

Escalonamento Multidimensional

Introdução

Dos livros A.D.M (D.Peña) e A.D.M. (J.Lattin, et al)

¹EACH-USP
Universidade de São Paulo

- 1 Introdução ao escalamento Multidimensional
 - Introducción
 - MDS Métrico
- 2 MDS não métrico
- 3 O modelos INDSCAL
- 4 Análise Multidimensional de preferência
 - Primeiras ideias
 - O método Clássico
- 5 Escalados Métricos
 - Coordenadas Principais
 - Alternativas

Outline

- 1 Introdução ao escalamento Multidimensional
 - Introducción
 - MDS Métrico
- 2 MDS não métrico
- 3 O modelos INDSCAL
- 4 Análise Multidimensional de preferência
 - Primeiras ideias
 - O método Clássico
- 5 Escalados Métricos
 - Coordenadas Principais
 - Alternativas

Introdução

- Na ACP e AF, focamos na compreensão do padrão de associação entre **variáveis**
- No Escalamento Multidimensional (MDS multidimensional scaling) focamos no estudo do padrão de similaridade entre **observações**
- MDS: conjunto de métodos para se obterem representações espaciais de similaridades ou proximidades entre entidades
- Usamos MDS quando nossa única informação é uma avaliação da proximidade ou similaridade relativa entre pares de objetos no conjunto de dados
- O objetivo é usar a informação sobre proximidade relativa para criar um mapa da dimensionalidade apropriada.
- No MDS, o método depende da natureza dos dados

Tipos de dados e métodos de MDS

- dados de proximidade, métricos em natureza (distâncias físicas, julgamentos de proximidade entre dos sujeitos, intervalos ou escalas de razão). **Escalonamento métrico ou clássico**
- Os dados de proximidade não tem propriedades de escala métrica (avaliação de similaridade entre produtos de marca, um respondente pode dizer que as marcas A e B são mais similares que as marcas A e C). **MDS não-métrico**. Neste tipo, os dados são fornecidos por um único indivíduo ou agrupados pelo conjunto de indivíduos para produzir uma classificação agregada

Tipos de dados e métodos de MDS

- Os dados envolvem proximidades entre objetos de conjuntos desconexos (análise de preferências: amostras de votantes que expressam preferências por candidatos presidenciais). Por ser resolvido utilizando MDS não métrico, a abordagem é chamada de **Desdobramento multidimensional**
- uma alternativa à abordagem de desdobramento é usar o **MDPREF (multidimensional Analysis of Preference)**, que usa um "modelo vetorial" em lugar de um "ponto ideal" para refletir as preferências dos indivíduos.

Aplicações potenciais

- Mapeamento Perceptual: escalonamento de distâncias psicológicas em uma tentativa de alcançar um "mapa mental" do indivíduo.
- Segmentação de mercado e posicionamento do produto: Precisamos além do mapa perceptual, as informações sobre preferências pelas dimensões subjetivas dos produtos.

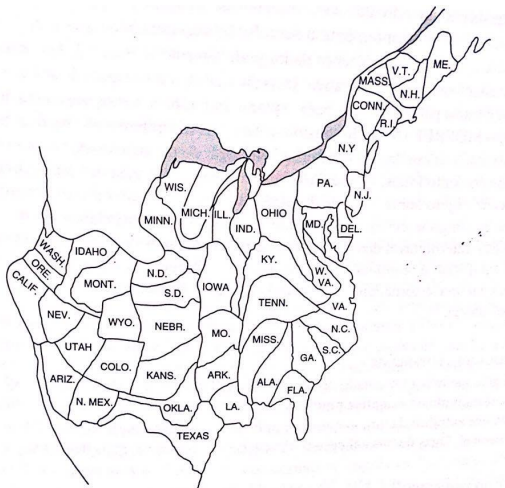
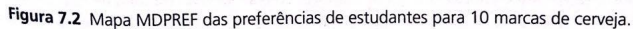


Figura 7.1 “Visão dos Estados Unidos pelos bostonianos” de Shepard.



Outline

1 Introdução ao escalamento Multidimensional

- Introducción
- MDS Métrico

2 MDS não métrico

3 O modelos INDSCAL

4 Análise Multidimensional de preferência

- Primeiras ideias
- O método Clássico

5 Escalados Métricos

- Coordenadas Principais
- Alternativas

MDS métrico e seu funcionamento

a intuição do funcionamento será dado observando a seguinte tabela

Tabela 7.1 Distância entre cidades na Europa (em quilômetros, "em linha reta")

	Atenas	Berlim	Dublin	Londres	Madri	Paris	Roma	Varsóvia
Atenas								
Berlim	1.800							
Dublin	2.859	1.314						
Londres	2.391	928	468					
Madri	2.373	1.865	1.458	1.260				
Paris	2.096	877	786	343	1.049			
Roma	1.039	1.184	1.902	1.443	1.377	1.116		
Varsóvia	1.630	526	1.826	1.454	2.386	1.382	1.350	

Intuição

- Para os dados acima, a localização de cada objeto é perdida
- No é possível recuperar a localização correta de cada continente
- a configuração pode ser rotacionada
- A abordagem analítica depende da fórmula que relaciona o produto escalar dos vetores desde a origem do sistema de coordenadas.
- usando a relação, podemos obter uma estimativa do produto escalar como uma função das distâncias observadas
- o procedimento é similar à de ACP, os componentes extraídos são as localizações na coordenada de cada objeto na análise (um componente para cada dimensão do mapa).

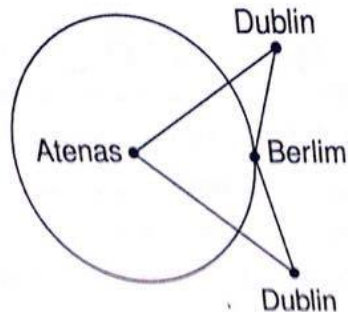


Figura 7.3 Ilustração do procedimento para a solução em lápis e papel

Mecânica

- temos inicialmente uma matriz de similaridades ou discrepâncias entre objetos. Podem ser interpretados como números em escala de intervalos linearmente relacionados às distâncias subjacentes via uma função linear (inclinação positiva para discrepâncias e negativa para similaridades).
- se os dados forem de similaridades, revertemos a escala (subtraindo cada valor do maior valor dos dados) para convertê-los em discrepâncias.
- para definir as distâncias (aproximadas), resolvemos o termo intercepto na função linear de modo a converter as discrepâncias de escala intervalar em estimativas de distância da escala de razão. (podemos assumir que o valor do coeficiente de inclinação linear é 1).

- A constante aditiva é o menor valor que garante que todas as estimativas de distância resultantes satisfaçam a desigualdade triangular.
- Assumimos que as diagonais são indefinidos, a constante a ser adicionada a todas as discrepâncias δ_{jk} para que sejam convertidas em distâncias estimadas (absolutas) d_{jk} é

$$c = \max_{i,j,k} (\delta_{ij} - \delta_{ik} - \delta_{jk})$$

- Se adicionamos c a todos os δ 's, de maneira que

$$d_{jk} = \delta_{jk} + c$$

teremos que

$$d_{ij} < d_{ik} + d_{jk}, \quad \forall i, j, k$$

- adicionar a constante, garante que todos os d são não negativos
- se as distâncias diagonais d_{ii} são iguais a zero para todo i (simetria: $\delta_{jk} = \delta_{kj}$), os axiomas métricos são satisfeitos para os d (quase euclidianos).

Conhecendo a matriz

Temos a matriz de distâncias estimadas $\mathbf{D} = [d_{ij}]$. Nosso objetivo é encontrar a localização na coordenada de cada objeto, dada pela matriz $\mathbf{X} = [x_{ir}]$, em que r é o número de dimensões da configuração. $\mathbf{x}_i' = (x_{i1}, \dots, x_{ir})$ é um vetor linha que representa as coordenadas do objeto i .

- Se a localização absoluta no mapa não é determinada, escolhemos arbitrariamente qualquer ponto para servir de origem

para quaisquer dois pontos j e k , a distância ao quadrado entre j e k (d_{jk}^2), pode ser expressa como uma função das distâncias de j e k a partir da origem (d_{ij} e d_{ik}) e o ângulo entre j e k (θ_{jik}):

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

rearranjando os termos :

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

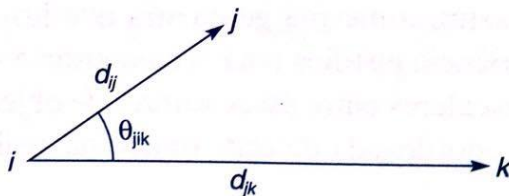


Figura 7.4 Relação geométrica entre distância e produto escalar.

- Sabemos que $d_{ij} = ||\mathbf{x}_j||$ e que $d_{ik} = ||\mathbf{x}_k||$.
- sabemos que

$$\mathbf{x}_j' \mathbf{x}_k = ||\mathbf{x}_j|| ||\mathbf{x}_k|| \cos \theta_{ijk} \quad (3)$$

assim, reescrevemos:

$$-\frac{1}{2}(d_{jk}^2 - d_{ij}^2 - d_{ik}^2) = \mathbf{x}_j' \mathbf{x}_k \quad (4)$$

Colocando o problema na forma matricial:

Seja $\mathbf{B}(i)_{(n-1) \times (n-1)}$, em que i indica que a localização do objeto i foi estabelecida como a origem. os elementos de $\mathbf{B}(i)$ são

$$b_{jk}(i) = -\frac{1}{2}(d_{jk}^2 - d_{ij}^2 - d_{ik}^2)$$

de onde

$$\mathbf{B}(i) = \mathbf{X}_i \mathbf{X}_i' \quad (5)$$

sendo $\mathbf{B}(i)$ simétrica, podemos utilizar a decomposição SVD como

$$\text{mat} \mathbf{B}(i) = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{U}_i' \quad (6)$$

onde \mathbf{U}_i é uma matriz de autovetores (mutuamente ortogonais) e $\mathbf{\Lambda}_i$ é uma matriz diagonal de autovalores:

$$\mathbf{X}_i = \mathbf{U}_i \mathbf{\Lambda}_i^{1/2} \quad (7)$$

No exemplo

Fixamos a origem arbitrariamente (por exemplo Atenas).

- A desvantagem dessa abordagem é que se a cidade estiver em algum lugar proximo da margem (ao invés do centro), quaisquer pequena discrepâncias nos dados (erros de arredondamento por exemplo), tendem a ser ampliadas na solução final
- para minimizar a distorção potencial, colocamos a origem no centroide da configuração. Para isto, criamos uma matriz $\mathbf{B}_{n \times n}$ em que os elementos b_{jk} são elementos da matriz duplamente centrada com elementos $-\frac{1}{2}d_{jk}^2$ ou

$$b_{jk} = -\frac{1}{2}d_{jk}^2 + \frac{1}{2} \sum_j \frac{d_{jk}^2}{n} + \frac{1}{2} \sum_k \frac{d_{jk}^2}{n} - \frac{1}{2} \sum_j \sum_k \frac{d_{jk}^2}{n^2} \quad (8)$$

Tabela 7.2 Matriz **B** para os dados das cidades europeias

1076014	-12160	-770338	-467392	-238364	-273582	438858	246968
-12160	151827	12688	8148	-284286	-35284	-85426	244495
-770338	12688	541039	326878	171543	188274	-318534	-151547
-467392	8148	326878	197399	103597	113331	-194096	-87863
-238364	-284286	171543	103597	622883	136205	54582	-566157
-273582	-35284	188274	113331	136205	74631	-93994	-109579
438858	-85426	-318534	-194096	54582	-93994	219018	-20406
246968	244495	-151547	-87863	-566157	-109579	-20406	444092

Tabela 7.3 Decomposição de valor único da matriz **B** para os dados das cidades europeias (primeiras duas dimensões)

	Autovalor	X_1	X_2
1	2240139	1011	239
2	1131445	77	-375
3	-54323	-715	-184
4	11084	-432	-114
5	-1652	-407	688
6	250	-274	28
7	-46	368	290
8	3	372	-573

Observe que somente dois dos oito autovalores são elevados, o que sugere que uma solução bidimensional é adequada ao mapeamento de cidades.

Nem todos os autovalores restantes são zeros, porque a configuração original não é exatamente bidimensional (a curvatura da terra entra em jogo quando se calcula a distância entre cidades em linha reta. As coordenadas das cidades são apresentadas na figura abaixo. Veja que se girarmos o mapa, ele coincide com o arranjo costumeiro das cidades. A correlação de 0,99 indica ajuste quase perfeito.

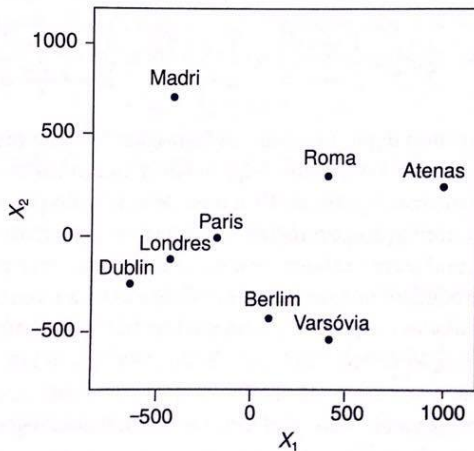


Figura 7.5 Solução clássica de MDS métrico para os dados das cidades europeias.

Estamos interessados em obter uma representação espacial de dados que são não métricos em sua natureza.

O exemplo que segue, trata sobre percepções de indivíduos sobre modelos de carros de preços comparáveis (entre \$ 30000 e \$35000). Foi entregue uma lista de 45 diferentes pares possíveis para cada indivíduo e lhes foi solicitado que classificassem os pares similares (classificação =1), aos menos similares (classificação =45) (Discrepâncias). A tabela abaixo mostra por exemplo $\delta_{3,5} = 1$ e $\delta_{6,9} = 45$.

Quando $d_{ij} = 4$ e $d_{ik} = 8$ podemos inferir que o objeto j é duas vezes mais distante do objeto i , assim como o objeto k o é do i .

Tabela 7.4 Discrepâncias percebidas para os diferentes modelos de carro

	BMW	Ford	Infiniti	Jeep	Lexus	Chrysler	Mercedes	Saab	Porsche	Volvo
BMW										
Ford	34									
Infiniti	8	24								
Jeep	31	2	25							
Lexus	7	26	1	27						
Chrysler	43	14	35	15	37					
Mercedes	3	28	5	29	4	42				
Saab	10	18	20	17	13	36	19			
Porsche	6	39	41	38	40	45	32	21		
Volvo	33	11	22	12	23	9	30	16	44	

Intuição

Joseph Kruskal desenvolveu um algoritmo numérico chamado de STRESS. A medida de desajuste reflete o afastamento da monotonicidade ordinal, entre a distância ajustada e as discrepâncias observadas. Quanto mais próxima a correspondência ordinal, menor o STRESS.

Abordagem iterativa

Passos

- 1. escolha r (dimensionalidade do espaço) para a solução escalonada, baseada em uma avaliação subjetiva da qualidade de ajuste e nos objetivos da análise. Quanto menor o número de objetos, menor o número de dimensões:

Regra prática:

Dimensões	objetos
$r = 1$	pelo menos 5
$r = 2$	pelo menos 9
$r = 3$	pelo menos 13

Abordagem iterativa

- 2. Escolha uma configuração inicial, colocando cada objeto em algum lugar no mapa. A localização de i na configuração bidimensional é $\mathbf{x}_i' = (x_{i1}, x_{i2})$.

Um bom candidato é a configuração obtida pela realização do MDS métrico sobre as discrepâncias.

- 3. Calcule a distância entre todos os pares de pontos no mapa:

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{ir} - x_{jr})^2} \quad (9)$$

- 4. Avalie a correspondência entre as distâncias d_{ij} e as discrepâncias δ_{ij} . A solução envolve encontrar uma transformação monotônica das discrepâncias (\hat{d}_{ij}) que minimize a soma das discrepâncias ao quadrado entre as atuais distâncias d_{ij} e as discrepâncias transformadas.

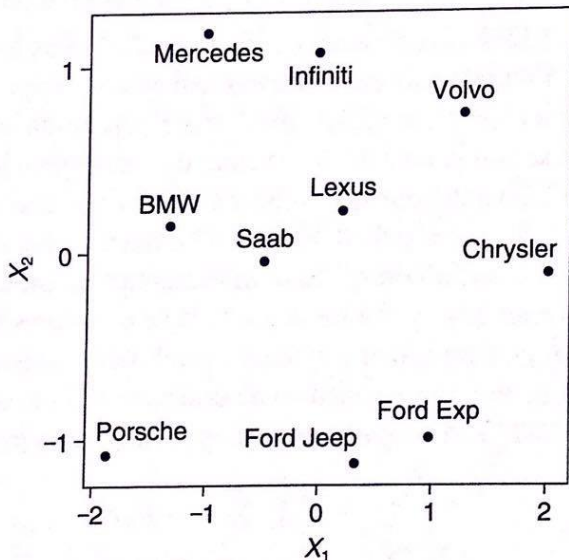


Figura 7.6 Configuração inicial arbitrária para os dados dos carros.

um gráfico de d_{ij} versus as discrepâncias δ_{ij} é às vezes chamado de diagrama de Shepard, pois Shepard sugeriu que o objetivo do MDS não métrico deve ser o de alcançar uma correspondência monotônica entre as dissimilaridades observadas e as distâncias.

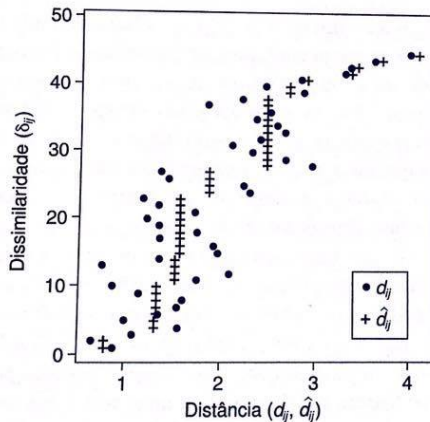


Figura 7.7 Diagrama de Shepard para a configuração inicial dos dados dos carros.

O gráfico acima mostra uma correspondência razoável (mas longe de ser perfeita), entre as discrepâncias observadas e as discrepâncias transformadas em nossa configuração inicial arbitrariamente escolhida.

EM geral, quando a classificação da similaridade percebida entre quaisquer dois objetos aumenta, é assim que também se comporta a distância entre os dois objetos na configuração.

Para captar a extensão geral dos afastamentos da monotonicidade Kruskal propôs uma medida que resume o desajuste de uma configuração particular, que a chamou de STRESS (Desajuste):

fórmula inicial para o STRESS:

$$STRESS = \sqrt{\frac{\sum_i \sum_j (d_{ij} - \hat{d}_{ij})^2}{\sum_i \sum_j d_{ij}^2}} \quad (10)$$

- 5. Utilizando um método de otimização numérico, busque uma alteração na configuração existente para reduzir o STRESS. O modo de Kruskal envolveu o método gradiente de busca, tomando derivada de STRESS em relação às localizações nas coordenadas para cada objeto na configuração. Por exemplo:

$$\partial STRESS / \partial x_{11}$$

o Vetor n dimensional das derivadas parciais é chamado de gradiente

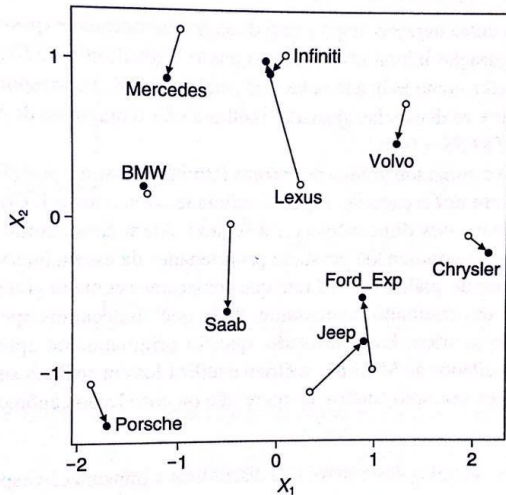


Figura 7.8 Gráfico apresentando as alterações após uma iteração.

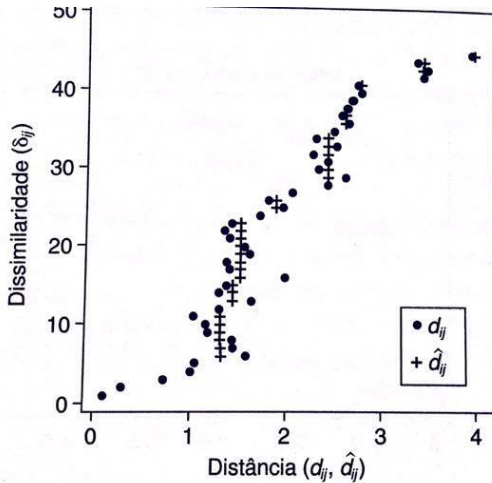


Figura 7.9 Diagrama de Shepard apresentando as melhorias após a iteração.

Exemplo: Mapa perceptual de automóveis

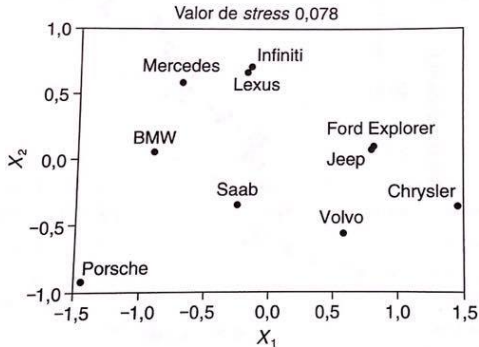


Figura 7.10 Dados dos carros: solução final a partir do ponto inicial aleatório $STRESS = 0,078$.

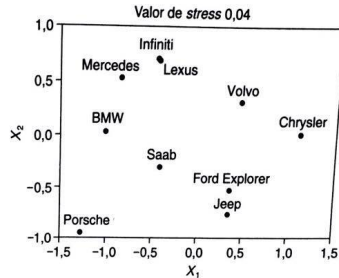


Figura 7.11 Dados dos carros: solução final a partir do ponto de partida de Torgerson $STRESS = 0,040$.

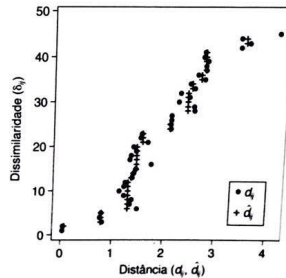


Figura 7.12 Diagrama de Shepard para uma configuração mais bem ajustada para os dados dos carros.

Questões sobre a aplicação do MDS não métrico

Quão bom é o ajuste?. não há padrões absolutos, mas existe a regra prática:

STRESS	Ajuste
0,20	Ruim
0,10	Regular
0,05	Bom
0,02	Excelente

De quantas dimensões necessito?

Quanto maior a dimensão, melhor o ajuste, mas, aumenta a complexidade. A decisão final sai de uma análise de custo-benefício.

Como interpretar as dimensões?

A configuração inicial fornecida pelo MDS não métrico é basicamente arbitrária. A solução pode ser alterada via rotação, reflexão.

Intuição

As pessoas são homogêneas com respeito às suas percepções, no entanto, não julgam da mesma forma. O modelo de escalonamento de diferenças individuais propõe desde a representação totalmente homogênea (único mapa para representar a similaridade perceptual para n indivíduos) e um completamente heterogêneo (mapas separados para cada indivíduo). As percepções individuais são modeladas:

$$y_{jt}^{(i)} = w_{it}^{1/2} x_{jt} \quad (11)$$

onde x_{jt} é a posição do objeto j na dimensão t do espaço de estímulo de grupo, $w_{it}^{1/2}$ é o peso que estende ou encolhe a dimensão t para o sujeito i e $y_{jt}^{(i)}$ é a posição do objeto j na dimensão t como percebido pelo sujeito i .

A distância percebida entre os objetos j e k é definida como a distância euclidiana ordinária nas coordenadas $y^{(i)}$ para o sujeito i :

$$d_{ijk} = \sqrt{\sum_{t=1}^r (y_{jt}^{(i)} - y_{kt}^{(i)})^2} \quad (12)$$

Substituindo na equação acima, temos:

$$d_{ijk} = \sqrt{w_{i1}(x_{j1} - x_{k1})^2 + \dots + w_{ir}(x_{jr} - x_{kr})^2} \quad (13)$$

Intuição

18 filmes foram criticados por 6 autores diferentes (de A a F) excluindo os filmes que a maioria dos críticos concordou ser incríveis (A) ou terríveis (abaixo de C). O objetivo é encontrar uma representação que explique as diferenças de preferências pelos filmes. Poderíamos utilizar o MDS não métrico, no entanto, os dados só fornecem uma fração das informações.

Modelo MDPREF

Objetivo: simplificar o modelo e fortalecer os pressupostos.

Outline

- 1 Introdução ao escalamento Multidimensional
 - Introducción
 - MDS Métrico
- 2 MDS não métrico
- 3 O modelos INDSCAL
- 4 **Análise Multidimensional de preferência**
 - **Primeiras ideias**
 - O método Clássico
- 5 Escalados Métricos
 - Coordenadas Principais
 - Alternativas

Introdução

- Na Análise de Componentes Principais (ACP) obtém-se um "mapa" dos dados em menor dimensão tal que as distâncias euclídeas entre observações sejam preservadas, no espaço original.
- O objetivo do Escalado Multidimensional (Multidimensional Scaling, MDS) é similar, pois pretende encontrar uma representação gráfica em menor dimensão mas a partir de uma matriz de distâncias entre os objetos ou indivíduos. Esta matriz de distâncias se pode calcular desde uma matriz de dados X sobre algumas variáveis medidas ou pode ser uma matriz de distâncias ou similaridades obtida diretamente como por exemplo, de um julgamento de expertos
- Existem muitos métodos de escalamento multidimensional, explicaremos apenas o método clássico.

Outline

- 1 Introdução ao escalamento Multidimensional
 - Introducción
 - MDS Métrico
- 2 MDS não métrico
- 3 O modelos INDSCAL
- 4 **Análise Multidimensional de preferência**
 - Primeiras ideias
 - **O método Clássico**
- 5 Escalados Métricos
 - Coordenadas Principais
 - Alternativas

Introdução

- O método clássico têm por objetivo representar uma matriz de distâncias (ou de similaridades) num mapa ou modelo geométrico.
- Um modelo geométrico é um conjunto de pontos $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ num espaço de q dimensões de forma que cada ponto represente um dos objetos ou indivíduos e que as distâncias entre os pontos se ajustem o máximo possível às distâncias entre os objetos.
- Desta forma o objetivo é duplo, primero: determinar a dimensão q do modelo e obter também as coordenadas (de dimensão q) dos pontos $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ que proporcionam um "bom"ajuste às distâncias observadas. Esse "bom"ajuste deverá ser medido com algum tipo de medida ou índice.

Método Clássico

- A idéia é que se dois objetos têm uma grande distância entre eles, os pontos que os representam estejam suficientemente separados entre si, pelo contrário, os objetos próximos sejam também representados por pontos sem muita separação.
- A solução não é única. O conjunto de coordenadas cujas distâncias se aproximem às distâncias originais podem-se trasladar, rotar ou refletir, sem que as distâncias fiquem distorsionadas, de forma que não se podem determinar completamente.
- O problema de localização se resolve tomando o ponto com as médias como a origem de coordenadas.
- Sobre as rotações, o algoritmo proporciona uma, mas qualquer outra pode servir para facilitar a interpretação das soluções.

Outline

1 Introdução ao escalamento Multidimensional

- Introducción
- MDS Métrico

2 MDS não métrico

3 O modelos INDSCAL

4 Análise Multidimensional de preferência

- Primeiras ideias
- O método Clássico

5 Escalados Métricos

- Coordenadas Principais
- Alternativas

Coordenadas Principais

Suponhamos que a matriz de dados \mathbf{X} seja tal que as distâncias euclídeas entre suas filas seja a solução desejada. Dada uma matriz como essa, calcular as distâncias euclídeas é simples. O problema é: Dada uma matriz de distâncias $\mathbf{D} \ n \times n$, como encontrar \mathbf{X} ?

- Primeiro, definimos a matriz $\mathbf{B} \ n \times n$ como

$$\mathbf{B} = \mathbf{X}\mathbf{X}'$$

- Os elementos de \mathbf{B} são

$$b_{ij} = \sum_{k=1}^q x_{ik}x_{jk}$$

Coordenadas Principais

Pode-se comprovar que o quadrado das distâncias euclídeas entre as filas de \mathbf{X} pode-se escrever em termos dos elementos de \mathbf{B} como

$$d_{ij}^2 = b_{ii} + b_{jj} - 2b_{ij}$$

Coordenadas Principais

Se em lugar das distâncias euclídeas colocamos as distâncias da matriz \mathbf{D} , então temos um sistema de equações por resolver. Este sistema de equações não é trivial, mas com algumas restrições os elementos de \mathbf{B} se podem encontrar em função das distâncias:

$$b_{ij} = -\frac{1}{2} [d_{ij}^2 - d_{i\cdot}^2 - d_{\cdot j}^2 + d_{\cdot\cdot}^2]$$

onde

$$d_{i\cdot}^2 = \frac{1}{n} \sum_{j=1}^n d_{ij}^2$$

$$d_{\cdot j}^2 = \frac{1}{n} \sum_{i=1}^n d_{ij}^2$$

$$d_{\cdot\cdot}^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2$$

Coordenadas Principais

Uma vez encontrados os elementos de \mathbf{B} , resta fatorizar para obter as coordenadas \mathbf{X} . A decomposição em valores singulares de \mathbf{B} se pode escrever

$$\mathbf{B} = \mathbf{U}\tilde{\mathbf{U}}'$$

onde $\tilde{\mathbf{U}} = \text{diag}(\lambda_1, \dots, \lambda_n)$ é a matriz diagonal com os autovalores de \mathbf{B} e \mathbf{U} seus correspondentes autovetores normalizados de forma que a soma dos quadrados dos elementos de cada coluna é 1. Os autovalores se ordenam de maior a menor de forma que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.

Coordenadas Principais

Quando \mathbf{D} vem de uma matriz $n \times q$ de posto máximo, então a matriz \mathbf{B} é de posto q e seus últimos $n - q$ autovalores são zero. Desta forma \mathbf{B} pode-se escrever como

$$\mathbf{B} = \mathbf{U}_q \tilde{\sim}_q \mathbf{U}_q'$$

onde $\tilde{\sim}_q$ contém os primeiros q autovalores e \mathbf{U}_q seus correspondentes autovetores. As coordenadas dos pontos procurados são

$$\mathbf{X} = \mathbf{U}_q \tilde{\sim}_q^{1/2}$$

onde $\tilde{\sim}_q^{1/2} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_q^{1/2})$. A melhor representação de dimensão k vem dada pelos primeiros k autovalores de \mathbf{B} e seus correspondentes autovetores.

O ajuste da representação em dimensão k aos dados originais pode-se medir com o critério

$$P_k = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^q \lambda_i}$$

Valores de P_k superiores a 0.8 sugerem um bom ajuste. Quando a matriz de dissimilaridades não é euclídea, a matriz \mathbf{B} não é definida positiva. Nestes casos alguns dos autovalores de \mathbf{B} serão negativos e, portanto, algumas coordenadas serão números complexos. Porém, se \mathbf{B} tiver um pequeno número de autovalores negativos e se estes são pequenos, a representação associada aos k maiores autovalores positivos é válida. Nestes casos o critério de ajuste pode ser

$$P_k^{(1)} = \frac{\sum_{i=1}^k \text{abs}(\lambda_i)}{\sum_{i=1}^n \text{abs}(\lambda_i)}$$

$$P_k^{(2)} = \frac{\sum_{i=1}^k \lambda_i^2}{\sum_{i=1}^n \lambda_i^2}$$

Outline

1 Introdução ao escalamento Multidimensional

- Introducción
- MDS Métrico

2 MDS não métrico

3 O modelos INDSCAL

4 Análise Multidimensional de preferência

- Primeiras ideias
- O método Clássico

5 Escalados Métricos

- Coordenadas Principais
- Alternativas

Alternativas

Algumas recomendações alternativas são: Critério do traço: Elegir o número de coordenadas de forma que a soma das que correspondam a autovalores positivos seja aproximadamente o mesmo que a soma de todos os autovalores (positivos e negativos). Critério da magnitude: Aceitar como genuinamente positivos os autovalores cuja magnitude exceda substancialmente a do maior autovalor negativo.

Matriz de correlação

