TD 11 - Analyse discriminante

But du TD : analyse en composantes principales et analyse discriminante.

On s'intéresse aux données iris. On cherche à différencier les différentes espèces à partir des autres variables disponibles.

Discrimination sur sépales et pétales

Décrire les données en jeu. Comprendre et effectuer les commandes suivantes :

```
par(mfrow=c(2,1));  
s.class(iris[,1 :2],iris[,5])  
s.class(iris[,3 :4],iris[,5]).
```

Quelle est la partie de la plante qui permet le mieux de différencier les espèces, à première vue? Calculer le pouvoir discriminant de la variable *longueur des pétales*. Pour cela, on utilisera le code suivant :

```
variance<-function(vect)  
{  
  l<-length(vect)  
  (l-1)/l*var(vect)  
}  
pl<-iris$Petal.Length; cla<-iris[,5];  
pl.lc<-tapply(pl,cla, length)  
pl.mc<-tapply(pl,cla,mean)  
pl.vinter<-1/length(pl)*sum(pl.lc*(pl.mc-mean(pl))^2)  
pl.v<-variance(pl)  
pl.vinter/pl.v
```

Analyse en composantes principales

On utilise les fonctions d'analyse en composantes principales de la librairie `ade4`.

- Effectuer une analyse en composantes principales sur les variables quantitatives grâce à la commande suivante : `iris.pca<-dudi.pca(iris[,1 :4],scanf=F)`.
- Interpréter les vecteurs propres obtenus dans l'ACP. Observer les valeurs propres.
- Visualiser le cercle des corrélations (exécuter `s.corcircle(iris.pca$co)`).
- Visualiser les groupes en projection sur les deux premiers axes principaux grâce à la commande : `s.class(iris.pca$li,iris[,5])`.

Analyse discriminante

- L'analyse discriminante se fait au moyen de la fonction `discrimin`. L'étudier et effectuer la commande : `iris.d<-discrimin(iris.pca,fac=iris[,5],scannf=F)`.
- Visualiser les données projetées sur les deux premiers axes : `s.class(iris.d$li,iris[,5])`.
- Afficher le cercle des corrélations grâce à la commande : `s.corcircle(iris.d$va)`.
- Calculer le pouvoir discriminant du premier axe obtenu par ACP, et enfin du premier axe obtenu par AFD.

Rappel : pouvoir discriminant d'une variable (notations du cours). On considère la variable X^j . La variance totale de cette variable est : $VT = \frac{1}{n} \sum_{i=1}^n (X_i^j - \bar{X}^j)^2$. La variance inter-classe est : (avec K classes) $V_{inter} = \frac{1}{n} \sum_{k=1}^K n_k (\bar{X}(k)^j - \bar{X}^j)^2$. Le pouvoir discriminant de la variable est : $\frac{V_{inter}}{VT}$.

TD 11 - Analyse discriminante

But du TD : analyse en composantes principales et analyse discriminante.

On s'intéresse aux données iris. On cherche à différencier les différentes espèces à partir des autres variables disponibles.

Discrimination sur sépales et pétales

Décrire les données en jeu. Comprendre et effectuer les commandes suivantes :

```
par(mfrow=c(2,1));  
s.class(iris[,1 :2],iris[,5])  
s.class(iris[,3 :4],iris[,5]).
```

Quelle est la partie de la plante qui permet le mieux de différencier les espèces, à première vue? Calculer le pouvoir discriminant de la variable *longueur des pétales*. Pour cela, on utilisera le code suivant :

```
variance<-function(vect)  
{  
  l<-length(vect)  
  (l-1)/l*var(vect)  
}  
pl<-iris$Petal.Length; cla<-iris[,5];  
pl.lc<-tapply(pl,cla, length)  
pl.mc<-tapply(pl,cla,mean)  
pl.vinter<-1/length(pl)*sum(pl.lc*(pl.mc-mean(pl))^2)  
pl.v<-variance(pl)  
pl.vinter/pl.v
```

Analyse en composantes principales

On utilise les fonctions d'analyse en composantes principales de la librairie `ade4`.

- Effectuer une analyse en composantes principales sur les variables quantitatives grâce à la commande suivante : `iris.pca<-dudi.pca(iris[,1 :4],scanf=F)`.
- Interpréter les vecteurs propres obtenus dans l'ACP. Observer les valeurs propres.
- Visualiser le cercle des corrélations (exécuter `s.corcircle(iris.pca$co)`).
- Visualiser les groupes en projection sur les deux premiers axes principaux grâce à la commande : `s.class(iris.pca$li,iris[,5])`.

Analyse discriminante

- L'analyse discriminante se fait au moyen de la fonction `discrimin`. L'étudier et effectuer la commande : `iris.d<-discrimin(iris.pca,fac=iris[,5],scannf=F)`.
- Visualiser les données projetées sur les deux premiers axes : `s.class(iris.d$li,iris[,5])`.
- Afficher le cercle des corrélations grâce à la commande : `s.corcircle(iris.d$va)`.
- Calculer le pouvoir discriminant du premier axe obtenu par ACP, et enfin du premier axe obtenu par AFD.

Rappel : pouvoir discriminant d'une variable (notations du cours). On considère la variable X^j . La variance totale de cette variable est : $VT = \frac{1}{n} \sum_{i=1}^n (X_i^j - \bar{X}^j)^2$. La variance inter-classe est : (avec K classes) $V_{inter} = \frac{1}{n} \sum_{k=1}^K n_k (\bar{X}(k)^j - \bar{X}^j)^2$. Le pouvoir discriminant de la variable est : $\frac{V_{inter}}{VT}$.