Pràctica 5: MDS - primera part.

• **Exercici 1.** Partint de la matriu de distàncies per carretera entre ciutats europees `eurodist` (paquet `datasets`, carregat per defecte), es vol construir un mapa mitjançant escalament multidimensional clàssic.

1. Utilitzem en primer lloc la funció `cmdscale()`. Els objectes de l'output: `points` i `eig`: Què són aquests objectes en relació al que hem vist a teoria?

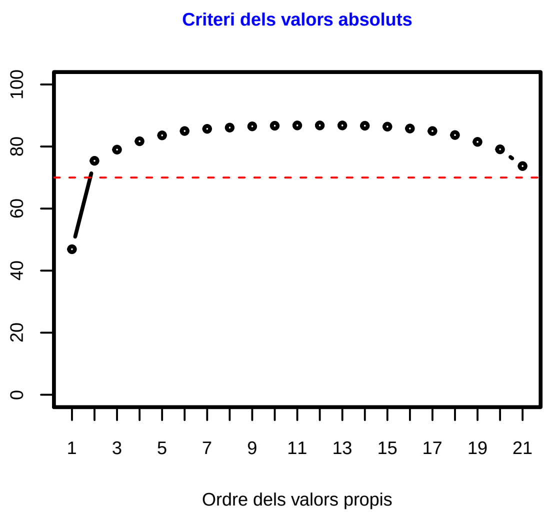
```
library(MASS)
eurodist; class(eurodist);
eudis <- as.matrix(eurodist)
rownames(eudis) <- abbreviate(rownames(eudis))
mds  <- cmdscale(eudis,eig=T)
loc  <- mds$points
vaps <- mds$eig
```

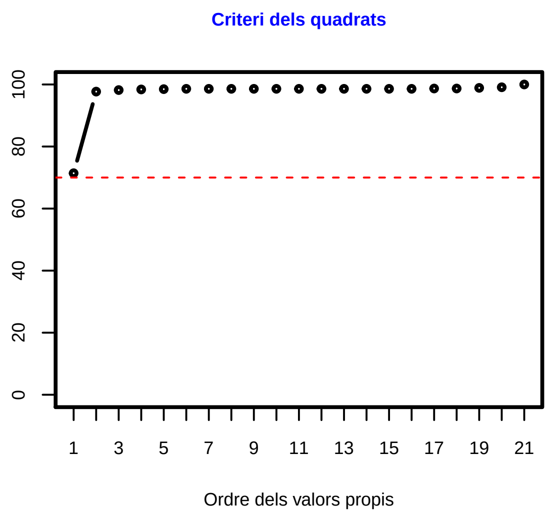
2. És semi-definida positiva la matriu de la qual s'estan calculant els vectors i valors propis? Quina matriu és? Quina matriu de distàncies podem assegurar que és semi-definida positiva sempre?

3. Per avaluar el nombre de dimensions escollides, relacionada amb la conservació de la inèrcia dels casos en dimensió k, aplicarem dos criteris:

$$\text{crit1: } \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{p}|\lambda_i|} \geq p_0 \quad \text{o bé} \quad \text{crit2: } \frac{\sum_{i=1}^{k}\lambda_i^2}{\sum_{i=1}^{p}\lambda_i^2} \geq p_0, \qquad \text{on } p_0 = 0.7 \text{ (per defecte, o un altre llindar).}$$

Per a cada criteri, calcula els qüocients per a tots els valors de k. Fes les gràfiques següents i respon en base a cada criteri: Dues dimensions són suficients? ((Pel color blau del títol: `col.main`=blue""))





4. Representa les ciutats en un mapa bidimensional (Fig. 1, costat esquerre). Pots fer `plot(....,type="n")` i després afegir `text()` per posar els noms a les coordenades dels punts.

5. Veient el mapa obtingut i tenint en compte la situació geogràfica real de les ciutats, fes una reflexió nord-sud ($y \to -y$), canvi que no altera la qualitat de representació 2D. Dibuixa de nou el mapa.

6. Busca ara una *rotació* en el pla que aproximi millor la posició geogràfica real (la rotació tampoc altera la qualitat de representació 2D): per "tempteig", troba un angle de rotació θ que et sembli addient. Recorda que una rotació al pla ve donada per una matriu G de sinus i cosinus, i que s'aplica a unes dades X multiplicant a la dreta per la matriu de rotació:

$$\mathbf{G} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \qquad\qquad \mathbf{Xrot} = \mathbf{X}\,\mathbf{G}$$

Fes proves amb els angles fins que trobis una represantació prou satisfactòria, ((pensa que no podràs recuperar la posició exacta als mapes, només aproximar-t'hi: **Per què?**)). Mostra la solucció rotada al mapa, comparada amb l'anterior solució, com a la Fig. 2. *Nota:* La unitat dels angles és *radians*, i R reconeix l'objecte `pi`.

Figure 1: Comparació de la solució inicial i la solució després de la reflexió.

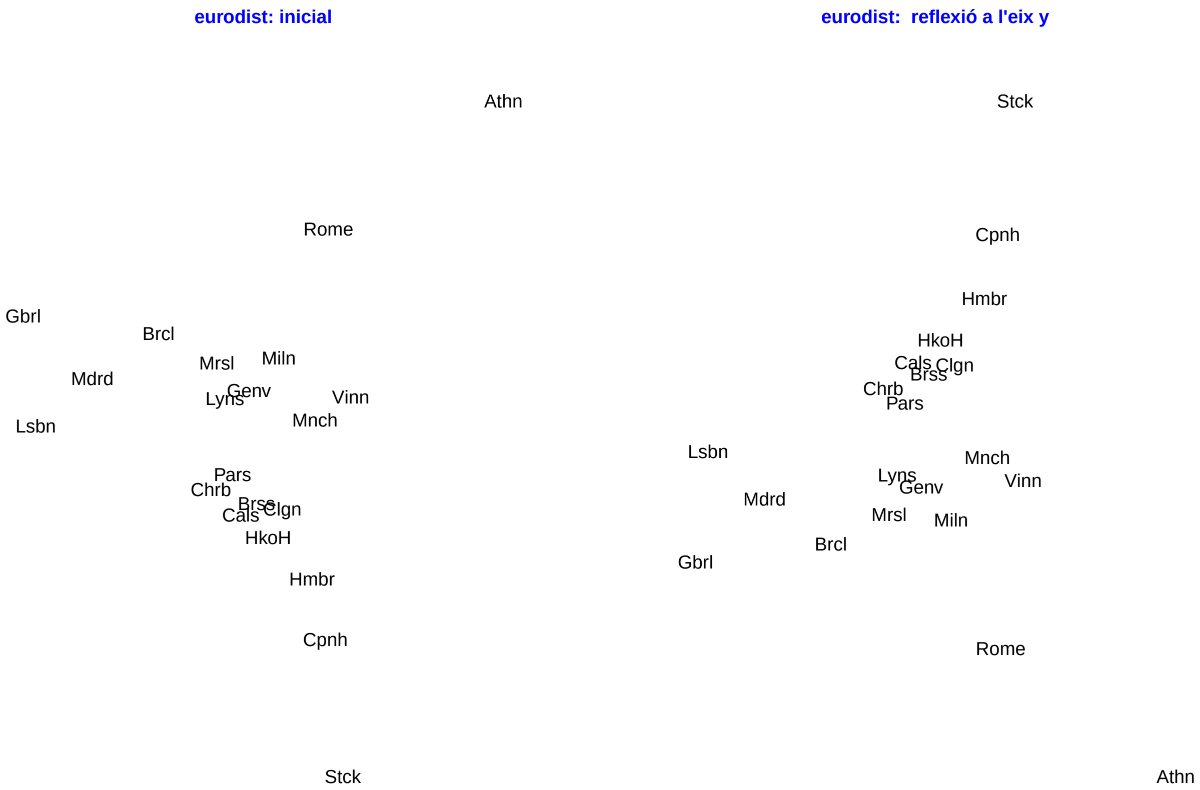
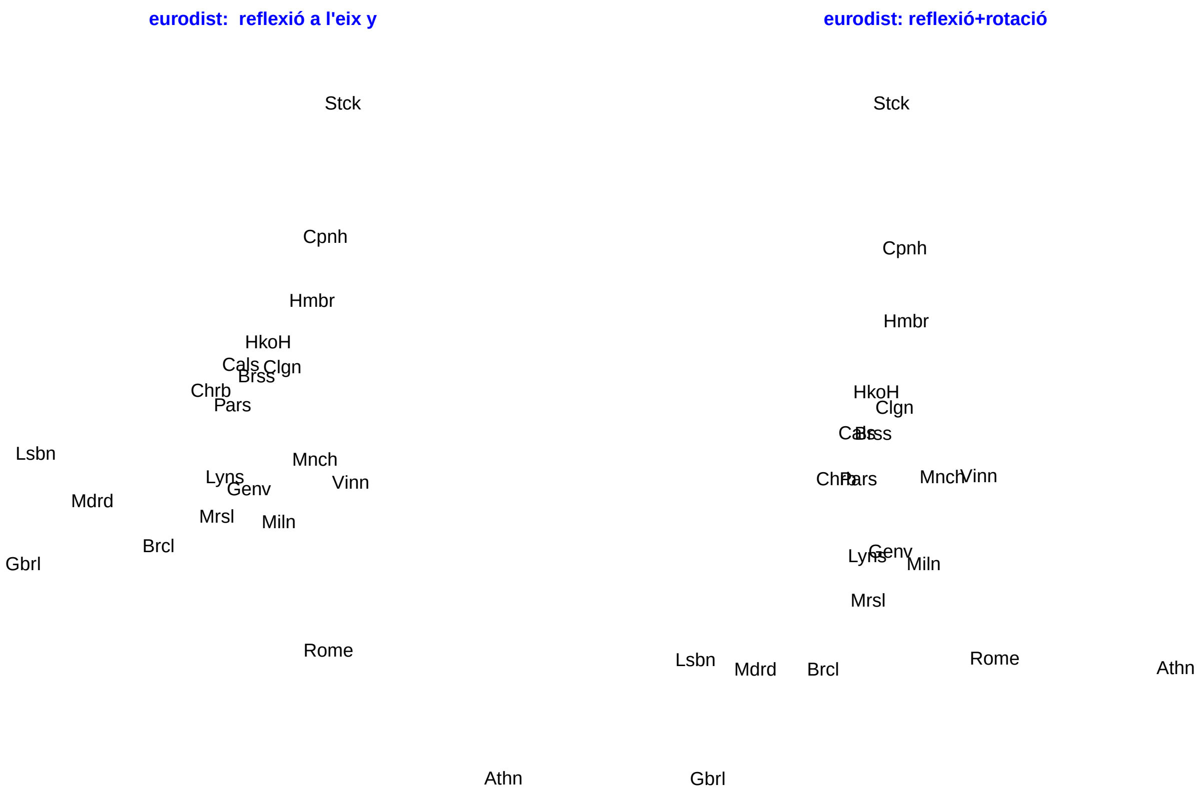


Figure 2: Amb reflexió (esquerra) i post-rotació (dreta).



7. Obté ara els mateixos resultats (valors propis i coordenades de R^2) seguint l'esquema explicat a les diapositives de teoria, sense utilitzar la funció `mdscale()`. *Ind:* Parteix de la matriu de distàncies al quadrat, i aplica:

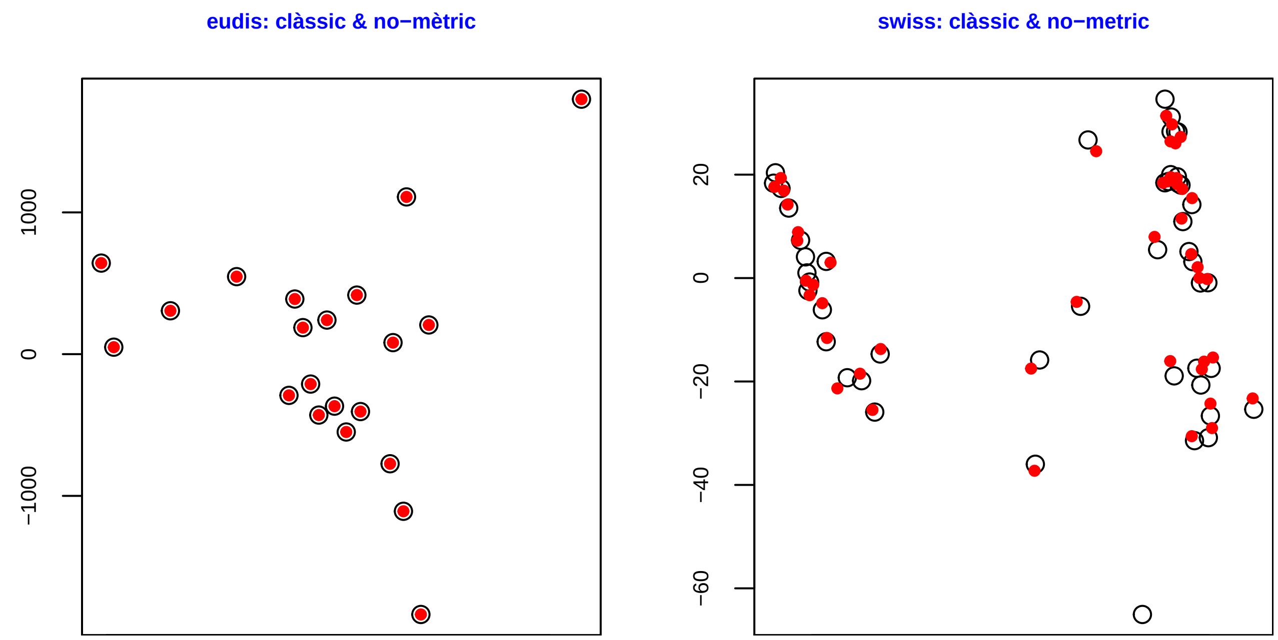
$$b_{ij} = -\frac{1}{2}\left(d_{ij}^2 - \overline{d_{i\bullet}^2} - \overline{d_{\bullet j}^2} + \overline{d_{\bullet\bullet}^2}\right)$$

8. A partir de la solució inicial sense rotar, aplica MDS **no-mètric** a les dades `eudis`, usant la funció `isoMDS()` de la llibreria `MASS`. Compara els punts resultants de l'escalament clàssic (• vermell) i no-mètric (∘ negre i més gran) en un mateix mapa. Comprova que, en aquest cas, els dos mètodes donen la mateixa solució (o aprox. igual). Has d'obtenir la Fig. 3 esquerra, aproximadament.

9. Comprova que amb l'exemple elaborat seguint els passos indicats a les dades `swiss` les dues solucions difereixen més:

   - Fes: `swiss.dist <- dist(swiss)` per obtenir les distàncies. Quina distància s'està aplicant (demana ajuda de la funció)? Quines altres distàncies es poden obtenir amb aquesta funció? Cerca la fórmula d'aquestes distàncies.
   - Aplica els MDS clásic i no-mètric a la matriu de distàncies obtinguda.
   - En aquest exemple, tot i que els casos són províncies, es treballa amb distàncies no geogràfiques: quin sentit li donaries al MDS?
   - Dibuixa les dues solucions sobre el mateix mapa (has d'obtenir la Fig. 3 dreta, aproximadament).

Figure 3: Dos exemples per comparar MDS clàssic i no-metric.



Referències
[1] B. Everitt and T. Hothorn (2011); *An Introduction to Applied Multivariate Analysis with R.* Springer.
[2] B. Everitt (2005); *An R and S-PLUS Companion to Multivariate Analysis.* Springer.
També, el llibre de Härdle-Simar (guia docent), i altres referències que trobareu a les ajudes de les funcions que apliquen MDS en R.
Nota: A la ref. [1] hi ha diversos exemples de MDS, i en un d'ells (pags. 121-122) s'esmenta un concepte interessant: *el mínim arbre recubridor* (*minimum spanning tree*), per complementar el MDS, amb el paquet `ape`.