

# Notes sur le calcul de distances en R

## 1 Calcul des distances à partir des produits scalaires

Si  $X$  est la matrice de données associée à un ensemble de  $n$  points de  $\mathbb{R}^p$ , on a établi dans le chapitre « Positionnement multidimensionnel » que les distances euclidiennes pouvaient être calculées en s'appuyant sur la matrice des produits scalaire en utilisant la relation

$$D_x^2 = \mathbf{h}_x \mathbb{1}_n' - 2W_x + \mathbb{1}_n \mathbf{h}_x'$$

où  $W_x = XX'$ ,  $\mathbf{h}_x = (\|\mathbf{x}_1\|^2, \dots, \|\mathbf{x}_n\|^2)'$  et  $\mathbb{1}_n$  est le vecteur colonne de dimension  $n$  rempli de 1.

Cette relation peut être généralisée au calcul des distances entre un ensemble de  $n$  points de  $\mathbb{R}^p$  et un ensemble de  $m$  points de  $\mathbb{R}^p$ . On obtient la relation

$$D_{xy}^2 = \mathbf{h}_x \mathbb{1}_m' - 2W_{xy} + \mathbb{1}_n \mathbf{h}_x'$$

où  $W_{xy} = XY'$ ,  $\mathbf{h}_x = (\|\mathbf{x}_1\|^2, \dots, \|\mathbf{x}_n\|^2)'$ ,  $\mathbf{h}_y = (\|\mathbf{y}_1\|^2, \dots, \|\mathbf{y}_m\|^2)'$ ,  $\mathbb{1}_n$  est le vecteur colonne de dimension  $n$  et  $\mathbb{1}_m$  est le vecteur colonne de dimension  $m$ .

Ces deux relations permettent de calculer les distances sans utilisation de boucles dans des logiciels matriciels comme Matlab et R.

## 2 Exemple de calcul en R

```
# Définition des données X et Y
n <- 4
m <- 3
p <- 2
X <- matrix(sample(1:20,n*p),nrow=n)
Y <- matrix(sample(1:20,m*p),nrow=m)

# Définition des vecteurs h.x, h.y, ones.n et ones.m
h.x <- rowSums(X^2)
h.y <- rowSums(Y^2)
ones.n <- rep(1,n)
ones.m <- rep(1,m)

# Calcul de la matrice des distances au carré entre les points de X
D2x <- h.x%*%t(ones.n) - 2*X %*% t(X) + ones.n%*%t(h.x)

# Calcul de la matrice des distances au carré entre les points de X et les points de Y
D2xy <- h.x%*%t(ones.m) - 2*X %*% t(Y) + ones.n%*%t(h.y)
```