^{1/2}$

Dimensionality

- Important issue with any MDS technique is number of dimensions required to represent the configuration of points
- Could express the (squared euclidean) distance between x_i and x_I in an (n-1) dimensional space as $\delta_{IJ} = \sum_{k=1}^{n-1}(v_{ik} - v_{jk})^2$
- Hence any "small" eigenvalue contribute little to the squared distance and so the d eigenvalues associated with the d "large" eigenvalue can be used to form the space representing the configuration of points
- Aim of MDS is to represent the observations graphically with d=2 or 3 dimensions
- To choose d we 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

Pseudo code:

1. Obtain the dissimilarities $\{d_{ij}\}$
2. Form B, each element of which is given by $b_{JK} = -1/2(d^2_{JK} - d^2_{ij} - d^2_{ik})$ with i representing the centroid/origin of all the observations
3. Create matrix Λ from the eigenvalues $\lambda_1, \ldots \lambda_{n-1}$ and the matrix V from the associated eigenvectors $v_1 \ldots v_{n-1}$ of 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 in $1 \ldots n$ and $j = 1 \ldots d$

- Produces a graph. Similar points placed based each other
- relative axis not clear result

How well did it work
- Construct a distance matrix from the points in the MDS plot
- Hope that its not very different from original dissimilarity matrix

- The STRESS of a MDS is defined to be $\sum_{i=2}^{n} \sum_{j<i} (\delta_{ij} - d_{ij})^2$
δ_{ij} is distance between i and j in the plot, d_{ij} is the distance between i and j in the dissimilarity matrix

Sammon Stress
- Sammon stress takes into account the size of the distances being approximated

$$\text{Sammon stress} = (d, \delta) \quad \frac{\sum_{i=1}^{n} \sum_{j \neq i} d_{ij}^{-1} (d_{ij} - \delta_{ij})^2}{\sum_{i=1}^{n} \sum_{j \neq i} d_{ij}}$$

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

Metric least Squares Scaling
- MLS 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 dissimilarity method, or to minimize an altered loss function such as Sammon Stress

MULTI DIMENSIONAL SCALING

Non metric multidimensional Scaling
- In non metric, MDS only the rank order of the dissimilarities must be preserved by the transformation of f of distances
- Stress^2 Kruskal = $\sum_{i=1}^{n} \sum_{j \neq i} [f(d_{ij}) - \delta_{ij}]^2 / \sum_{i=1}^{n} \sum_{i<j} \delta_{ij}^2$

- Kruskal's approach - choose a d-dimensional configuration to minimize the 'stress'
- The input distances are allowed to be subject to a monotonic transformation
- Similar to Sammon mapping, an iterative algorithm used to minimize the stress
- For 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.
  (Stress should always decrease as dimensionality increases)

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 X and Y respectively
- There is a one to one mapping from the i^{th} point in X to j^{th} point in Y
- The sum of squared distance between corresponding points in the configuration is

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

- To match the configurations, one DF (Item) is kept constant (the reference configuration), whilst the other is transformed

- Assume that Y is the reference configuration, whilst X is to be transformed to achieve the best match with Y
- The new coordinates of the points in the X space will be

$$X_i' = pA^T x_i + b$$

p = a dilation matrix
A = orthogonal matrix causing a rotation/and or reflection
b = a translation factor

- New sum of squared distance is given:

$$R^2 = \sum_{i=1}^{n} (y_i - pAx_i - b)^T (y_i - pAx_i - b)$$

- By minimizing this we can estimate optimal p, A and B
- The measure of match between the two configurations is the minimized value of R^2. known as <u>procrustes Sum of Squares</u>
- It can also examine plots of particular 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 vectors
2. Compute optimal values for p, A and b and apply it to non-reference configuration
3. Calculate the procrustes Sum of Squares (smaller values indicate likeness)

Standardised full data together

(connection) with PCA:

- PCA is performed by eigen decomposition of the data covariance matrix to provide new variables that are found 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 distance matrix to find a low dimensional configuration of points 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 principle co-ordinates that would be obtained from PCA

New sum of squared distances between points is the

$$R^2 = \sum_{i=1}^{N} (y_i - pA^T x_i - b)^T (y_i - pA^T x_i - b).$$

By minimizing this we can obtain the optimal p, A and b

- The measure of the match between the 2 configs is the minimized value of R^2. Known as the Procrustes Sum of Squares.
- Can also examine plots of pointwise residuals between the final transformed config and the 'ref' config

Procrustes Code.

1. Translate the configs to the origin by subtracting the average vector for each config from its coordinate values
2. Compute optimal values for p A and b and apply to non-ref config
3. Calculate Procrustes Sum of Squares - Smaller value indicates likeness

Connection with PCA.

- PCA performed by eigen decomposition of the data cov matrix to produce new variables that are formed from linear combos of the original variables. The new variables are uncorrelated and account for max var in the original variables.
- Classical MDS performs e-decomp of the data distance matrix to find a lower dimen config of the entities such that distances are preserved as closely as possibly in a least squares sense
- When euclidean dist is used within classical MDS, the resulting low dimen co-ordinates are the same as the principal components that would be obtained from PCA.