L'ANALYSE DISCRIMINANTE

Pierre-Louis GONZALEZ

ANALYSE DISCRIMINANTE

Prédire une variable qualitative à *k* classes
à l'aide de *p* prédicteurs

Deux aspects

- **Descriptif:** Quelles sont les combinaisons linéaires de variables qui permettent de séparer le mieux possible les ***k*** catégories ?

- **Décisionnel:** Un nouvel individu se présente pour lequel on connaît les valeurs des prédicteurs. Décider dans quelle catégorie il faut l'affecter

ANALYSE DISCRIMINANTE

Ensemble des méthodes utilisées pour prédire une variable qualitative à k catégories à l'aide de p prédicteurs.

EXEMPLES

Médecine Connaissant les symptômes présentés par un patient, peut-on porter un diagnostic sur sa maladie ?

Finance

- A partir des bilans d'une société, est-il possible d'estimer son risque de faillite à 2 ans ou 3 ans (scoring financier) ?

- Au moment d'une demande de prêt par un client, peut-on prévoir en fonction des caractéristiques du client, le risque de contentieux (credit scoring) ?

Pétrole

Au vu des analyses des carottes issues d'un forage, est-il possible de présumer de l'existence d'une nappe de pétrole ?

Télédétection

A partir de mesures par satellite des ondes réfléchies ou absorbées par le sol dans différentes longueurs d'onde, peut-on reconstituer automatiquement la nature du terrain étudié (forêt, sable, ville, mer...) ?

Marketing direct

Connaissant les caractéristiques d'un client, peut-on prévoir sa réponse à une offre de produit par courrier ?

Étude de textes

Interprétation d'une typologie

Quelques dates:

- Mahalanobis 1927
- Hotelling 1931
- Fisher 1936
- Rao 1950
- Anderson 1951
- Vapnik 1998

MÉTHODES GÉOMÉTRIQUES

Recherche des meilleures fonctions discriminantes

$$g\left(X_1, X_2 \ldots X_p\right)$$

LINÉAIRES

A.C.P. sur le nuage des centres de gravité des groupes munis de différentes métriques

NON LINÉAIRES

- **quadratique**
- **création de nouvelles variables** $f\left(X_1, X_2 \ldots X_p\right)$ **et application d'une méthode linéaire**
- **découpage en variables qualitatives et application d'une méthode sur variables qualitatives**

MÉTHODES PROBABILISTES

Estimation directe des probabilités d'appartenance aux groupes définis par *Y*

PARAMÉTRIQUE

- **modèle gaussien**
- **maximum de vraisemblance avec autre modèle paramétrique**

SEMI PARAMETRIQUE

- **régression logistique**

NON PARAMÉTRIQUE

- **Estimation de densité :**
- **Fonctions orthogonales**
- **Noyaux de Parzen**
- ***K* plus proches voisins**

Autres approches

- **Méthodes de type « boite noire » induisant le minimum d'erreurs de classement**
  - **Réseaux de neurones**
  - **SVM (Support Vecteur Machine)**

I. MÉTHODES GÉOMÉTRIQUES

1. Données - Notations

Les n individus $\underline{e_i}$ de l'échantillon constituent un nuage E, de $\mathbf{R}^p$ partagé en k sous-nuages : $E_1, E_2 \dots E_k$ de centres de gravité $\underline{g}_1, \underline{g}_2 \dots \underline{g}_k$

de matrices de variances $V_1, V_2 \dots V_k$

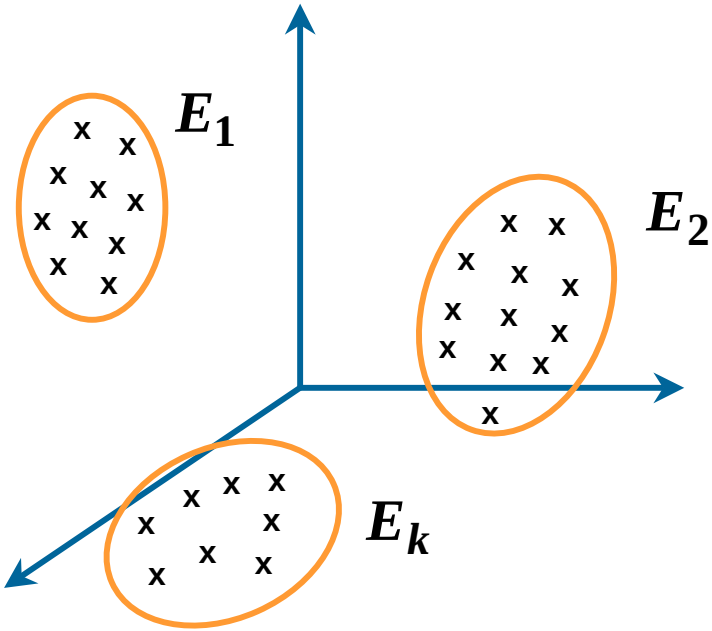
g = centre de gravité de E

V = matrice de variance de E

n individus e_i affectés des poids

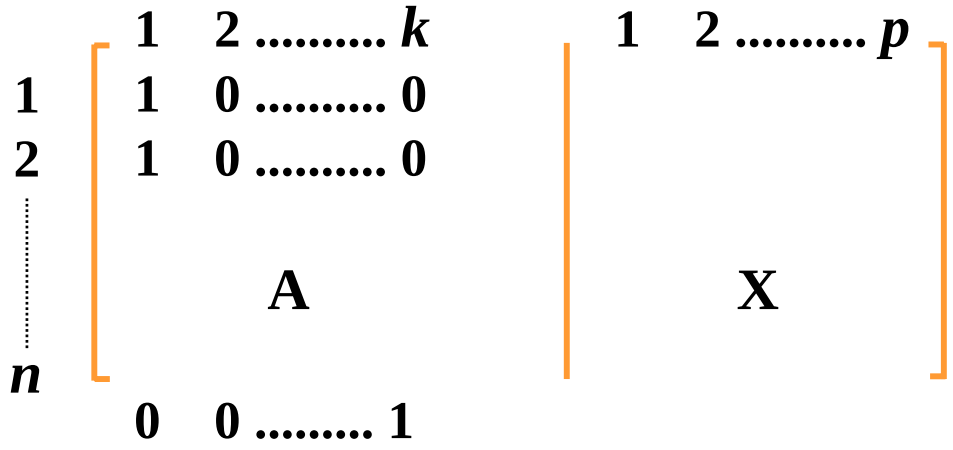
$p_1, p_2 \dots p_n$

rangés dans une matrice diagonale D



Notations matricielles

tableau de données



A Matrice des indicatrices de la variable qualitative à prédire

X Matrice des prédicteurs

$D_q = A'DA$ matrice diagonale des poids q_j des sous-nuages.

$(A'DA)^{-1}(A'DX)$ ses lignes sont les coordonnées des k centres de gravité $\underline{g}_1, \underline{g}_2 \cdots \underline{g}_k$

poids de la classe *j*

$$q_j = \sum_{e_i \in E_j} p_i$$

Centres de gravité

$$\underline{g}_j = \frac{1}{q_j} \sum_i p_i \underline{e}_i \quad \text{pour} \quad e_i \in E_j \qquad \underline{g} = \sum_{j=1}^{k} q_j \underline{g}_j$$

Matrice de variance-covariances de la classe Ej

$$V_j = \frac{1}{q_j} \sum_{\underline{e}_i \in E_j} p_i \left(\underline{e}_i - \underline{g}_j\right)\left(\underline{e}_i - \underline{g}_j\right)'$$

Matrice de variance interclasse : matrice de variance $\boldsymbol{B}$ des $\boldsymbol{k}$ centres de gravité affectés des poids $\boldsymbol{q_j}$:

$$\boldsymbol{B} = \sum_{j=1}^{k} q_j \left(\underline{g}_j - \underline{g}\right)\left(\underline{g}_j - \underline{g}\right)'$$

Matrice de variance intra-classe :

$$\boldsymbol{W} = \sum_{j=1}^{k} q_j V_j$$

En règle générale ***W*** inversible

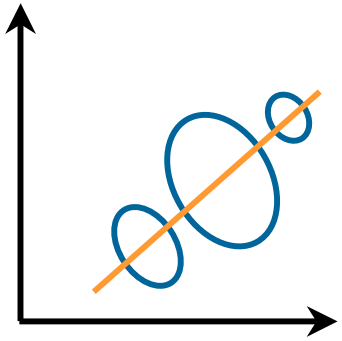
B non inversible (***k*** centres de gravité dans un sous-espace de dimension ***k-1*** de $\mathbf{R}^{p}$)

$$V = W + B$$

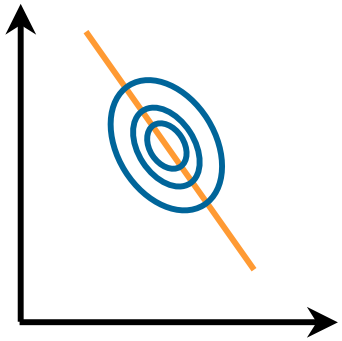
Variance totale = Moyenne des variances + Variance des moyennes

En analyse discriminante, on considère trois types de matrices

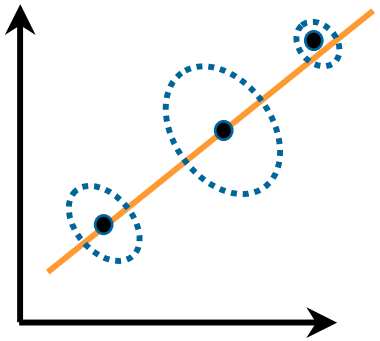
de variances-covariances et donc **trois types de corrélations**.



corrélation totale



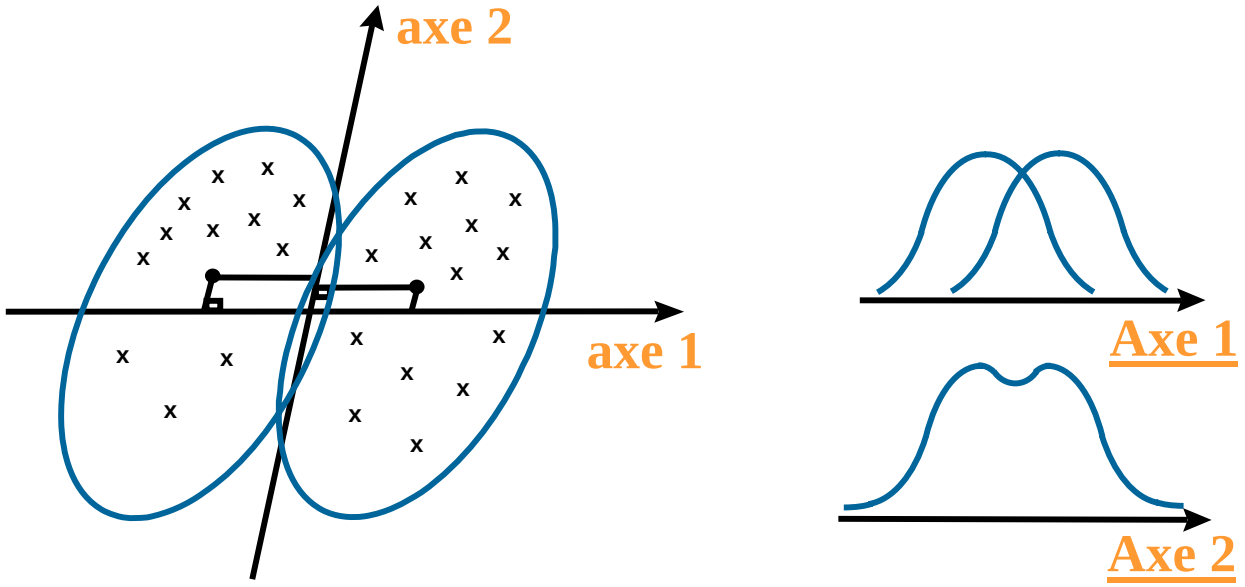
corrélation intra-classes



corrélation inter-classes

2. L'analyse factorielle discriminante (A.F.D.)

Elle consiste à chercher de nouvelles variables (les **variables discriminantes**) correspondant à des directions de $\mathbf{R}^p$ qui séparent le mieux possible en projection les k groupes d'observations.



L'axe 1 possède un bon pouvoir discriminant
L'axe 2 ne permet pas de séparer en projection les 2 groupes.

Supposons $\mathbf{R}^p$ muni d'une métrique M (calcul des distances)

$\underline{a}$ = **axe discriminant**

$\underline{u}$ = **facteur associé** $\qquad \underline{u} = M\underline{a}$

$X\underline{u}$ = **variable discriminante**

L'inertie du nuage des g_j projetés sur $\underline{a}$ doit être maximale.

La matrice d'inertie du nuage des $\underline{g}$ est MBM, l'inertie du nuage projeté sur $\underline{a}$ est $\underline{a}'MBM\underline{a}$ si $\underline{a}$ est M-normé à 1.

Il faut aussi qu'en projection sur $\underline{a}$, chaque sous-nuage reste bien groupé

donc que $\underline{a}'MV_JM\underline{a}$ soit faible pour $j = 1, 2 \ldots k$.

On cherchera donc à minimiser :

$$\sum_{j=1}^{k} q_j \underline{a}'MV_jM\underline{a} \quad \text{soit} \quad \underline{a}'MWM\underline{a}$$

Critère

La relation $\boldsymbol{V = B + W}$ entraîne que $\boldsymbol{MVM = MBM + MWM}$

donc : $\boldsymbol{\underline{a}' MVM \underline{a} = \underline{a}' MBM \underline{a} + \underline{a}' MWM \underline{a}}$

Maximiser le rapport de l'inertie inter-classe à l'inertie totale

$$\max_{\underline{a}} \frac{\underline{a}' MBM \underline{a}}{\underline{a}' MVM \underline{a}}$$

Ce maximum est atteint si $\boldsymbol{\underline{a}}$ est vecteur propre de $\boldsymbol{(MVM)^{-1}(MBM)}$ associé à sa plus grande valeur propre $\boldsymbol{\lambda_1}$

$$\boldsymbol{M^{-1}V^{-1}BM\underline{a} = \lambda \underline{a}}$$

A l'axe discriminant $\underline{a}$ est alors associé le facteur discriminant $\underline{u}$ tel que :

$$\underline{u} = M\,\underline{a}$$

On a alors : $V^{-1}B\underline{u} = \lambda\underline{u}$

Les facteurs discriminants, donc les variables discriminantes $X\,\underline{u}$ sont indépendantes de la métrique M.

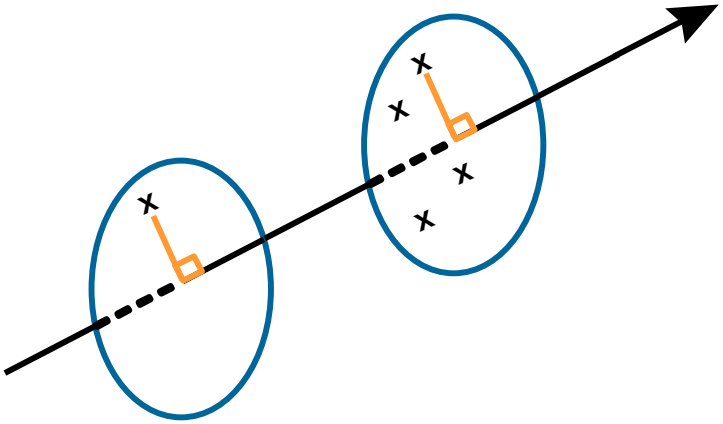
On choisira par commodité $M = V^{-1}$

$$\begin{cases} BV^{-1}\underline{a} = \lambda\underline{a} \\ V^{-1}B\underline{u} = \lambda\underline{u} \end{cases}$$

On a toujours $0 \le \lambda_1 \le 1$ car λ_1 est la quantité à maximiser.

Cas particuliers

Cas $\lambda_1 = 1$



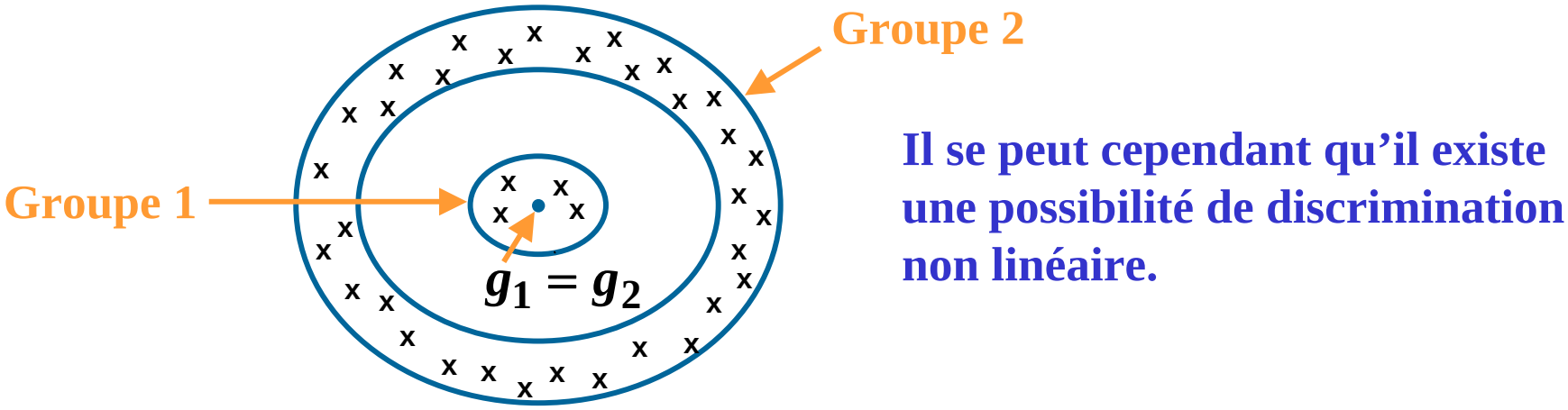
En projection sur $\underline{\boldsymbol{a}}$ les dispersions intra-classes sont nulles. Les $\boldsymbol{k}$ nuages sont donc chacun dans un hyperplan orthogonal à $\underline{\boldsymbol{a}}$.

Il y a discrimination parfaite si les centres de gravité se projettent en des points différents.

Cas $\lambda_1 = 0$

Le meilleur axe ne permet pas de séparer les centres de gravité g_i , c'est le cas où ils sont confondus.

Les nuages sont donc concentriques et aucune séparation linéaire n'est possible.

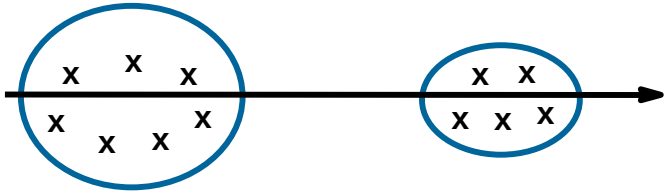


Il se peut cependant qu'il existe une possibilité de discrimination non linéaire.

La distance au centre permet ici de séparer les groupes, mais il s'agit d'une fonction quadratique des variables.

Autres propriétés

La valeur propre λ est une mesure pessimiste du pouvoir discriminant d'un axe.



$\lambda < 1$ mais les groupes sont bien séparés

Le nombre des valeurs propres non nulles, donc d'axes discriminants est égal à $k-1$ dans le cas habituel où $n > p > k$ et où les variables ne sont pas liées par des relations linéaires.

Remarque: Le cas de deux groupes

Il n'y a qu'une seule variable discriminante puisque $k-1=1$.

L'axe discriminant est alors nécessairement la droite reliant les deux centres de gravité $\underline{g}_2$ et $\underline{g}_1$: $\underline{a}=\underline{g}_1-\underline{g}_2$

Le facteur discriminant $\underline{u}$ vaut donc : $\underline{u}=V^{-1}\left(\underline{g}_1-\underline{g}_2\right)$

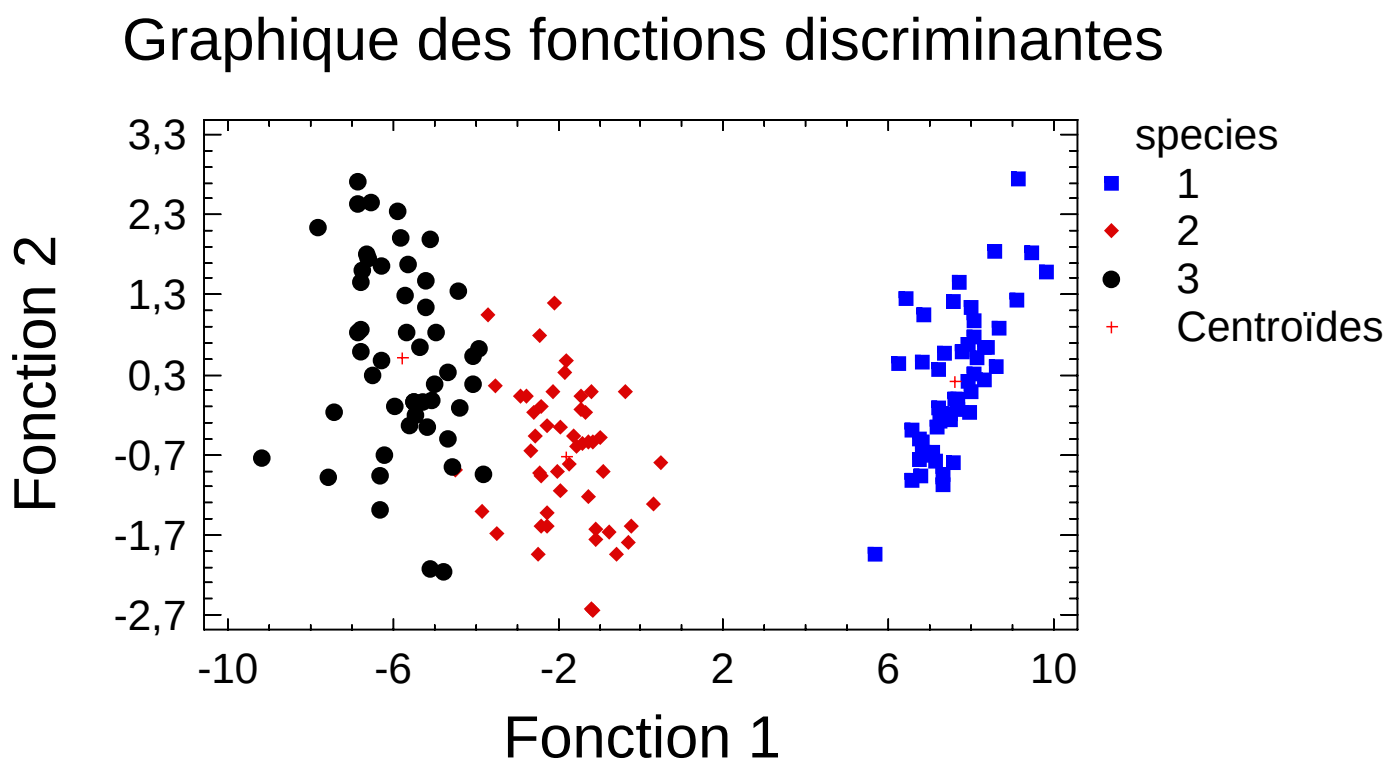
ou $\underline{u}=W^{-1}\left(\underline{g}_1-\underline{g}_2\right)$ qui lui est proportionnel

$W^{-1}\left(\underline{g}_1-\underline{g}_2\right)$ **est la fonction de Fisher (1936).**

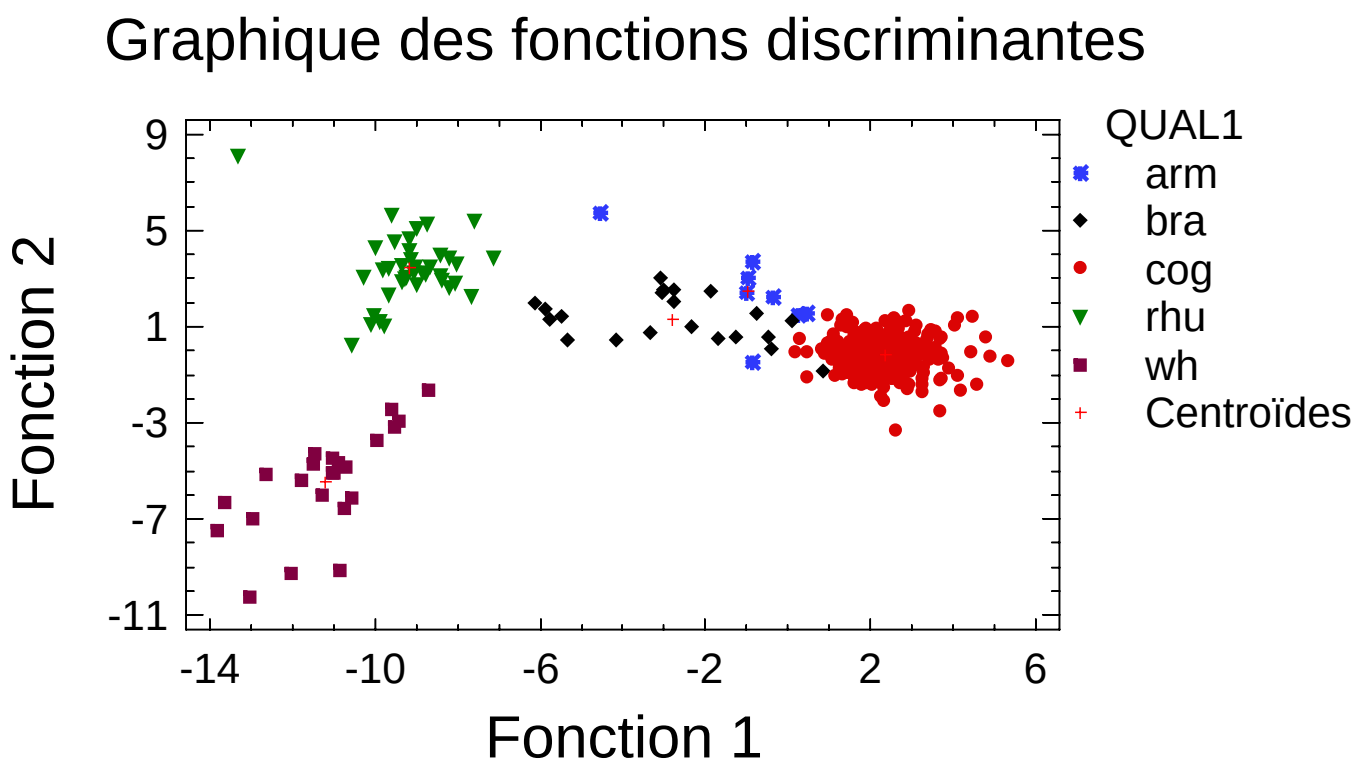
3. Exemples: Les iris de Fisher

		
Iris setosa	Iris versicolor	Iris virginica

Les iris de Fisher



Discrimination entre divers spiritueux à l'aide de dosages moléculaires



4. Une A.C.P. particulière

D'après les équations précédentes, l'analyse factorielle discriminante n'est autre que l'A.C.P. du nuage des k centres de gravité avec la métrique V^{-1} .

On en déduit que les variables discriminantes sont non corrélées deux à deux.

Dans le cas où il existe plusieurs axes discriminants $(k > 2)$ on peut utiliser les représentations graphiques usuelles de l'A.C.P. : cercle des corrélations...

5. Règles géométriques d'affectation

Ayant trouvé la meilleure représentation de la séparation en k groupes des n individus, on peut alors chercher à affecter une observation $\underline{e}$ à l'un des groupes.

La règle naturelle consiste à calculer les distances de l'observation à classer à chacun des k centres de gravité et à affecter selon la distance la plus faible. Métrique à utiliser ?

Règle de Mahalanobis Fisher

On utilise W^{-1}

$$d^2(\underline{e};\underline{g}_i) = (\underline{e}-\underline{g}_i)' W^{-1} (\underline{e}-\underline{g}_i)$$

6. Exemple: Qualité des vins de Bordeaux

Les données

	Température	Soleil	Chaleur	Pluie	Qualité
1	3064	1201	10	361	2
2	3000	1053	11	338	3
3	3155	1133	19	393	2
4	3085	970	4	467	3
5	3245	1258	36	294	1
6	3267	1386	35	225	1
7	3080	966	13	417	3
8	2974	1189	12	488	3
9	3038	1103	14	677	3
10	3318	1310	29	427	2
11	3317	1362	25	326	1
12	3182	1171	28	326	3
13	2998	1102	9	349	3
14	3221	1424	21	382	1
15	3019	1230	16	275	2
16	3022	1285	9	303	2
17	3094	1329	11	339	2
18	3009	1210	15	536	3
19	3227	1331	21	414	2
20	3308	1366	24	282	1
21	3212	1289	17	302	2
22	3361	1444	25	253	1
23	3061	1175	12	261	2
24	3478	1317	42	259	1
25	3126	1248	11	315	2
26	3458	1508	43	286	1
27	3252	1361	26	346	2
28	3052	1186	14	443	3
29	3270	1399	24	306	1
30	3198	1259	20	367	1
31	2904	1164	6	311	3
32	3247	1277	19	375	1
33	3083	1195	5	441	3
34	3043	1208	14	371	3

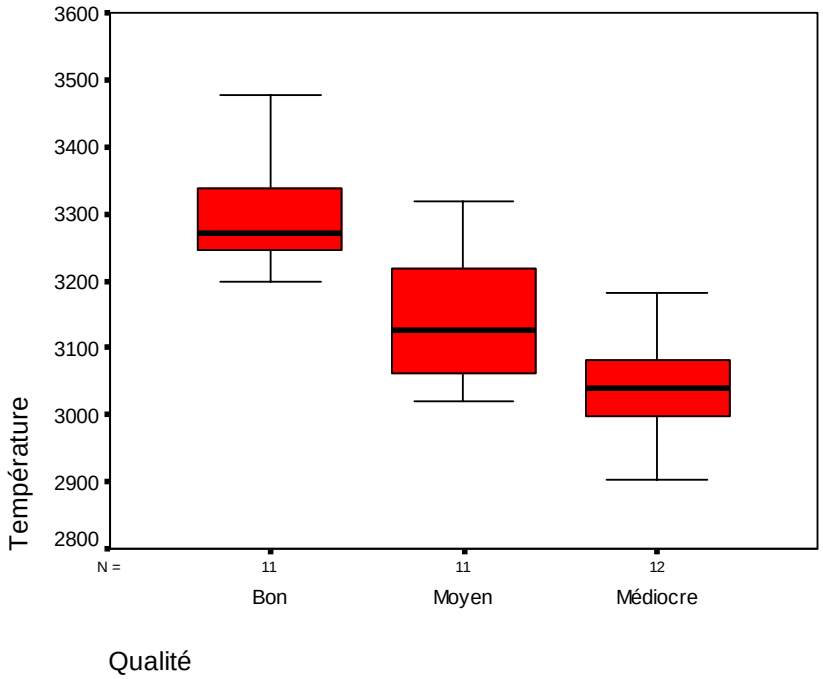
Analyse préalable

Température

Report

Température

Qualité	Mean	N	Std. Deviation
1	3306.36	11	92.06
2	3140.91	11	100.05
3	3037.33	12	69.34
Total	3157.88	34	141.18



Measures of Association

	Eta	Eta Squared
Température * Qualité	.799	.639

$$\text{Rapport de corrélation} = \eta^2 = \frac{\text{Between Groups Sum of Squares}}{\text{Total Sum of Squares}}$$

ANOVA Table

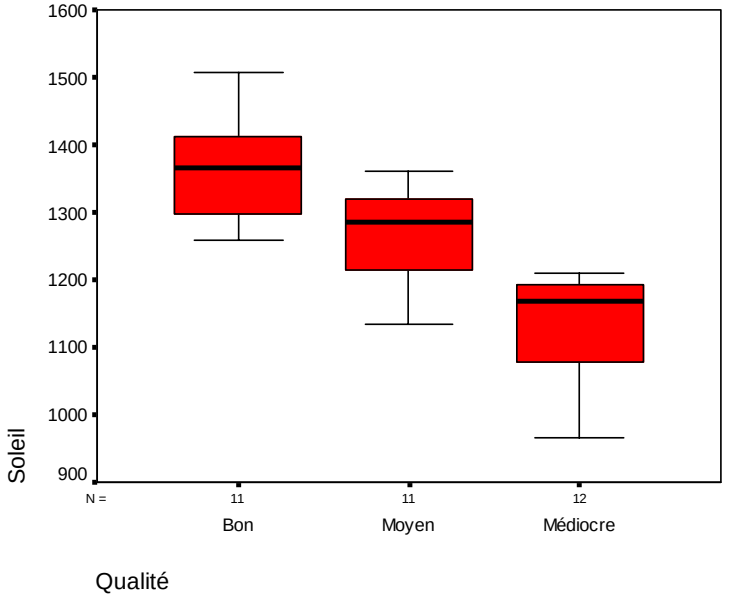
			Sum of Squares	df	Mean Square	F	Sig.
Température * Qualité	Between Groups	(Combined)	420067.4	2	210033.704	27.389	.000
	Within Groups		237722.1	31	7668.456		
	Total		657789.5	33			

Soleil

Report

Soleil

Qualité	Mean	N	Std. Deviation
Bon	1363.64	11	80.31
Moyen	1262.91	11	71.94
Médiocre	1126.42	12	88.39
Total	1247.32	34	126.62



Measures of Association

	Eta	Eta Squared
Soleil * Qualité	.786	.618

ANOVA Table

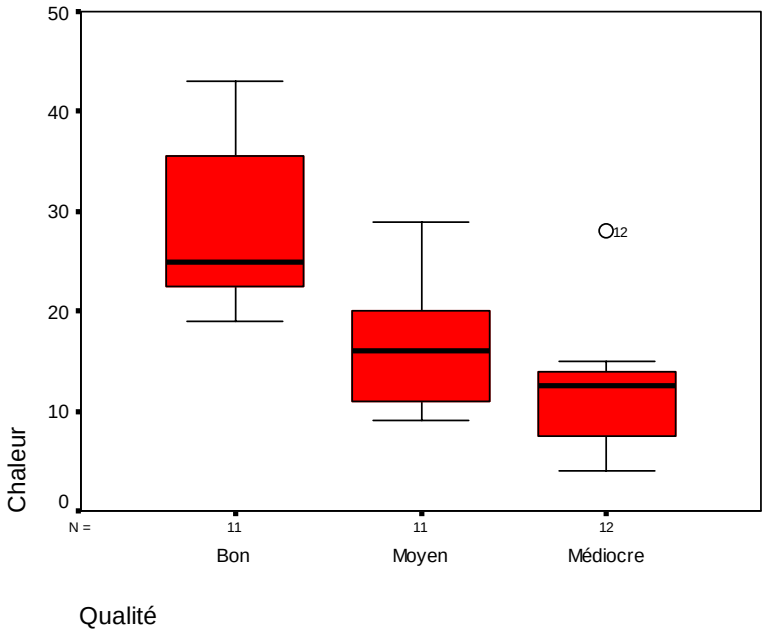
			Sum of Squares	df	Mean Square	F	Sig.
Soleil * Qualité	Between Groups	(Combined)	326909.1	2	163454.535	25.061	.000
	Within Groups		202192.4	31	6522.335		
	Total		529101.4	33			

Chaleur

Report

Chaleur

Qualité	Mean	N	Std. Deviation
Bon	28.55	11	8.80
Moyen	16.45	11	6.73
Médiocre	12.08	12	6.30
Total	18.82	34	10.02



Measures of Association

	Eta	Eta Squared
Chaleur * Qualité	.705	.497

ANOVA Table

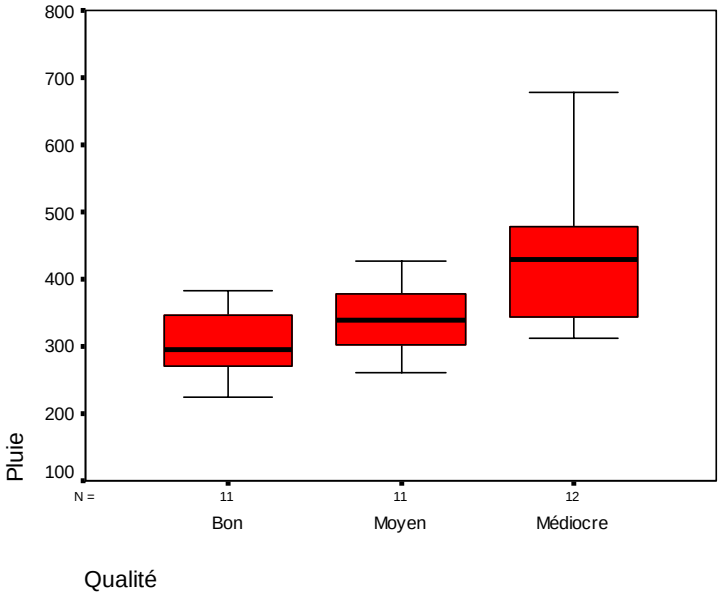
			Sum of Squares	df	Mean Square	F	Sig.
Chaleur * Qualité	Between Groups	(Combined)	1646.570	2	823.285	15.334	.000
	Within Groups		1664.371	31	53.689		
	Total		3310.941	33			

Pluie

Report

Pluie

Qualité	Mean	N	Std. Deviation
Bon	305.00	11	52.29
Moyen	339.64	11	54.99
Médiocre	430.33	12	104.85
Total	360.44	34	91.40



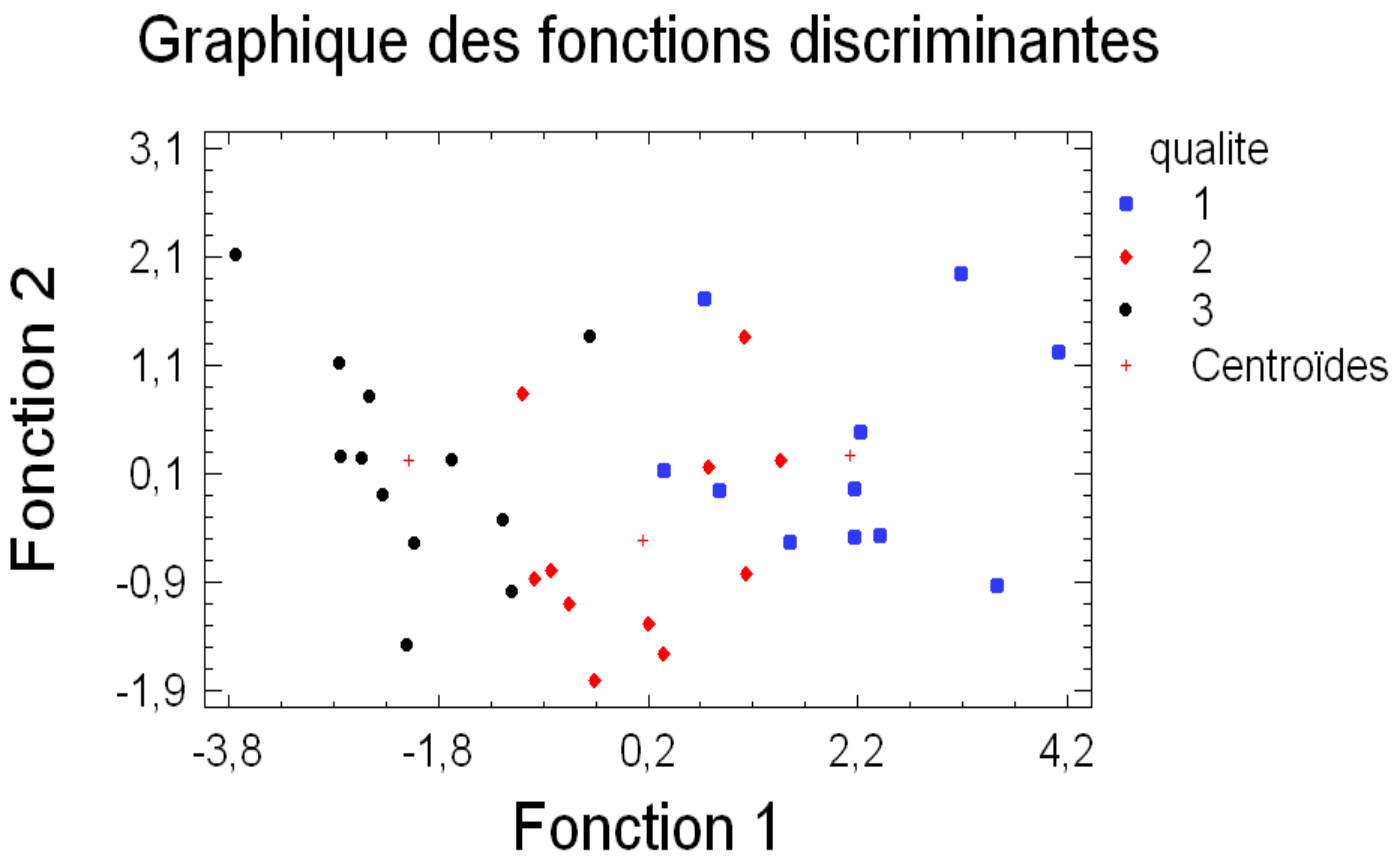
Measures of Association

	Eta	Eta Squared
Pluie * Qualité	.594	.353

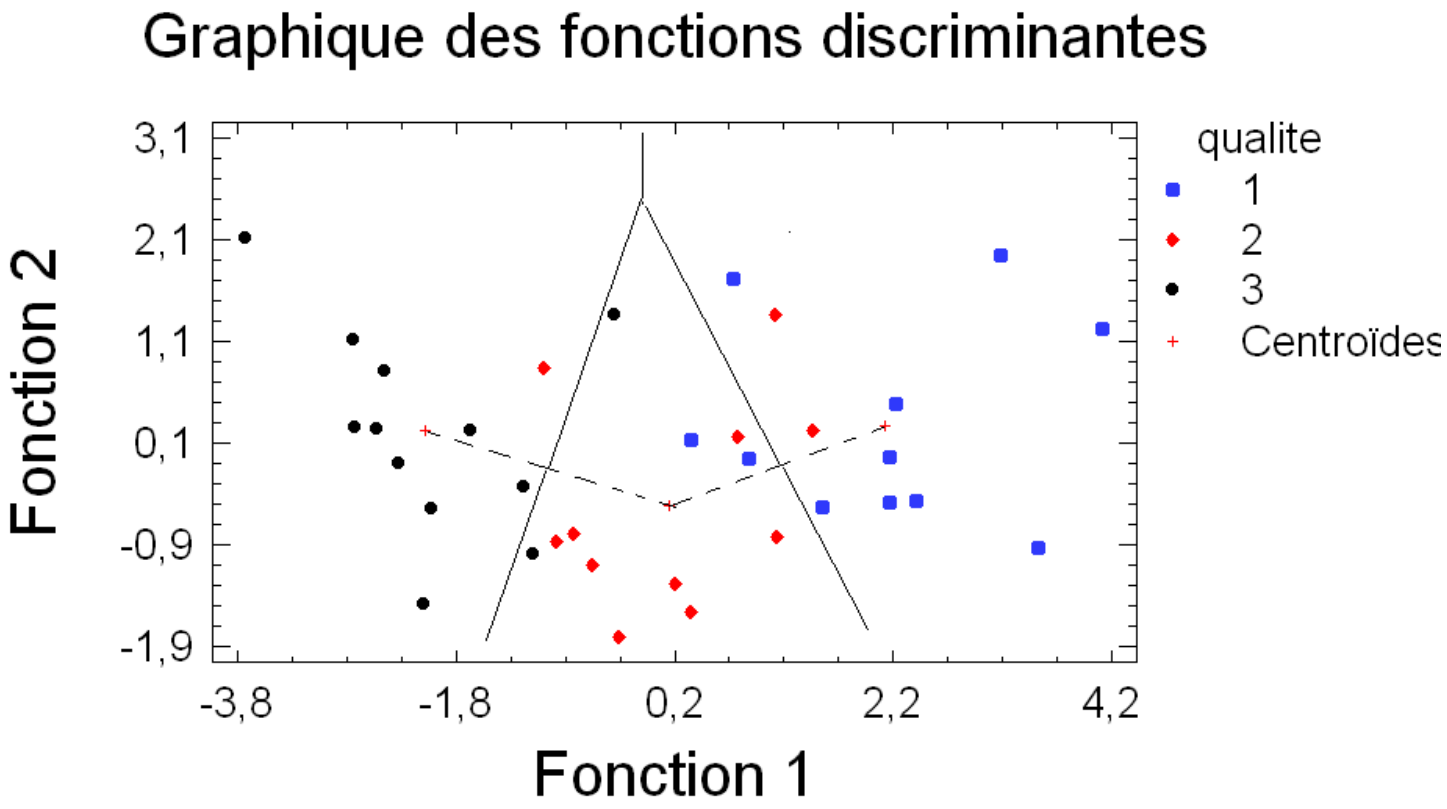
ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Pluie * Qualité	Between Groups	(Combined)	97191.170	2	48595.585	8.440	.001
	Within Groups		178499.2	31	5758.039		
	Total		275690.4	33			

Qualité des vins de Bordeaux



Qualité des vins de Bordeaux



Qualité des vins de Bordeaux: Pourcentage de bien classés

Tableau de classement

Observé qualite	Taille	Groupe Prévu qualité 1	2	3
1	11	9	2	0
		(81,82%)	(18,18%)	(0,00%)
2	11	2	8	1
		(18,18%)	(72,73%)	(9,09%)
3	12	0	2	10
		(0,00%)	(16,67%)	(83,33%)

Pourcentage d'observations bien classées: 79,41%

Qualité des vins de Bordeaux
Fonctions discriminantes

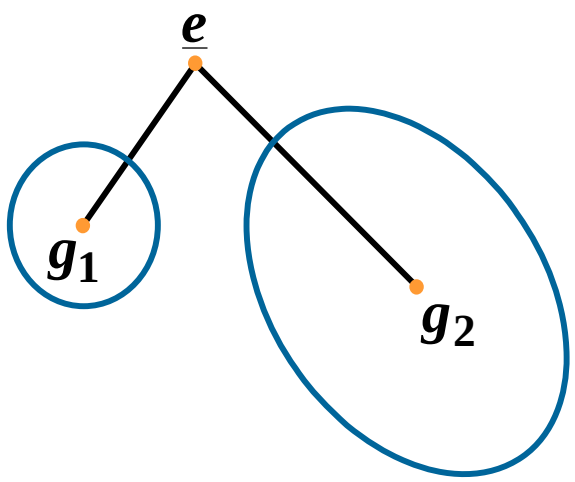
Coefficients des fonctions discriminantes pour qualité

Coefficients standardisés

	1	2
stemp	0,750126	-0,00405015
insol	0,547064	-0,430399
chaleur	-0,198237	0,935229
hpluies	-0,445097	0,468536

7. Insuffisance des règles géométriques

L'utilisation de la règle précédente conduit à des affectations incorrectes lorsque les dispersions des groupes sont très différentes entre elles : rien ne justifie alors l'usage de la même métrique pour les différents groupes.



$\underline{e}$ plus proche de g_1 que de g_2 au sens habituel.

Pourtant, il est plus naturel d'affecter $\underline{e}$ à la deuxième classe qu'à la première dont le pouvoir d'attraction est moindre.

Solution : métriques locales $\boldsymbol{M_i}$

Dans la plupart des cas, on choisit $\boldsymbol{M_i}$ proportionnel à $\boldsymbol{V_i^{-1}}$.

La question de l'optimalité d'une règle de décision géométrique ne peut cependant être résolue sans référence à un **modèle probabiliste**.

8. Remarques concernant la présentation de l'analyse discriminante dans les logiciels « américains »

8.1. Par ses liens avec l'analyse canonique, les auteurs de langue anglaise utilisent le terme : « ***ANALYSE DISCRIMINANTE CANONIQUE*** ».

On cherche la combinaison linéaire des variables qui a le plus grand coefficient de corrélation multiple avec la variable de classe.

- **Ce coefficient de corrélation est appelé première corrélation canonique.**

La valeur propre λ_1 (équation $\boldsymbol{V^{-1}B\underline{u} = \lambda_1 \underline{u}}$) est égale au carré de ce coefficient de corrélation.

- La variable définie par la combinaison linéaire est appelée la première composante canonique ou première variable canonique.

La deuxième variable canonique répond à deux critères :

- ne pas être corrélée avec la première,

- avoir le plus grand coefficient de corrélation multiple possible avec la variable de classe.

Ce processus peut être répété jusqu'au moment où le nombre de variables canoniques est égal au nombre de variables de départ ou au nombre de classes moins 1 s'il est plus petit.

8.2. Analyse de variance et métrique W^{-1}

S'il n'y avait qu'une seule variable explicative, on mesurerait l'efficacité de son pouvoir séparateur sur la variable de groupe au moyen d'une analyse de variance ordinaire à 1 facteur :

$$F = \frac{\text{Variance inter} / k-1}{\text{Variance intra} / n-k}$$

Comme il y a $\boldsymbol{p}$ variables, on peut rechercher la combinaison linéaire définie par des coefficients $\boldsymbol{u}$ donnant la valeur maximale pour la statistique de test, ce qui revient à maximiser :

$$\frac{\underline{u}' B \underline{u}}{\underline{u}' W \underline{u}}$$

La solution est donnée par l'équation :

$$W^{-1} B \underline{u} = \mu \underline{u} \quad \text{avec } \mu \text{ maximal}$$

Les vecteurs propres de $\boldsymbol{W^{-1}B}$ sont les mêmes que ceux de $\boldsymbol{V^{-1}B}$

avec $\mu = \dfrac{\lambda}{1-\lambda} \Leftrightarrow \lambda = \dfrac{\mu}{1+\mu}$

Les logiciels « américains » fournissent cette valeur propre μ :

si : $\mathbf{0 \leq \lambda \leq 1}$

on a en revanche : $\mathbf{0 \leq \mu \leq \infty}$

A ce point près, l'utilisation de V^{-1} ou de W^{-1} comme métrique est indifférente.

9. Analyse canonique discriminante et régression

L'analyse canonique discriminante, se réduit dans le cas de deux groupes à une régression multiple.

En effet après avoir centré, l'espace engendré par les deux indicatrices de la variable des groupes est de dimension ***1***.

Il suffit donc de définir une variable centrée $\boldsymbol{Y}$ ne prenant que les deux valeurs $\boldsymbol{a}$ et $\boldsymbol{b}$ sur les groupes ***1*** et ***2***.

$$\left(\boldsymbol{n_1 a + n_2 b = 0}\right)$$

On obtiendra alors un vecteur des coefficients de régression proportionnel à la fonction de Fisher pour un choix quelconque de $\boldsymbol{a}$.

IMPORTANT

On prendra garde au fait que les hypothèses habituelles de la régression ne sont pas vérifiées, bien au contraire :

Ici Y est non aléatoire

X l'est.

Ne pas utiliser, autrement qu'à titre indicatif, les statistiques usuelles fournies par un programme de régression.

10. Analyse discriminante sur variables qualitatives

$\boldsymbol{Y}$: variable de groupe

$\boldsymbol{\chi_1, \chi_2, \ldots \chi_p}$ variables explicatives à $\boldsymbol{m_1, m_2, \ldots m_p}$ modalités.

Exemples

- **Solvabilité d'emprunteurs auprès de banques**

Y → bon payeur
Y → mauvais payeur

χ_1 : sexe χ_2 : catégorie professionnelle

- **Risque en assurance automobile**

Y → bon conducteur
Y → mauvais conducteur

χ_1 : sexe χ_2 : tranche d'âge χ_3 : véhicule sportif ou non

- **Reclassement dans une typologie**

Y : classes

Caractéristiques du problème

- Grand nombre de prédicteurs qualitatifs
- Échantillons volumineux

Méthodes classiques inadaptées

- Analyse discriminante classiques : variables quantitatives
- Modèle log linéaire : trop de variables

D I S Q U A L

Méthode de discrimination fondée sur l'analyse factorielle

Prédicteurs qualitatifs

Estimer $P\left(Y = y / \chi_1 = x_1 \quad \chi_2 = x_2 \quad ...\right)$

- Approche multinomiale irréaliste

  P estimé par la fréquence

  $\prod_{i=1}^{k} m_i$ cases !

- Approche modèle

  Log-linéaire, linéaire, on néglige certaines interactions.

  Ex : $\left| n \;\; P\left(Y = y \,\middle|\, \chi_1 = i, \chi_2 = j, \chi_3 = k\right)\right.$

  $= \alpha_0 + \alpha_i + \beta_j + \sigma_k + \delta_{ij} + \varepsilon_{ik}$

Une méthode de discrimination sur variables qualitatives : la méthode DISQUAL

Les p prédicteurs sont p variables qualitatives $\chi_1\ \chi_2 \cdots \chi_p$ à $m_1\ m_2 \ldots m_p$ modalités.

1ère étape A.C.M. des variables $\chi_1\ \chi_2 \cdots \chi_m$

$\Leftrightarrow$ Analyse des correspondances du tableau disjonctif

$$X = \left(X_1 \middle| X_2 \middle| \ldots \middle| X_p\right)$$

2ème étape

On remplace les p variables qualitatives par les q coordonnées sur les axes factoriels

→ analyse discriminante sur ces q variables numériques Z_1 Z_2 ... Z_q

Facteur discriminant d = combinaison linéaire des Z_j qui sont des combinaisons linéaires des indicatrices.

3ème étape

Expression de ***d*** comme combinaison linéaire des indicatrices

$\Leftrightarrow$ attribuer à chaque catégorie de chaque variable une valeur numérique ou score.

Ceci revient donc à transformer chaque variable qualitative en une variable discrète à ***m*** valeurs (associées à chaque modalité).

L'ANALYSE DISCRIMINANTE

MÉTHODES DÉCISIONNELLES

MÉTHODES PROBABILISTES

1. La règle bayésienne

- k groupes en proportion $p_1, p_2 \ldots p_k$
- La distribution de probabilité du vecteur observation $\underline{x} = (x_1, \ldots x_p)$ est donnée pour chaque groupe j par une densité (ou une loi discrète) $f_j(\underline{x})$.

Observation $(x_1, x_2 \ldots x_p)$

probabilité qu'elle provienne du groupe j
formule de Bayes

$$P(G_j|\underline{x}) = \frac{P(\underline{x}|G_j)P(G_j)}{\sum_{j=1}^{k} P(\underline{x}|G_j)P(G_j)}$$

$$P(G_j|x) = \frac{p_j f_j(\underline{x})}{\sum_{j=1}^{k} p_j f_j(\underline{x})}$$

Règle bayésienne

Affecter $\underline{x}$ au groupe qui a la probabilité a posteriori maximale.

$\Rightarrow$ chercher le maximum de $p_j f_j(x)$

Il est nécessaire de connaître ou d'estimer $f_j(\underline{x})$

- méthodes non paramétriques
- méthodes paramétriques : cas gaussien p-dimensionnel, discrimination logistique.

2. Le modèle normal multidimensionnel

Hypothèse de travail

$\underline{x}$ **distribués selon** $N_p(\underline{\mu}, \Sigma_j)$ pour chaque groupe

Densité:

$$f_j(\underline{x}) = \frac{1}{(2\pi)^{p/2}\left(\det \Sigma_j\right)^{1/2}} \exp\left[-\frac{1}{2}\left(\underline{x}-\underline{\mu}_j\right)'\Sigma_j^{-1}\left(\underline{x}-\underline{\mu}_j\right)\right]$$

2-1. Cas général

Règle bayésienne : $\max_j p_j \; f_j(\underline{x})$ devient par passage aux logarithmes

(de l'expression $-2\,Log\,p_j f_j$) :

$$\min : (\underline{x} - \underline{\mu}_j)' \sum_j^{-1} (\underline{x} - \underline{\mu}_j) - 2\,Log\; p_j + Log\,(\det \Sigma_j)$$

Lorsque les Σ_j sont différents, cette règle est donc **quadratique** $\rightarrow$ il faut comparer k fonctions quadratiques de $\underline{x}$

Σ_j est en général estimé par $\frac{n}{n-1} V_j$

$\underline{\mu}_j$ par $\underline{g}_j$

2-2. Cas d'égalité des matrices de variance covariance

Si $\Sigma_1 = \Sigma_2 = \ldots \Sigma$, la règle devient linéaire car :

Alors $\begin{cases} \log\left(\det \Sigma_j\right) = \text{ constante} \\ \left(\underline{x} - \underline{\mu}_j\right)' \Sigma^{-1} \left(\underline{x} - \underline{\mu}_j\right) = \Delta^2\left(\underline{x}, \underline{\mu}_j\right) \end{cases}$

$=$ distance de Mahalanobis de $\underline{x}$ à $\underline{\mu}$

En développant, en éliminant $\underline{x}' \Sigma^{-1} \underline{x}$, on obtient :

d'où en divisant par *-2* ➔ $\max\left[\underline{x}' \Sigma^{-1} \underline{\mu}_j - \frac{1}{2} \underline{\mu}'_j \Sigma^{-1} \underline{\mu}_j + \log p_j \right]$

Si Σ est estimé par $\frac{n}{n-k}W$:

règle Bayésienne $\Leftrightarrow$ règle géométrique (si égalité des p_j)

La règle géométrique est alors optimale.

La probabilité a posteriori d'appartenance au groupe j est proportionnelle à :

$$p_j \exp\left[-\frac{1}{2}\Delta^2\left(\underline{x},\underline{\mu}_j\right)\right]$$

2-3. Cas de deux groupes avec égalité de Σ_1 et Σ_2

On affectera $\underline{x}$ au groupe 1 si :

$$\left[\underline{x}'\Sigma^{-1}\underline{\mu}_1 - \frac{1}{2}\underline{\mu}'_1\Sigma^{-1}\underline{\mu}_1 + Log\ p_1\right] > \left[\underline{x}'\Sigma^{-1}\underline{\mu}'_2 - \frac{1}{2}\underline{\mu}'_2\Sigma^{-1}\underline{\mu}_2 + Log\ p_2\right]$$

$$\underline{x}'\Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) > \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) + Log\ \frac{p_2}{p_1}$$

Si $p_1 = p_2 = 0{,}5$, on retrouve la règle de Fisher en estimant Σ par $\frac{n}{n-2}W$.

Soit $$s(\underline{x}) = \underline{x}' \sum^{-1} (\mu_1 - \mu_2) - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)' \sum^{-1} (\underline{\mu}_1 - \underline{\mu}_2) Log \frac{p_2}{p_1}$$

On affectera $\underline{x}$ au groupe 1 si $s(\underline{x}) > 0$

au groupe 2 si $s(\underline{x}) < 0$

$s(\underline{x})$ **appelée « score » ou statistique d'Anderson.**

Propriété

$S(\underline{x})$ est liée simplement à la probabilité a posteriori d'appartenance au groupe 1.

Démonstration :

$$P(G_1|\underline{x}) = \frac{p_1 f_1(\underline{x})}{p_1 f_1(\underline{x}) + p_2 f_2(\underline{x})} = P$$

$$\Rightarrow \quad \frac{1}{P} = 1 + \frac{p_2}{p_1}\frac{f_2(\underline{x})}{f_1(\underline{x})}$$

$$= 1 + \frac{p_2}{p_1}\exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu}_2)'\sum^{-1}(\underline{x}-\underline{\mu}_2) + \frac{1}{2}(\underline{x}-\underline{\mu}_1)'\sum^{-1}(\underline{x}-\underline{\mu}_1)\right]$$

$$\Rightarrow \quad \frac{1}{P} - 1 = \frac{p_2}{p_1} \exp\left[\frac{1}{2}\Delta^2\left(\underline{x}, \underline{\mu}_1\right) - \frac{1}{2}\Delta^2\left(\underline{x}, \underline{\mu}_2\right)\right]$$

d'où $Log\left(\frac{1}{P} - 1\right) = -S\left(\underline{x}\right)$

$$P = \frac{1}{1 + e^{-s\left(\underline{x}\right)}} = \frac{e^{s\left(\underline{x}\right)}}{1 + e^{s\left(\underline{x}\right)}}$$

***P* « fonction logistique du score »**

Remarque : Lorsque $p_1 = p_2 = \frac{1}{2}$

$$P = \frac{1}{1 + e^{-\frac{1}{2}\left[\Delta^2\left(\underline{x}, \underline{\mu}_1\right) - \Delta^2\left(\underline{x}, \underline{\mu}_2\right)\right]}}$$

2-4. A propos de certains tests :

Test d'égalité des matrices Σ_i : test de Box

Si l'hypothèse $\Sigma_1 = \Sigma_2 = ... = \Sigma_k$ est vraie, la quantité :

$$1 - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \left[\left(\Sigma \frac{1}{n_i - 1} - \frac{1}{n-k} \right) (n-k) Log \left| \frac{n}{n-k} W \right| - \Sigma (n_i - 1) Log \left| \frac{n_i}{n_i - 1} V_i \right| \right]$$

suit approximativement une loi $\chi^2_{\frac{p(p+1)(k-1)}{2}}$

Si on rejette l'hypothèse d'égalité, doit-on utiliser les règles quadratiques ?

Ce n'est pas sûr :

- Test de Box pas parfaitement fiable

- Règle quadratique $\Rightarrow$ estimation de chaque Σ_j (donc de plus de paramètres).

  Lorsque les échantillons sont de petite taille, les fonctions obtenues sont très peu robustes.

➡ il vaut mieux choisir une règle linéaire.

Nombre de paramètres à estimer

- **Exemple:**
  - Avec p = 10 variables
  - Avec k= 4 groupes

  L'analyse discriminante linéaire demande l'estimation de 95 paramètres et l'analyse discriminante quadratique l'estimation de 260 paramètres

2.5. Qualité de la discrimination

a. Cas de 2 groupes

Soit un ensemble de ℓ variables parmi les p composantes de $\underline{x}$

Supposons que $\Delta_p^2 = \Delta_l^2$: en d'autres termes les $(p-\ell)$ autres variables n'apportent aucune information pour séparer les deux populations, alors :

$$\frac{(n_1+n_2-p-1)n_1n_2\left(D_p^2-D_l^2\right)}{(p-l)(n_1+n_2)(n_1+n_2-2)+n_1n_2D_l^2} = F(p-l, n_1+n_2-p-1)$$

On peut ainsi tester :

- l'apport d'une nouvelle variable en prenant $\ell = p-1$
- l'apport de toutes les variables $\left(\Delta_p^2 = 0\right)$

b. Plus de 2 groupes :

On utilise le Λ de Wilks

Sous $\boldsymbol{H_0 : \mu_1 = \mu_2 = \ldots = \mu_k}$

$$\Lambda = \frac{|W|}{|V|} = \frac{|W|}{|W+B|} = \frac{1}{|W^{-1}B + I|}$$

suit la loi de Wilks de paramètres $\boldsymbol{(p, n-k, k-1)}$

Justification : $\boldsymbol{nV}$, $\boldsymbol{nW}$, $\boldsymbol{nB}$ suivent des lois de Wishart à

$\boldsymbol{n-1, n-k, k-1}$ d.d.l.

c. Remarque dans le cas de deux groupes

Le test de Wilks et le test de la distance de Mahalanobis $\left(H_0 \;\; \Delta_p^2 = 0\right)$ sont identiques :

$\boldsymbol{B}$ étant de rang $\mathbf{1}$, on a :

$$\Lambda = \frac{1}{1 + D_p^2 \dfrac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 2)}} = \frac{1}{1+\mu} = 1 - \lambda$$

$$\begin{cases} \mu = \text{valeur propre de } W^{-1}B \\ \lambda = \text{valeur propre de } V^{-1}B \\ \mu = \dfrac{\lambda}{1-\lambda} \end{cases}$$

d. Paramètres usuels fournis par les logiciels

- **Wilks:** $$\Lambda = \prod_{i=1}^{k-1}(1-\lambda_i)$$

λ_i = corrélation canonique au carré

Plus le Λ (Wilks) est faible, meilleure est la discrimination

- **Trace de Pillaï** = $$\text{Trace}(V^{-1}B) = \sum_{i=1}^{k}\lambda_i$$

- **Trace de Hotelling-Lawley**

$$\text{Trace}(W^{-1}B) = \sum_{i=1}^{k-1}\frac{\lambda_i}{1-\lambda_i} = \sum_{i=1}^{k-1}\mu_i$$

- **Plus grande valeur propre de Roy** : μ_1

2.6. Sélection de variables pas à pas

En discriminante à k groupes, on utilise souvent le test de variation de Λ mesuré par :

$$\frac{n-k-p}{k-1}\left(\frac{\Lambda_p}{\Lambda_{p+1}}-1\right)$$ que l'on compare à un $F_{(k-1,n-k-p)}$

La plupart des logiciels présentent des techniques de **sélection ascendante,descendante ou mixte des variables**. SAS propose la procédure STEPDISC.

Sélection ascendante (option **Forward**)

• A l'étape initiale aucune variable n'est présente.

• A chaque étape on fait entrer la variable qui contribue le plus au pouvoir discriminant du modèle, mesuré par le lambda de Wilks.

• La sélection s'arrête quand aucune des variables non sélectionnées ne convient au sens du seuil de probabilité choisi pour le F de Fisher.

Sélection descendante (option **Backward**)

• On démarre avec le modèle complet (construit avec toutes les variables)

• A chaque étape, la variable contribuant le moins au pouvoir discriminant du modèle est éliminée.

• La sélection s'arrête quand on ne peut plus éliminer de variables étant donné le seuil de probabilité choisi pour le F de Fisher.

Sélection mixte (option **Stepwise**)

• On démarre comme dans la procédure ascendante.

• Dès qu'une variable entre dans le modèle, on vérifie compte tenu de cette entrée si l'une des variables déjà présentes est susceptible d'être éliminée.

• La sélection s'arrête quand on ne plus ajouter ou éliminer de variables.

3. Mesures d'efficacité des règles de classement

Critère usuel	**Probabilité de bien classer une observation quelconque. Les diverses méthodes sont comparées en fonction de leurs taux d'erreur.**

3.1 Taux d'erreur théorique pour deux groupes avec $\Sigma_1 = \Sigma_2$ et distribution normale

Quand $p_1 = p_2$, on affecte l'individu au groupe 1 si :

$$S(\underline{x}) = \left[\underline{x}' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) - \frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) \right] > 0$$

La probabilité d'erreur de classement est donc :

$$P\left[S(\underline{x}) > 0 \,|\, \underline{x} \in N_p(\underline{\mu}_2, \Sigma) \right]$$

La loi de $S(\underline{x})$ est une loi de Gauss à une dimension comme combinaison linéaire des composantes de $\underline{x}$.

$$E(S(\underline{x})) = \underline{\mu}'_2 \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$$

$$= -\frac{1}{2}(\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$$

$$= -\frac{1}{2}\Delta_p^2$$

$$V(S(\underline{x})) = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \Sigma \Sigma^{-1}(\underline{\mu}_1 - \underline{u}_2) = \Delta_p^2$$

$S(\underline{x})$ **suit** N $\left(-\frac{1}{2}\Delta_p^2 ; \Delta_p\right)$ si $\underline{x} \in G_2$

La probabilité de classer dans le groupe 1 une observation du groupe 2 est :

$$P(1|2) = P\left(\frac{S(\underline{x}) + \frac{1}{2}\Delta_p^2}{\Delta_p} > \frac{0 + \frac{1}{2}\Delta_p^2}{\Delta_p} \right)$$

$$P(1|2) = P\left(U > \frac{\Delta_p}{2} \right) \quad \text{où } U \text{ suit } \mathrm{N}\ (0\ ;\ 1)$$

Elle est égale à $P(2|1)$.

Cette relation donne une interprétation concrète à la distance de Mahalanobis.

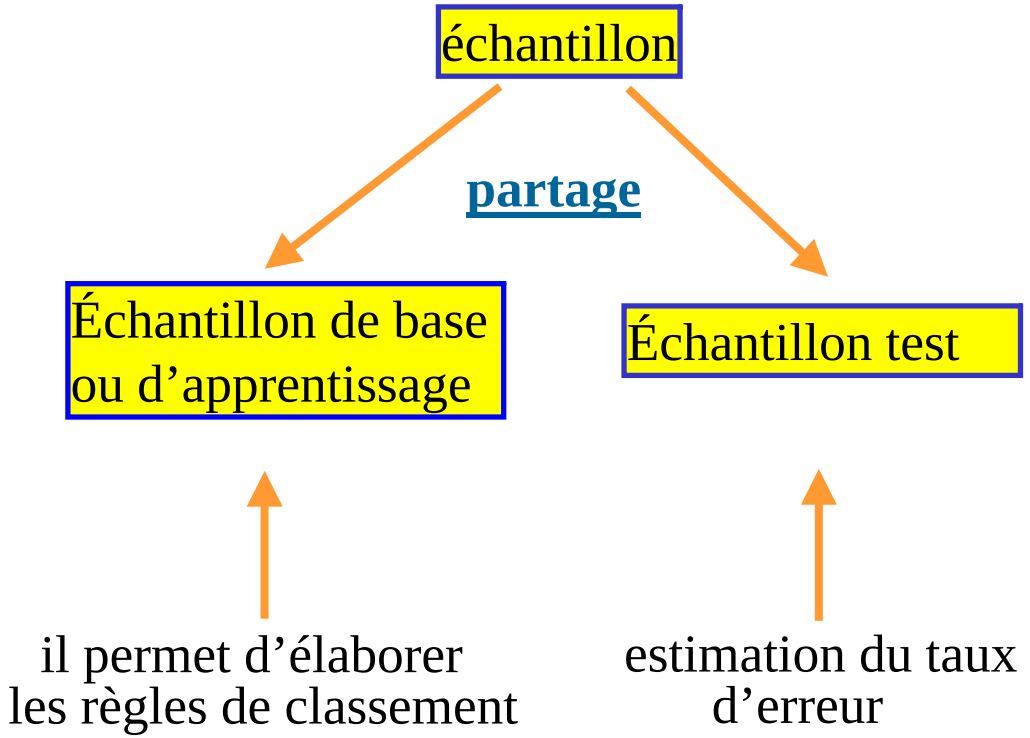
Remarque Estimations biaisées → sous-estimation du taux d'erreur.

3-2. Méthode de resubstitution

Réaffectation des observations selon les fonctions discriminantes trouvées.

Inconvénient On sous-estime le taux d'erreur.

3-3. Échantillon d'apprentissage; Échantillon test



3.4 Validation croisée

- Pour i = 1 à n, on construit la règle de décision sur la base privée de son $i^{ème}$ élément et on affecte ce dernier à l'un des groupes suivant cette règle.

- **Le taux d'erreur estimé est alors la fréquence de points mal classés de la sorte. L'estimation du taux d'erreur ainsi obtenu est pratiquement sans biais**

- La variance de l'estimation est d'autant plus importante que n est grand, puisque dans ce cas, les différentes règles de décision construites à partir de n-2 observations communes ont tendance à se ressembler.

4. Méthodes non paramétriques

Les méthodes non paramétriques consistent à estimer la densité de probabilité en chaque point de l'échantillon.

Deux méthodes sont souvent utilisées :

- méthode du noyau
- méthode des ***k*** plus proches voisins.

4.1 La méthode des noyaux

La méthode des noyaux généralise la notion d'histogramme. Dans le cas unidimensionnel, pour estimer la densité en un point x, on centre l'intervalle de longueur R de l'histogramme en ce point. La densité est alors le rapport de la probabilité de l'intervalle sur la longueur de l'intervalle.

Dans le cas multidimensionnel, considérons l'ellipsoïde centré sur $\underline{x}$:

$$\Omega_{r,j}(\underline{X}) = \left\{ y \,/\, (y-x)' V_j^{-1} (y-x) \leq r^2 \right\}$$

Notons $I_j(z)$ la variable indicatrice de l'ellipsoïde $\left\{z / z' V_j^{-1} z \leq r^2\right\}$

La densité de probabilité estimée s'écrit : $f_j(x) = \dfrac{\sum_{y \in G_j} I_j(y - x)}{n_j v_r(j)}$

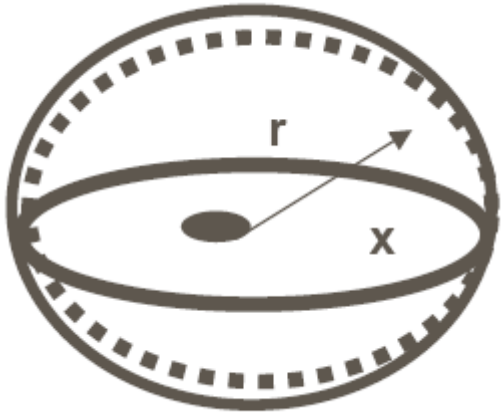
avec $n_j =$ nombre d'éléments du groupe j

$v_r(j) =$ volume de l'ellipsoïde

La méthode du noyau consiste à utiliser une fonction (le noyau) plus « lisse » que la variable indicatrice $I_j(z)$.

On trouve dans la littérature (et les logiciels) **différents types de noyaux** :

- uniforme: On compte le nombre d'observations appartenant à la boule de rayon R. Ce nombre est aléatoire.



- normal
- Epanechnikov
- biweight kernel
- triweight kernel

La difficulté d'utilisation de ces méthodes réside dans le choix du noyau et le choix de *r*.

4.2. Méthode des *k* plus proches voisins

On cherche les ***k*** points les plus proches de l'individu $\underline{x}$ et on classe $\underline{x}$ dans le groupe le plus représenté : la probabilité a posteriori d'appartenir au groupe ***j*** est égale au quotient entre le nombre d'individus du groupe ***j*** parmi les ***k*** points, et le nombre de voisins (***k***).

Le choix de ***k*** est moins crucial que le choix de r dans la méthode des noyaux. On peut choisir ***k*** optimisant une proportion de bien classés en validation croisée.

5. La régression logistique

Lorsqu'il n'y a pas que deux groupes, sous l'hypothèse de normalité et d'égalité des matrices de variance, la probabilité a posteriori est une fonction logistique du score, lui-même fonction linéaire des variables explicatives.

$$Log\left(\frac{1}{P}-1\right)=Log\,\frac{p_2}{p_1}\frac{f_2(\underline{x})}{f_1(\underline{x})}=\underbrace{-S(\underline{x})}$$

Donc :

$$Log\,\frac{f_2(\underline{x})}{f_1(\underline{x})} = Log\,\frac{p_1}{p_2} + \underbrace{\alpha + \underline{\beta}'\,\underline{x}}$$

$$(*)\quad \textbf{Log}\ \frac{f_2(x)}{f_1(x)} = \beta_0 + \underline{\beta}'\,\underline{x} \qquad \text{où} \quad \underline{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$$

Ceci amène à définir la régression logistique à partir de l'expression (*).

Hypothèse de travail

$$\text{Log}\ \frac{f_2(x)}{f_1(x)} = \beta_0 + \beta' \underline{x} \quad \text{où} \quad \underline{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$$

Le modèle de la régression logistique consiste à estimer les ***(p+1)*** paramètres selon le maximum de vraisemblance.

$$P(G_1|x) = \frac{p_1 f_1(\underline{x})}{p_1 f_1(\underline{x}) + p_2 f_2(\underline{x})} = \frac{\frac{p_1}{p_2}\frac{f_1(\underline{x})}{f_2(\underline{x})}}{1 + \frac{p_1}{p_2}\frac{f_1(\underline{x})}{f_2(\underline{x})}}$$

$$P(G_1|x) = \frac{\exp\left(Log \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x} \right)}{1 + \exp\left(Log \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x} \right)} \qquad (1)$$

$$P(G_2|x) = \frac{1}{1 + \exp\left(Log \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x} \right)} \qquad (2)$$

On montre que les expressions (1) et (2) sont conservées pour la famille de distributions :

$$f_i(\underline{x}) = c_i \exp\left[-\frac{1}{2} (\underline{x} - \underline{\mu}_i)' \sum^{-1} (\underline{x} - \mu_i) \right] h(\underline{x})$$

où h est une fonction arbitraire de $\underline{x}$ intégrable non négative et c_i une constante telle que f_i soit une densité de probabilité.

En effet, $\boldsymbol{h}$ n'intervient pas dans le calcul de (1) et (2) :

- si $\boldsymbol{h(x) \equiv 1}$ on retombe sur la loi multinormale
- **on peut faire intervenir des variables binaires dans le modèle**
- **on peut appliquer le modèle au cas où un groupe de la population est dissymétrique (*h(x)* constante dans la population normale, croissante ailleurs)**

→ **méthode générale**

Expression de la vraisemblance des β (n$_1$ et n$_2$ fixés)

$$L = \prod_{i \in G_1} f_1(\underline{x}_i) \prod_{i \in G_2} f_2(\underline{x}_i) \qquad \text{avec :} \quad f(\underline{x}) = p_1 f_1(\underline{x}) + p_2 f_2(\underline{x})$$

On a
$$\begin{cases} f_1(\underline{x}) = \dfrac{P(G_1|\underline{x}) f(\underline{x})}{p_1} \\ f_2(\underline{x}) = \dfrac{P(G_2|\underline{x}) f(\underline{x})}{p_2} \end{cases}$$

D'où $$L = \frac{1}{p_1^{n_1} p_2^{n_2}} \prod_{i \in G_1} P(G_1|\underline{x}_i) \prod_{i \in G_2} P(G_2|\underline{x}_i) \prod_{i=1}^{n_1+n_2} f(\underline{x}_i)$$

soit : $$L = \frac{L_1 L_2}{p_1^{n_1} p_2^{n_2}}$$

L_1 = vraisemblance conditionnelle des paramètres connaissant les $\underline{x}_i$

L_2 = densité (incondionnelle) des $\underline{x}_i$

f non connue, on estime $\beta_0, \beta_1 \ldots \beta_p$ par une méthode de maximum de vraisemblance conditionnelle :

$$\max_{\beta} \prod_{i \in G_1} \frac{\exp\left(Log \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x}_i \right)}{1 + \exp\left(Log \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x}_i \right)} \prod_{i \in G_2} \frac{1}{1 + \exp\left(Log \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x}_i \right)}$$

Nécessité d'utiliser une méthode numérique.

(Pas de solution analytique à l'équation de vraisemblance).

Les β étant estimés, la règle Bayésienne peut être appliquée pour les classements.

$$Log \frac{P(G_1 | \underline{x})}{P(G_2 | \underline{x})} = Log \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x}$$

On affectera au groupe 1 si $Log \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x} > 0$

Avantages - Inconvénients de la régression logistique

Résultats meilleurs que la règle géométrique, pour :

- des populations non gaussiennes
- des populations où Σ_1 très différent de Σ_2

  mais procédure de calcul plus complexe.

Lorsque les données proviennent de deux populations normales avec $\Sigma_1 = \Sigma_2$ la régression logistique est moins performante que l'analyse discriminante. Seul ($\frac{f_1}{f_2}$ supposé connu).

BIBLIOGRAPHIE CONCERNANT LES METHODES D'ANALYSE DISCRIMINANTE ET DE SEGMENTATION

Références générales en statistique

G. GOVAERT (Editeur) **" Analyse des données "** Hermés Lavoisier (2003)

L. LEBART, A. MORINEAU, M.PIRON
" Statistique exploratoire multidimensionnelle " 3ème édition Dunod (2000)

G. SAPORTA **"Probabilités, analyse des données et statistique"** 2ème édition Technip (2006).

S. TUFFERY **"Data mining et statistique décisionnelle"** Technip 2010

S.TUFFERY **« Étude de cas en statistique décisionnelle » Technip 2009**

M. TENENHAUS **"Statistique: Méthode pour décrire,expliquer et prévoir ".** Dunod (2006).

Analyse discriminante et Segmentation

BARDOS M. **« Analyse discriminante : Application au risque et scoring financier »** Dunod (2001)

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. **“ Classification And Regression Trees**. **”** Monterey, California, Wadsworth & Brooks (1984)

CELEUX G. (Editeur scientifique) **“ Analyse discriminante sur variables continues ”** Collection didactique INRIA (1990)

CELEUX G ;, NAKACHE J.P.**“ Analyse discriminante sur variables qualitatives ”** Polytechnica (1994)

DROESBEKE J-J., LEJEUNE M., SAPORTA G. (Editeurs) **“ Modèles statistiques explicative pour données qualitatives ”** Technip (2005)

HUBERTY C. **“Applied discriminant analysis”** Wiley (1994)

NAKACHE J-P., CONFAIS J. **“ Statistique explicative appliquée ”** Technip (2003)

TOMASSONE R., DANZART M., DAUDIN J.J., MASSON J.P. **“ Discrimination et classement ”** Masson (1988)

ZIGHED D.A., RAKOTOMALALA R. **“ Graphes d’induction ”** Hermès (2000)

Sites INTERNET

Le site de la Société Française de Statistique : www.sfds.asso.fr

L’aide en ligne du logiciel SAS : http://support.sas.com/documentation/online.doc

Le site de Statsoft sur la statistique et le data mining : www.statsoft.com

Liste de méthodes de segmentation : www.recursive-partitioning.com/classification_trees