Classification des données de grande dimension: application à la vision par ordinateur

Stéphane Girard, Charles Bouveyron, Cordelia Schmid

Classification des données de grande dimension. Application à la vision par ordinateur

Stéphane Girard

INRIA Rhône-Alpes, projet Mistis
`http ://mistis.inrialpes.fr/~girard`

en collaboration avec Charles Bouveyron et Cordelia Schmid

Plan

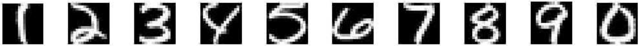
Plan

Introduction

La classification est un problème récurrent :

- qui intervient généralement dans les applications nécessitant une prise de décision,
- les données modernes sont souvent de grande dimension.

Exemple 1 : reconnaissance optique de caractères



$\downarrow$ 1 $\downarrow$ 2 $\downarrow$ 3 $\downarrow$ 4 $\downarrow$ 5 $\downarrow$ 6 $\downarrow$ 7 $\downarrow$ 8 $\downarrow$ 9 $\downarrow$ 0

Exemple 2 : reconnaissance d'objets à partir d'images

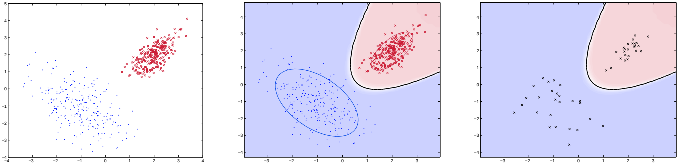
$\downarrow$ **Vélo** $\downarrow$ **Humain** $\downarrow$ **Voiture** $\downarrow$ **Moto**

Le problème de la classification

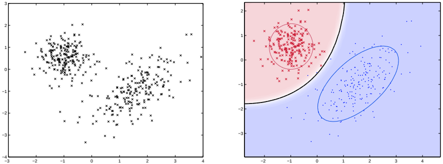
Le problème de la classification est :

- organiser des données $x_1, ..., x_n \in \mathbb{R}^p$ en k classes,
- les labels des données sont notés $z_1, ..., z_n \in \{1, ..., k\}$.

Approche supervisée : jeu de données complètes $(x_1, z_1), ...,$ (x_n, z_n) disponible pour l'apprentissage



Approche non-supervisée : uniquement les observations $x_1, ..., x_n$



Le modèle de mélange

On suppose classiquement que

- les observations $x_1, ..., x_n$ sont des réalisations indépendantes d'un vecteur aléatoire $X \in \mathbb{R}^p$,
- les labels $z_1, ..., z_n$ sont issus d'une variable aléatoire Z,

où :

- Z suit une loi multinomiale de paramètres $\pi_1, ..., \pi_k$ appelés proportions du mélange, *i.e.* $\mathbb{P}(Z = i) = \pi_i$, $i = 1, ..., k$.
- sachant $Z = i$, X suit une loi multidimensionnelle de densité $f_i(x)$.

En résumé, la densité de X s'écrit :

$$f(x) = \sum_{i=1}^{k} \pi_i f_i(x).$$

Règle de Bayes et modèle de mélange

La classification vise donc construire une règle de décision δ :

$$\begin{aligned} \delta : \mathbb{R}^p &\rightarrow \{1, ..., k\}, \\ x &\rightarrow z. \end{aligned}$$

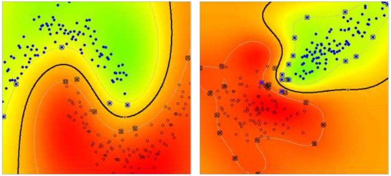
La règle optimale δ^* (pour un coût 0-1), dite règle de Bayes ou du MAP (Maximum A Posteriori), est :

$$\begin{aligned} \delta^*(x) &= \underset{i=1,...,k}{\operatorname{argmax}} \, \mathbb{P}(Z = i | X = x) \\ &= \underset{i=1,...,k}{\operatorname{argmax}} \, \mathbb{P}(X = x | Z = i) \mathbb{P}(Z = i) \\ &= \underset{i=1,...,k}{\operatorname{argmin}} \, K_i(x), \end{aligned}$$

où la fonction de coût K_i est telle que $K_i(x) = -2\log(\pi_i f_i(x))$.
Remarque *: la construction de la règle de décision consiste à estimer f_i ou de façon équivalente K_i.*

Fléau de la dimension en classification (1)

Classification non-paramétrique : On ne choisit pas de modèle *a priori* pour f_i. Estimateur de type histogramme ou noyau.



Fléau de la dimension [Bel57] en classification non-paramétrique :

- Point de vue pratique : il faut beaucoup d'observations pour estimer correctement une fonction de plusieurs variables.
- Point de vue théorique : erreur d'estimation d'une densité de p variables de l'ordre de $n^{-1/(p+2)}$.

Modèle gaussien Full-GMM (QDA en supervisé) :

$$K_i(x) = (x - \mu_i)^t \Sigma_i^{-1} (x - \mu_i) + \log(\det \Sigma_i) - 2\log(\pi_i) + C^{te}.$$

Modèle gaussien Com-GMM qui suppose que $\forall i,\ \Sigma_i = \Sigma$ (LDA en supervisé) :

$$K_i(x) = \mu_i^t \Sigma^{-1} \mu_i - 2\mu_i^t \Sigma^{-1} x - 2\log(\pi_i) + C^{te}.$$

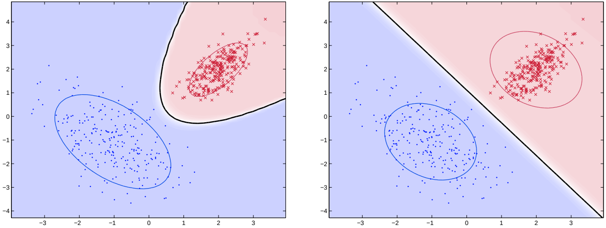


Fig. 1. Règles de décision de Full-GMM (gauche) et Com-GMM (droite).

Problème *: il est nécessaire d'inverser Σ_i ou Σ.*

Fléau de la dimension en classification (2)

Fléau de la dimension dans le cas du mélange gaussien :

- le nombre de paramètres croît avec le carré de la dimension,

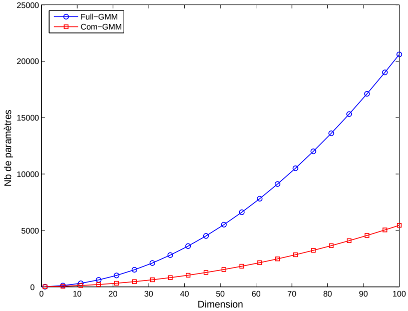


Fig. 2. Nombre de paramètres à estimer des modèles Full-GMM et Com-GMM en fonction de la dimension et ce pour 4 classes.

- si n est faible, les estimations des matrices de covariance sont mal conditionnées ou singulières,
- il est alors difficile ou impossible de les inverser et la règle de décision en est d'autant perturbée.

Solutions existantes : réduction de dimension

Réduction de dimension :

- de façon globale (ACP, sélection de variables, ...),
- liée au but de classification (FDA, ...).

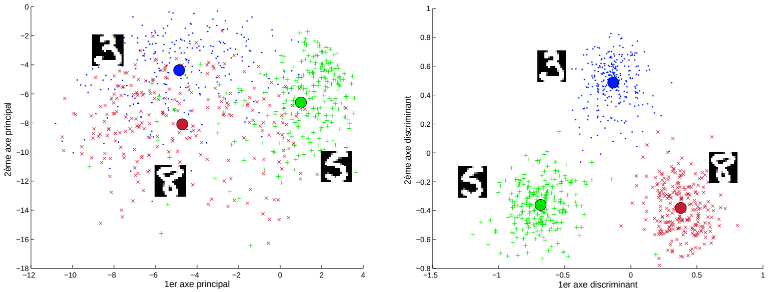


Fig. 3. Projection des données USPS $\in \mathbb{R}^{256}$ sur les 2 premiers axes principaux (gauche) et sur les 2 premiers axes discriminants (droite).

Solutions existantes : régularisation

Régularisation des estimateurs des matrices de covariance :

- type *ridge* : $\tilde{\Sigma}_i = \hat{\Sigma}_i + \sigma_i I_p$,
- PDA [Hast95] : $\tilde{\Sigma}_i = \hat{\Sigma}_i + \sigma_i \Omega$,
- RDA [Frie89] fournit un classifieur variant entre QDA et LDA.

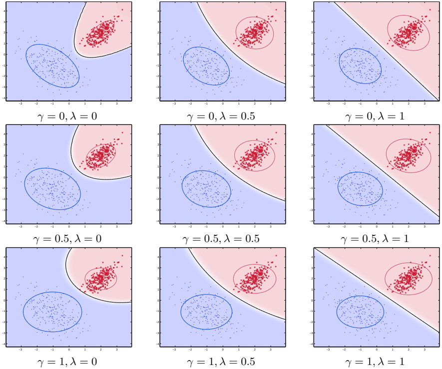


Fig. 4. Influence des paramètres γ et λ sur le classifieur RDA.

Solutions existantes : modèles parcimonieux

Modèles parcimonieux :

- diag-GMM : $\Sigma_i = \mathrm{diag}(\sigma_{i1}, ..., \sigma_{ip})$,
- sphe-GMM : $\Sigma_i = \sigma_i I_p$,
- re-paramétrisation de Celeux *et al.* [Cel95] : 14 modèles allant du plus général au plus parcimonieux ($\rightarrow$EDDA).

Classification dans des sous-espaces :

- modèle DSM de Flury *et al.* [Flur97],
- mélange de PPCA [Tip99].

Plan

Les « bienfaits » de la dimension

Le phénomène de l'espace vide [Scot83] met en évidence que :

- les espaces de grande dimension sont quasiment vides,
- les données vivent dans des sous-espaces de dimensions intrinsèques inférieures à la dimension de l'espace p.

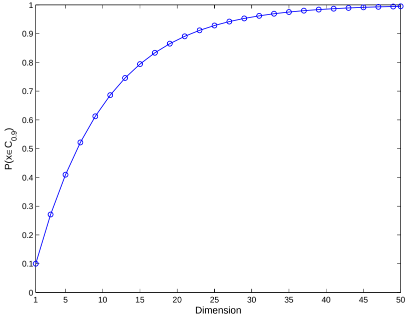


Fig. 5. Probabilité que $X \sim U_{B_p(0,1)}$ ***soit dans la coquille comprise entre les boules de rayon*** 0.9 ***et*** 1***, en fonction de la dimension :*** $\mathbb{P}(X \in C_{[0.9,1]}) = 1 - 0.9^p$***.***

Les « bienfaits » de la dimension

Un autre phénomène intervient en grande dimension :

- les espaces de grande dimension étant quasiment vides,
- il est plus facile de séparer les groupes en grande dimension avec un classifieur adapté.

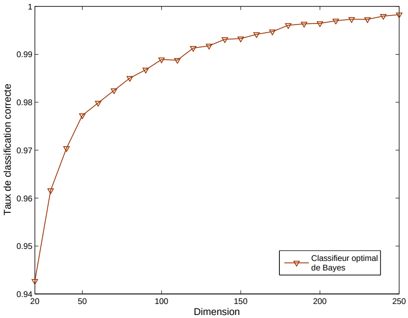


Fig. 6. Taux de classification correcte du classifieur optimal de Bayes en fonction de la dimension (données simulées).

L'idée de notre modélisation

Il est possible d'adapter ces postulats au cadre de la classification :

- les données de chaque classe vivent dans des sous-espaces différents de dimensions intrinsèques différentes,
- le fait de conserver toutes les dimensions permet de discriminer plus facilement les données.

Nous proposons donc une paramétrisation du modèle gaussien :

- qui exploite ces caractéristiques des données de grande dimension,
- au lieu de pallier les problèmes dus à la grande dimension des données.

Modélisation

Nous nous plaçons dans le cadre du modèle de mélange gaussien :

$$f(x) = \sum_{i=1}^{k} \pi_i f(x, \theta_i), \text{ avec } f(x, \theta_i) \sim \mathcal{N}(\mu_i, \Sigma_i).$$

En se basant sur la décomposition spectrale de Σ_i, on peut écrire :

$$\Sigma_i = Q_i \Delta_i Q_i^t,$$

où :

- Q_i est la matrice orthogonale des vecteurs propres de Σ_i,
- Δ_i est la matrice diagonale des valeurs propres de Σ_i.

Modélisation

Nous proposons de paramétrer la matrice Δ_i de la façon suivante :

$$\Delta_i = \left(\begin{array}{ccc|ccc} a_{i1} & & 0 & & & \\ & \ddots & & & \mathbf{0} & \\ 0 & & a_{id_i} & & & \\ \hline & & & b_i & & 0 \\ & \mathbf{0} & & & \ddots & \\ & & & & \ddots & \\ & & & 0 & & b_i \end{array} \right) \begin{array}{l} \left.\vphantom{\begin{array}{c} a \\ \ddots \\ a \end{array}}\right\} d_i \\ \left.\vphantom{\begin{array}{c} b \\ \ddots \\ \ddots \\ b \end{array}}\right\} (p - d_i) \end{array}$$

où $a_{ij} \geq b_i$, pour $j = 1, ..., d_i$.

Remarque *: cette paramétrisation est toujours possible car si on prend $d_i = p - 1$, pour $i = 1, ..., k$, alors on a le modèle Full-GMM.*

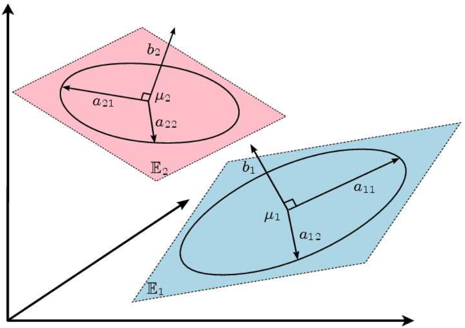


Fig. 7. Notre paramétrisation du modèle de mélange gaussien.

Nous définissons en outre :

- $\mathbb{E}_i$ l'espace engendré par les vect. prop. associés aux a_{ij},
- $\mathbb{E}_i^\perp$ son supplémentaire dans $\mathbb{R}^p$,
- P_i et $P_i^\perp$ les opérateurs de projection sur $\mathbb{E}_i$ et $\mathbb{E}_i^\perp$.

Le modèle $[a_{ij}b_iQ_id_i]$ et ses sous-modèles

Ainsi, nous obtenons une paramétrisation du modèle gaussien :

- qui est fonction de a_{ij}, b_i, Q_i et d_i,
- dont la complexité est contrôlée par les dimensions d_i des sous-espaces,
- que nous noterons $[a_{ij}b_iQ_id_i]$ dans la suite.

En forçant certains paramètres à être communs dans une même classe ou entre les classes :

- nous obtenons des modèles de plus en plus régularisés,
- qui vont du modèle général au modèle le plus parcimonieux.

Notre famille contient 28 modèles répartis de la façon suivante :

- 14 modèles à orientations libres,
- 12 modèles à orientation commune,
- 2 modèles à matrice de covariance commune.

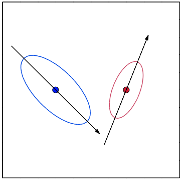
Le modèle $[a_{ij}b_iQ_id_i]$ et ses sous-modèles

Modèle	Nb de prms, $k = 4$ $d = 10$, $p = 100$	Type de classifieur
$[a_{ij}b_iQ_id_i]$	4231	Quadratique
$[a_{ij}b_iQd_i]$	1396	Quadratique
$[a_jbQd]$	1360	Linéaire
Full-GMM	20603	Quadratique
Com-GMM	5453	Linéaire

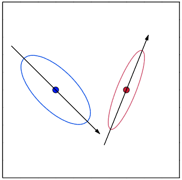
Table 1. Propriétés des modèles de la famille de $[a_{ij}b_iQ_id_i]$

Remarque *: le modèle* $[a_{ij}b_iQ_id_i]$ *qui engendre un classifieur quadratique requiert l'estimation de moins de paramètres que le modèle Com-GMM qui engendre un classifieur linéaire.*

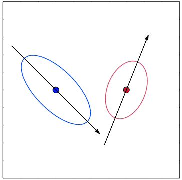
Le modèle $[a_{ij}b_iQ_id_i]$ et ses sous-modèles



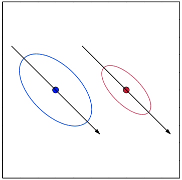
modèle $[a_ib_iQ_id]$



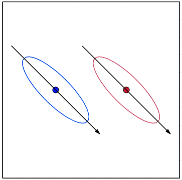
modèle $[ab_iQ_id]$



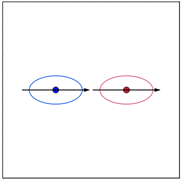
modèle $[a_ibQ_id]$



modèle $[a_ib_iQd]$



modèle $[abQd]$



modèle $[abI_2d]$

Fig. 8. Influence des paramètres a_i, b_i et Q_i sur les densités de 2 classes en dimension 2 et avec $d_1 = d_2 = 1$.

Plan

Construction du classifieur HDDA

En supervisé, l'estimation des paramètres par MV est directe :

$$\hat{\pi}_i = \frac{n_i}{n},\ \ \hat{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n} s_{ij}x_j,$$

$$\hat{\Sigma}_i = W_i = \frac{1}{n_i}\sum_{j=1}^{n} s_{ij}(x_j - \hat{\mu}_i)(x_j - \hat{\mu}_i)^t,$$

où $n_i = \sum_{j=1}^{n} s_{ij}$ avec $s_{ij} = 1_{\{z_j = i\}}$.

Calcul des probabilités conditionnelles :

$$\mathbb{P}(Z = i|X = x_j, \theta) = 1/\sum_{\ell=1}^{k} \exp\left(\frac{1}{2}(K_i(x_j) - K_\ell(x_j))\right),$$

où la fonction de coût K_i est telle que $K_i(x) = -2\log(\pi_i f(x, \theta_i))$.

Expression de la fonction de coût K_i

Dans le cas du modèle $[a_i b_i Q_i d_i]$:

$$K_i(x) = \frac{1}{a_i}\|\mu_i - P_i(x)\|^2 + \frac{1}{b_i}\|x - P_i(x)\|^2 + d_i \log(a_i) + (p-d_i)\log(b_i) - 2\log(\pi_i).$$

Points forts *:*

- *pas besoin d'inverser la matrice de covariance,*
- *ni d'estimer les dernières colonnes de la matrice Q_i.*

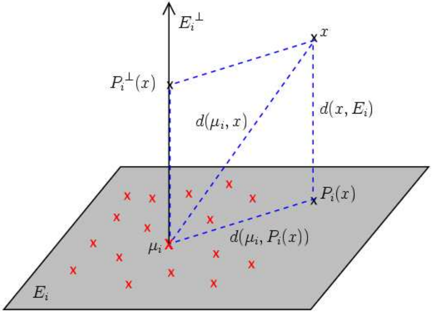


Fig. 9. Les sous-espaces $\mathbb{E}_i$ et $\mathbb{E}_i^\perp$ de la $i$ème composante.

Construction du classifieur HDDC

En non supervisé, les paramètres sont estimés par l'algorithme EM :

- Étape E : cette étape calcule à l'itération q les probabilités conditionnelles $t_{ij}^{(q)} = \mathbb{P}(Z = i | X = x_j, \theta^{(q)})$:

$$t_{ij}^{(q)} = 1 / \sum_{\ell=1}^{k} \exp\left(\frac{1}{2}(K_i^{(q-1)}(x_j) - K_\ell^{(q-1)}(x_j))\right).$$

- Étape M : cette étape calcule les estimateurs des θ_i en maximisant la vraisemblance conditionnellement aux $t_{ij}^{(q)}$:

$$\hat{\pi}_i^{(q)} = \frac{n_i^{(q)}}{n}, \ \hat{\mu}_i^{(q)} = \frac{1}{n_i^{(q)}} \sum_{j=1}^{n} t_{ij}^{(q)} x_j,$$

$$\hat{\Sigma}_i^{(q)} = W_i^{(q)} = \frac{1}{n_i^{(q)}} \sum_{j=1}^{n} t_{ij}^{(q)} (x_j - \hat{\mu}_i^{(q)})(x_j - \hat{\mu}_i^{(q)})^t,$$

où $n_i^{(q)} = \sum_{j=1}^{n} t_{ij}^{(q)}$.

Estimations des a_{ij}, b_i et Q_i

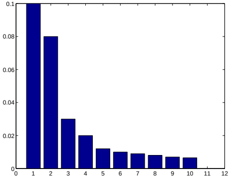
Les estimateurs du MV des paramètres du modèle $[a_{ij}b_iQ_id_i]$ sont explicites :

- Sous-espace $\mathbb{E}_i$: les d_i premières colonnes de Q_i sont estimées par les vecteurs propres associés aux d_i plus grandes valeurs propres λ_{ij} de W_i.
- Estimateur de a_{ij} : les paramètres a_{ij} sont estimés par les d_i plus grandes valeurs propres λ_{ij} de W_i.
- Estimateur de b_i : le paramètre b_i est estimé par :

$$\hat{b}_i = \frac{1}{(p - d_i)} \left(\mathrm{trace}(W_i) - \sum_{j=1}^{d_i} \lambda_{ij} \right).$$

Remarque *: 16 des modèles de notre famille ont des estimateurs du MV explicites. Les autres requièrent une méthode itérative.*

Estimation des paramètres discrets



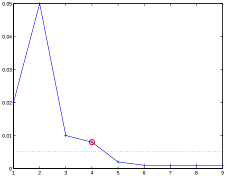


Fig. 10. Le scree-test de Cattell : éboulis des valeurs propres (gauche) et différences entre valeurs propres consécutives (droite).

Estimation des dimensions intrinsèques d_i :

- nous utilisons la méthode du *scree-test* de Cattell [Catt66],
- cela permet d'estimer de façon commune les k paramètres d_i,
- en supervisé, le seuil s est choisi par validation croisée,
- en non supervisé, s est choisi grâce au critère BIC [Schw78].

Estimation du nombre de groupes k :

- en supervisé, k est connu,
- en non supervisé, k est choisi grâce au critère BIC.

Considérations numériques

- Stabilité numérique : la règle de décision des classifieurs HDDA et HDDC ne dépend pas des vecteurs propres associés aux plus petites valeurs propres de W_i dont la détermination est instable.

- Réduction de la durée de calcul : pas besoin de déterminer les derniers vecteurs propres de W_i $\rightarrow$ réduction des temps de calcul avec une procédure adaptée ($\times 60$ pour $p = 1000$).

- Cas particulier où $n < p$: il est alors préférable, d'un point de vue numérique, de calculer les vecteurs propres de $U_i U_i^t$ au lieu de $W_i = U_i^t U_i$ où U_i contient les données centrées de C_i ($\times 500$ pour $n = 13$ et $p = 1000$).

Plan

HDDA : influence de la dimension

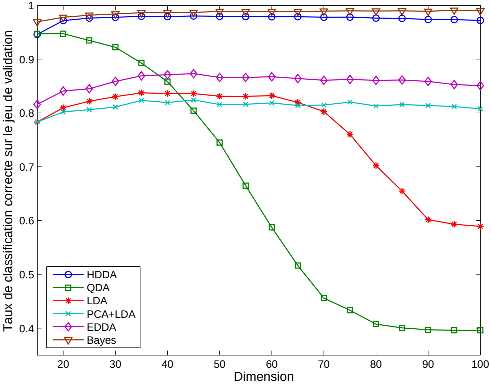


Fig. 11. Taux de classification correcte en fonction de la dimension de données (données simulées selon $[a_{ij}b_iQ_id_i]$***).***

Il apparaît que :

- l'HDDA est peu sensible à la dimension des données,
- l'HDDA est aussi performante en grande dimension que le classifieur quadratique (QDA) l'est en dimension faible.

HDDA : influence de la taille du jeu d'apprentissage

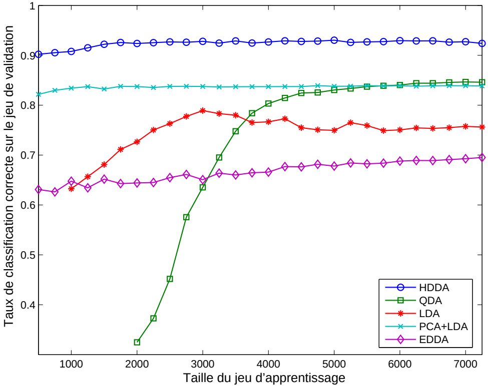


Fig. 12. Taux de classification correcte en fonction de la taille du jeu d'apprentissage (données réelles USPS $\in \mathbb{R}^{256}$***).***

Il apparaît que :

- l'HDDA est peu sensible à la taille du jeu d'apprentissage,
- l'HDDA est plus performante que les autres méthodes sur ce jeu de données réelles.

HDDA : comparaison avec les méthodes classiques

Méthode	Taux de classif. correcte	Temps d'app. (sec.)
HDDA $[a_{ij}bQ_id]$	**0.948**	~ 1
RDA ($\gamma = 0.3,\ \lambda = 0$)	0.935	~ 1
QDA (full-GMM)	0.846	~ 1
LDA (com-GMM)	0.757	~ 1
EDDA $[\lambda_k B_k]$	0.696	~ 1
SVM (linéaire)	0.926	~ 12

Table 2. Résultats de classification obtenus sur les données USPS ($p = 256$, $n_{app} = 7250$).

Il apparaît que :

- l'HDDA est plus performante que les autres méthodes sur ce jeu de données réelles,
- l'HDDA est aussi rapide que les autres méthodes basées sur le modèle de mélange (hors choix de modèles).

HDDC : sélection de modèles

Modèle de simulation	Modèle de classification					
	$[a_{ij}b_iQ_id_i]$	$[a_{ij}bQ_id_i]$	$[a_ib_iQ_id_i]$	$[a_ibQ_id_i]$	$[ab_iQ_id_i]$	$[abQ_id_i]$
$[a_{ij}b_iQ_id_i]$	96.7	82.8	97.3*	91.9	**97.5***	90.3
$[a_{ij}bQ_id_i]$	73.0	72.7	77.9	**78.2***	75.8	75.1
$[a_ib_iQ_id_i]$	97.9	87.1	98.3*	92.9	**98.6***	91.7
$[a_ibQ_id_i]$	82.6	80.0	**88.2***	86.3*	87.5	86.5
$[ab_iQ_id_i]$	96.5	82.5	**98.0***	84.4	95.2	82.2
$[abQ_id_i]$	71.2	75.2	**79.7**	79.3*	71.1	70.7

Table 3. Taux de classification correcte (en %) obtenus par les modèles de l'HDDC sur différents jeux de données simulées. Le modèle choisi par le critère BIC est noté d'une étoile.

Il apparaît que :

- le modèle $[a_ib_iQ_id_i]$ semble être particulièrement efficace,
- l'hypothèse que Δ_i n'a que deux valeurs propres différentes semble être un moyen efficace de régulariser son estimation.

HDDC : estimation des paramètres discrets

Nb de classes k	Seuil choisi s	Dimensions d_i	Valeur BIC
2	0.18	2,16	414
3	0.21	**2,5,10**	407
4	0.25	2,2,5,10	414
5	0.28	2,5,5,10,12	416
6	0.28	2,5,6,10,10,12	424

Table 4. Sélection du nombre de classes et des dimensions grâce au critère BIC. Les données contiennent 3 groupes dont les dimensions d_i sont 2, 5 et 10.

Il apparaît que :

- l'estimation du nombre de classes k par BIC est efficace,
- l'estimation des dimensions d_i grâce au *scree-test* de Cattell et au critère BIC est satisfaisante.

HDDC : comparaison avec la sélection de variables

Modèle	Variables originales	Avec réduction de dimension (ACP)
Sphe-GMM	0.340	0.340
Diag-GMM	0.355	0.535
Com-GMM	0.625	0.635
Full-GMM	0.640	0.845
VS-GMM [Raft05]	0.925	/
HDDC $[a_i b_i Q_i d_i]$	**0.950**	/

Table 5. Taux de classification correcte sur les données « Crabes ».

Il apparaît que :

- notre approche est plus efficace que la réduction de dimension et la sélection de variables sur ce jeu de données réelles,
- l'HDDC est efficace même en dimension faible sur des données complexes.

HDDC : les étapes de l'algorithme

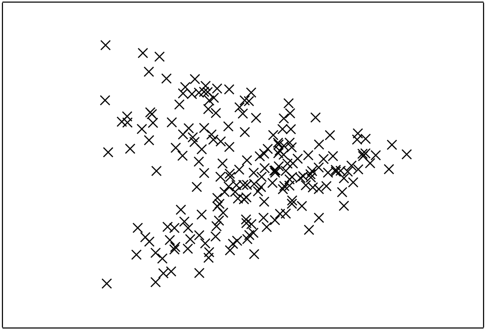


Fig. 13. Projection des données « Crabes » sur les axes principaux.

Données « Crabes » :

- 200 individus en dimension $p = 5$ (5 caractéristiques morphologiques des crabes),
- répartis en 4 classes (MB, FB, MO et FO).

HDDC : les étapes de l'algorithme

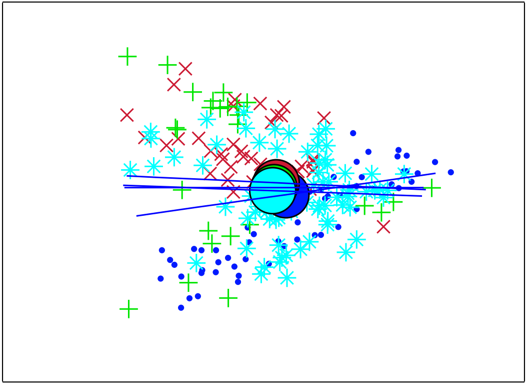


Fig. 14. Etape n° 1 de l'HDDC sur les données « Crabes ».

HDDC : les étapes de l'algorithme

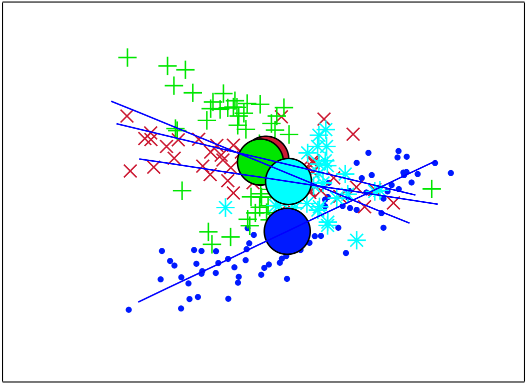


Fig. 14. Etape n° 4 de l'HDDC sur les données « Crabes ».

HDDC : les étapes de l'algorithme

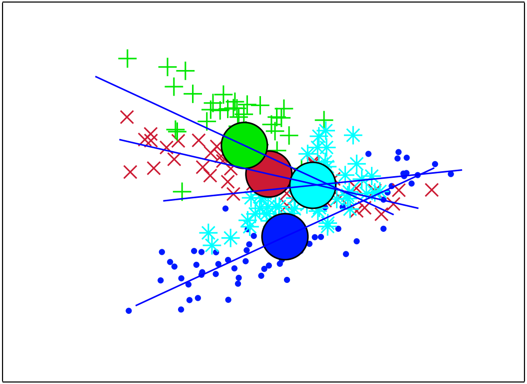


Fig. 14. Etape n° 7 de l'HDDC sur les données « Crabes ».

HDDC : les étapes de l'algorithme

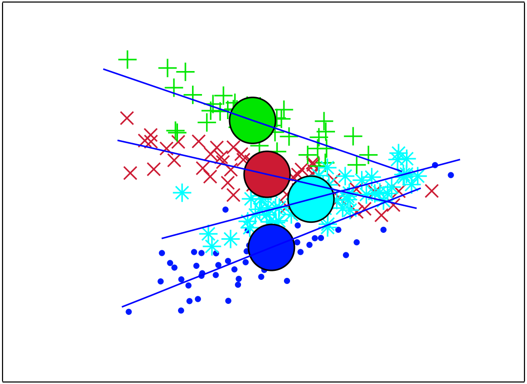


Fig. 14. Etape n˚10 de l'HDDC sur les données « Crabes ».

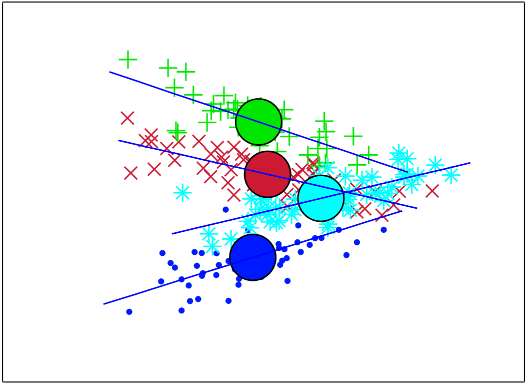


Fig. 14. Etape n˚12 de l'HDDC sur les données « Crabes ».

Application à la caractérisation du sol de Mars

- ANR (2008–2010) avec le Laboratoire de Planétologie de Grenoble.
- Images hyperspectrales de la planète Mars.
- En chaque pixel $i = 1, \ldots, n$, on mesure un spectre de $p = 256$ longueurs d'ondes. On a des centaines de milliers d'observations ($n = 200.000$, pour une image de taille 200×1000).
- Classification en $k = 5$ catégories : eau, CO_2, glace, minéraux, et poussière.
- On se limite à une classification non-supervisée, ne prenant en compte ni la nature spatiale de l'image, ni la structure fonctionnelle des observations (courbes).

Application à la caractérisation du sol de Mars

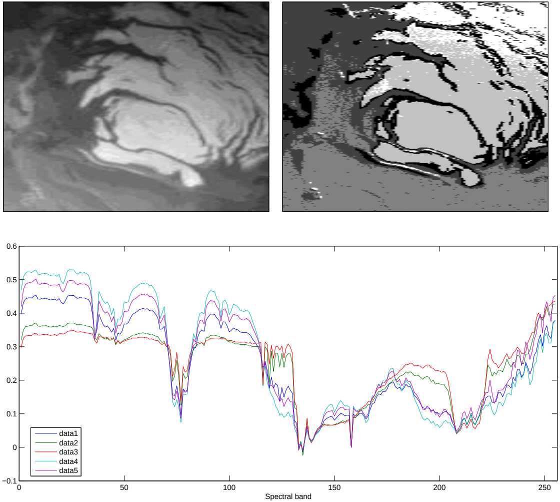


Fig. 15. Analyse de données hyperspectrales de Mars.

Application à la localisation d'objets en image

- ACI (2004–2006) avec le projet LEAR (INRIA Rhône-Alpes).
- Chaque image est représentée par environ 250 descripteurs - vecteurs de 128 caractéristiques locales (gradient, histogramme local des niveaux de gris, ...) - calculés en des points d'intérêt détectés automatiquement [Mik03].
- Au total, des milliers d'observations en dimension $p = 128$ à classer en $k = 2$ catégories : objet/fond.
- Classification semi-supervisée : on sait si l'objet est présent ou non dans l'image, mais on ne connait pas la classe des points d"intérêt.

Application à la localisation d'objets en image

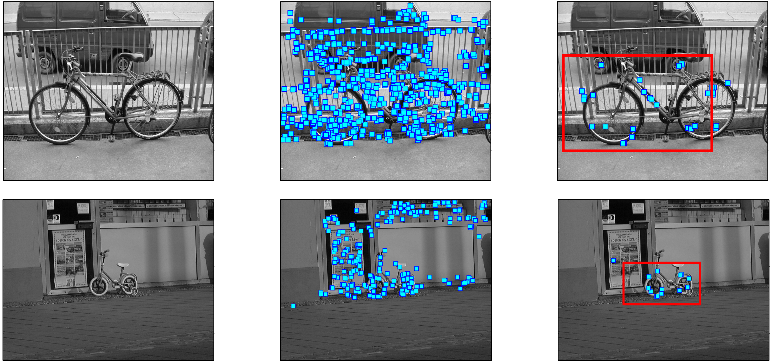


Fig. 16. Localisation de l'objet "vélo" sur des images de test.

Plan

Conclusion

Une famille de modèles gaussiens pour la grande dimension :

- qui prend en compte le fait que les données de grande dimension vivent dans sous-espaces de dimensions faibles,
- dont la complexité est contrôlée par les dimensions des sous-espaces spécifiques,
- qui va du modèle le plus général au modèle le plus parcimonieux,
- qui donne naissance aux méthodes HDDA et HDDC dont les règles de décision sont interprétables.

Travaux en cours et perspectives

- intégration de l'HDDA dans le logiciel MIXMOD,
- classification des données de grande dimension spatialement corrélées (avec J. Blanchet),
- catégorisation automatique du sol de la planète Mars (avec le Laboratoire de Planétologie de Grenoble).

References

- L. Bergé, C. Bouveyron and S. Girard. HDclassif : An R package for model-based clustering and discriminant analysis of high-dimensional data, *Journal of Statistical Software*, 46, 1–29, 2012.
- C. Bouveyron, G. Celeux and S. Girard Intrinsic dimension estimation by maximum likelihood in isotropic probabilistic PCA, *Pattern Recognition Letters*, 32, 1706–1713, 2011.
- C. Bouveyron and S. Girard. Robust supervised classification with mixture models : Learning from data with uncertain labels, *Pattern Recognition*, 42, 2649–2658, 2009.
- C. Bouveyron, S. Girard and C. Schmid. High Dimensional Data Clustering, *Computational Statistics and Data Analysis*, 52, 502–519, 2007.
- C. Bouveyron, S. Girard and C. Schmid. High-dimensional discriminant Analysis, *Communications in Statistics - Theory and Methods*, 36(14), 2607–2623, 2007.
- C. Bouveyron, S. Girard and C. Schmid. *Class-specific subspace discriminant analysis for high-dimensional data*, In C. Saunders et al., eds, LNCS, vol. 3940, p. 139-150. Springer-Verlag, 2006.