

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

RÉSUMÉ

A l'instar de la régression linéaire, l'analyse discriminante conduit à des modèles instables en présence de prédicteurs quasi-colinéaires. Une approche continuum basée sur la discrimination à partir de variables latentes est proposée. Elle recouvre aussi bien l'analyse factorielle discriminante que l'analyse discriminante de type *PLS* et de type *Ridge*. L'approche est illustrée sur la base d'exemples.

Mots-clés : *Analyse discriminante, approche continuum, discrimination PLS, discrimination Ridge*

ABSTRACT

Similarly to linear regression, discriminant analysis leads to unstable models in presence of quasi-collinearity among predictors. A continuum approach based on discrimination using latent variables is discussed. It encompasses Fisher's linear discriminant analysis, *PLS* discriminant analysis and *Ridge* discriminant analysis. The methods of analysis are illustrated on the basis of real and simulated data.

Keywords : *Discriminant analysis, continuum approach, PLS discrimination, Ridge discrimination.*

1. Introduction

L'analyse discriminante et la régression linéaire sont parmi les méthodes statistiques les plus utilisées pour la prédiction à partir d'un ensemble de variables. La régression linéaire s'attache à prédire une ou plusieurs variables quantitatives, tandis que l'analyse discriminante vise à prédire une variable qualitative. Cependant, il est bien connu que l'une comme l'autre de ces deux méthodes conduit à des modèles instables en présence de prédicteurs quasi-colinéaires.

Parmi les méthodes proposées comme alternatives à la régression linéaire, nous pouvons citer la régression de type *Ridge* (Hoerl et Kennard, 1970; Cazes, 1975), la

régression *PLS* (Stahle et Wold, 1987; Wold, 1995; Tenenhaus, 1998) et l'approche continuum (Stone et Brooks, 1990; De Jong *et al.*, 2001). Le problème engendré par la quasi-colinéarité dans le cadre de l'analyse discriminante a été relativement moins étudié que dans le cadre de la régression linéaire. Parmi les méthodes proposées dans ce contexte, nous pouvons citer l'analyse discriminante régularisée (Friedman, 1989), l'analyse discriminante *PLS* (Sjöström *et al.*, 1986; Kemsley, 1996; Barker et Rayens, 2003; Nocairi *et al.*, 2005) et l'analyse discriminante de type *Ridge* (Smid et McDonald, 1976).

À partir d'une nouvelle formulation de l'analyse factorielle discriminante et de l'analyse discriminante *PLS*, nous proposons une approche continuum qui, outre le fait qu'elle présente un aspect synthétique, offre tout un éventail de modèles. Nous pouvons également souligner la simplicité de la méthode du point de vue conceptuel et de la mise en œuvre. Des propriétés liées à la méthode sont démontrées permettant de jeter un éclairage supplémentaire sur l'intérêt de l'approche continuum. Enfin, nous illustrons la méthode de discrimination proposée sur la base de données réelles et de données simulées.

2. Méthode

2.1. Analyse factorielle discriminante

Nous disposons d'un tableau de données X formé de p variables quantitatives x_1, \dots, x_p et d'une variable qualitative C qui indique l'appartenance de n individus à q groupes connus *a priori*. Dans la suite, les variables quantitatives sont supposées centrées.

Afin de discriminer les groupes induits par la variable qualitative C , nous cherchons, dans un premier temps, une variable latente $z = Xa$ de sorte à maximiser la corrélation entre z et $\mathbb{E}(z/C)$ qui est l'espérance conditionnelle de z par rapport à C . Le fondement de cette approche a été justifié par Nocairi *et al.* (2005). Nous avons :

$$\text{Cor}^2(z, \mathbb{E}(z/C)) = \frac{a^T B a}{a^T T a},$$

où T et B désignent respectivement les matrices de variances-covariances totales et inter-classes. Cette quantité, connue sous l'appellation rapport de corrélation, est un indicateur de discrimination. Le vecteur qui maximise ce rapport est donné par le vecteur propre de $T^{-1}B$ associé à la plus grande valeur propre.

Afin de déterminer une deuxième variable discriminante, nous considérons la matrice X^* formée par les résidus de la régression des variables x_j , ($j = 1, \dots, p$) sur la variable z . Par la suite, la même démarche explicitée précédemment est effectuée sur la base de X^* . Ce procédé peut être répété plusieurs fois afin de déterminer d'autres variables latentes. Nous pouvons remarquer que du fait de la non corrélation en analyse factorielle discriminante des variables discriminantes, l'approche proposée correspond à l'analyse factorielle discriminante usuelle (annexe 1).

Les variables discriminantes peuvent être utilisées dans une perspective de description ou dans une perspective de classement de nouveaux individus. Pour ce

dernier cas, il est préconisé de considérer les coordonnées factorielles de tout nouvel individu et de l'affecter au groupe dont le centre de gravité est le plus proche au sens de la métrique de Mahalanobis. En fait, comme les variables discriminantes sont orthogonales, le choix de la métrique de Mahalanobis revient en définitive à considérer les coordonnées factorielles standardisées.

Pour le choix du nombre de variables latentes à inclure dans le modèle de discrimination, nous pouvons effectuer une procédure de validation croisée du type Jackknife en cherchant le nombre de variables latentes correspondant au maximum d'individus bien classés.

2.2. Analyse discriminante PLS

La détermination des variables discriminantes implique l'inversion de la matrice T , ce qui est susceptible de causer des problèmes d'instabilité dans le modèle d'affectation et des difficultés d'interprétation. L'objectif de ce paragraphe est de pallier ce problème en utilisant une approche de type PLS.

Au lieu de maximiser la corrélation entre $z = Xa$ et $\mathbb{E}(z/C)$, nous proposons de maximiser leur covariance. Nous avons :

$$\text{Cov}(z, \mathbb{E}(z/C)) = a^T B a.$$

Il s'ensuit qu'en imposant la contrainte de détermination $a^T a = 1$, la solution optimale pour a est donnée par le vecteur propre de la matrice B associé à la plus grande valeur propre. Etant donné la similarité de cette approche avec celle poursuivie en régression PLS, nous désignons la démarche de discrimination basée sur B par l'analyse discriminante PLS (AD-PLS). D'autres variables discriminantes peuvent être déterminées selon le même schéma que précédemment en considérant les résidus de la régression des variables de X sur les variables latentes déterminées dans les étapes précédentes.

Il est à noter que la diagonalisation de la matrice B au lieu de la matrice $T^{-1}B$ dans le cadre de l'analyse discriminante a été préconisée par Cazes (1981). L'intérêt du travail présenté dans ce paragraphe est d'apporter de nouvelles justifications à cette idée et de la placer dans un contexte, plus récent, qui est celui de démarche PLS. Il faut remarquer, à ce propos, qu'une autre version de l'analyse discriminante PLS a été proposée (Sjöström *et al.*, 1986; Kemsley, 1996). Elle est basée sur la régression PLS2 en considérant pour matrice Y à prédire, le tableau formé par les variables indicatrices associées aux groupes. Comme cela a été souligné par Barker et Rayens (2003), cette démarche conduit à une forme illogique de la matrice à diagonaliser en ce sens qu'elle ressemble à une matrice de variances-covariances inter-classes, excepté que chaque centre de gravité est pondéré par le carré de l'effectif du groupe associé.

Comme pour l'analyse factorielle discriminante, les variables latentes déterminées à partir de l'AD-PLS peuvent servir à des fins de description ou de classement de nouveaux individus. Le nombre de variables à retenir pour le classement des individus dans les différents groupes peut être déterminé par validation croisée.

2.3. Discrimination continuum de type Ridge

Entre l'*AFD* qui est basée sur l'analyse spectrale de $T^{-1}B$ et l'*AD-PLS* basée sur la diagonalisation de B , nous définissons une famille de méthodes de discrimination basée sur la matrice :

$$\begin{cases} (T_\alpha)^{-1}B, \\ T_\alpha = (1 - \alpha).T + \alpha.I, \end{cases}$$

où α est un réel compris entre 0 et 1 et I est la matrice identité.

Lorsque α croît, la méthode de discrimination correspond à une contraction progressive de T^{-1} vers l'identité; les deux situations extrêmes étant l'*AFD* ($\alpha = 0$) et l'*AD-PLS* ($\alpha = 1$).

Cette procédure s'apparente à une stratégie de discrimination de type *Ridge* (Smid et McDonald, 1976) car la matrice à diagonaliser $[(1 - \alpha).T + \alpha.I]^{-1}B$ est proportionnelle à une matrice qui peut s'écrire sous la forme :

$$[T + k(\alpha).I]^{-1}B,$$

où $k(\alpha) = \frac{\alpha}{1 - \alpha}$ est une fonction qui constitue une bijection de $[0, 1[$ sur \mathbb{R}^+ .

De manière concrète, la démarche préconisée consiste à déterminer une première variable discriminante $z = Xa$, où a est un vecteur propre normé de $T_\alpha^{-1}B$ associé à la plus grande valeur propre. Par la suite et de manière similaire à ce qui a été décrit précédemment, nous recommandons la même procédure en considérant les résidus de la régression des variables de X sur z . Ainsi, de proche en proche, nous pouvons déterminer plusieurs variables discriminantes.

Les avantages de cette approche par rapport à une démarche de discrimination *Ridge* classique sont :

- le paramètre α est compris entre 0 et 1, ce qui facilite l'exploration des différentes solutions offertes par la méthode;
- la démarche discutée ici procure une stratégie pour définir plusieurs variables discriminantes qui sont orthogonales;
- la démarche procure une approche synthétique qui regroupe l'analyse factorielle discriminante et l'analyse discriminante *PLS*.

Une autre manière de justifier l'approche continuum pour la discrimination consiste à remarquer que nous pouvons reformuler les différentes méthodes de discrimination en considérant la même fonction à optimiser, à savoir $a^T B a$, et différentes contraintes de détermination. Pour l'*AFD*, la contrainte est $a^T T a = 1$. Pour l'*AD-PLS*, la contrainte est $a^T a = 1$. Avec la première contrainte (*AFD*), il y a le risque, bien connu, qu'une variable latente de faible variance et probablement liée au bruit émerge comme ayant un bon pouvoir discriminant. Avec la deuxième contrainte (*AD-PLS*), il y a le risque de déterminer des variables latentes ayant une variance relativement grande et un pouvoir discriminant relativement réduit. Une manière de

procéder consiste à considérer un compromis entre les deux approches en cherchant à maximiser le même critère ($a^T B a$) sous la contrainte $(1 - \alpha)a^T T a + \alpha a^T a = 1$, où α est un réel compris entre 0 et 1. De la sorte, il est possible d'explorer un continuum de solutions entre l'*AFD* et l'*AD-PLS*. Ce problème d'optimisation conduit à l'analyse spectrale de l'opérateur $T_\alpha^{-1} B$.

Les propriétés suivantes vont jeter davantage de lumière sur le fondement de la méthode.

2.4. Propriétés

Soit a_α ($a_\alpha^T a_\alpha = 1$) le vecteur des coefficients de la première variable discriminante obtenue par la discrimination continuum. De manière plus précise, a_α est un vecteur propre normé de $T_\alpha^{-1} B$ associé à la plus grande valeur propre.

Soit $V_T(\alpha) = a_\alpha^T T a_\alpha$, la variance totale associée à la variable $z_\alpha = X a_\alpha$. Soit $V_B(\alpha) = a_\alpha^T B a_\alpha$, la variance inter-classes de z_α et, enfin, soit $I(\alpha) = \frac{a_\alpha^T B a_\alpha}{a_\alpha^T T a_\alpha} = \frac{V_B(\alpha)}{V_T(\alpha)}$ le rapport de corrélation de la variable z_α par rapport à la variable qualitative C .

Nous avons les propriétés suivantes :

V_T et V_B sont des fonctions croissantes de α , alors que I est une fonction décroissante de α .

La démonstration de ces propriétés se trouve en annexe 2. L'interprétation des propriétés est sans équivoque : lorsque α croît, la démarche explore des directions « plus stables » (plus grande inertie) mais le pouvoir discriminant se dégrade. De là, il apparaît que la notion de compromis prend tout son sens.

De fait, la discrimination continuum de type *Ridge* dépend de deux paramètres à savoir le paramètre d'ajustement α et le nombre de variables latentes à retenir. Dans le cadre d'une procédure de classement des individus, les deux paramètres peuvent être déterminés par validation croisée. Cela consiste à fixer des valeurs de α entre 0 et 1 et déterminer, pour chaque valeur, le nombre optimal r_α de variables latentes par validation croisée. Une méthode de validation consiste à mettre à l'écart, à tour de rôle, chacun des individus, établir un modèle de discrimination à l'aide des individus restants et prédire l'appartenance de l'individu écarté à une des classes. C'est cette méthode que nous avons adoptée. Les performances des modèles associés aux différents couples (α, r_α) sont comparées sur la base du pourcentage d'individus bien classés et le couple de paramètres associé à la meilleure performance est retenu.

3. Applications

3.1. Données réelles

Nous considérons une expérience dans laquelle l'objectif est de discriminer neuf variétés d'huiles d'olives provenant de neuf régions d'Italie. Les données concernent 572 individus (huiles d'olives) sur lesquels ont été mesurées huit variables

correspondant aux teneurs en huit acides gras. Ces données ont été présentées et analysées par Forina et Tiscornia (1982) et Forina *et al.*, (1983).

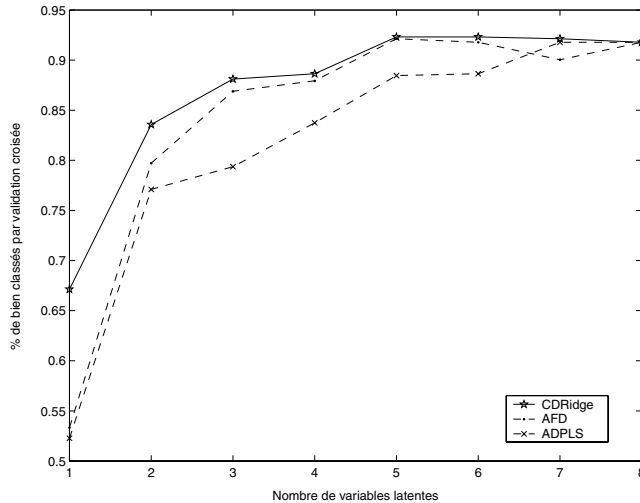


FIGURE 1

Pourcentages d'individus bien classés par validation croisée pour l'approche par validation croisée (CD-Ridge), l'AFD et l'AD-PLS en fonction du nombre de variables latentes introduites dans le modèle

Afin de discriminer les huiles des différentes régions, nous avons appliqué l'approche continuum de discrimination, l'AFD et l'AD-PLS. Nous avons fait varier le nombre de variables latentes (discriminantes) et pour chaque nombre, nous avons déterminé la valeur de α optimale en cherchant à maximiser le pourcentage d'individus bien classés selon la procédure de validation croisée définie ci-dessus. Les résultats sont reportés sur la figure 1. Nous avons également reporté sur cette figure, le pourcentage d'individus bien classés en fonction du nombre de variables discriminantes introduites dans le modèle de classement déterminé soit par AFD soit par AD-PLS. Il ressort que si, *in fine*, les trois modèles donnent la même performance lorsque huit variables latentes sont introduites, il est clair que dès les premières variables latentes l'approche continuum a une meilleure performance que les autres méthodes. Ceci signifie, en particulier, qu'une représentation graphique basée sur les premières variables latentes de l'approche continuum est à même de mieux refléter le phénomène de discrimination que les mêmes graphiques obtenus à l'aide des autres méthodes.

3.2. Données simulées

Nous avons considéré un tableau de données, X , contenant 40 lignes et 27 colonnes. Nous avons effectué une analyse en composantes principales sur X . Sur la base de la représentation graphique du premier plan principal, nous avons constitué quatre groupes correspondant aux quatre quadrants du plan factoriel. Par la suite, nous

avons essayé de retrouver ces quatre groupes en effectuant l'approche continuum de discrimination sur X , l'*AFD* et, enfin, l'*AD-PLS*.

Dans un deuxième temps, nous avons simulé une deuxième partition des individus en quatre groupes sur la base, cette fois-ci, d'un plan principal associé à des composantes principales de faibles inerties (7^{ième} et 8^{ième} composantes). Par la suite, nous avons essayé de retrouver ces groupes à l'aide de l'approche continuum de discrimination, de l'*AFD* et de l'*AD-PLS*. Sur la figure 2-a (resp. 2-b) nous avons représenté les résultats concernant le premier (resp. deuxième) tableau de données simulé. Ces résultats concernent les pourcentages d'individus bien classés obtenus par validation croisée en fonction du nombre de variables latentes introduites dans le modèle. Dans la première étude de simulation, l'approche continuum de discrimination et l'*AD-PLS* ont une performance largement meilleure que celle de l'*AFD*. L'approche continuum présente une performance légèrement supérieure à celle de l'*AD-PLS* (fig. 2-a). Pour la deuxième étude de simulation, c'est l'*AD-PLS* qui présente une performance médiocre et l'approche continuum est nettement meilleure que les deux autres méthodes surtout pour un nombre de variables latentes inférieur à 8 (fig. 2-b).

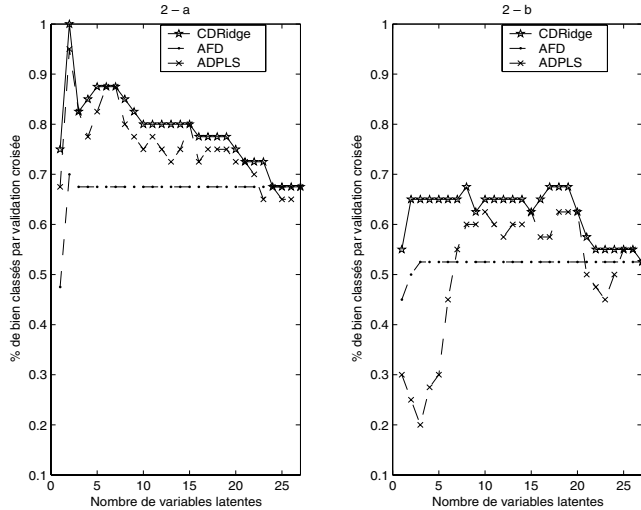


FIGURE 2

Pourcentages d'individus bien classés par validation croisée.

2-a : groupes simulés à partir de composantes principales de fortes inerties.

2-b : groupes simulés à partir de composantes principales de faibles inerties.

Dans la figure 3, nous avons représenté l'évolution de la variance totale, de la variance inter-classes et du rapport de corrélation associés à la première variable discriminante en fonction du paramètre α . De manière plus précise, nous avons considéré les logarithmes des variances totale et inter-classes car cela permet de mieux visualiser le comportement de ces fonctions au voisinage de zéro. Dans la figure 3-a, qui correspond au premier tableau de données simulées, il apparaît que les logarithmes des variances inter-classes et totales croissent assez rapidement sans que

cela se traduise par une nette dégradation du pouvoir discriminant comme cela est reflété par l'évolution du rapport de corrélation. Le graphique semble indiquer qu'une valeur de α autour de 0.2 est un bon compromis (la valeur optimale obtenue par validation croisée est égale à 0.179). Dans la figure 3-b, qui correspond au deuxième tableau de données simulées, il apparaît que les logarithmes des variances inter-classes et totale ne présentent pas une croissance rapide. Par contre le rapport de corrélation décroît assez rapidement et atteint une valeur relativement faible pour $\alpha = 1$ (*AD-PLS*). Ce graphique semble indiquer qu'une valeur α autour de 0.3 constitue un bon compromis (la valeur optimale obtenue par validation croisée est égale à 0.246).

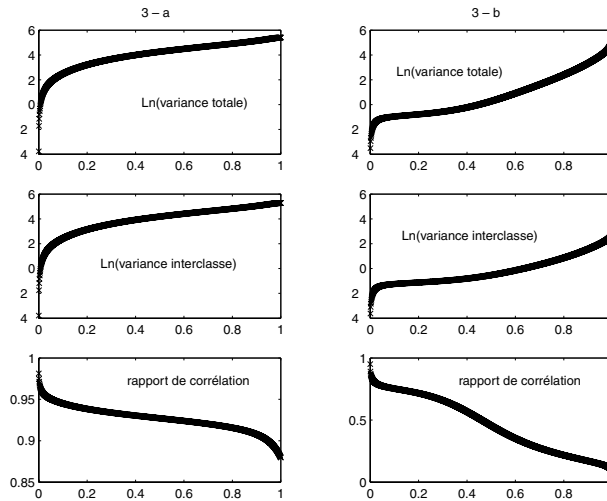


FIGURE 3

Évolution des logarithmes des variances totale et inter-classes et du rapport de corrélation de la première variable latente en fonction des valeurs de α pour les deux tableaux de données simulées

4. Conclusion

L'approche continuum de discrimination discutée dans ce papier présente un aspect synthétique car elle permet de regrouper plusieurs méthodes : *AFD*, *AD-PLS* et discrimination Ridge. Sur un plan conceptuel et un plan pratique, elle présente une grande simplicité. Les propriétés liées à cette approche montrent clairement qu'elle explore tout un éventail de solutions qui constituent des compromis entre avoir une bonne discrimination et une bonne stabilité des résultats comme cela a été illustré à l'aide des applications.

L'idée du continuum est, au départ, intuitive car elle vise à établir un pont entre l'*AFD* et l'*AD-PLS*. Nous avons essayé de justifier cette démarche de différentes manières. Cependant, cet aspect mérite d'être davantage exploré en considérant des approches de discrimination sous contraintes ou une approche bayésienne, à l'instar

de ce qui a été fait dans le cadre de la régression de type Ridge (Cazes, 1975; Gibbons et McDonald, 1984).

Références

- BARKER M. et RAYENS W. (2003), PLS for discrimination. *Journal of Chemometrics*, vol. 17, 3, 153-197.
- CAZES P. (1975), Protection de la régression par utilisation de contraintes linéaires et non linéaires. *Revue de Statistique Appliquée*, vol. XXIII, 3, 37-57.
- CAZES P. (1981), L'analyse de certains tableaux rectangulaires décomposés en blocs : généralisation des propriétés rencontrées dans l'étude des correspondances multiples. III. Codage simultané de variables qualitatives et quantitatives [ANA BLOCS III]. *Cahiers de l'analyse des données*, vol. 6, 1, 9-18.
- DE JONG S., WISE B. M. et RICKER N. L. (2001), Canonical partial least squares and continuum power regression. *Journal of Chemometrics*, vol. 15, 85-100.
- FORINA M. et TISCORNIA E. (1982), Pattern recognition methods in the prediction of italian olive oil origin by their fatty acid content. *Annali di Chimica*, 72, 143-155.
- FORINA M., ARMANINO C., LANTERIS S. et TISCORNIA E. (1983), Classification of olive oils from their acid composition in food research and data analysis, *Martens H. and Russwurm H. Jr., eds, Applied Science Publishers Ltd, Barking, England*, 189-214.
- FRIEDMAN J. H. (1989), Regularized discriminant analysis. *JASA*, vol. 84, 405, 165-175.
- GIBBONS D. I. et MCDONALD G. C. (1984), A rational interpretation of the Ridge trace. *Technometrics*, vol. 26, 4, 339-346.
- HOERL A. E. et KENNARD R. W. (1970), Ridge Regression : biased estimation for nonorthogonal Problems. *Technometrics*, vol. 12, 1, 55-67.
- KEMSLEY E. K. (1996), Discriminant analysis of high-dimensional data : a comparison of principal components analysis and partial least squares data reduction methods. *Chemometrics and Intelligent Laboratory systems*, 33, 47-61.
- NOCAIRI H., QANNARI E. M., VIGNEAU E. et BERTRAND D. (2005), Discrimination on latent variables with respect to patterns with multi-collinear data. *Computational Statistics and Data Analysis*, 48, 139-147.
- SJÖSTRÖM M., WOLD S. et SÖDERSTRÖM B. (1986), PLS discriminant plots. In : E. S. Gelsema and L. N. Kanal Eds., *Pattern recognition in practice II*, Elsevier, Amsterdam, p. 486.
- SMIDT R. K. et MCDONALD L. L. (1976), Ridge discriminant analysis, *University of Wyoming department of Statistics. Technical Report No, 108, Laramie, Wyoming*.

- STAHLE L. et WOLD S. (1987), Partial least squares analysis with cross-validation for the two-class problem : a Monte Carlo study. *Journal of Chemometrics*, 1, 185-196.
- STONE M. et BROOKS R. J. (1990), Continuum regression : Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *J. R. Statist Soc.*, B 52, 2, 237-269.
- TENENHAUS M. (1998), *La régression PLS : théorie et pratique*. Editions Technip, Paris.
- WOLD S. (1995), PLS for multivariate linear modelling. In : H. van de Waterbeemd Ed., *QSAR : chemometric methods in molecular design, methods and principles in medicinal chemistry, vol. 2*, Verlag Chemie, Weinheim, Germany.

Annexe 1

Soit $z_k = Xa_k$ ($k = 1, 2, \dots, r = \min(q - 1, p)$) les variables discriminantes standardisées de l'AFD. Il est bien connu que les vecteurs a_k ($k = 1, 2, \dots, r$) sont les vecteurs propres de la matrice $T^{-1}B$:

$$Ba_k = \lambda_k Ta_k, \quad \text{où } \lambda_k = \frac{a_k^T Ba_k}{a_k^T Ta_k} \text{ pour } k = 1, 2, \dots, r.$$

Dans la présente annexe, nous montrons que la variable discriminante z_k ($k = 2, \dots, r$) correspond à la première variable discriminante de l'AFD effectuée sur le tableau $X_{(k-1)}$ formé des résidus des régressions des variables de X sur z_1, z_2, \dots, z_{k-1} . Nous désignons par $T_{(k-1)}$ et $B_{(k-1)}$ les matrices de variances covariances totales et inter-classes associées à $X_{(k-1)}$. Nous allons montrer que pour $m = 1, \dots, r$:

$$B_{(k-1)}a_m = \begin{cases} 0 & \text{si } m = 1, \dots, k - 1, \\ \lambda_m T_{(k-1)}a_m & \text{si } m = k, k + 1, \dots, r. \end{cases}$$

et

$$X_{(k-1)}a_m = \begin{cases} 0 & \text{si } m = 1, \dots, k - 1, \\ z_m & \text{si } m = k, k + 1, \dots, r \end{cases}$$

Ceci montre en particulier que z_k est la première variable discriminante associée à $X_{(k-1)}$.

Soit S la matrice ($n \times n$) définie par :

$$S_{ii'} = \begin{cases} \frac{1}{n_h} & \text{si les individus } i \text{ et } i' \text{ appartiennent au même groupe } h \text{ d'effectif } n_h, \\ 0 & \text{si les individus } i \text{ et } i' \text{ appartiennent à des groupes différents.} \end{cases}$$

Nous pouvons facilement vérifier que :

$$B = \frac{1}{n} X^T S X \quad \text{et} \quad B_{(k-1)} = \frac{1}{n} X^T_{(k-1)} S X_{(k-1)}.$$

Dans la suite, nous utilisons la propriété stipulant que les variables z_m ($m = 1, 2, \dots, r$) sont orthogonales et que pour $j \neq m$, nous avons $a_j^T B a_m = \lambda_j a_j^T T a_m = 0$.

Nous avons :

$$\begin{aligned} X_{(k-1)} a_m &= \left(I - \sum_{j=1}^{k-1} \frac{z_j z_j^T}{z_j^T z_j} \right) X a_m \\ &= z_m - \sum_{j=1}^{k-1} z_j \frac{z_j^T z_m}{z_j^T z_j} = \begin{cases} 0 & \text{si } m = 1, \dots, k-1 \\ z_m & \text{si } m = k, k+1, \dots, r. \end{cases} \end{aligned}$$

Par suite :

$$B_{(k-1)} a_m = 0, \quad \text{pour } m = 1, \dots, k-1.$$

Pour $m = k, k+1, \dots, r$, nous avons :

$$\begin{aligned} T_{(k-1)} a_m &= \frac{1}{n} X^T \left(I - \sum_{j=1}^{k-1} \frac{z_j z_j^T}{z_j^T z_j} \right) \left(I - \sum_{j=1}^{k-1} \frac{z_j z_j^T}{z_j^T z_j} \right) X a_m \\ &= \frac{1}{n} X^T \left(I - \sum_{j=1}^{k-1} \frac{z_j z_j^T}{z_j^T z_j} \right) X a_m = \frac{1}{n} X^T X a_m = T a_m. \end{aligned}$$

et

$$\begin{aligned} B_{(k-1)} a_m &= \frac{1}{n} X^T \left(I - \sum_{j=1}^{k-1} \frac{z_j z_j^T}{z_j^T z_j} \right) S \left(I - \sum_{j=1}^{k-1} \frac{z_j z_j^T}{z_j^T z_j} \right) X a_m \\ &= \frac{1}{n} X^T \left(I - \sum_{j=1}^{k-1} \frac{z_j z_j^T}{z_j^T z_j} \right) S X a_m \\ &= B a_m - X^T \sum_{j=1}^{k-1} \frac{z_j}{z_j^T z_j} a_j^T B a_m = B a_m. \end{aligned}$$

Il en résulte que :

$$B_{(k-1)} a_m = B a_m = \lambda_m T a_m = \lambda_m T_{(k-1)} a_m, \quad \text{pour } m = k, k+1, \dots, r.$$

Annexe 2

Dans cette annexe, nous présentons les démonstrations des propriétés énoncées dans le paragraphe 2.4.

Soit a_α le vecteur propre normé associé à la plus grande valeur propre de $T_\alpha^{-1}B$, où $T_\alpha = (1 - \alpha).T + \alpha.I$:

$$a_\alpha = \arg\left\{ \max_{a^T a=1} f_\alpha(a) \right\} \quad \text{avec} \quad f_\alpha(a) = \frac{a^T B a}{a^T T_\alpha a}.$$

Dans un premier temps, nous montrons que la fonction définie sur $[0, 1]$ par $V_T(\alpha) = a_{\alpha}^T T a_\alpha$ est croissante.

Notons que nous avons pour tout vecteur a ($a^T a = 1$) $a^T T_\alpha a = (1 - \alpha)a^T T a + \alpha$ et donc pour α_i ($0 \leq \alpha_i \leq 1$) :

$$a_{\alpha_i}^T T_\alpha a_{\alpha_i} = (1 - \alpha)V_T(\alpha_i) + \alpha \quad (1)$$

Soit α_1 et α_2 deux valeurs fixées telles que $0 \leq \alpha_1 \leq \alpha_2 \leq 1$. Par définition de a_{α_1} et a_{α_2} , nous avons :

$$a_{\alpha_1} = \arg\left\{ \max_{a^T a=1} f_{\alpha_1}(a) \right\} \implies \frac{a_{\alpha_2}^T B a_{\alpha_2}}{a_{\alpha_2}^T T_{\alpha_1} a_{\alpha_2}} \leq \frac{a_{\alpha_1}^T B a_{\alpha_1}}{a_{\alpha_1}^T T_{\alpha_1} a_{\alpha_1}} \quad (2)$$

$$a_{\alpha_2} = \arg\left\{ \max_{a^T a=1} f_{\alpha_2}(a) \right\} \implies \frac{a_{\alpha_1}^T B a_{\alpha_1}}{a_{\alpha_1}^T T_{\alpha_2} a_{\alpha_1}} \leq \frac{a_{\alpha_2}^T B a_{\alpha_2}}{a_{\alpha_2}^T T_{\alpha_2} a_{\alpha_2}} \quad (3)$$

La multiplication de l'inégalité (2) par l'inégalité (3) conduit à l'inégalité suivante :

$$\frac{a_{\alpha_2}^T T_{\alpha_2} a_{\alpha_2}}{a_{\alpha_1}^T T_{\alpha_2} a_{\alpha_1}} \leq \frac{a_{\alpha_2}^T T_{\alpha_1} a_{\alpha_2}}{a_{\alpha_1}^T T_{\alpha_1} a_{\alpha_1}} \quad (4)$$

Comme les deux vecteurs a_{α_1} et a_{α_2} sont normés, le développement direct de l'inégalité (4) compte tenu de l'égalité (1) conduit à :

$$\frac{(1 - \alpha_2)V_T(\alpha_2) + \alpha_2}{(1 - \alpha_2)V_T(\alpha_1) + \alpha_2} \leq \frac{(1 - \alpha_1).V_T(\alpha_2) + \alpha_1}{(1 - \alpha_1).V_T(\alpha_1) + \alpha_1}. \quad (5)$$

De là, il est facile de déduire en considérant le produit en croix et en simplifiant que $V_T(\alpha_1) \leq V_T(\alpha_2)$.

Nous allons maintenant montrer que la fonction $V_B(\alpha) = a_\alpha^T B a_\alpha$ est également croissante. Partant de l'inégalité (3), nous pouvons écrire :

$$\frac{a_{\alpha_1}^T B a_{\alpha_1}}{a_{\alpha_2}^T B a_{\alpha_2}} \leq \frac{a_{\alpha_1}^T T_{\alpha_2} a_{\alpha_1}}{a_{\alpha_2}^T T_{\alpha_2} a_{\alpha_2}}. \quad (6)$$

Si de plus, nous considérons que $\alpha_1 \leq \alpha_2$, nous avons compte tenu de l'égalité (1) :

$$V_T(\alpha_1) \leq V_T(\alpha_2) \implies \frac{a_{\alpha_1}^T T_{\alpha_2} a_{\alpha_1}}{a_{\alpha_2}^T T_{\alpha_2} a_{\alpha_2}} \leq 1,$$

il en résulte que :

$$\frac{a_{\alpha_1}^T B a_{\alpha_1}}{a_{\alpha_2}^T B a_{\alpha_2}} \leq 1, \quad (7)$$

d'où $V_B(\alpha_1) \leq V_B(\alpha_2)$.

Finalement pour montrer la décroissance de la fonction $I(\alpha) = \frac{a_\alpha^T B a_\alpha}{a_\alpha^T T a_\alpha}$, nous utilisons le fait que la fonction V_T est croissante. Soit $0 \leq \alpha_1 \leq \alpha_2 \leq 1$, nous avons $V_T(\alpha_1) \leq V_T(\alpha_2)$, il en résulte alors l'inégalité suivante :

$$(1 - \alpha_1)V_T(\alpha_1)V_T(\alpha_2) + \alpha_1 \cdot V_T(\alpha_1) \leq (1 - \alpha_1)V_T(\alpha_1)V_T(\alpha_2) + \alpha_1 \cdot V_T(\alpha_2). \quad (8)$$

Cette inégalité implique que :

$$\frac{(1 - \alpha_1)V_T(\alpha_1) + \alpha_1}{(1 - \alpha_1)V_T(\alpha_2) + \alpha_1} \geq \frac{V_T(\alpha_1)}{V_T(\alpha_2)}. \quad (9)$$

De l'inégalité (2), nous déduisons facilement l'inégalité suivante :

$$\frac{V_B(\alpha_1)}{V_B(\alpha_2)} \geq \frac{(1 - \alpha_1)V_T(\alpha_1) + \alpha_1}{(1 - \alpha_1)V_T(\alpha_2) + \alpha_1} \quad (10)$$

La combinaison des deux inégalités (9) et (10) conduit directement à l'inégalité :

$$I(\alpha_2) \leq I(\alpha_1).$$