Une revue des méthodes de discrimination pour la spectrométrie PIR (centrée sur les méthodes linéaires)

JM Roger

jean-michel.roger@cemagref.fr

Plan

- Qu'est ce que la discrimination ?
- Le modèle de l'analyse discriminante
- Application à la spectrométrie
- Méthodes alternatives

Les principes de l'analyse discriminante

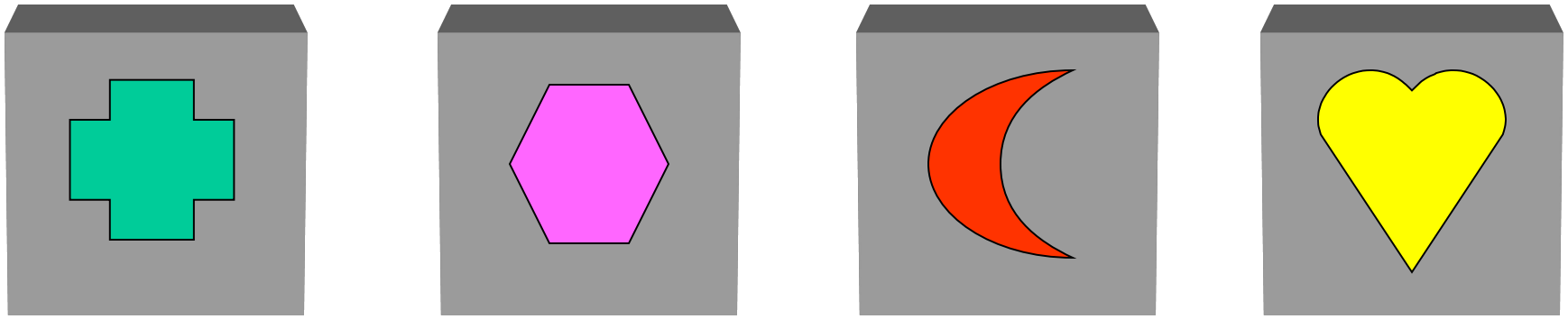
Qu'est ce que la discrimination?

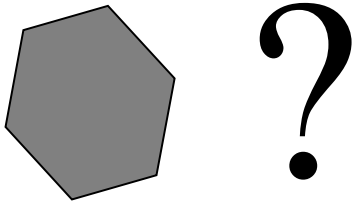
Affecter un objet à une classe (un groupe), parmi une collection de classes (de groupes) prédéfinies

≠

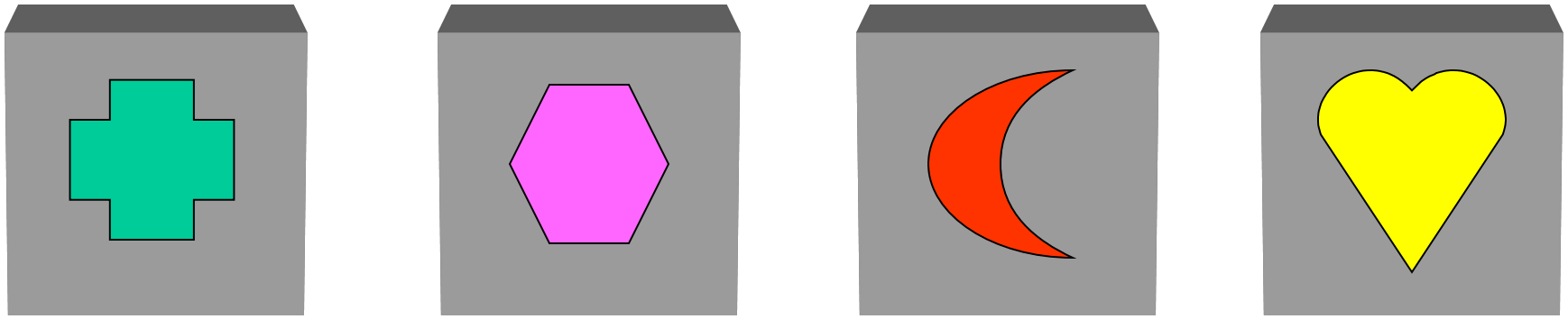
classification

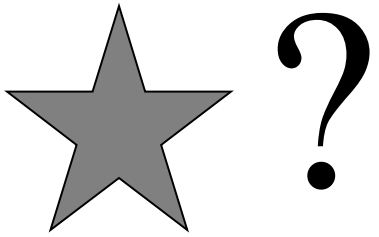
La discrimination : un exemple





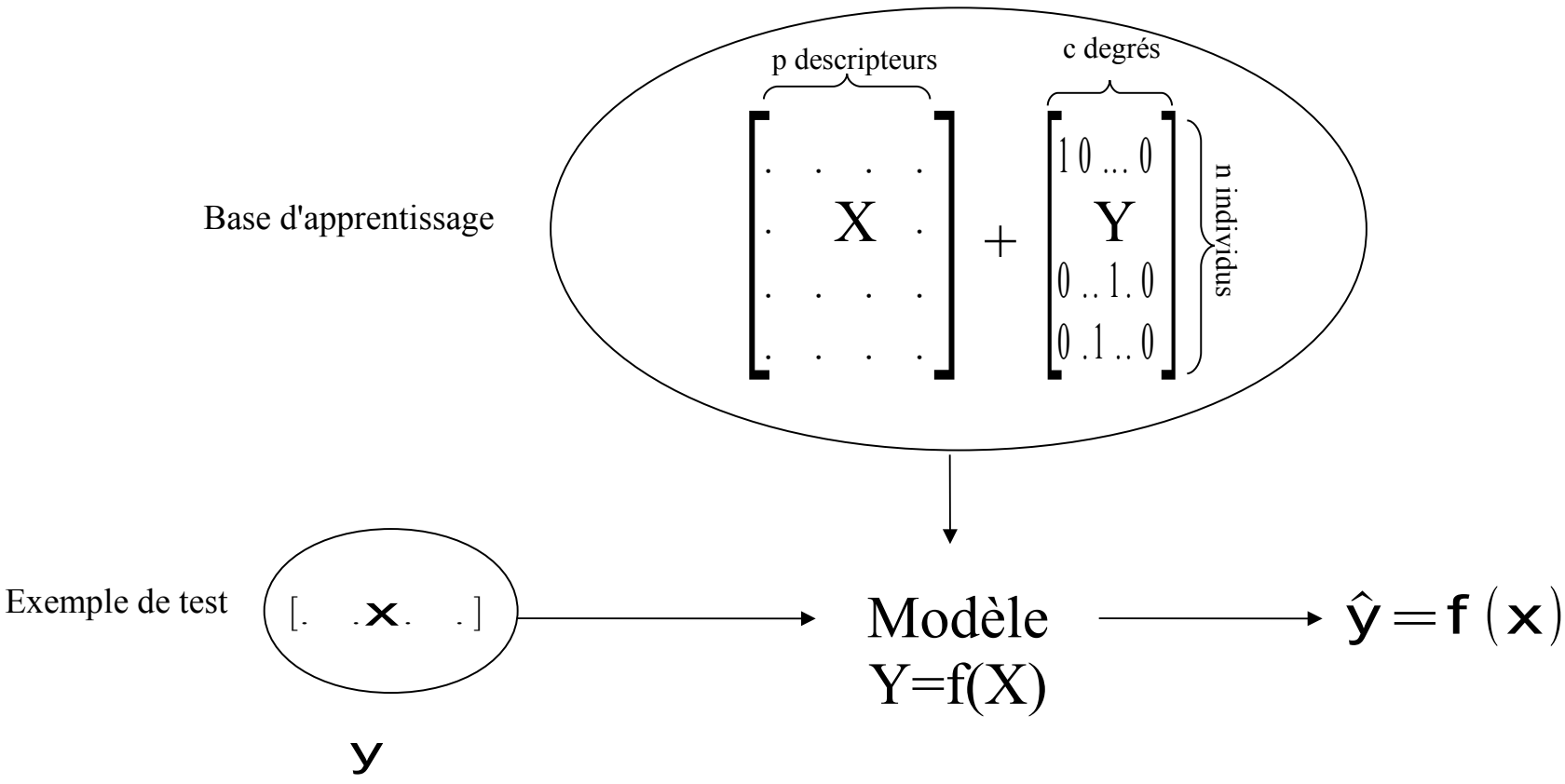
La discrimination : un autre exemple





Avant tout...

... un peu de formalisme



Matrice de Confusion

Pour chaque individu : $y \stackrel{?}{=} \hat{y}$

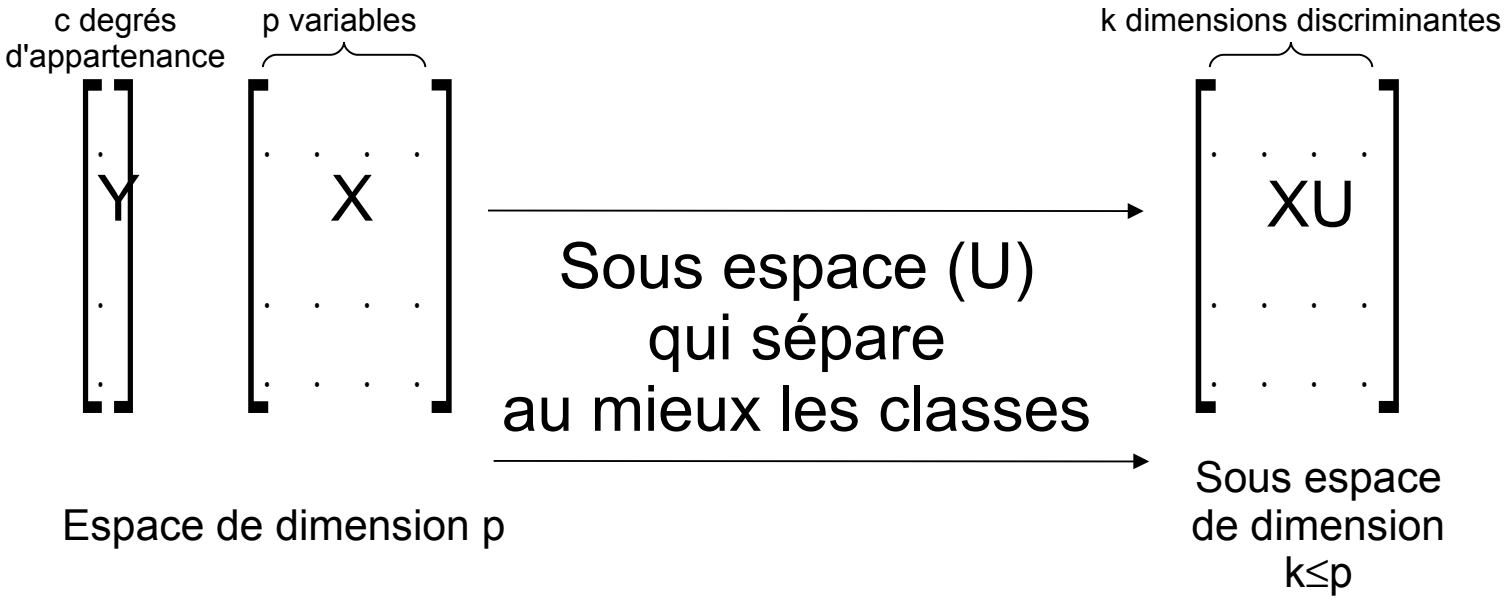
	Actual classes						Total
Number of individuals	14	12	18	21	14	19	98
Predicted classes	C1	C2	C3	C4	C5	C6	
C1	10	0	1	0	0	0	11
C2	0	12	0	2	0	0	14
C3	3	0	15	0	1	1	20
C4	0	0	0	18	0	3	21
C5	1	0	0	0	13	0	14
C6	0	0	2	1	0	15	18
% good assignments							
Error ratio	**29%**	**0%**	**17%**	**14%**	**7%**	**21%**	**15%**

Erreur

Puisque Y contient des degrés d'appartenance, on a directement : $Cf = \hat{Y}^T Y$

Schéma général des méthodes d'analyse discriminante linéaire

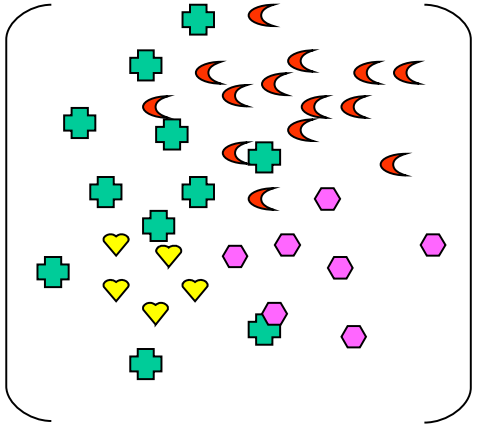
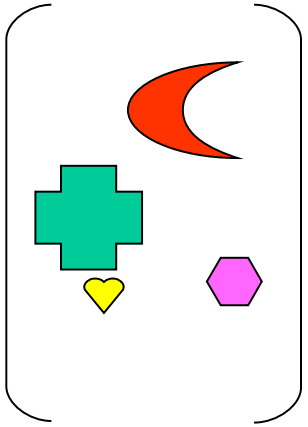
1ère étape : projection



2ème étape : décision

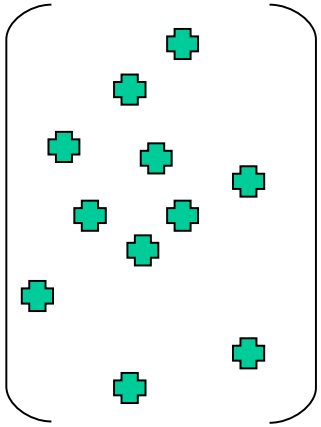
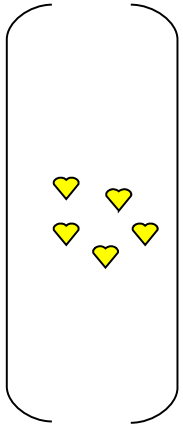
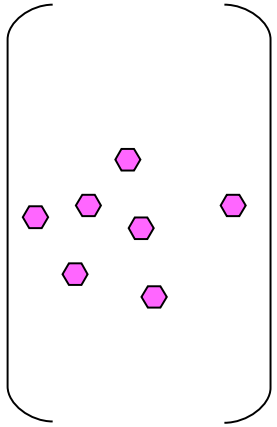
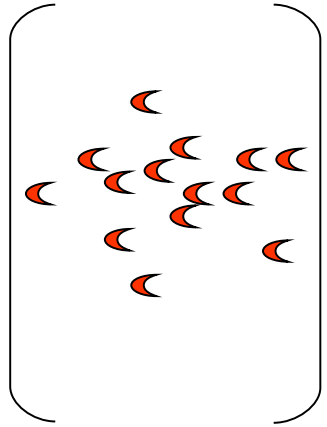
Quelle est la classe la plus proche de $x^T U$?

La décomposition des variances

Var  = Var 

Variance totale (T)

Variance inter classes (B)

+Var  +Var  +Var  +Var 

Variance intra classes (W)

Variances…

- Théorème de Huygens : T = W + B

- Donc, maximiser B relativement à T revient à maximiser :

  - B/W « english school » (Fisher, LDA)

ou

  - B/T « french school » (Benzecri, AFD)

Les 2 sont appelés Wilk's Lambda Λ

Deux méthodes équivalentes, car :
B/T = (B/W) / (1 + B/W)

0 ← - - - - - - - - - - B/T - - - - - - - - - - → 1

0 ← - - - - - - - - - - B/W - - - - - - - - - - → +∞

Le modèle de base de l'Analyse Discriminante (point de vue intuitif)

- On cherche le vecteur u tel que Xu maximise le ratio B/T

- Une fois trouvé, on projette les données orthogonalement à u

- On répète le process ...

Le modèle de base de l'Analyse Discriminante (point de vue mathématique)

- On cherche le vecteur u tel que : B(Xu) / T(Xu) maxi
- Or B(Xu) = u^TBu et T(Xu) = u^TTu
- Donc Λ(Xu) = u^TBu / u^TTu est une fonction de R^p to R

- Maximier Λ(Xu) $\Leftrightarrow$ Trouver le zéro de grad(Λ(Xu))

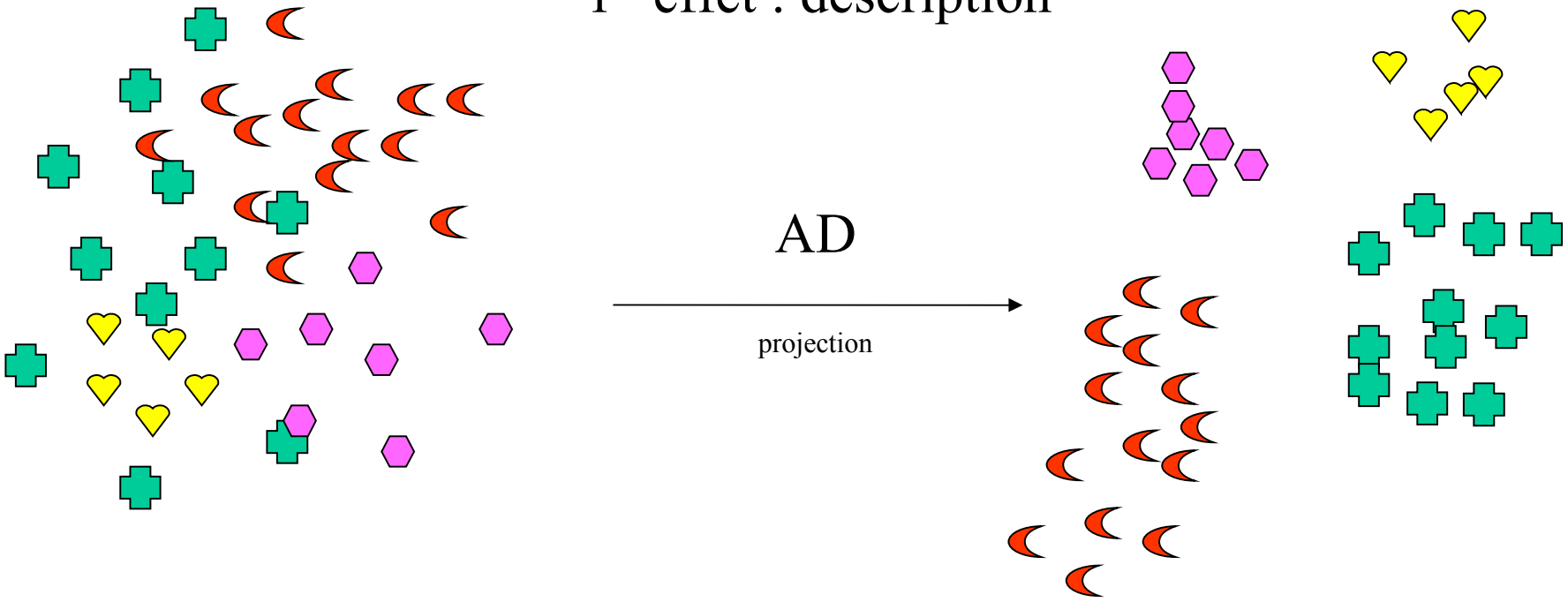
$$\text{grad}(\Lambda(Xu)) = 2((u^TTu)Bu - (u^TBu)Tu) / (u^TTu)^2$$

$$\text{Max}(\Lambda(Xu)) \Leftrightarrow (B - \lambda T)u = 0 \Leftrightarrow T^{-1}B\,u = \lambda u$$

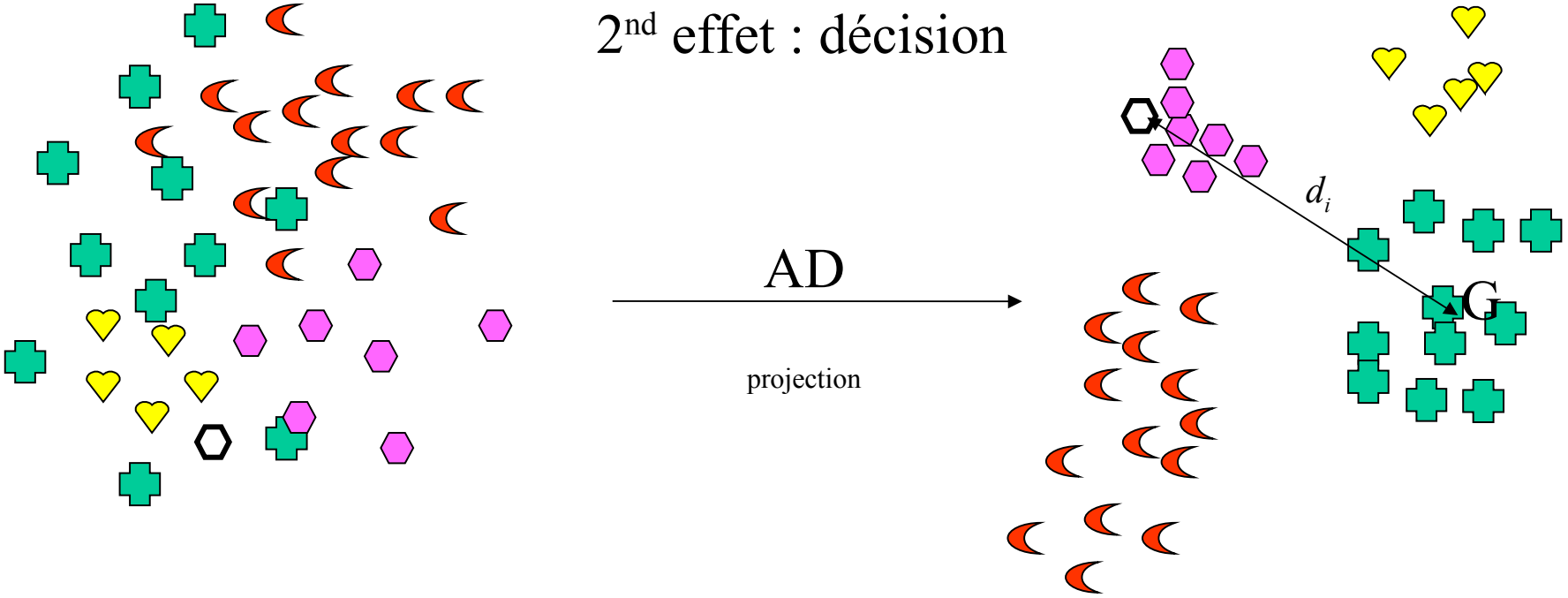
U contient les (c-1) premiers vecteurs propres de $T^{-1}B$

Le modèle de base de l'Analyse Discriminante (point de vue pratique)

1er effet : description



Le modèle de base de l'Analyse Discriminante (point de vue pratique)

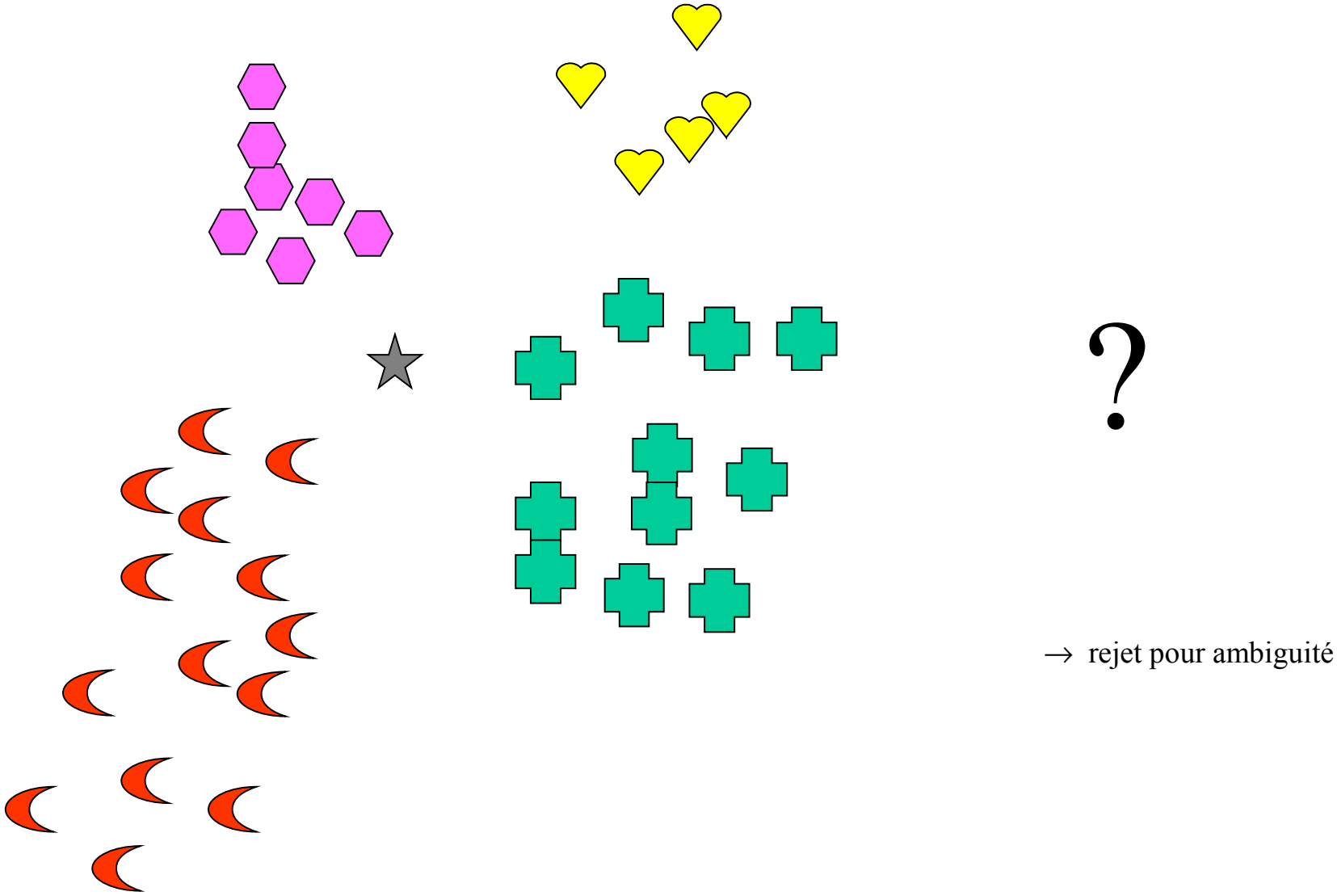


On calcule un degré d'appartenance sur la base de la distance aux centres de gravité des classes

On adopte la classe qui présente le degré maxi

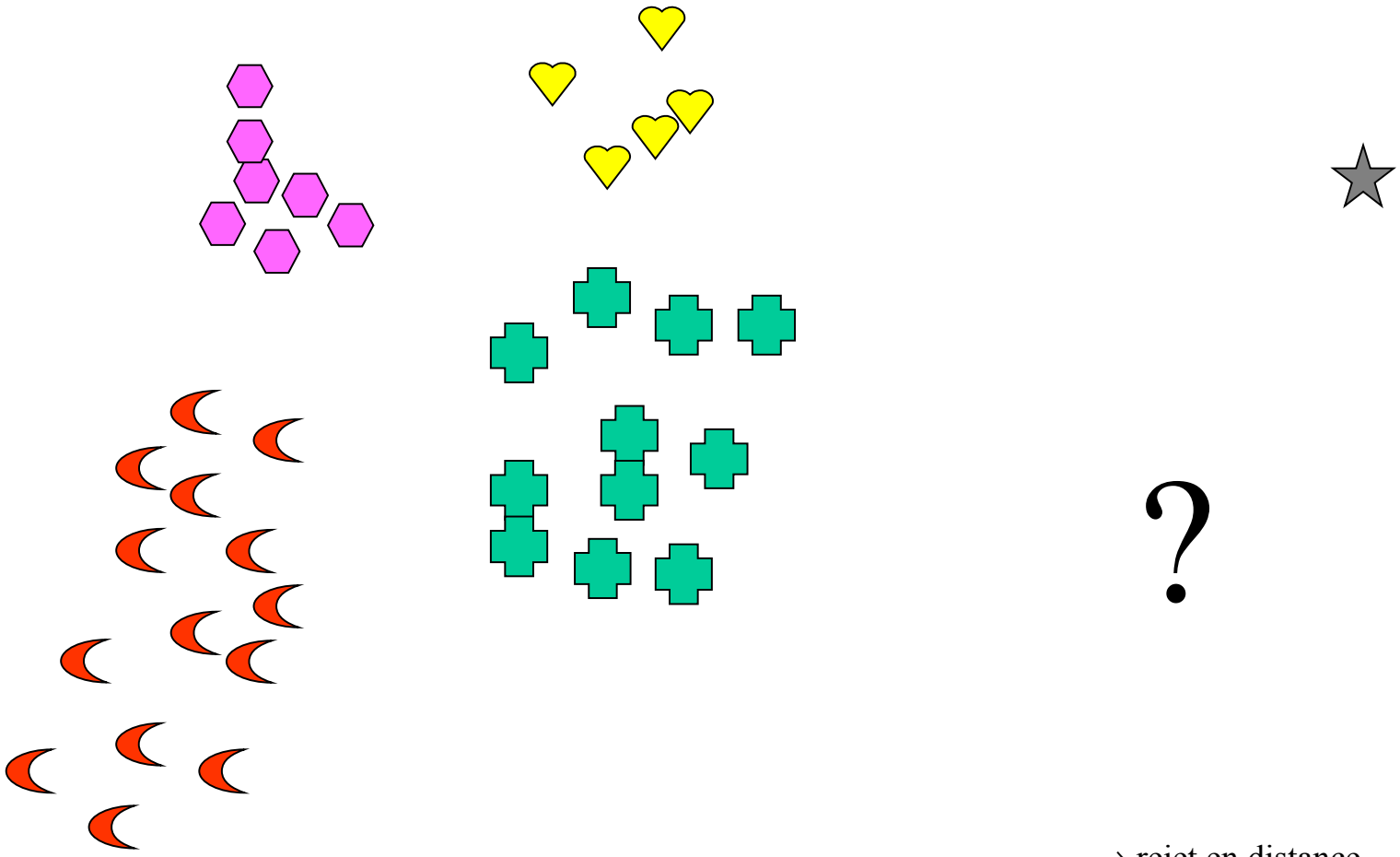
$$\alpha_i = \frac{e^{-d_i}}{\sum_{j=1}^{c} e^{-d_j}} \quad \text{ou} \quad \alpha_i = \frac{\frac{1}{d_i}}{\sum_{j=1}^{c} \frac{1}{d_j}}$$

Le modèle de base de l'Analyse Discriminante (point de vue pratique)



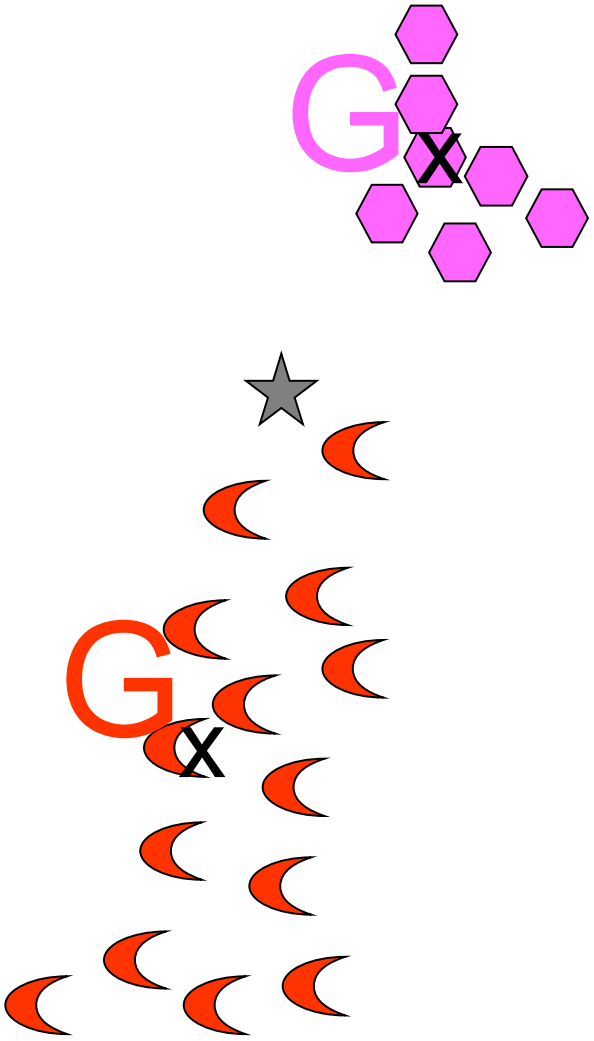
→ rejet pour ambiguité

Le modèle de base de l'Analyse Discriminante (point de vue pratique)



→ rejet en distance

Le modèle de base de l'Analyse Discriminante (point de vue pratique)



?

→ la forme des classes doit être prise en compte

La distance de Mahalonobis

- **But** : Calculer la distance d'un point à une classe
- **L'idée** : pondérer chaque dimension inversement à la largeur de la classe
- **Le formalisme** :

Distance Euclidienne : $d(x,g_i)^2=(x-g_i)^T(x-g_i)$

Distance de Mahalanobis : $d(x,g_i)^2=(x-g_i)^T S_i^{-1} (x-g_i)$

Avec S_i la matrice de variance covariance de la classe i

Application de l'analyse discriminante à la spectrométrie

Application à la spectrométrie

Les variables spectrales sont très corrélées

Donc, le rang réel des données est très inférieur au rang mathématique

Donc, T n'est pas inversible

L'analyse discriminante ne peut donc pas être appliquée directement aux données spectrales

Application à la spectrométrie

- Deux solutions :
  - Réduire la dimension pour rendre T inversible
    - Sélection de variables
    - Construction de variables latentes
  - Appliquer une méthode alternative
    - PLS2 directe, Multi PLS binaire, SIMCA, Méthode continuum, ...

ACP + Analyse Discriminante

PCA-DA

- On construit un premier sous espace par ACP sur X, sur lequel T sera inversible
- On applique l'analyse discriminante sur cet espace

Inconvénient : l'espace de l'ACP n'est pas optimal pour la discrimination

NB : certains programmes le font sans le dire, lorsque la matrice X est mal conditionnée

PLS2 + Analyse Discriminante

PLS-DA

JOURNAL OF CHEMOMETRICS
J. Chemometrics 2003; 17: 166–173
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cem.785

Partial least squares for discrimination

Matthew Barker[1] and William Rayens[2]*

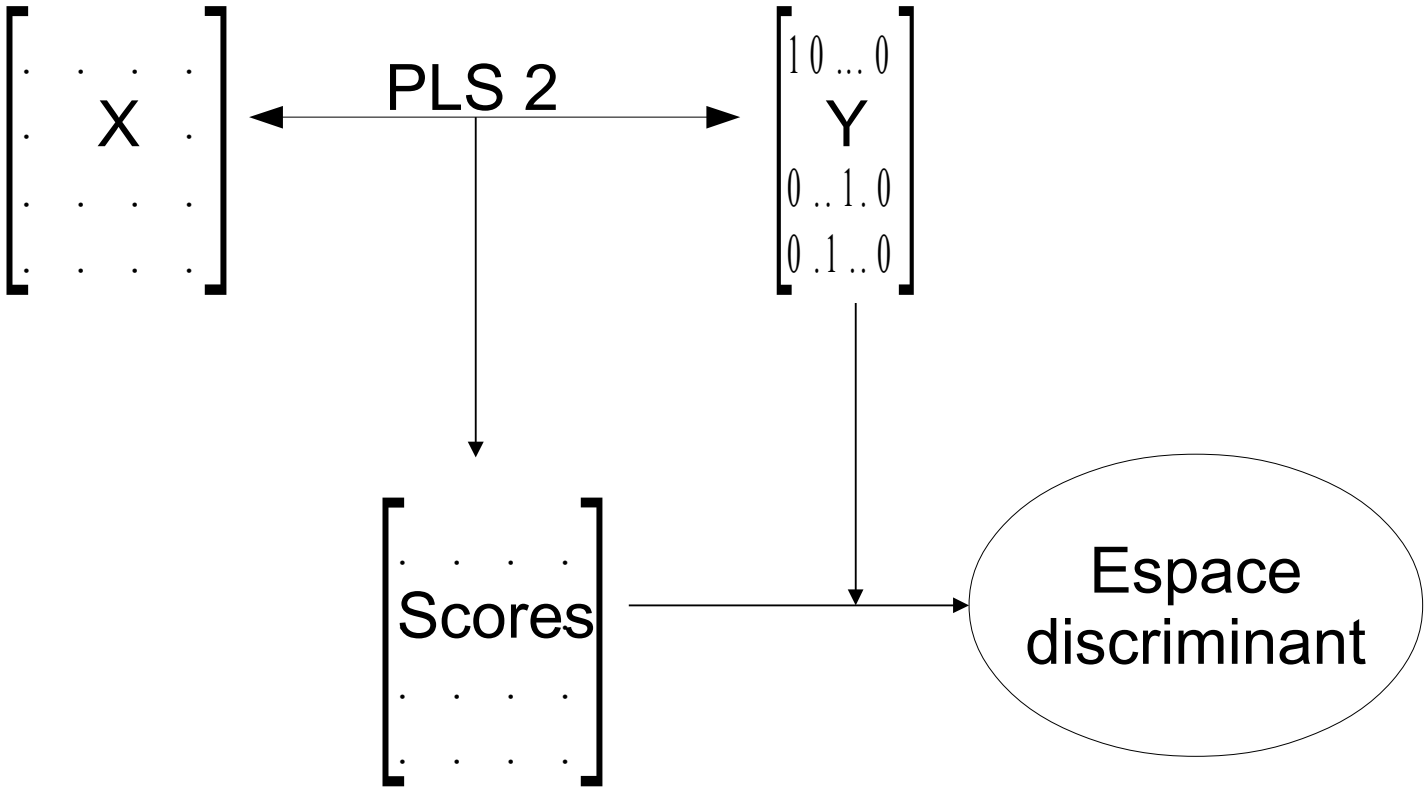
[1]Biometrics and Statistical Sciences, The Procter and Gamble Company, Mason, OH 45040, USA
[2]Department of Statistics, University of Kentucky, Lexington, KY 40506, USA

Received 25 April 2000; Revised 29 October 2002; Accepted 18 November 2002

Partial least squares (PLS) was not originally designed as a tool for statistical discrimination. In spite of this, applied scientists routinely use PLS for classification and there is substantial empirical evidence to suggest that it performs well in that role. The interesting question is: why can a procedure that is principally designed for overdetermined regression problems locate and emphasize group structure? Using PLS in this manner has heurestic support owing to the relationship between PLS and canonical correlation analysis (CCA) and the relationship, in turn, between CCA and linear discriminant analysis (LDA). This paper replaces the heuristics with a formal statistical explanation. As a consequence, it will become clear that PLS is to be preferred over PCA when discrimination is the goal and dimension reduction is needed. **Copyright © 2003 John Wiley & Sons, Ltd.**

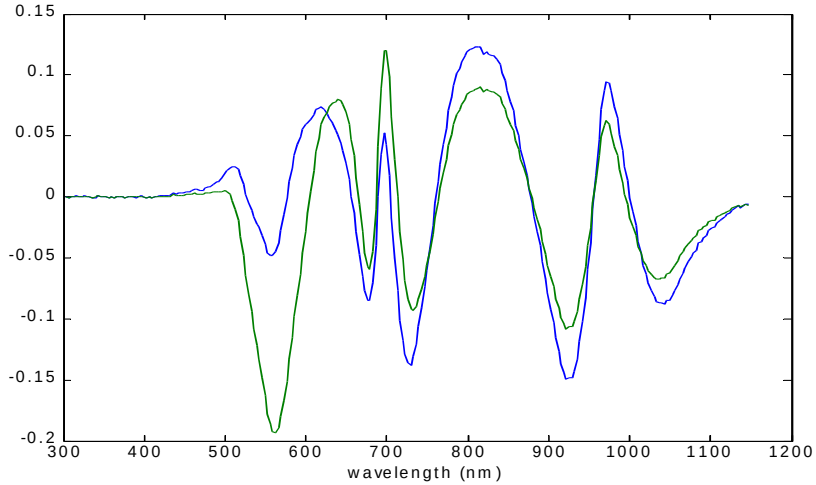
PLS2 + Analyse Discriminante

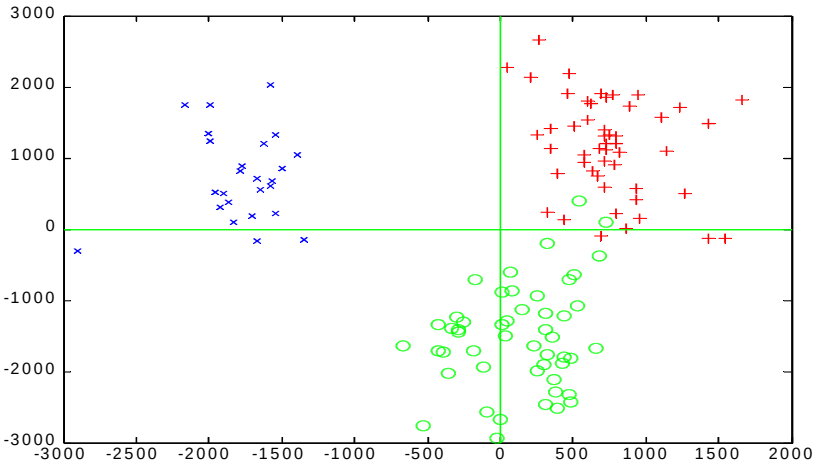
PLS-DA

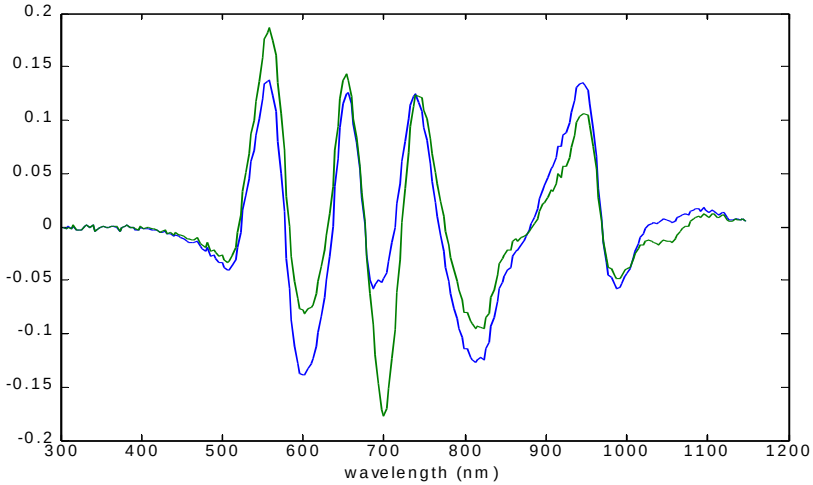


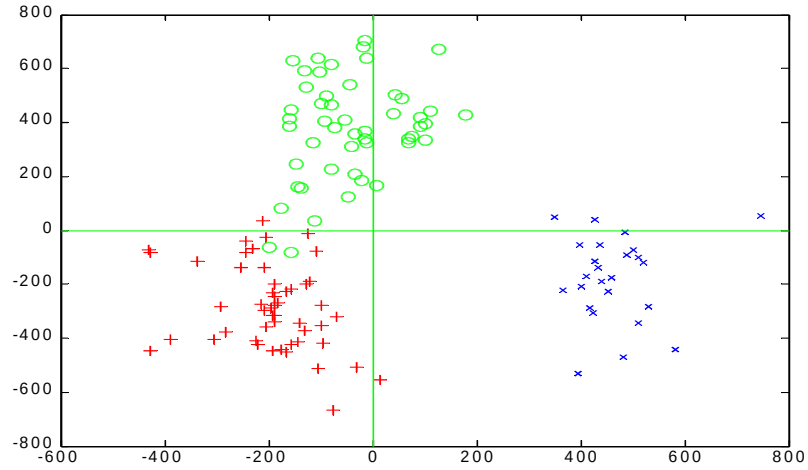
Variables latentes + Analyse Discriminante

exemple sur 3 cépages









PCA-DA 8 var : Λ=0,82

PLS-DA 8 var : Λ=0,84

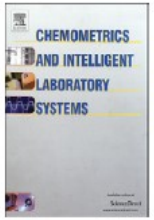
Sélection de variables + Analyse Discriminante

Chemometrics and Intelligent Laboratory Systems xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab



Model calibration and feature selection for orange juice authentication by ^{1}H NMR spectroscopy

Evelyne Vigneau [a,b,*], Freddy Thomas [c]

[a] L'UNAM University, Oniris, "Sensometrics and Chemometrics Laboratory", Nantes, F-44322, France
[b] INRA, Nantes, F-44316, France
[c] EUROFINS Analytics, Nantes, F-44323, France

ARTICLE INFO

Article history:
Received 25 February 2011
Received in revised form 29 April 2011
Accepted 9 May 2011
Available online xxxx

Keywords:
Orange juice
Authentication
^{1}H NMR spectroscopy
Data pretreatment
Classification model
Feature selection

ABSTRACT

A ^{1}H NMR spectroscopic profiling approach has been investigated to discriminate between authentic and adulterated juices. An experimental database of 150 samples of authentic or adulterated orange juices, with a known percentage of clementine juice, was prepared. A repeated stratified cross-validation process was adopted for the validation of PLS regression models and classification rules. The choice of a type of statistical data pre-treatment was discussed. The result was that logarithmic transformation combined with Pareto scaling was the most relevant. The selection of spectral variables has also proven to lead to better results than using the whole spectral range. Various feature selection procedures were compared. The *CovSel* approach appeared to be the most efficient. However, for a better understanding of the features of atypical profiles, it would be wise to use more than one selection procedure like the ones based on Backward Interval PLS regression approach or on the genetic algorithms.

© 2011 Published by Elsevier B.V.

Sélection de variables + Analyse Discriminante

- Méthodes de type « filter »
- On applique un filtre $\phi(x_i)$ sur chaque variable de X et on ne garde que celles qui sont au dessus d'un seuil
- Par analogie avec la régression et le R^2, on peut employer le Λ de Wilks

Peut être utile pour « visualiser » les zones d'intérêt

N'est pas efficace car ignore la structure des données

Sélection de variables + Analyse Discriminante

- Méthodes de type « wrapper »
  - On parcours l'espace des sélections
  - On soumet le modèle à ces sélections
  - On cherche celle(s) qui donne le(s) meilleur(s) résultats

  - Différentes heuristiques de parcours
    - Algorithmes génétiques
    - Recuit simulé
    - Descente de gradient
    - ...

Sélection de variables + Analyse Discriminante

- Méthodes de type « embedded » : la sélection se fait en même temps que l'optimisation

  - Par exemple, Stepwise :

    - Introduction pas à pas des variables
    - Optimisation du Λ de Wilks

Sélection de variables

adaptée à la spectrométrie

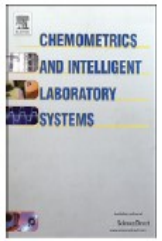
et à la discrimination

Chemometrics and Intelligent Laboratory Systems 106 (2011) 216–223

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab



CovSel: Variable selection for highly multivariate and multi-response calibration
Application to IR spectroscopy

J.M. Roger [a,*], B. Palagos [a], D. Bertrand [b], E. Fernandez-Ahumada [a]

[a] Cemagref BP 5095, 34033 Montpellier Cedex 1, France
[b] INRA BP 71627, 44316 Nantes Cedex 3, France

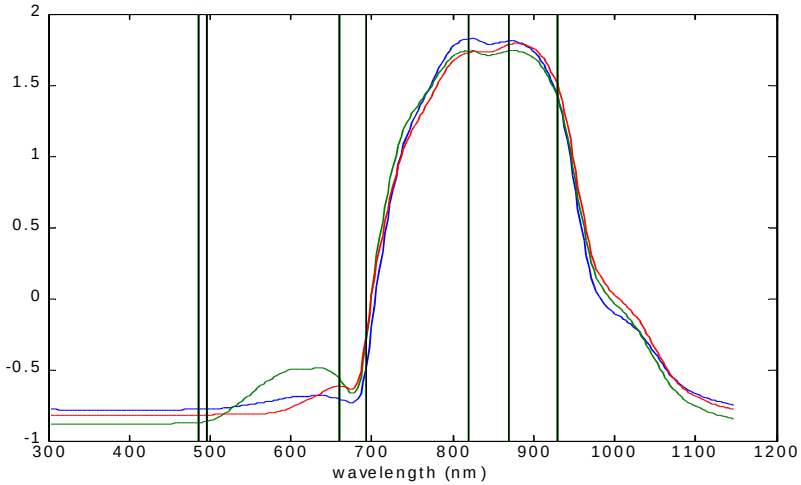
Sélection PLS-like

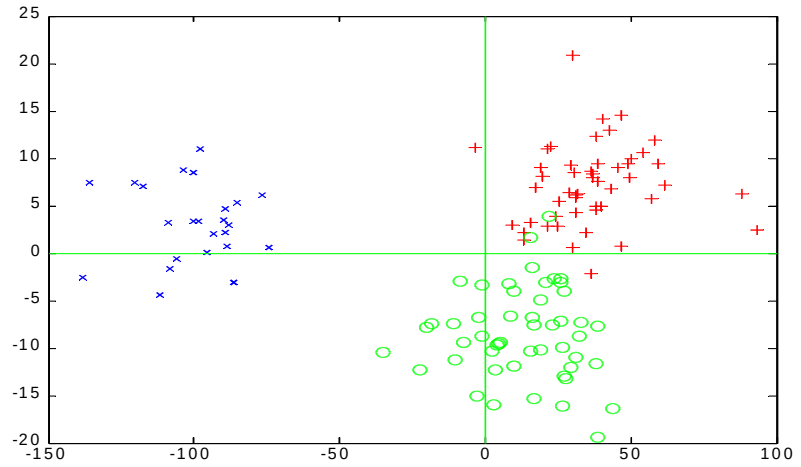
Optimisation de la covariance

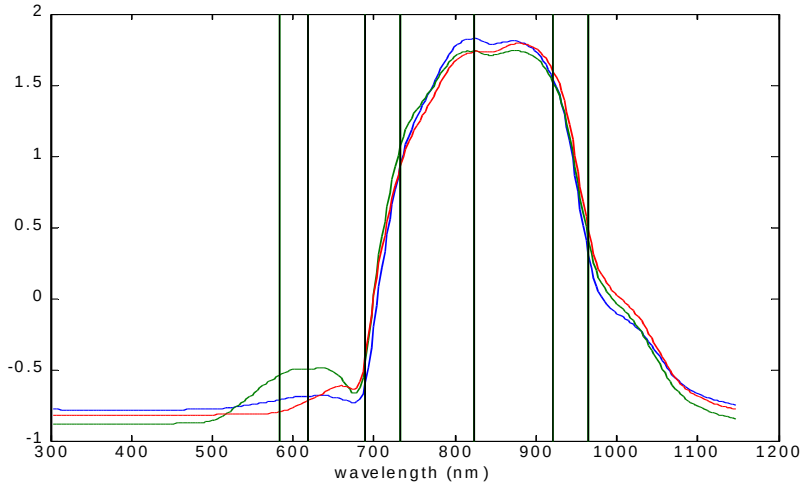
Garantie de sélection répartie

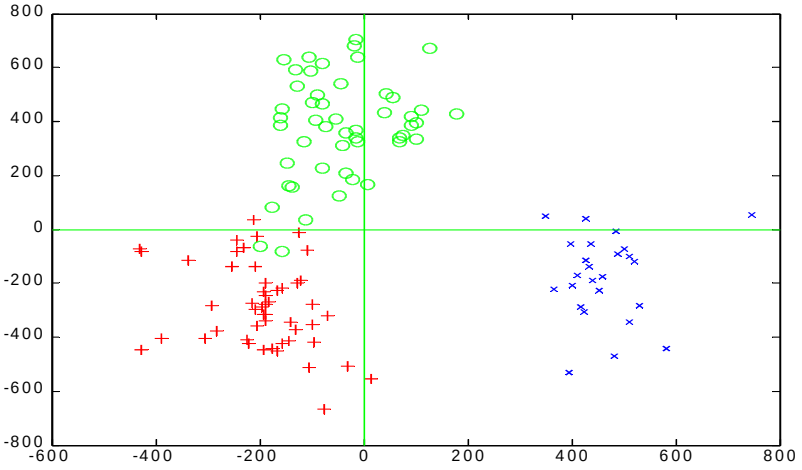
Sélection de variables + Analyse Discriminante

exemple sur 3 cépages









Stepwise 7 var : Λ=0,89

CovSel 7 var : Λ=0,81

Des méthodes alternatives

Prédiction directe des degrés d'appartenance par régression linéaire

$$\hat{y} = x^T B$$

+ Tire profit de la capacité de la PLS2 à prédire plusieurs réponses
+ Modèle simple

- Ne gère pas (directement) la contrainte de somme, ni de limite (0 – 1)
- Ce qui est prédit n'est pas un degré d'appartenance

À la base de la PLS-DA de « the unscrambler » ?

Méthodes continuum

à la recherche directe de l'espace discriminant

Rev. Statistique Appliquée, 2005, LIII (2), 29-41

APPROCHE CONTINUUM DE LA DISCRIMINATION DE TYPE RIDGE

H. NOCAIRI, M. HANAFI, E.M. QANNARI

ENITIAA / INRA,
Unité de Sensométrie et de Chimiométrie,
rue de la Géraudière, BP 82228,
44322 NANTES Cedex 03, France
hanafi@ enitiaa-nantes.fr

Vecteurs propres de : $(\mathrm{T}+k\mathrm{I})^{-1}\mathrm{B}$; $k\in[0,+\infty]$

Méthodes continuum

à la recherche directe de l'espace discriminant

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Chemometrics and Intelligent Laboratory Systems 79 (2005) 31 – 41

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemolab

Discriminating from highly multivariate data by Focal Eigen Function discriminant analysis; application to NIR spectra

J.M. Roger*, B. Palagos, S. Guillaume, V. Bellon-Maurel

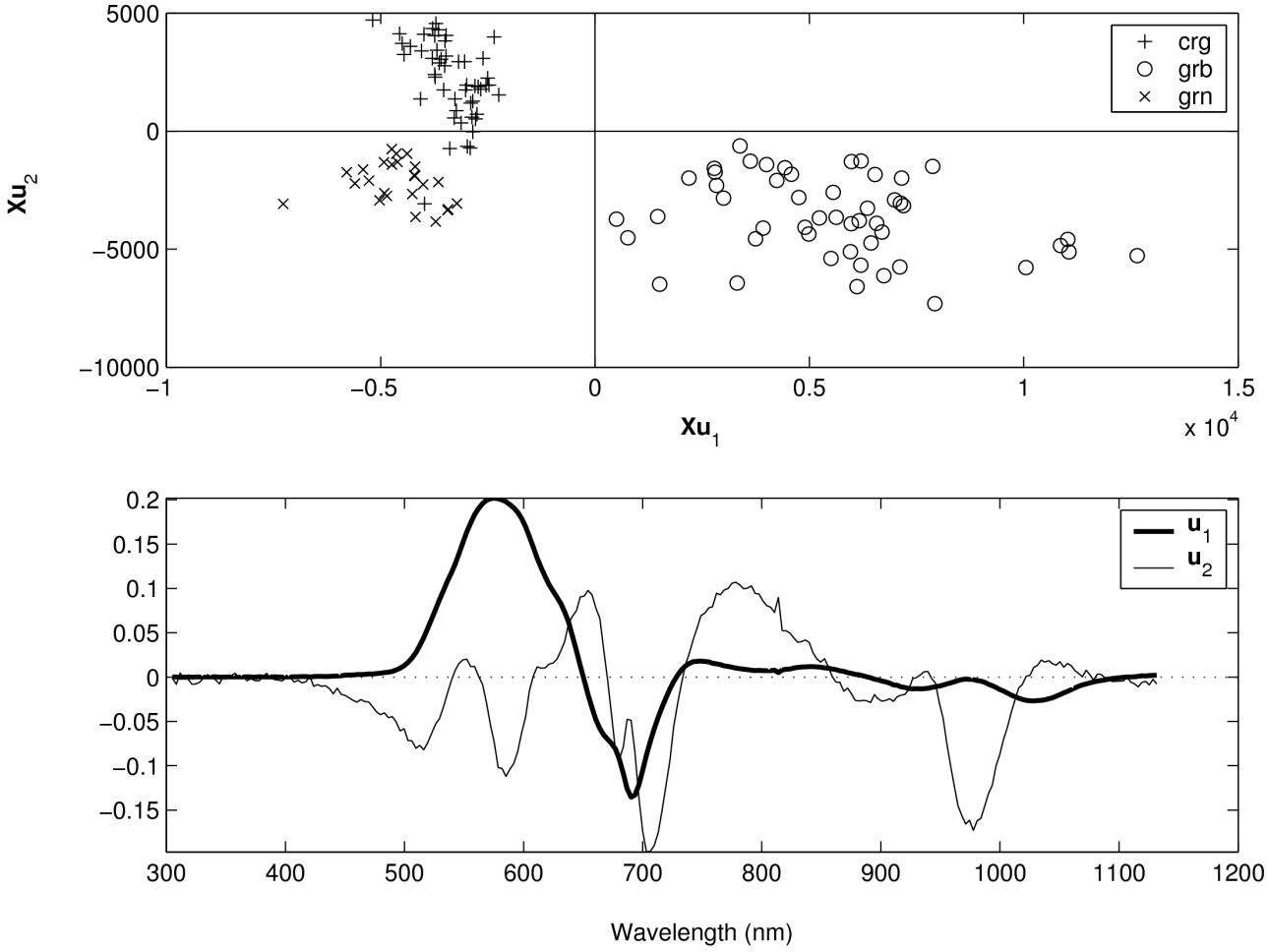
Cemagref BP 5095-34033 Montpellier Cedex1, France

Received 15 September 2004; received in revised form 28 January 2005; accepted 7 March 2005
Available online 21 June 2005

Vecteurs propres de : $(\mathrm{B} - k\mathrm{T})$; $k \in [0,1]$

Méthodes continuum

FEF-DA



Méthodes utilisant

un modèle par classe

PLS binaire : chaque classe est discriminée de toutes les autres classes

Pour un nouvel individu, tous les modèles sont testés, fournissant autant de degrés d'appartenance

SIMCA : un modèle ACP est calculé pour chaque classe

Chaque nouvel individu est projeté dans l'espace local de chaque classe et l'examen des résidus et des qualités de projection détermine le meilleur modèle

Un autre point de vue,

compatible avec la spectrométrie

S'intéresser aux VP de $T^{-1}B$ ou de $W^{-1}B$ revient à dire que l'on s'intéresse à la variance inter-classes **SANS** la variance intra-classes

T^{-1} ou W^{-1} réalise une « pondération » qui **réduit** l'effet de ce qui est responsable de la dispersion des classes

En poussant le raisonnement à l'extrême, il suffit d'enlever le sous espace qui porte W, puis de diagonaliser B …

Ce point de vue résonne avec les notions d'espaces utile et nuisible et de hard / soft projections
(thèse de JC Boulet)

Conclusion

L'analyse discriminante linéaire offre un cadre formel intéressant :

Sous espace discriminant de dimension réduite

Possibilité de mettre en oeuvre des règles d'affectation

Son utilisation sur les données spectrales requiert des adaptations

Variables latentes, sélection de variables

Des alternatives pour la construction de l'espace discriminant sont encore recherchées.