



HAL
open science

Deux méthodes statistiques pour la classification et la régression en grande dimension

Stéphane Girard

► **To cite this version:**

Stéphane Girard. Deux méthodes statistiques pour la classification et la régression en grande dimension. 2019. hal-02156948

HAL Id: hal-02156948

<https://hal.science/hal-02156948v1>

Preprint submitted on 14 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Deux méthodes statistiques pour la classification et la régression en grande dimension

Stéphane Girard

Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

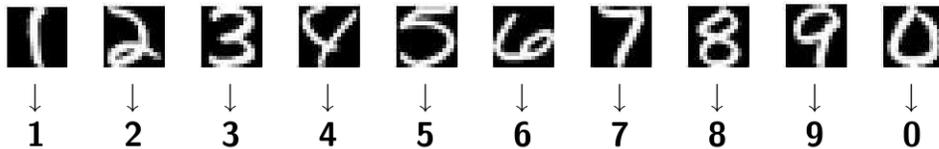
1 Classification en grande dimension (HDDA & HDDC)

Introduction

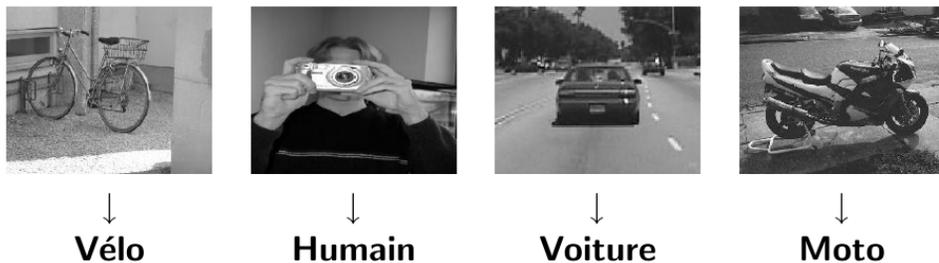
La classification est [un problème récurrent](#) :

- 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 (USPS)



[Exemple 2](#) : reconnaissance d'objets à partir d'images

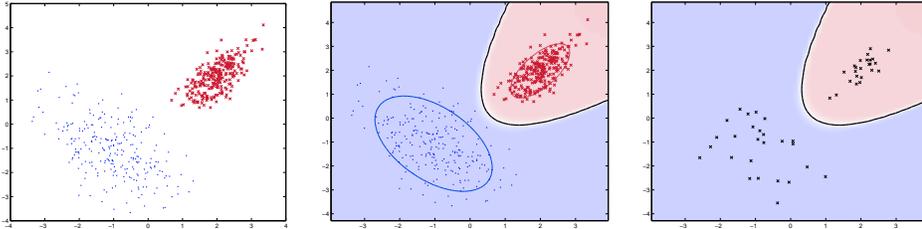


Le problème de la classification

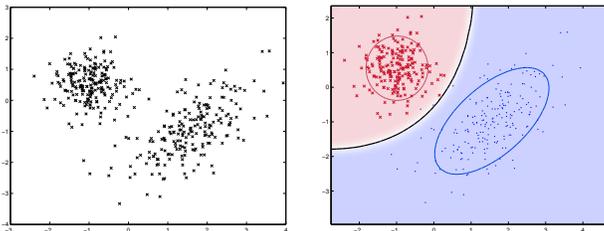
Le [problème de la classification](#) est d'organiser des données en classes :

- observations quantitatives $x_1, \dots, x_n \in \mathbb{R}^p$,
- labels $z_1, \dots, z_n \in \{1, \dots, k\}$.

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



Approche non-supervisée : uniquement les observations x_1, \dots, x_n (clustering)



Le modèle de mélange

On suppose classiquement que

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

où :

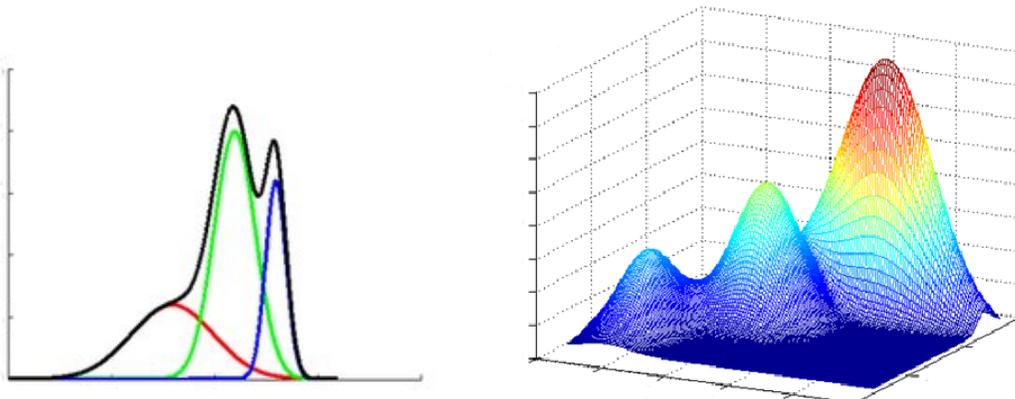
- Z suit une loi multinomiale de paramètres π_1, \dots, π_k appelés proportions du mélange ou probabilités a priori, *i.e.* $\mathbb{P}(Z = i) = \pi_i, i = 1, \dots, k$.
- sachant $Z = i, X$ suit une loi multidimensionnelle de densité $f_i(x)$.

En résumé, la densité de X s'écrit :

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

Modèles de mélange

Exemples de densités obtenues avec $k = 3$ composantes gaussiennes en dimension $p = 1$ (gauche) et $p = 2$ (droite).



Règle de Bayes et modèle de mélange

La classification vise donc construire une **règle de décision** δ :

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

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

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

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

Modèles gaussiens

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):

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

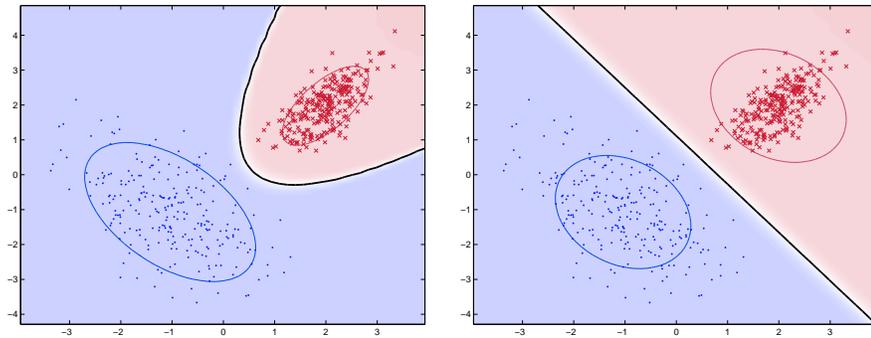


Fig. 1. Règles de décision de Full-GMM (gauche) et Com-GMM (droite).

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

Fléau de la dimension en classification

Fléau de la dimension dans le cas du mélange gaussien :

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

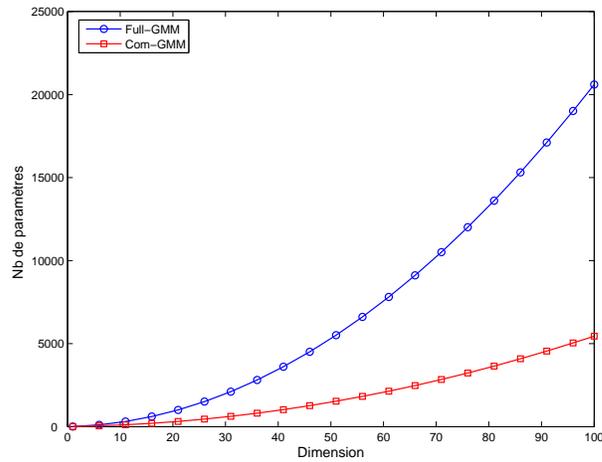


Fig. 2. Nombre de paramètres à estimer des modèles Full-GMM et Com-GMM en fonction de la dimension et ce pour 4 classes.

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 .

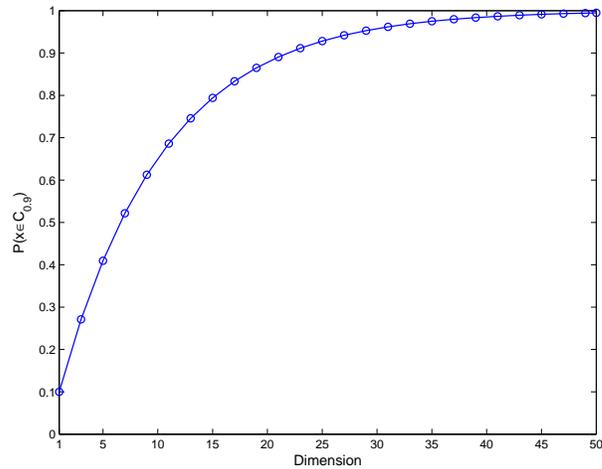


Fig. 5. Probabilité que $X \sim U_{B_p(0,1)}$ soit dans la coquille comprise entre les boules de rayon 0.9 et 1, en fonction de la dimension : $\mathbb{P}(X \in C_{[0.9,1]}) = 1 - 0.9^p$.

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 classes en grande dimension avec une méthode adaptée.

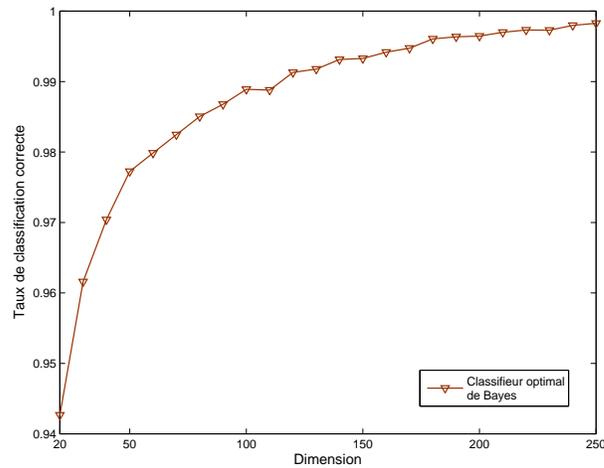


Fig. 6. Taux de classification correcte du classifieur optimal de Bayes en fonction de la dimension (données simulées).

Idée de modélisation

Il est possible d'adapter ces idées 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.

Introduction d'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

On se place 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 \$\Sigma_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

Il est possible de paramétrer la matrice Δ_i de la façon suivante :

$$\Delta_i = \begin{pmatrix} \begin{matrix} a_{i1} & & 0 \\ & \ddots & \\ 0 & & a_{id_i} \end{matrix} & & \mathbf{0} \\ & & & \begin{matrix} b_i & & 0 \\ & \ddots & \\ 0 & & b_i \end{matrix} \end{pmatrix} \left. \begin{array}{l} \} \\ \} \end{array} \right\} \begin{array}{l} d_i \\ (p - d_i) \end{array}$$

où $a_{ij} \geq b_i$, pour $j = 1, \dots, d_i$, la dimension intrinsèque de la classe.

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

Modélisation

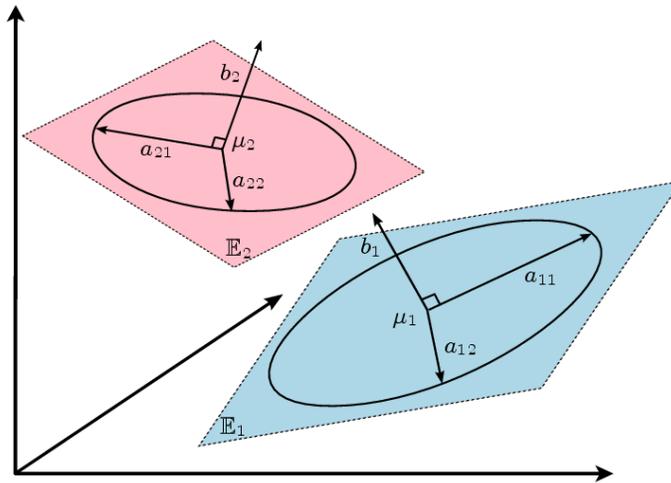


Fig. 7. Paramétrisation du modèle de mélange gaussien.

On définit également :

- \mathbb{E}_i l'espace intrinsèque 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, on obtient 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,
- noté $[a_{ij}b_iQ_id_i]$ dans la suite.

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

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

Cette 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

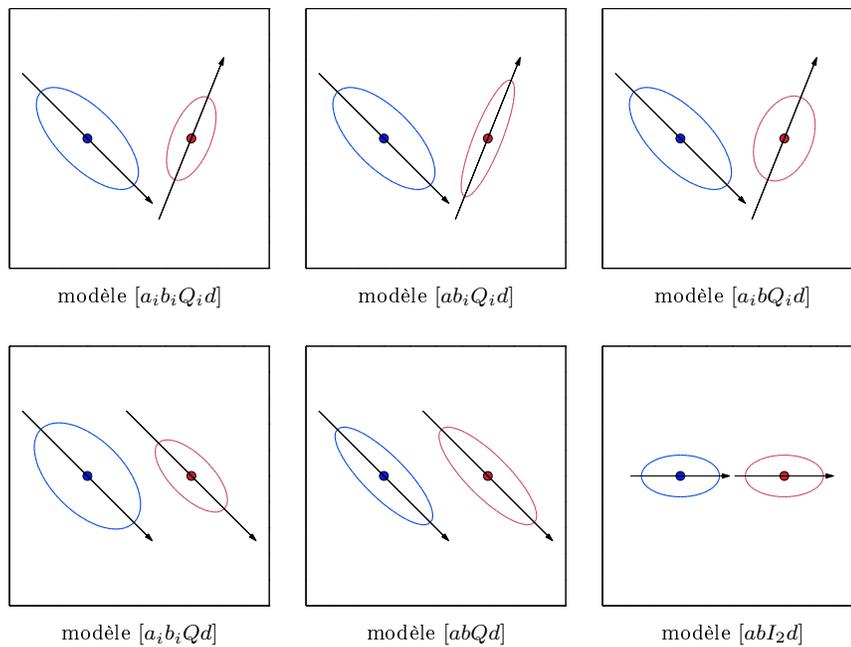


Fig. 8. Influence des paramètres a_i , b_i et Q_i sur les densités de 2 classes en dimension 2 et avec $d_1 = d_2 = 1$.

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
$[a_{ij}b_iQ_d]$	1396	Quadratique
$[a_j b Q d]$	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.

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 numériquement la matrice de covariance,
- ni d'estimer les dernières colonnes de la matrice Q_i .

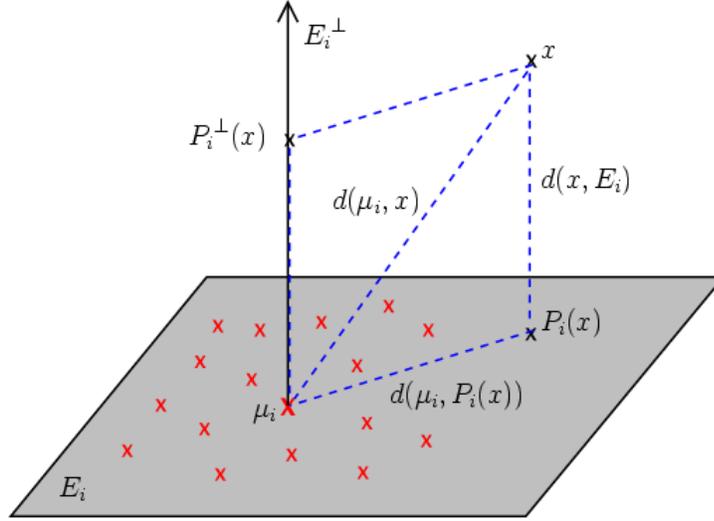


Fig. 9. Les sous-espaces E_i et E_i^\perp de la i ème composante.

Construction du classifieur HDDA

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

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

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

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

Calcul des probabilités conditionnelles :

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

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

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

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

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

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

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

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 **explicités** :

- **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 :

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

Remarque : 16 modèles ont des estimateurs du MV explicites. Les 12 autres requièrent une méthode itérative.

Estimation des paramètres discrets

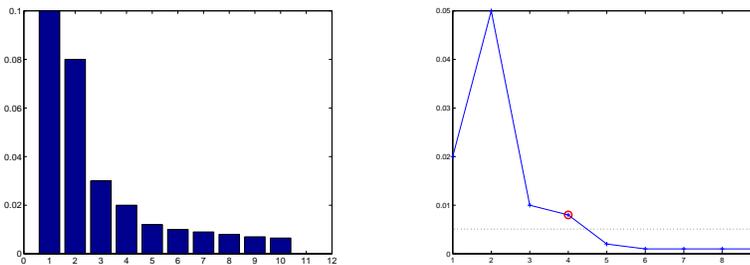


Fig. 10. Le scree-test de Cattell : éboulis des valeurs propres (gauche) et différences entre valeurs propres consécutives (droite).

Estimation des **dimensions intrinsèques d_i** :

- On utilise 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 classes 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 → 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 $p = 1000$).

HDDA : influence de la taille du jeu d'apprentissage

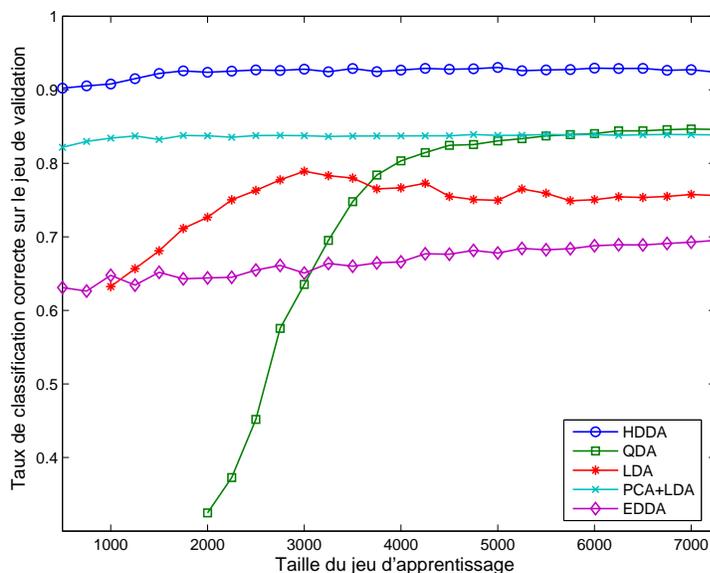


Fig. 12. Taux de classification correcte en fonction de la taille du jeu d'apprentissage (données réelles USPS $\in \mathbb{R}^{256}$).

Il apparaît que HDDA est peu sensible à la taille du jeu d'apprentissage,

HDDA : comparaison avec les méthodes classiques

Méthode	Taux de classif. correcte	Temps d'app. (sec.)
HDDA $[a_{ij} b Q_i d]$	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 :

- HDDA est plus performante que les autres méthodes sur ce jeu de données réelles,
- 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 : les étapes de l'algorithme

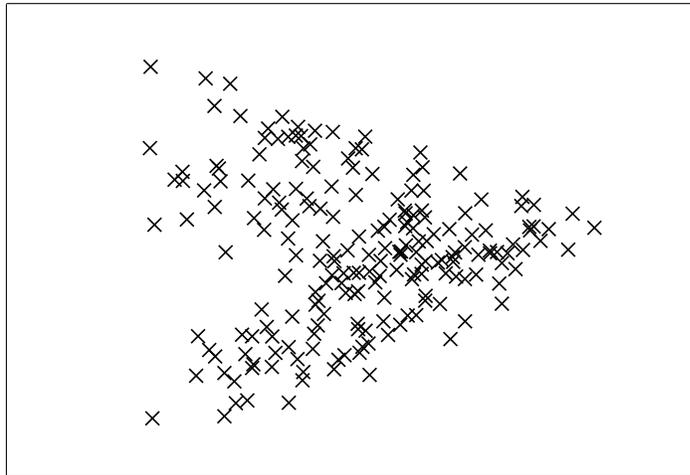


Fig. 13. Projection des données "Crabes" sur les axes principaux.

Données "Crabe" :

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

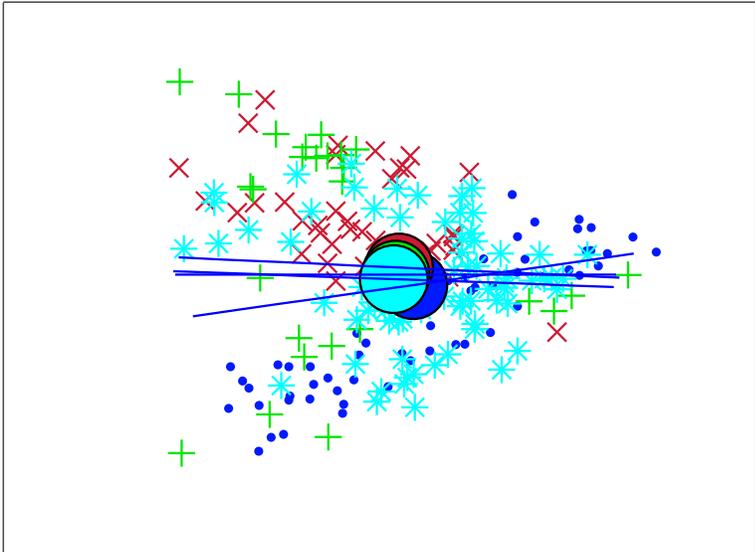


Fig. 14. Itération 1 de HDDC sur les données "Crabs".

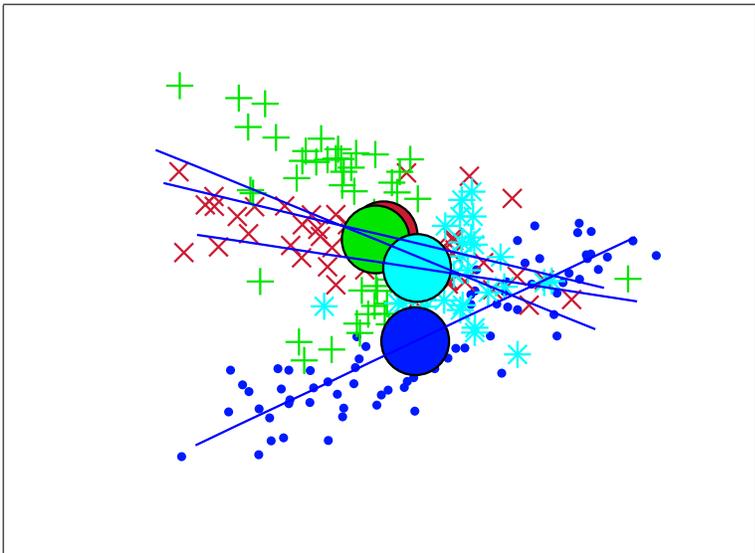


Fig. 14. Itération 4 de HDDC sur les données "Crabs".

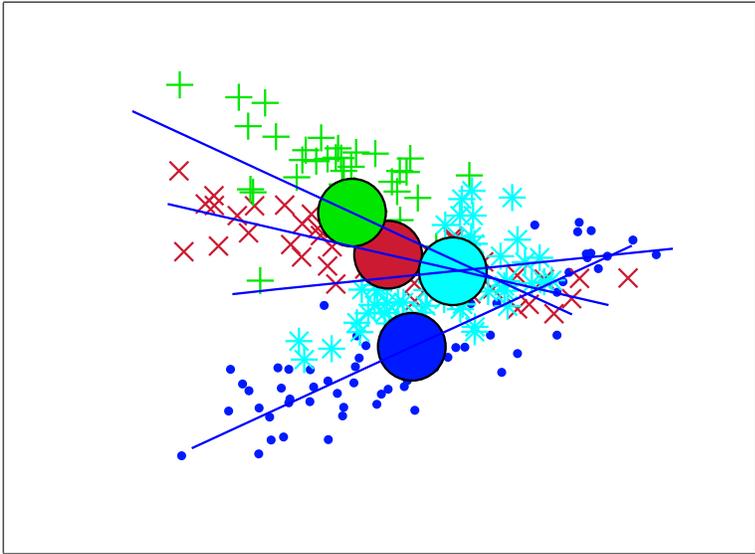


Fig. 14. Itération 7 de HDDC sur les données "Crabes".

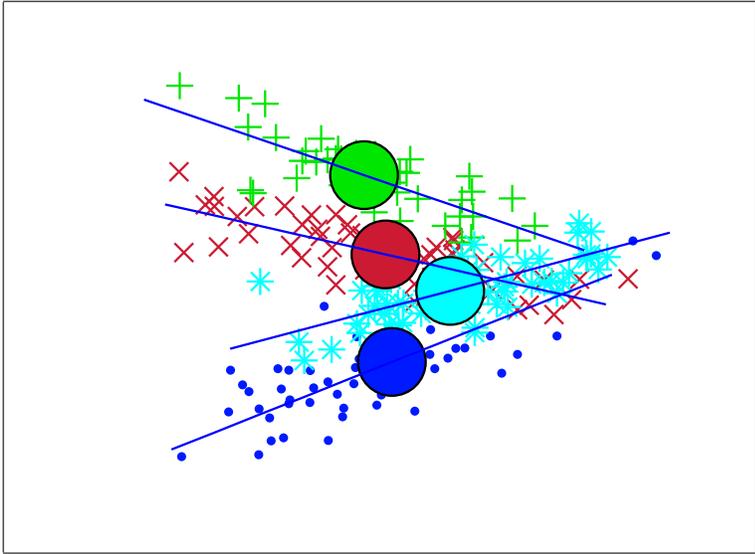


Fig. 14. Itération 10 de HDDC sur les données "Crabes".

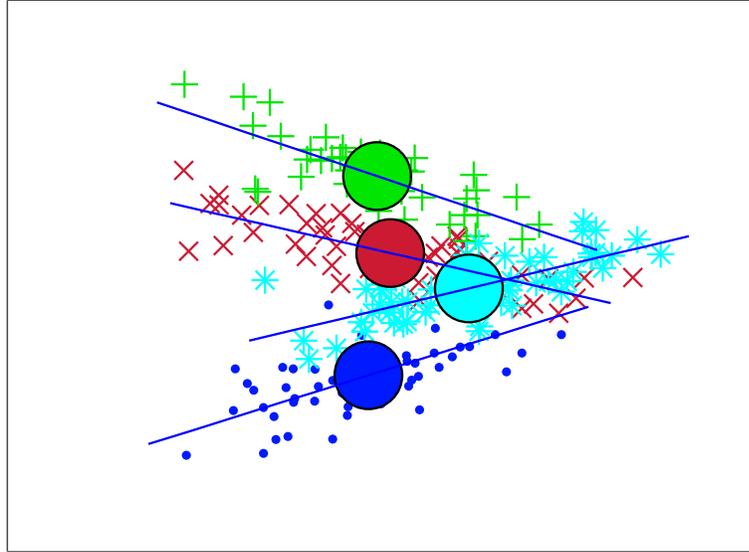


Fig. 14. Itération 12 de HDDC sur les données "Crabs".

2 Régression en grande dimension (SIR)

Régression multivariée

Soient $Y \in \mathbb{R}$ et $X \in \mathbb{R}^p$. Le but est d'estimer $G : \mathbb{R}^p \rightarrow \mathbb{R}$ telle que

$$Y = G(X) + \xi \quad \text{où } \xi \text{ est indépendant de } X.$$

- Irréaliste quand p est grand (*curse of dimensionality*).
- **Réduction de dimension supervisée** : Remplacer X par sa projection sur un sous-espace de dimension inférieure sans perte d'information sur la loi de Y sachant X .
- **Central subspace** : plus petit sous-espace S tel que, conditionnellement à la projection de X sur S , Y et X sont indépendants. Analogue des **espaces intrinsèques** en HDDA.

Réduction de dimension

- Supposons (pour simplifier) que $\dim(S) = 1$ i.e. $S = \text{span}(b)$, avec $b \in \mathbb{R}^p \implies$ **Single index model**:

$$Y = g(b^t X) + \xi$$

où ξ est indépendant de X .

- L'estimation de la fonction p -variée G est remplacée par l'estimation de la fonction univariée g et de la direction b .
- **But de SIR = Sliced Inverse Regression** [Li, 1991] : Estimer une base du *central subspace* (i.e. b dans ce cas particulier.)

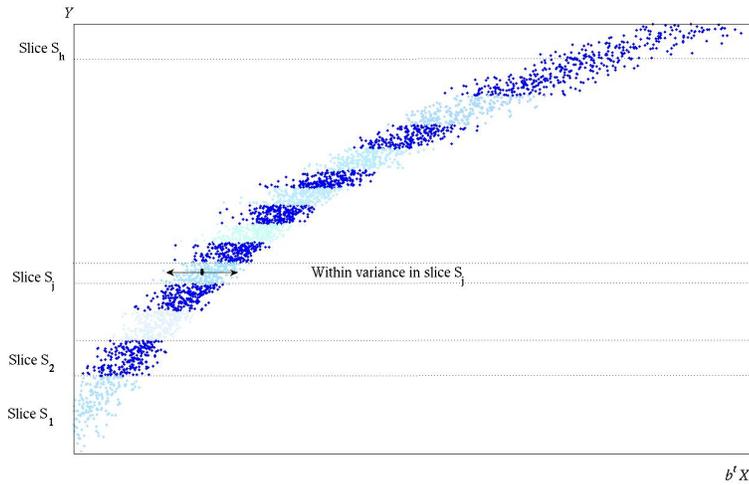
Idée de la méthode SIR

- Trouver la direction b telle que $b^t X$ explique au mieux Y .
- Inversement, quand Y est fixé, $b^t X$ ne doit pas varier.
- Trouver la direction b minimisant les variations de $b^t X$ sachant Y .

En pratique :

- Le support de Y est divisé en h tranches S_j .
- Minimisation de la variance intra-tranches $b^t X$ sous la contrainte $\text{var}(b^t X) = 1$.
- Equivalent à maximiser la variance inter-tranches sous la même contrainte.

Illustration



Procédure d'estimation

Etant donné un échantillon $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$, la direction b est estimée par

$$\hat{b} = \underset{b}{\operatorname{argmax}} b^t \hat{\Gamma} b \text{ sous la contrainte } b^t \hat{\Sigma} b = 1. \quad (1)$$

où $\hat{\Sigma}$ est la matrice de covariance empirique et $\hat{\Gamma}$ est la matrice de covariance inter-tranches définie par

$$\hat{\Gamma} = \sum_{j=1}^h \frac{n_j}{n} (\bar{X}_j - \bar{X})(\bar{X}_j - \bar{X})^t, \quad \bar{X}_j = \frac{1}{n_j} \sum_{Y_i \in S_j} X_i,$$

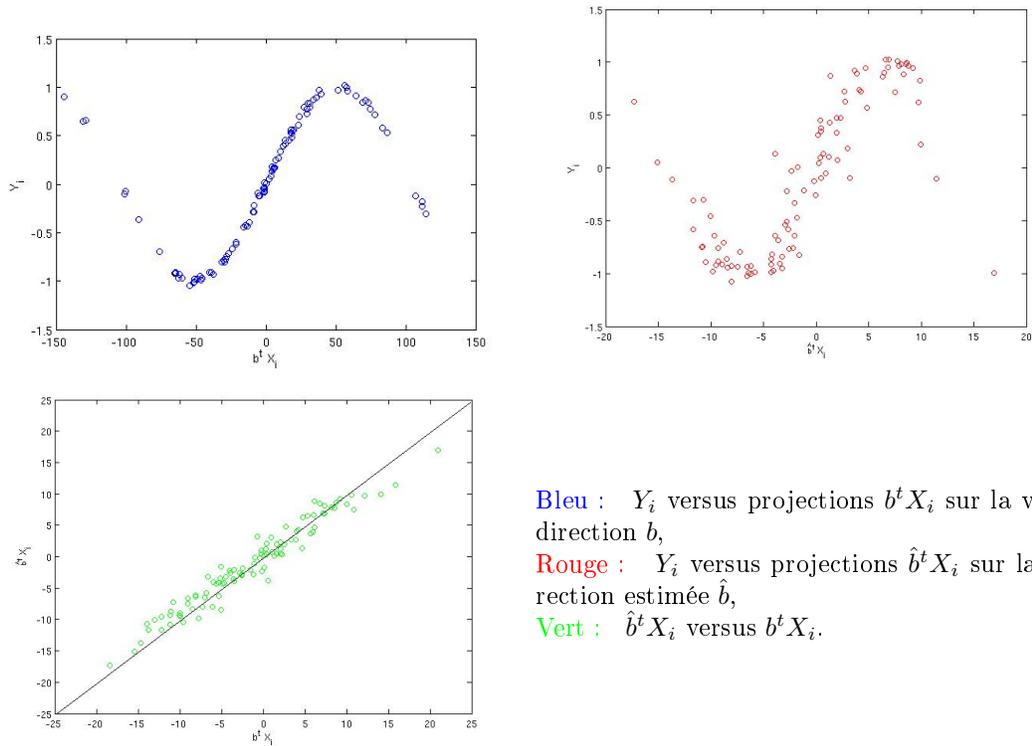
où n_j est le nombre d'observations dans la tranche S_j .

Le problème d'optimisation (1) a une solution explicite : \hat{b} est solution du problème de diagonalisation (généralisé) $\hat{\Gamma} b = \lambda \hat{\Sigma} b$.

Procédure d'estimation

- Pour estimer un *Central subspace* de dimension $d = \dim(S) > 1$, il suffit de considérer les plus d vecteurs propres associés aux d plus grandes valeurs propres.
- SIR est une méthode de réduction de dimension plus efficace que l'ACP en régression car elle est supervisée : elle utilise à la fois X et Y .

Exemple en dimension $p = 10$ avec $n = 100$ données simulées



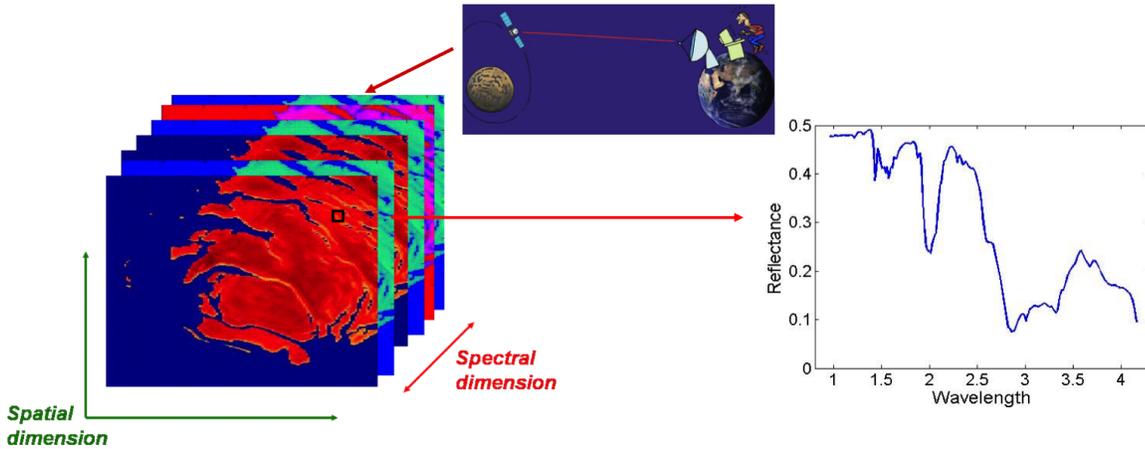
Estimation de propriétés physiques de la planète Mars à partir d'images hyperspectrales

Contexte :

- Observation du pôle sud de Mars par le spectromètre OMEGA à bord de la mission Mars Express.
- Image 3D : En chaque pixel, un spectre de $p = 184$ longueurs d'onde est enregistré.
- Cette partie de Mars est composée principalement de glace (d'eau et de CO_2) et de poussière.

But : Pour chaque spectre $X \in \mathbb{R}^p$, estimer le paramètre physique associé $Y \in \mathbb{R}$ (ici la taille des grains de glace de CO_2).

Images hyperspectrales



Un problème inverse

Problème direct.

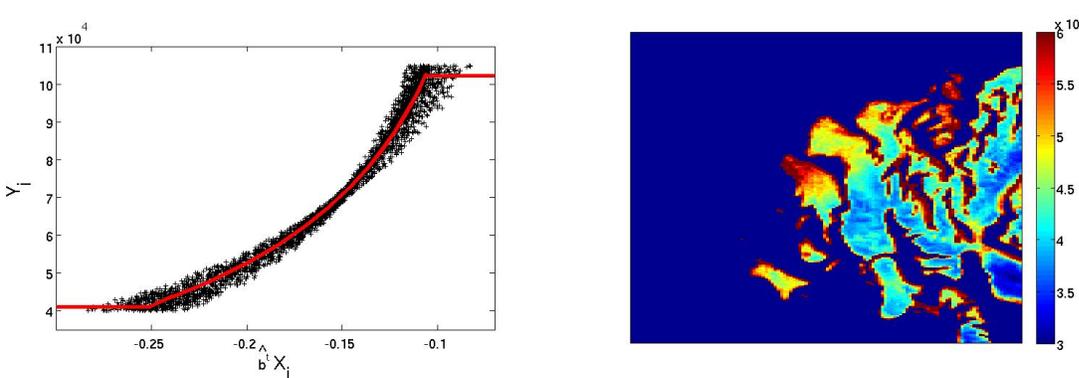
- Modélisation physique d'un spectre à partir d'un modèle de réflectance surfacique.
- A partir d'un paramètre physique Y , simulation de $X = F(Y)$.
- Génération de $n = 12.000$ spectres synthétiques (accompagnés de leurs paramètres physiques associés).

⇒ Base d'apprentissage.

Problème inverse.

- Estimation de la relation fonctionnelle inverse $Y = G(X)$.
- Hypothèse de réduction de dimension $G(X) = g(b^t X)$.
- b est estimé par SIR, g est estimé par régression nonparamétrique unidimensionnelle.

Carte de CO₂ estimée



Taille des grains de glace de CO₂ estimée par SIR à partir d'une image hyperspectrale de la planète Mars.

Extensions de SIR

- Régularisation (pour des X de très grande dimension),
- Adaptation à des sorties Y multivariées,
- Version séquentielle (pour des flux de données),
- Version robuste (aux outliers sur X), ...

References

- [1] R. Cattell. The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2):245–276, 1966.
- [2] K. Li. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 316–327, 1991.
- [3] D. Scott and J. Thompson. Probability density estimation in higher dimensions. In *Fifteenth Symposium in the Interface*, pages 173–179, 1983.
- [4] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.
- [5] C. Bouveyron, S. Girard, and C. Schmid. High dimensional data clustering. *Computational Statistics and Data Analysis*, 52:502–519, 2007.
- [6] C. Bouveyron, S. Girard, and C. Schmid. High dimensional discriminant analysis. *Communication in Statistics - Theory and Methods*, 36(14):2607–2623, 2007.
- [7] J. Jacques, C. Bouveyron, S. Girard, O. Devos, L. Duponchel, and C. Ruckebusch. Gaussian mixture models for the classification of high-dimensional vibrational spectroscopy data. *Journal of Chemometrics*, 24:719–727, 2010.
- [8] C. Bouveyron, G. Celeux, and S. Girard. Intrinsic dimension estimation by maximum likelihood in isotropic probabilistic PCA. *Pattern Recognition Letters*, 32(14):1706–1713, 2011.
- [9] C. Bouveyron, M. Fauvel, and S. Girard. Kernel discriminant analysis and clustering with parsimonious Gaussian process models. *Statistics and Computing*, 25:1143–1162, 2015.
- [10] C. Bernard-Michel, L. Gardes, and S. Girard. Gaussian regularized sliced inverse regression. *Statistics and Computing*, 19:85–98, 2009.
- [11] C. Bernard-Michel, L. Gardes, and S. Girard. A note on sliced inverse regression with regularizations. *Biometrics*, 64:982–986, 2008.
- [12] C. Bernard-Michel, S. Douté, M. Fauvel, L. Gardes, and S. Girard. Retrieval of Mars surface physical properties from Omega hyperspectral images using regularized sliced inverse regression. *Journal of Geophysical Research - Planets*, 114, 2009.
- [13] M. Chavent, S. Girard, V. Kuentz, B. Liquet, T.M.N. Nguyen, and J. Saracco. A sliced inverse regression approach for data stream. *Computational Statistics*, 29:1129–1152, 2014.
- [14] R. Coudret, S. Girard, and J. Saracco. A new sliced inverse regression method for multivariate response. *Computational Statistics and Data Analysis*, 77:285–299, 2014.
- [15] A. Chiancone, F. Forbes, and S. Girard. Student sliced inverse regression. *Computational Statistics and Data Analysis*, 113:441–456, 2017.