Estimation de densités conditionnelles par sélection de modèles et application à la segmentation d'images hyperspectrales

E. Le Pennec
(SELECT - INRIA Saclay / Université Paris Sud)
et
S. Cohen (IPANEMA - Soleil)

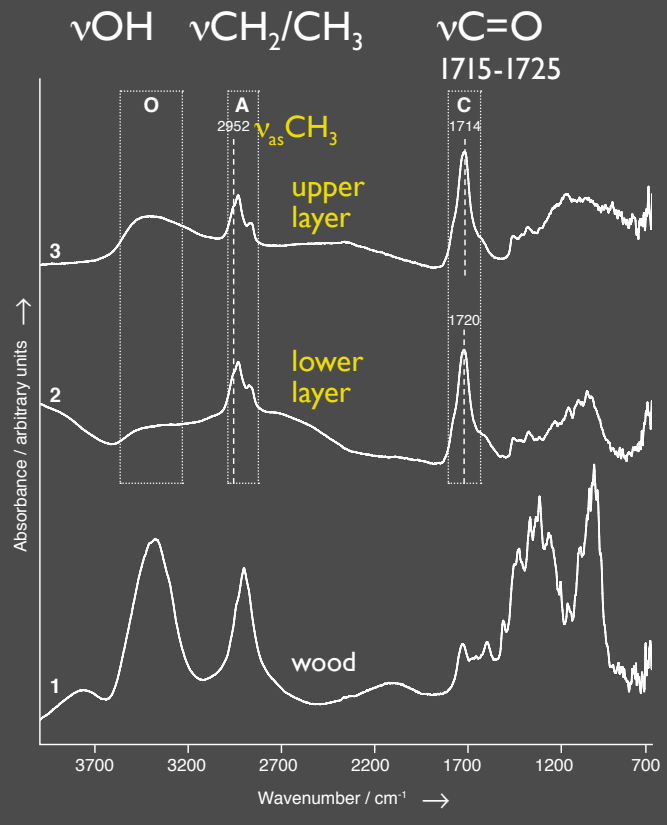
IECN
07 janvier 2011

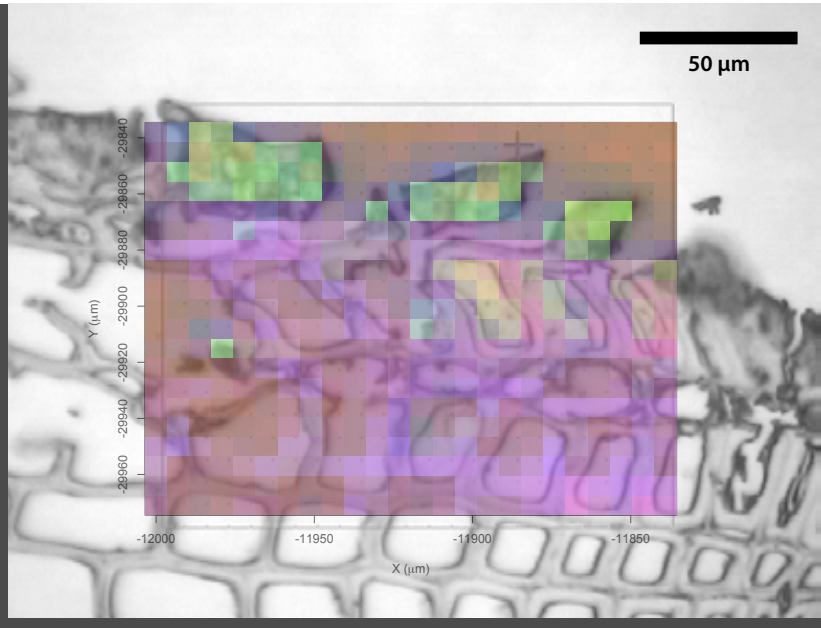
INRIA

A. Stradivari (1644 - 1737)

Provigny (1716)

A. Giordan © Cité de la Musique





SOLEIL SYNCHROTRON

4 / 8 cm^{-1} resolution
64 / 128 scans
typ. 1 min/sp, 400sp

very simple process
no protein (amide I, amide II)
no gums, nor waxes
@SOLEIL: SMIS

cité de la musique

SOLEIL SYNCHROTRON

CRCC

CENTRE DE RECHERCHE ET DE RESTAURATION DES MUSÉES DE FRANCE

Laboratoire de Dynamique, Interactions et Réactivité

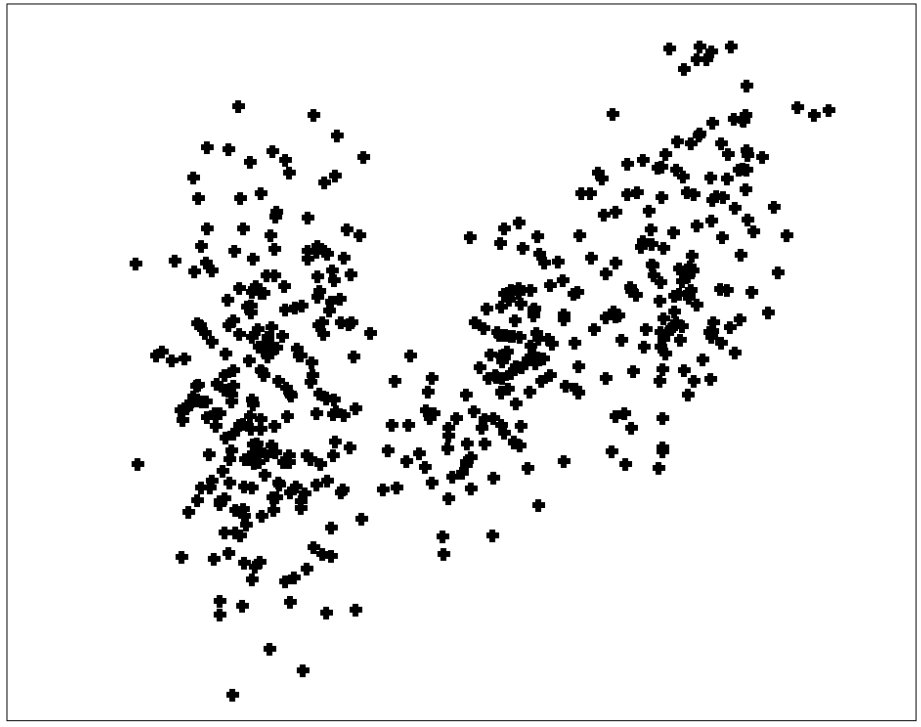
J.-P. Echard, L. Bertrand, A. von Bohlen, A.-S. Le Hô, C. Paris, L. Bellot-Gurlet, B. Soulier, A. Lattuati-Derieux, S. Thao, L. Robinet, B. Lavédrine, and S. Vaiedelich. *Angew. Chem. Int. Ed.*, 49(1), 197-201, 2010.

Segmentation d'images hyperspectrales

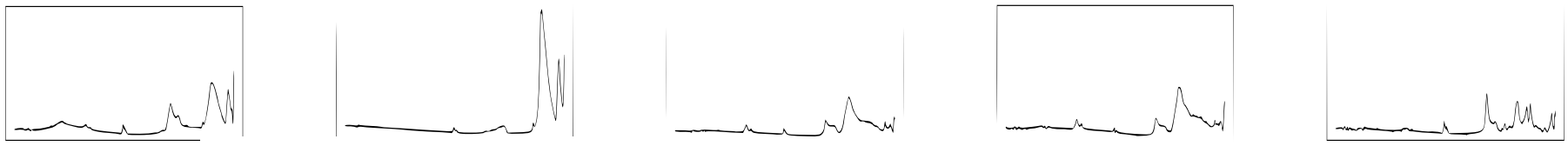
- Données :
  - image de taille n comprise entre ~ 1000 et ~ 100000 pixels,
  - spectres $\mathcal{S}$ de ~ 1024 points,
  - résolution $\sim 4/8$ cm^{-1} (10 fois meilleure dans le visible),
  - possibilité de mesurer de très nombreux spectres par minute...
- Objectifs immédiats :
  - segmentation automatique de ces images,
  - sans intervention humaine,
  - aide à l'analyse des résultats.
- Objectifs lointains :
  - classification automatique,
  - interprétation...

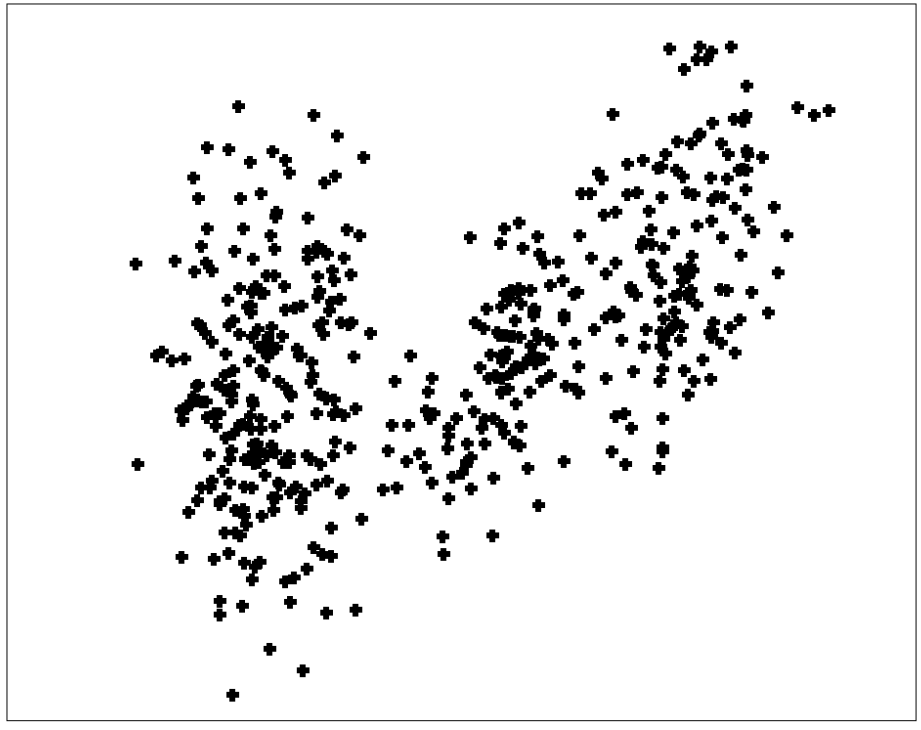
Un problème “jouet”

Un problème “jouet”

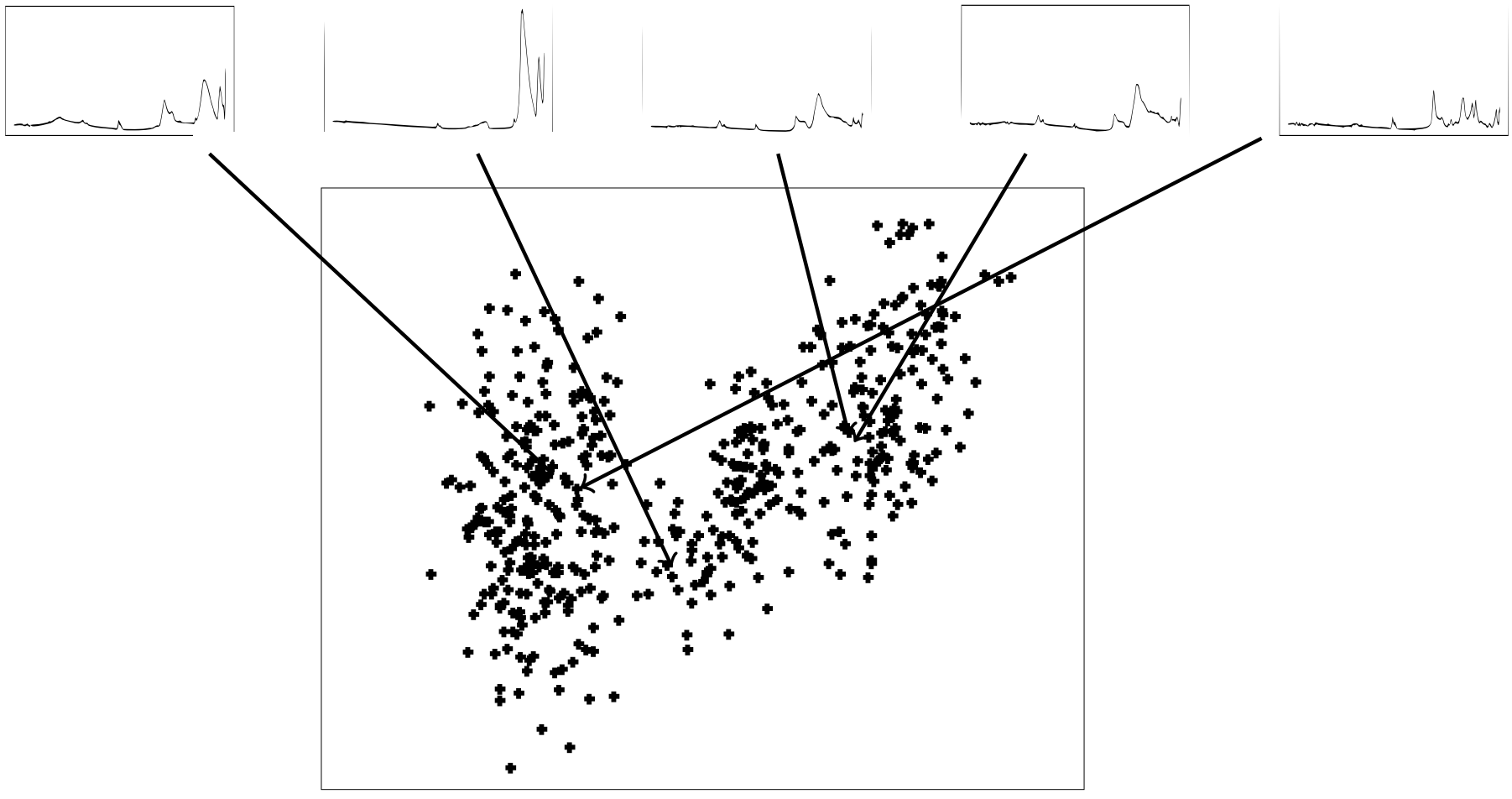


Un problème “jouet”

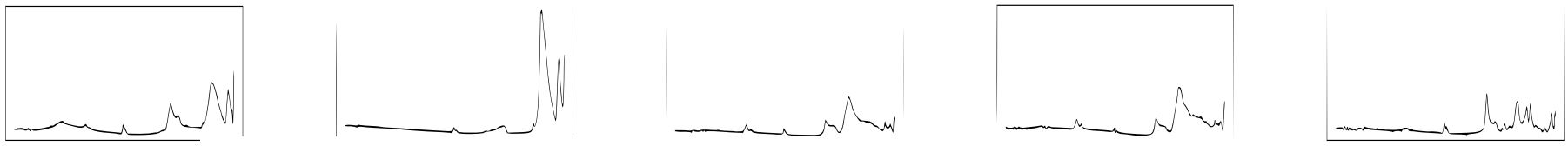


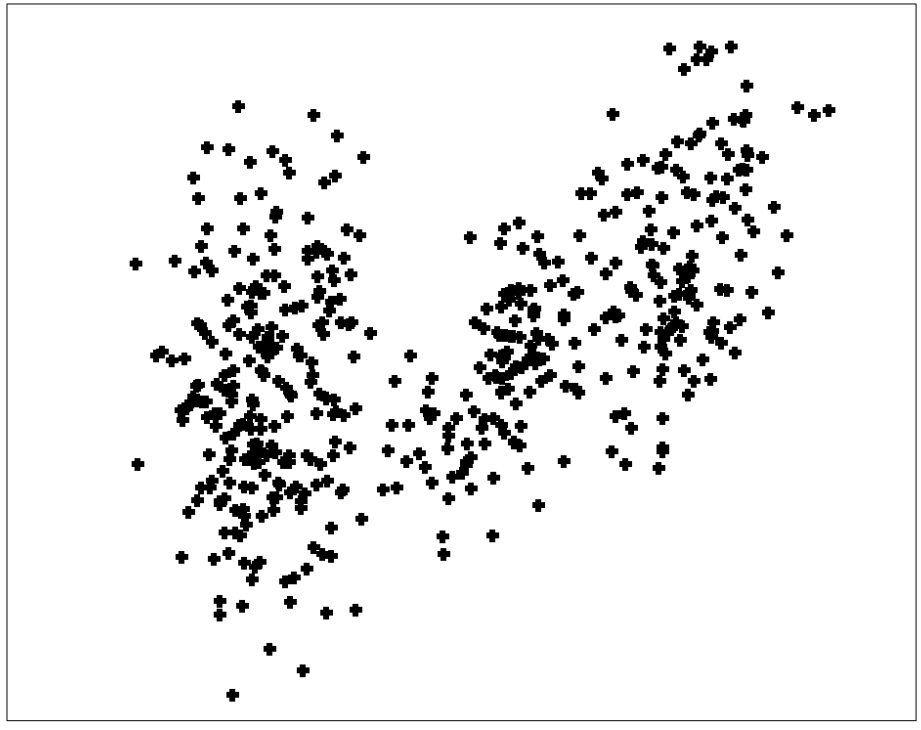


Un problème “jouet”

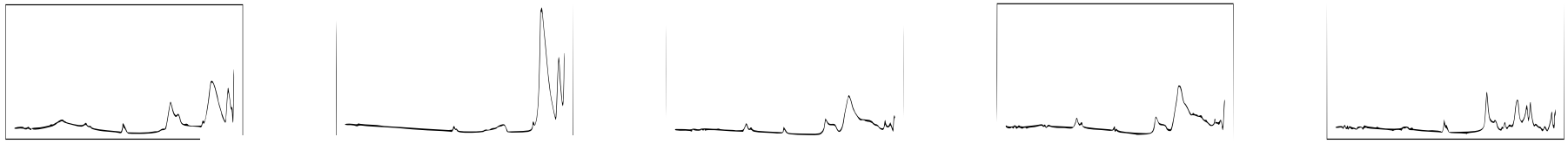


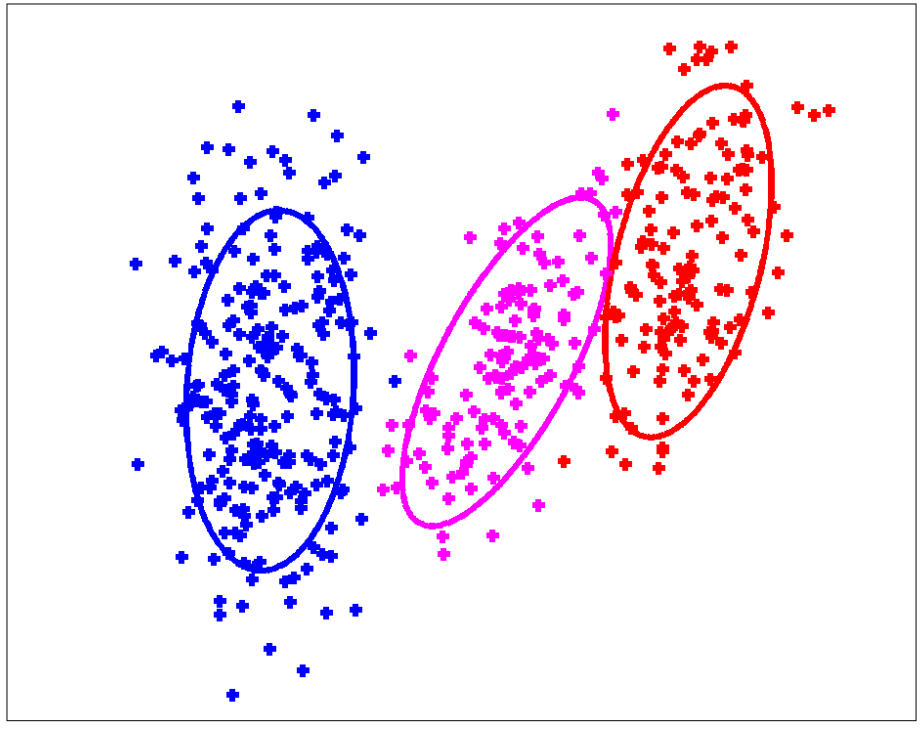
Un problème “jouet”



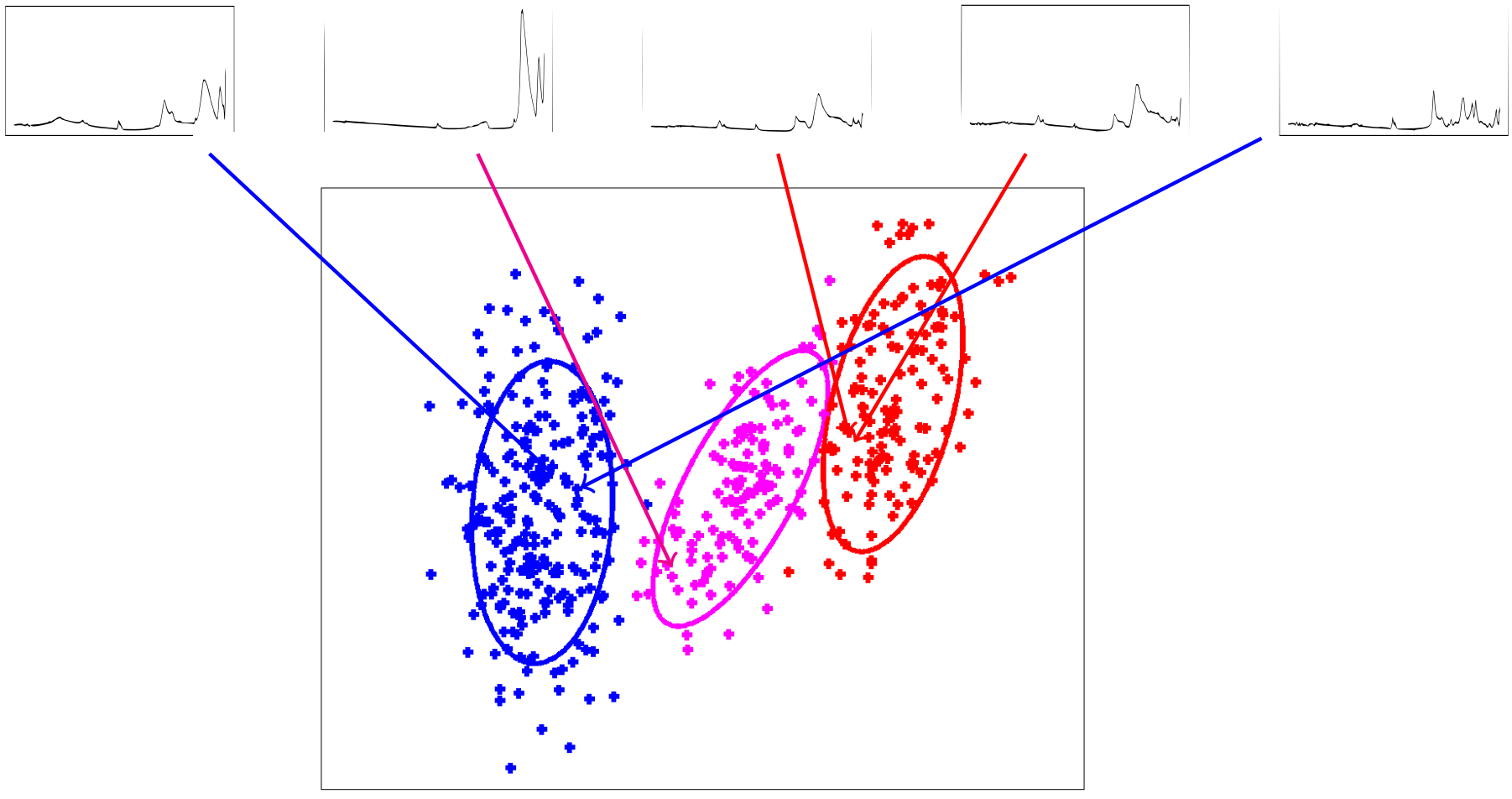


Un problème “jouet”

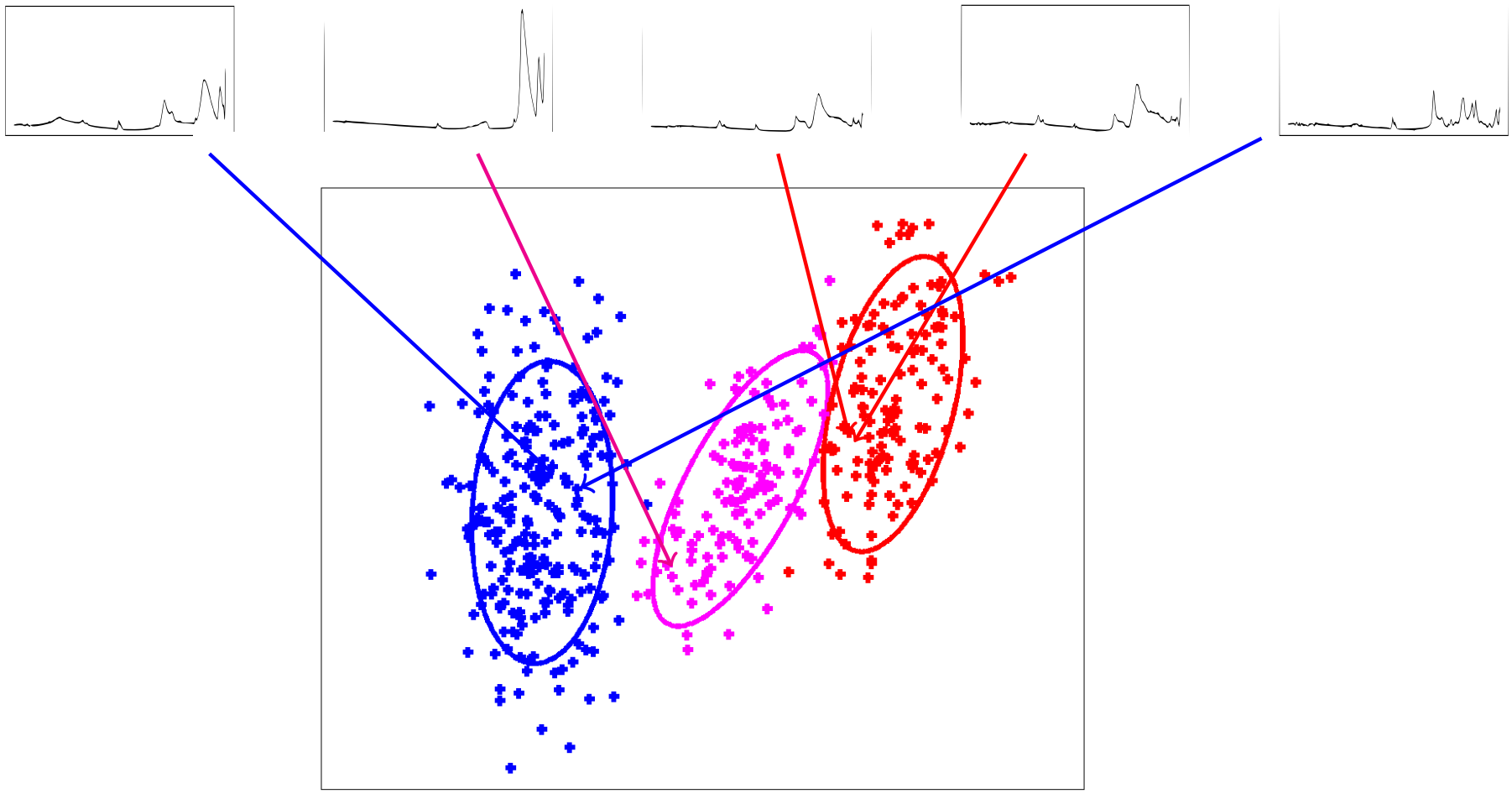




Un problème “jouet”



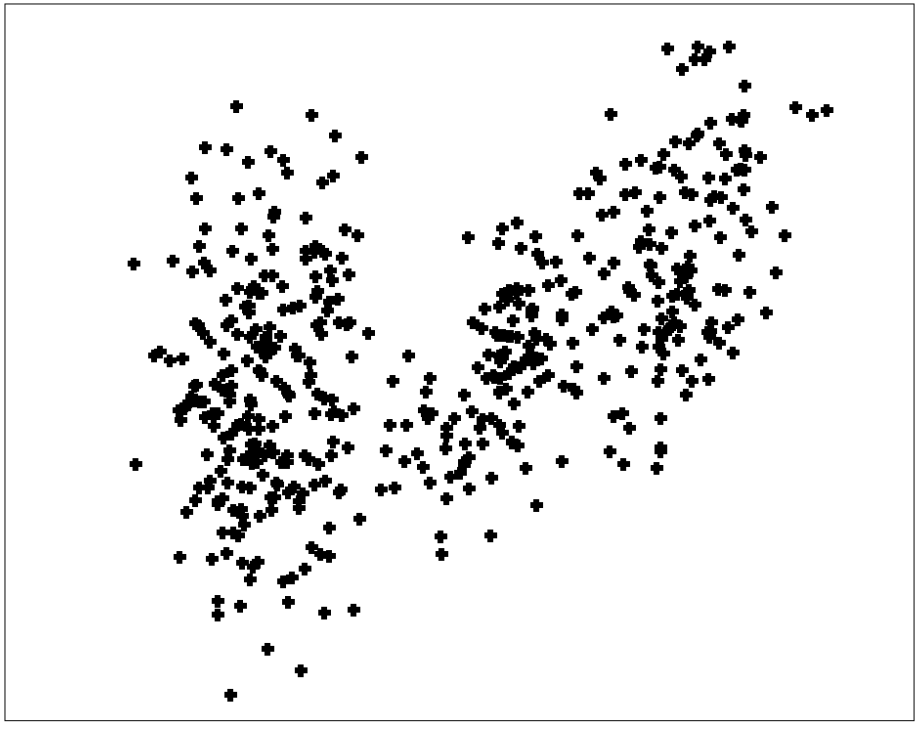
Un problème “jouet”



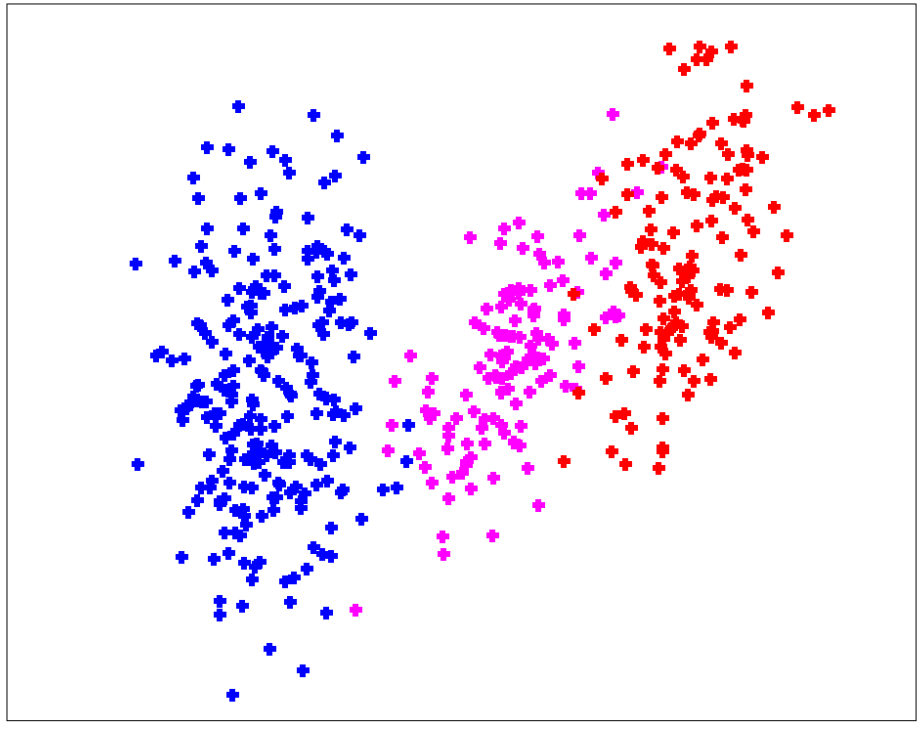
- Représentation : utilisation d’une correspondance entre les objets et des points dans un espace de grande dimension.

Modélisation “stochastique”

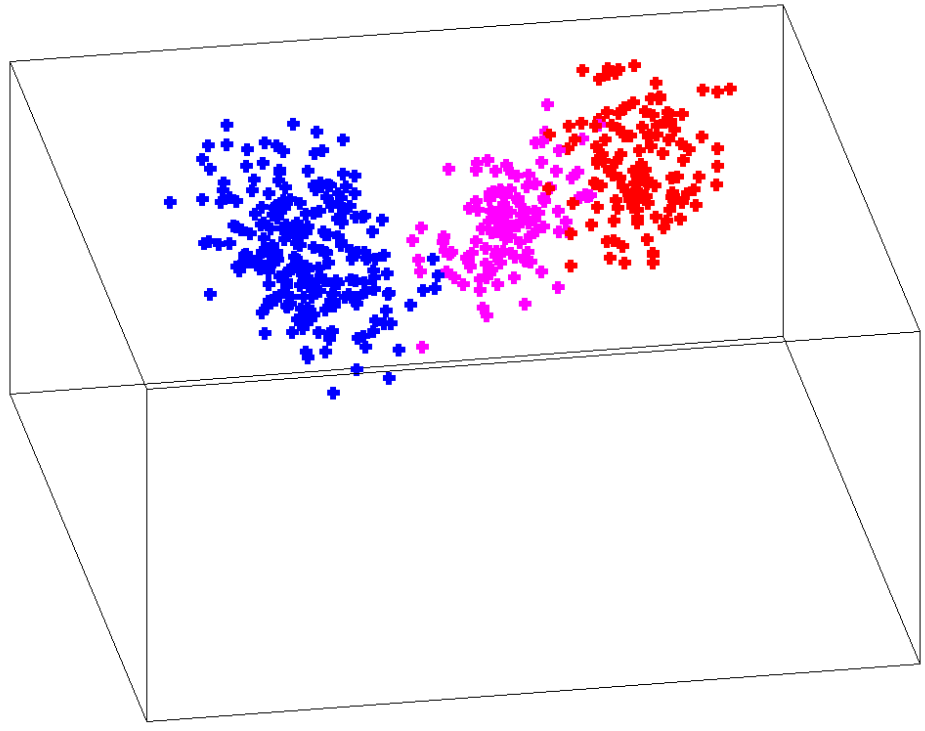
Modélisation “stochastique”



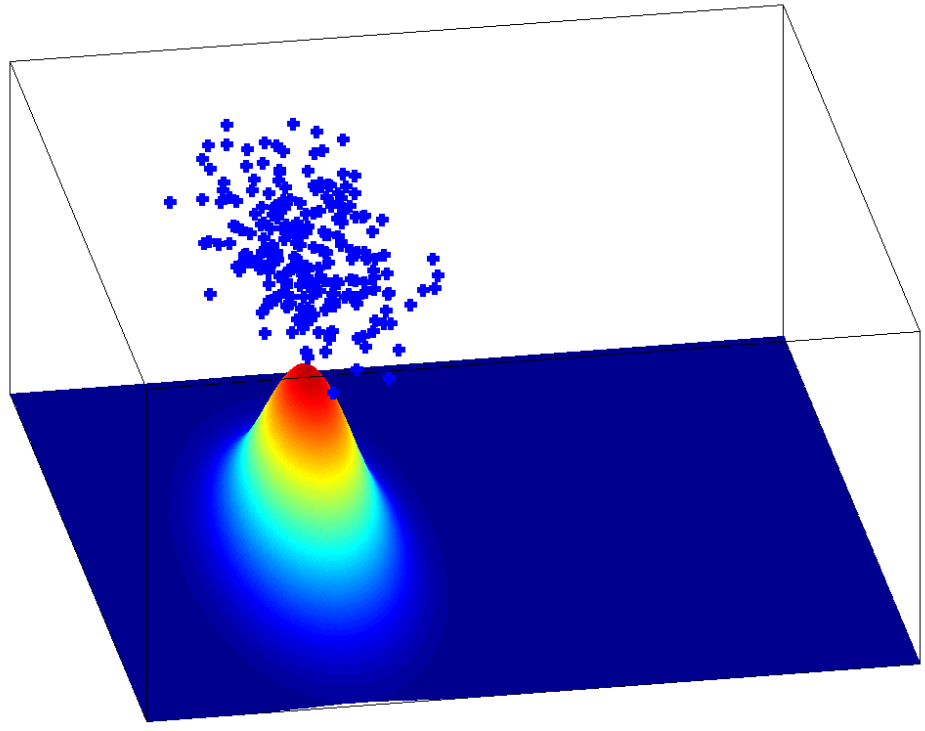
Modélisation “stochastique”



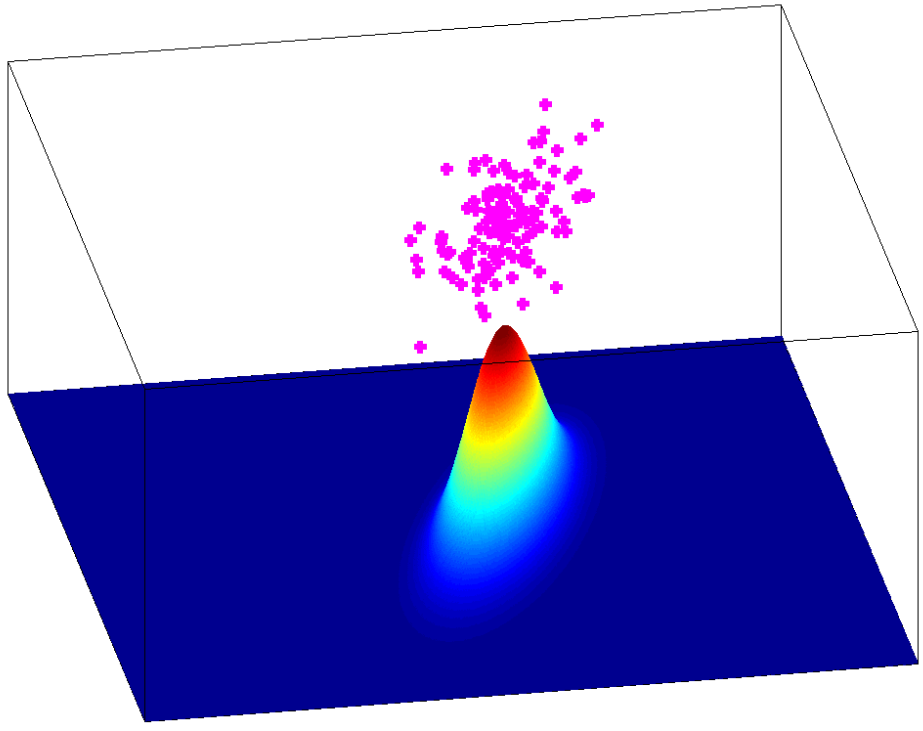
Modélisation “stochastique”



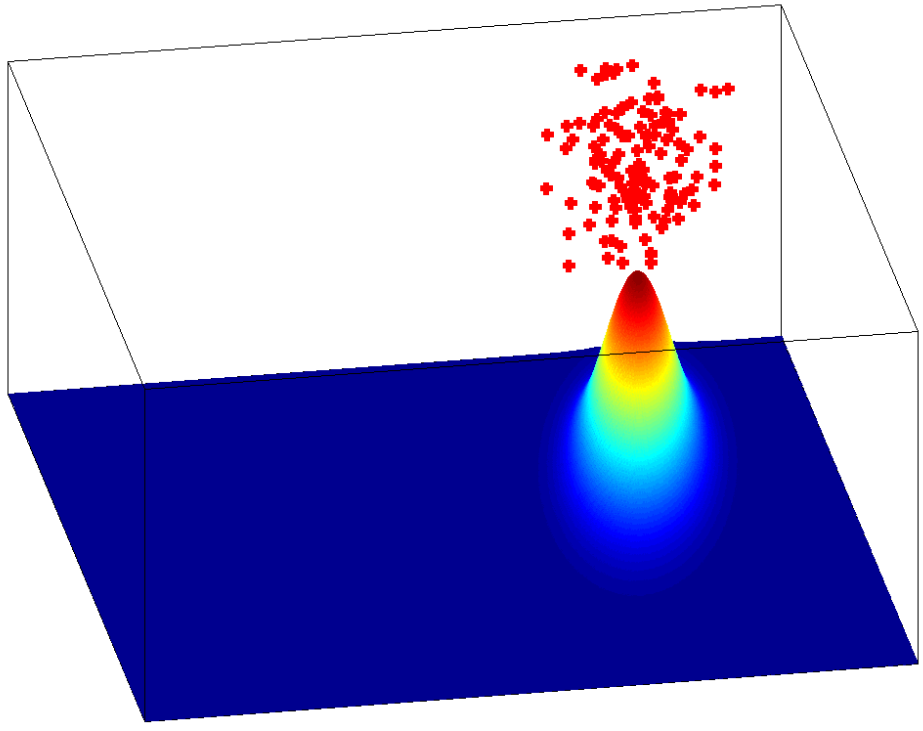
Modélisation “stochastique”



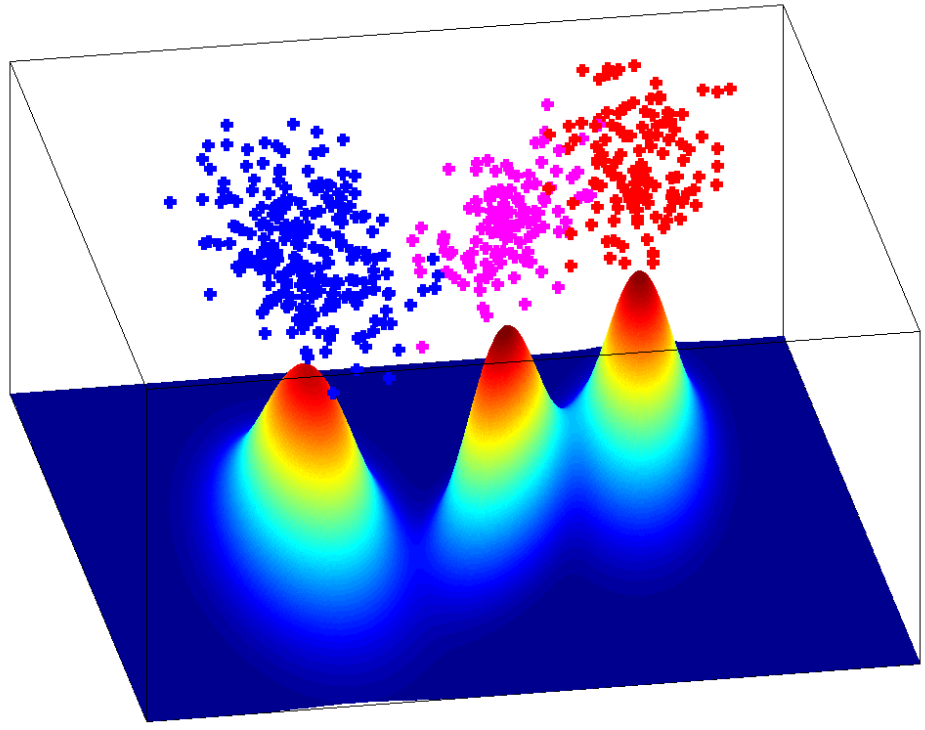
Modélisation “stochastique”



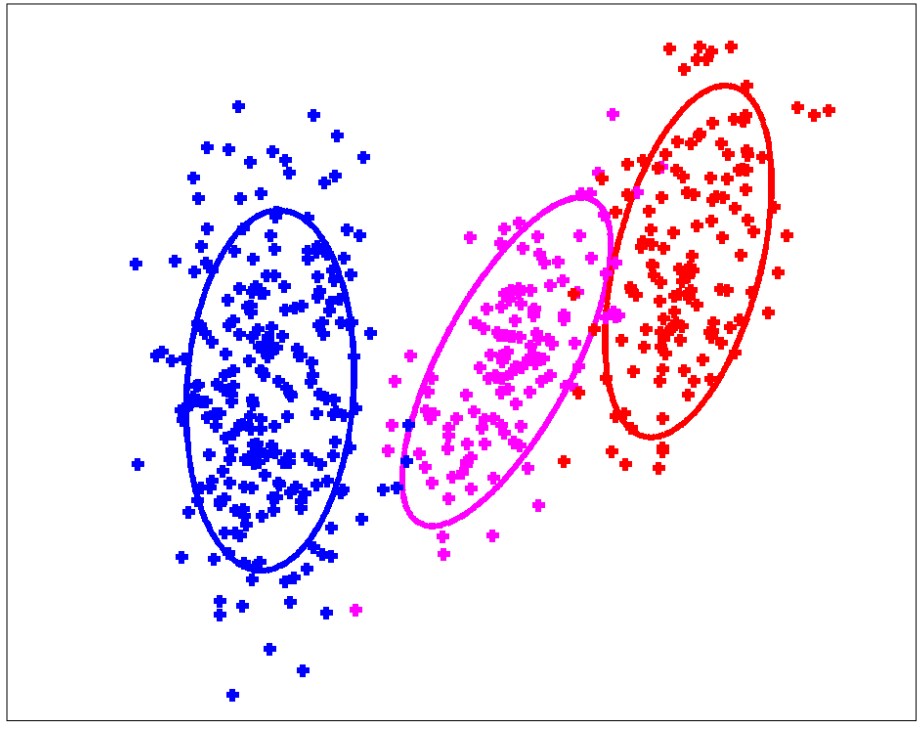
Modélisation “stochastique”



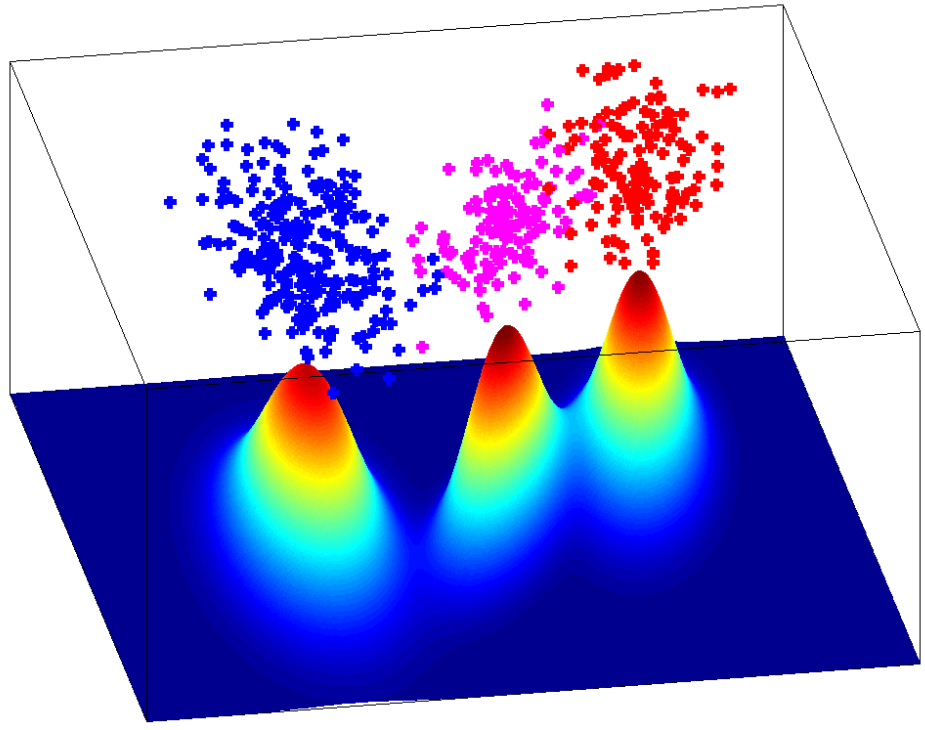
Modélisation “stochastique”



Modélisation “stochastique”



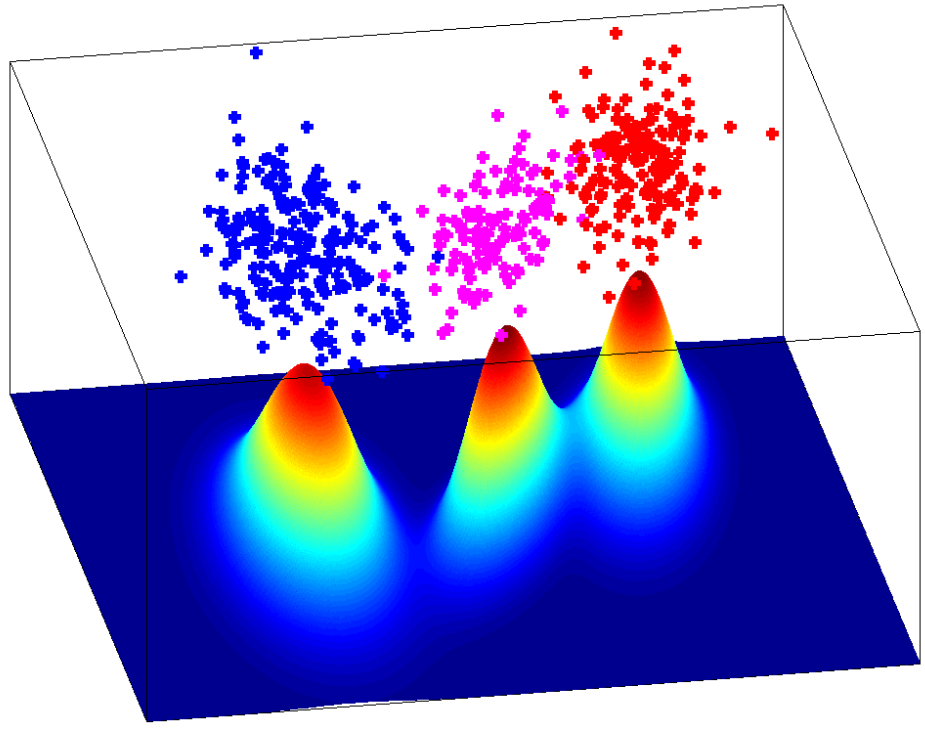
Modélisation “stochastique”



- Modèle : mélange de gaussiennes à K classes.
- Densité du mélange :

$$s_{K,(\pi_k,\mu_k,\Sigma_k)}(x) = \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} e^{-\frac{1}{2}(x-\mu_k)^t \Sigma_k^{-1}(x-\mu_k)} dx$$

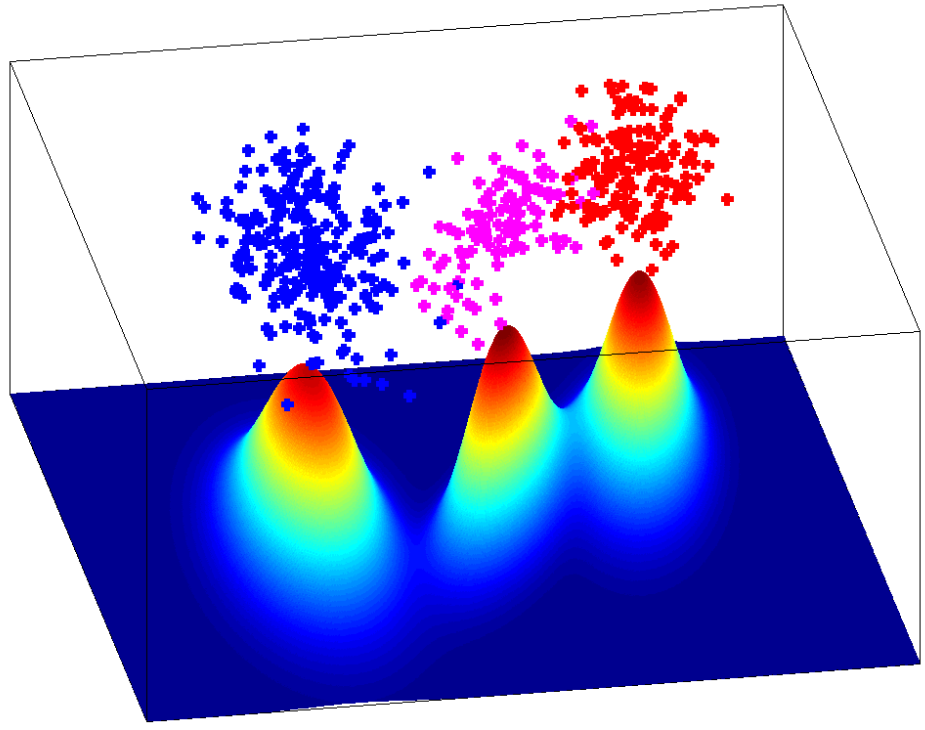
Modélisation “stochastique”



- Modèle : mélange de gaussiennes à K classes.
- Densité du mélange :

$$s_{K,(\pi_k,\mu_k,\Sigma_k)}(x) = \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} e^{-\frac{1}{2}(x-\mu_k)^t \Sigma_k^{-1}(x-\mu_k)} dx$$

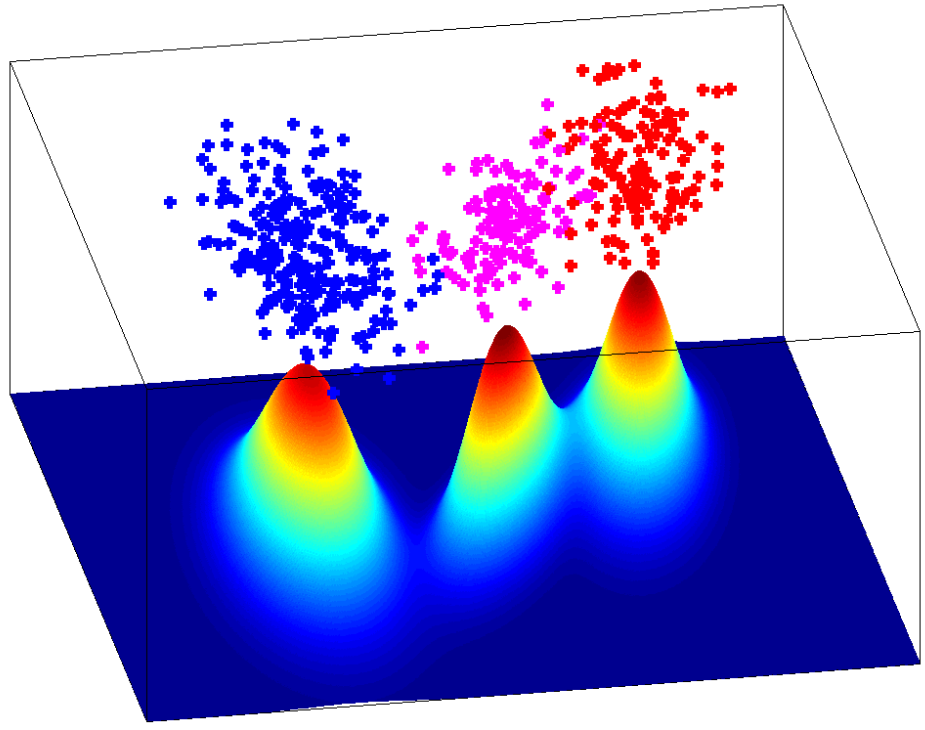
Modélisation “stochastique”



- Modèle : mélange de gaussiennes à K classes.
- Densité du mélange :

$$s_{K,(\pi_k,\mu_k,\Sigma_k)}(x) = \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} e^{-\frac{1}{2}(x-\mu_k)^t \Sigma_k^{-1}(x-\mu_k)} dx$$

Modélisation “stochastique”

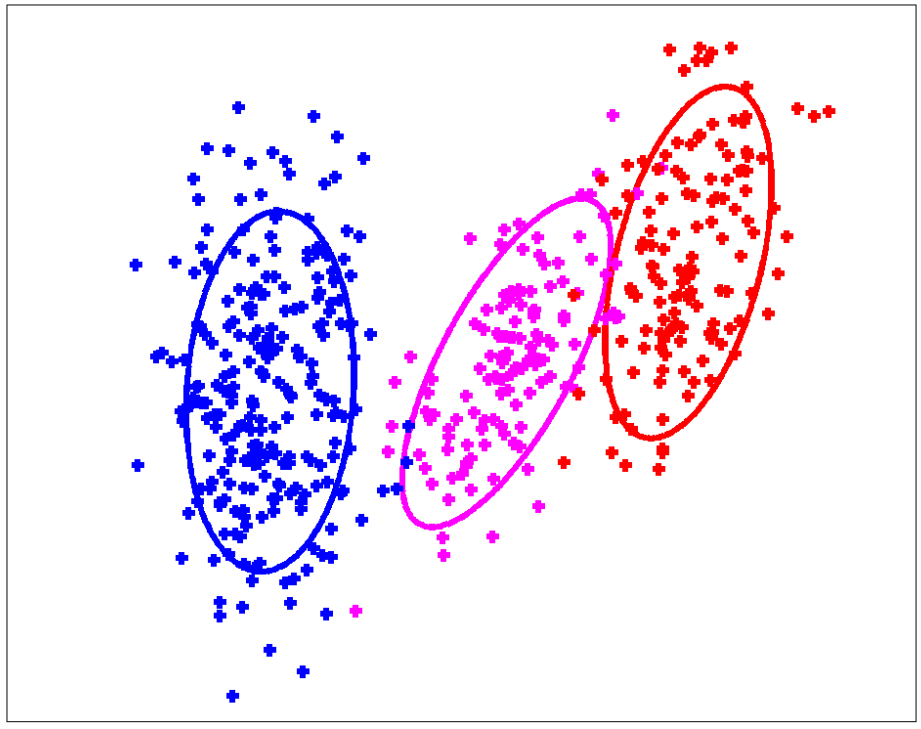


- Modèle : mélange de gaussiennes à K classes.
- Densité du mélange :

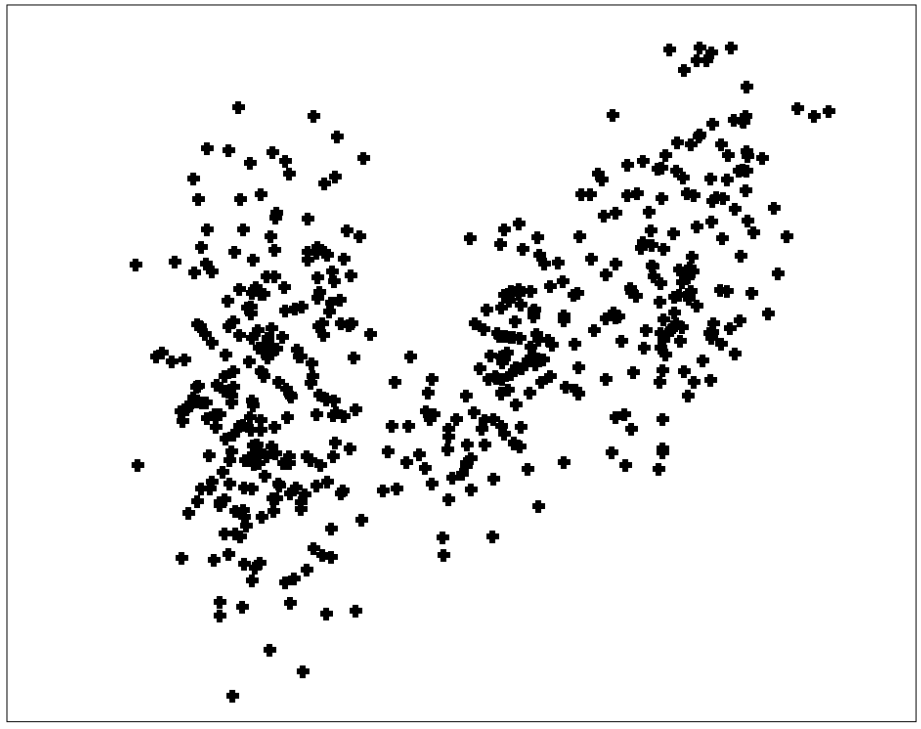
$$s_{K,(\pi_k,\mu_k,\Sigma_k)}(x) = \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} e^{-\frac{1}{2}(x-\mu_k)^t \Sigma_k^{-1}(x-\mu_k)} dx$$

Estimation “statistique”

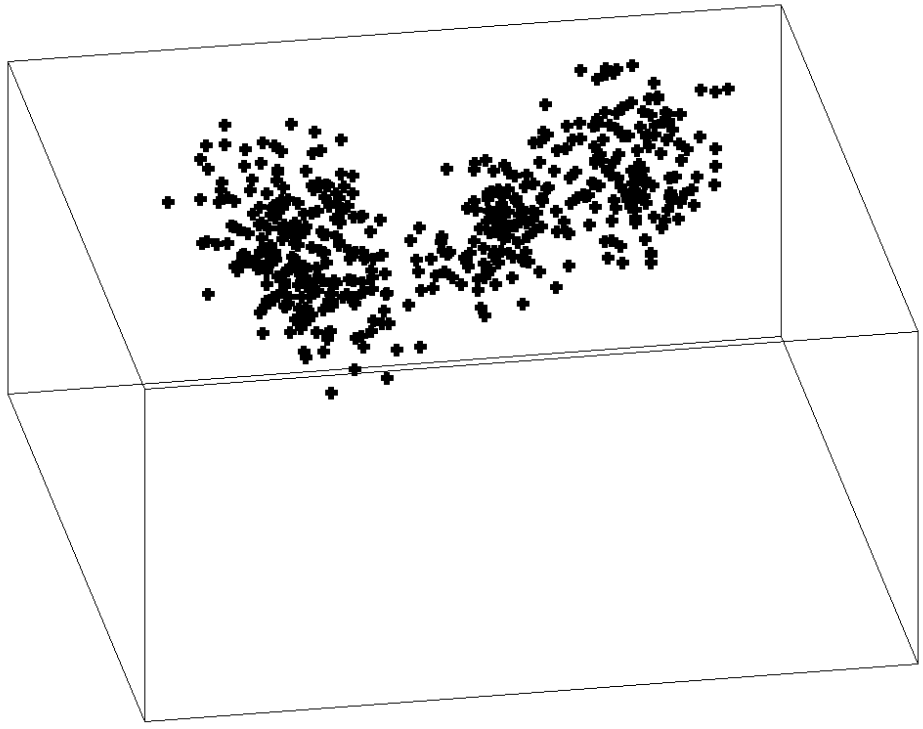
Estimation “statistique”



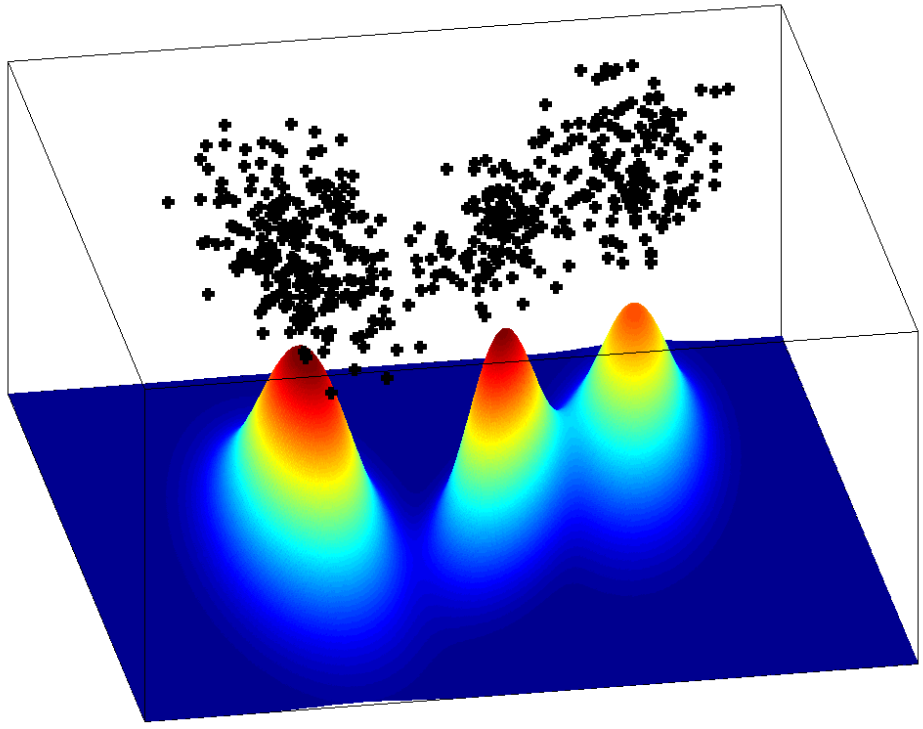
Estimation “statistique”



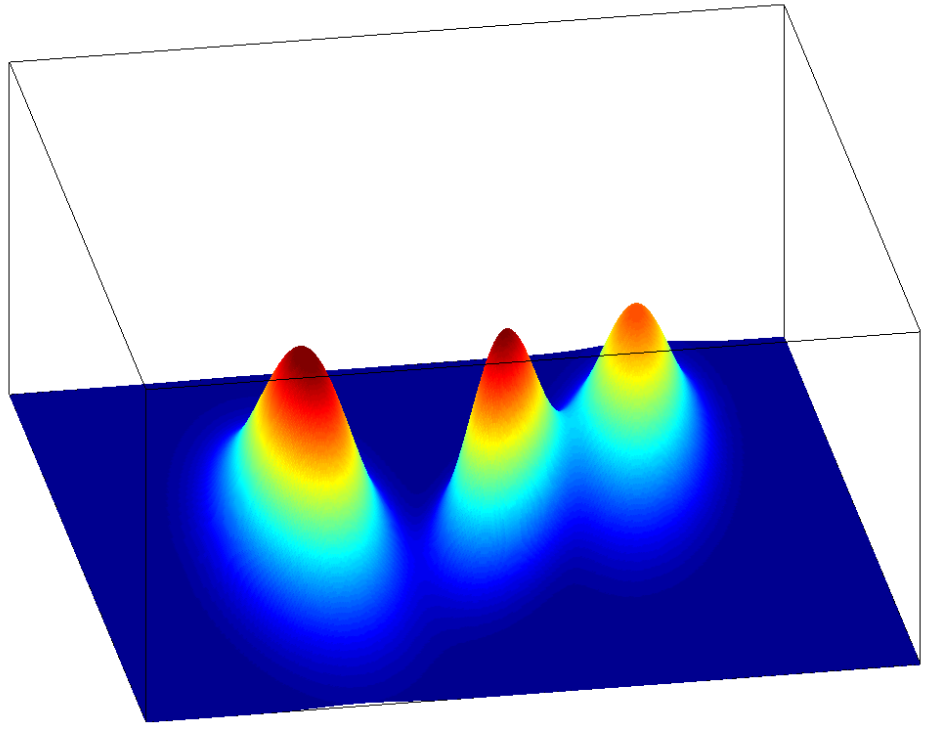
Estimation “statistique”



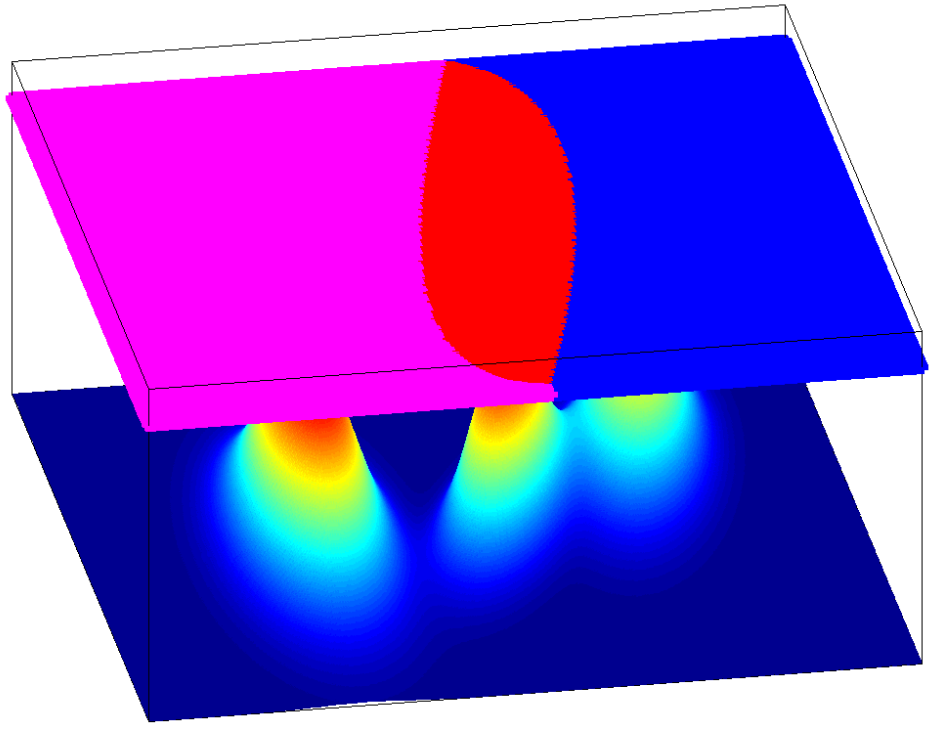
Estimation “statistique”



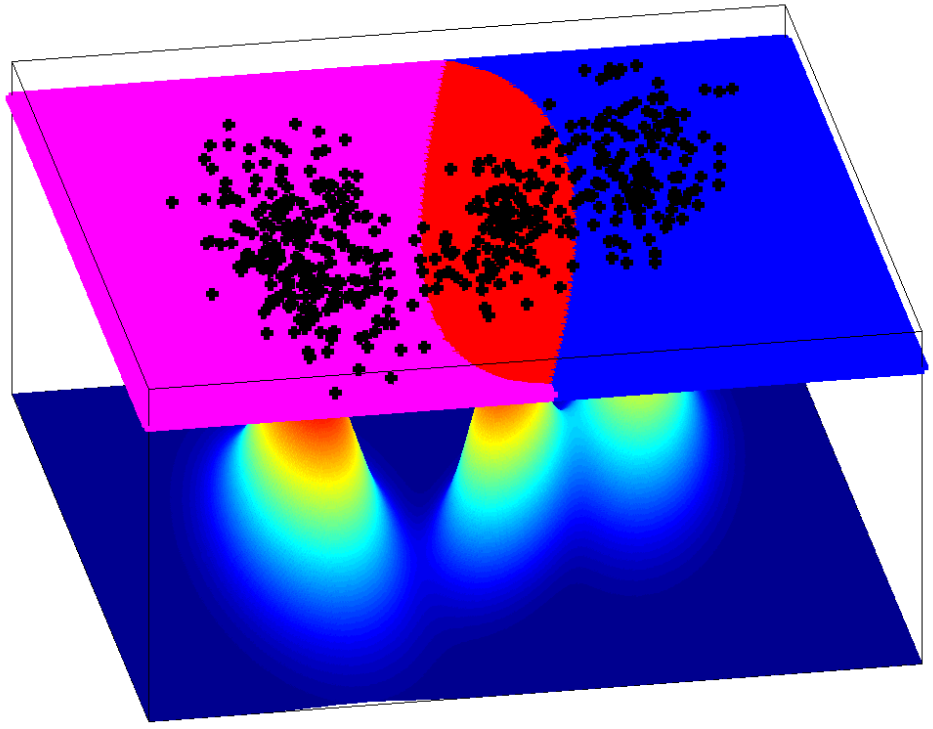
Estimation “statistique”



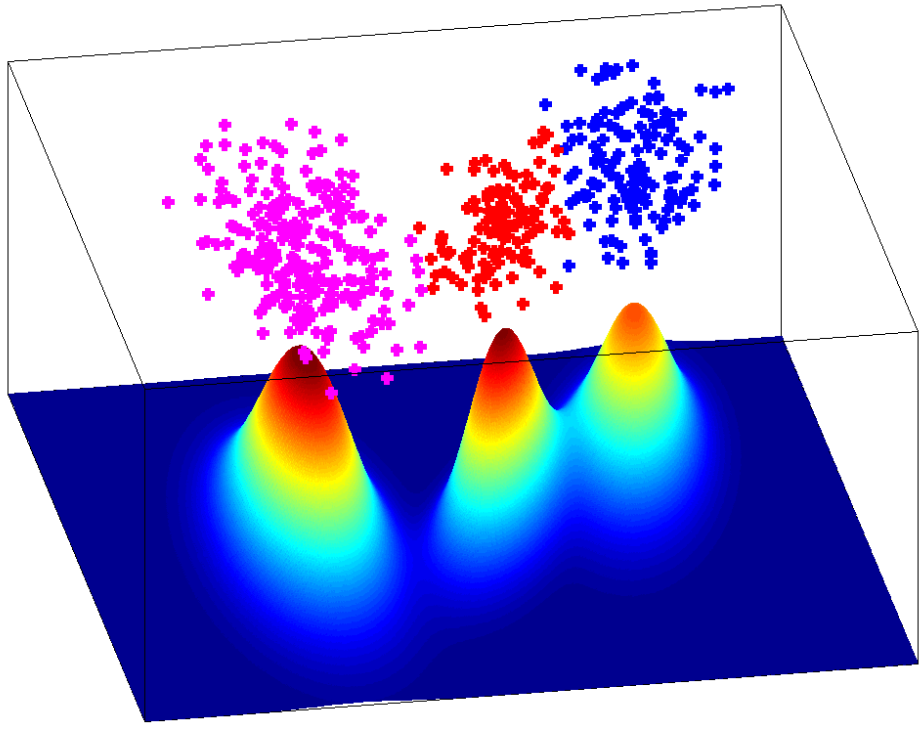
Estimation “statistique”



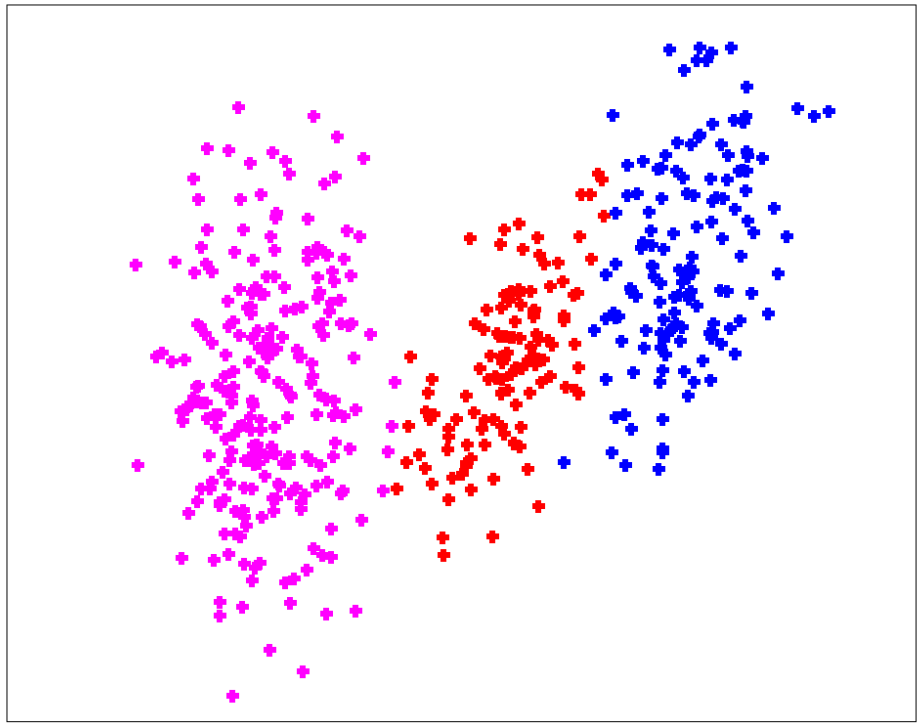
Estimation “statistique”



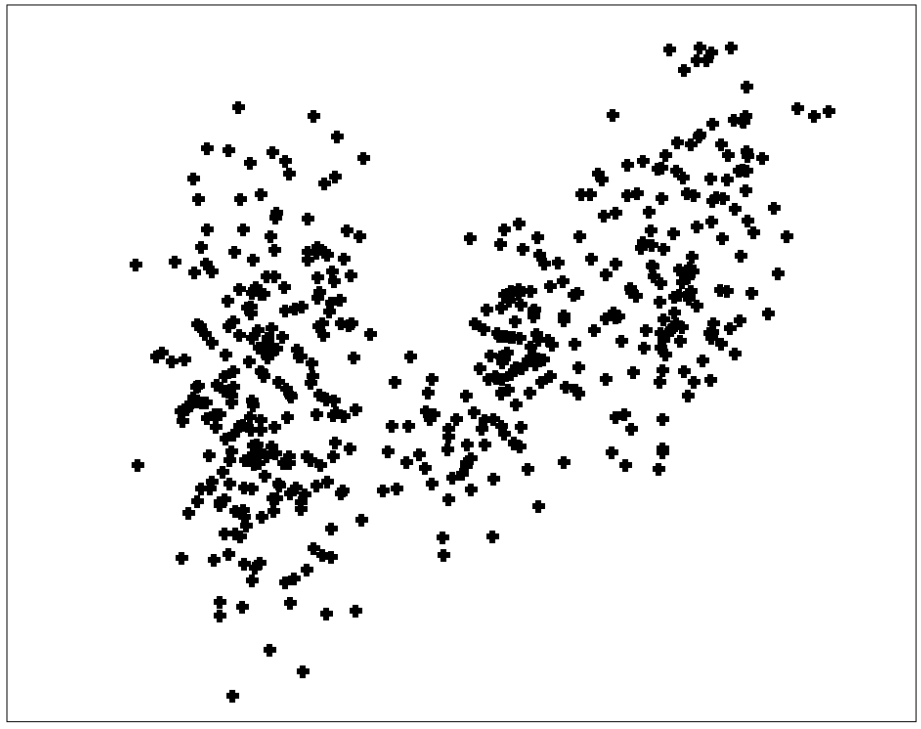
Estimation “statistique”



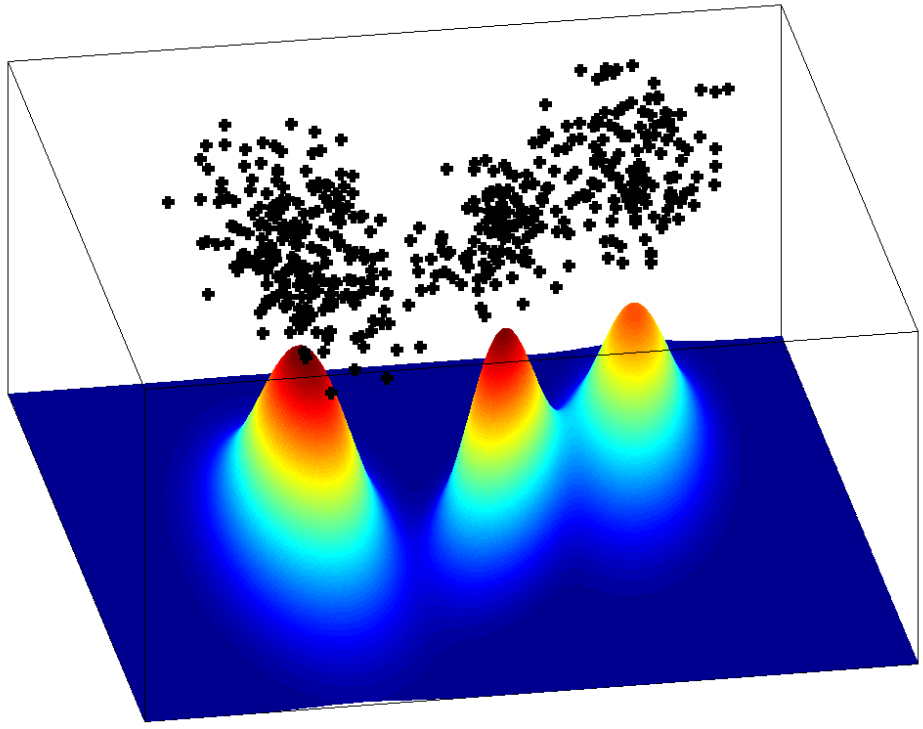
Estimation “statistique”



Estimation “statistique”



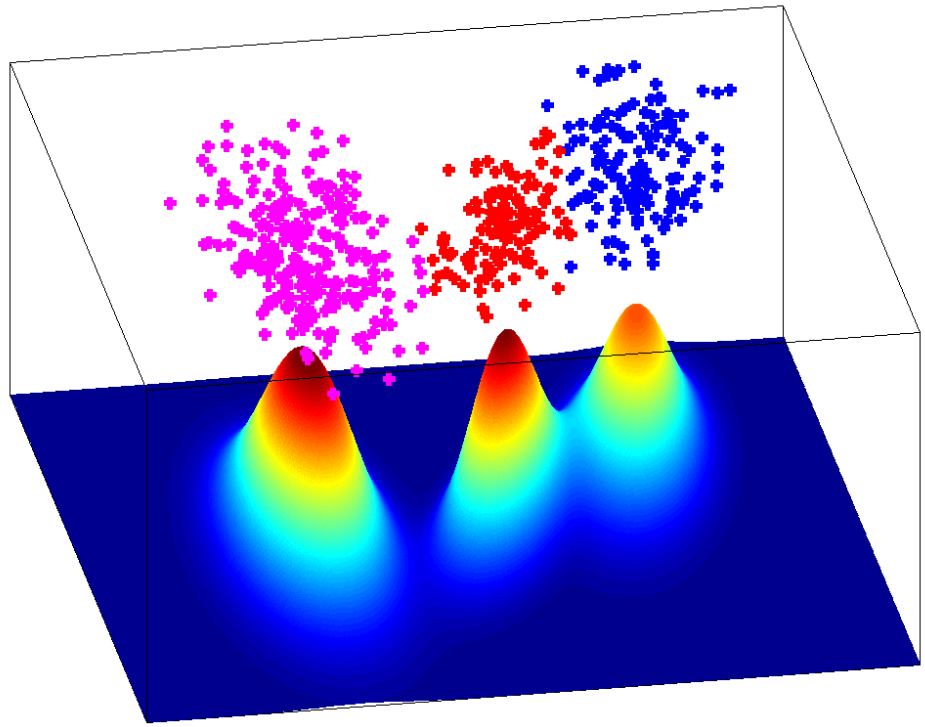
Estimation "statistique"



- Estimation des π_k, $\widehat{\mu_k}$ et $\widehat{\Sigma_k}$ par maximum de vraisemblance :

$$(\widehat{\pi_k}, \widehat{\mu_k}, \widehat{\Sigma_k}) = \operatorname{argmax} \sum_{i=1}^{n} \log s_{K,(\pi_k,\mu_k,\Sigma_k)}(X_i)$$

Estimation "statistique"



- Estimation des π_k, $\widehat{\mu_k}$ et $\widehat{\Sigma_k}$ par maximum de vraisemblance :

$$(\widehat{\pi_k}, \widehat{\mu_k}, \widehat{\Sigma_k}) = \operatorname{argmax} \sum_{i=1}^{n} \log s_{K,(\pi_k,\mu_k,\Sigma_k)}(X_i)$$

- Estimation de $\widehat{k}(x)$ par maximum à posteriori :

$$\widehat{k}(x) = \operatorname{argmax} \widehat{\pi_k} \frac{1}{\sqrt{(2\pi)^d |\widehat{\Sigma_k}|}} e^{-\frac{1}{2}(x-\widehat{\mu_k})^t \widehat{\Sigma_k}^{-1}(x-\widehat{\mu_k})}$$

Modélisation par un mélange de gaussiennes

- Modélisation stochastique des spectres $\mathcal{S}$:
  - existence de K classes de spectres,
  - proportion π_k pour chacune des classes ($\sum_{k=1}^{K} \pi_k = 1$),
  - loi gaussienne $\mathcal{N}(\mu_k, \Sigma_k)$ sur chacune des classes (restriction forte !)
- Densité s_0 de $\mathcal{S}$ proche de

$$s(\mathcal{S}) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(\mu_k, \Sigma_k)(\mathcal{S}).$$

- Objectif : estimer les paramètres K, π_k, μ_k, Σ_k à partir des données.
- Pourquoi ? : possibilité d'assigner ensuite une classe à une observation par maximum de vraisemblance

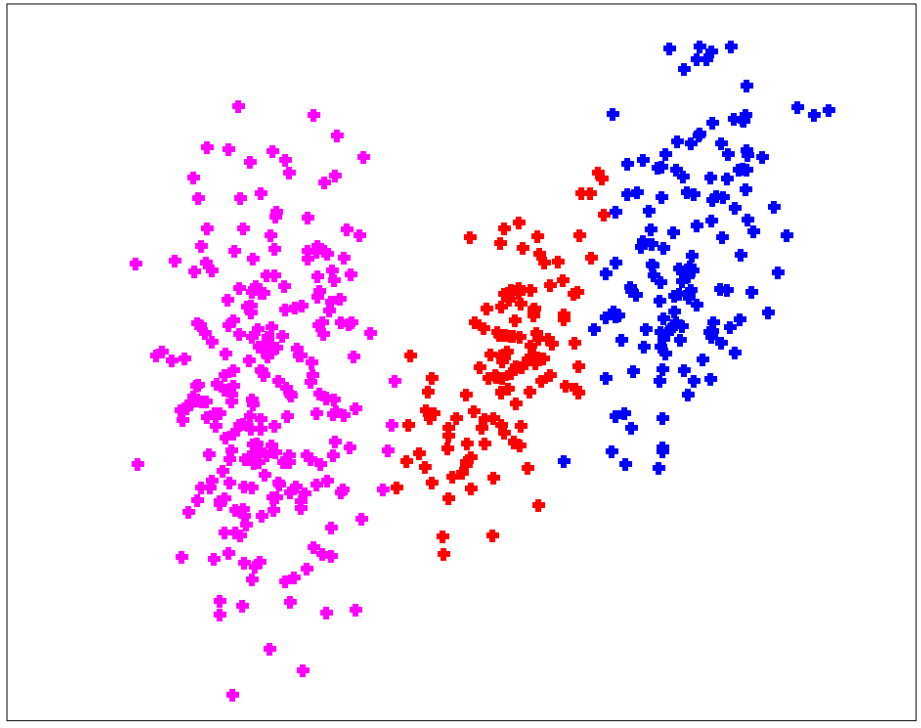
$$\widehat{k}(\mathcal{S}) = \operatorname{argmax} \pi_k \, \mathcal{N}(\mu_k, \Sigma_k)(\mathcal{S})$$

Modèle de mélange de gaussiennes

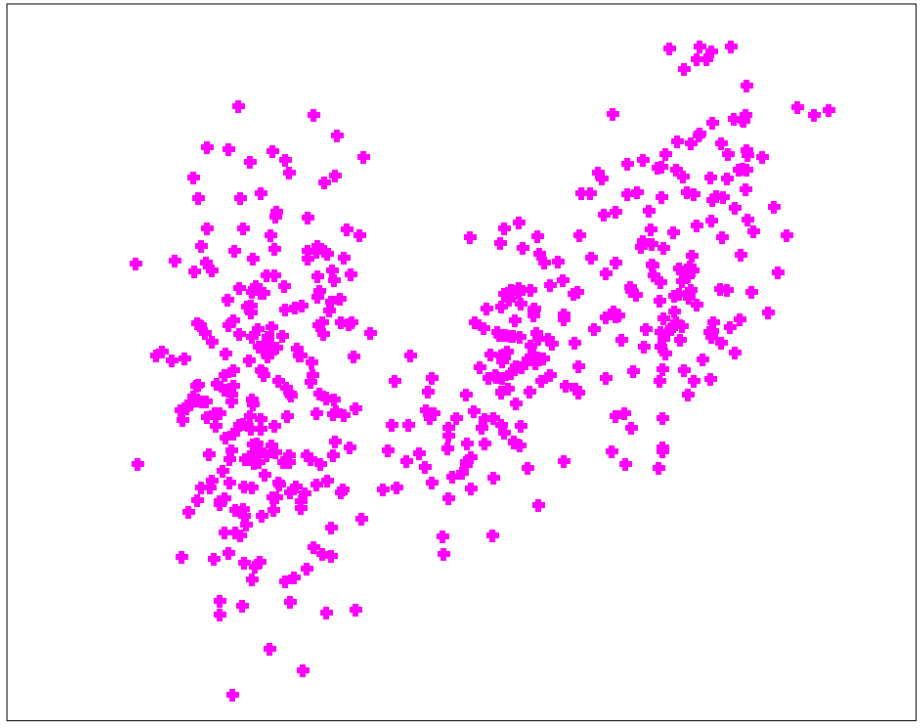
- Densité s_0 de $\mathcal{S}$ proche de $s_m(\mathcal{S}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k, \Sigma_k)(\mathcal{S})$.
- Modèle $S_m = \{s_m\}$:
  - choix d'un nombre de classe K,
  - choix d'une structure pour les moyennes μ_k et les covariances $\Sigma_k = L_k D_k A_k D_k'$
- Modèles $[\mu\, L\, D\, A]^K$: contraintes (valeurs connues, communes ou libres...) sur les moyennes μ_k, les volumes L_k, les bases de diagonalisation D_k et les valeur propres A_k.
- Modèle S_m : modèle paramétrique de dimension $(K-1) + \dim([\mu\, L\, D\, A]^K)$ dans un espace de dimension p.
- Estimation par maximum de vraisemblance des paramètres :
  - pour chaque classe, la moyenne μ_k et la covariance $\Sigma_k = L_k D_k A_k D_k'$
  - les proportions π_k du mélange.
- Technique classique avec algorithme (EM) efficace disponible.

Combien de classes ?

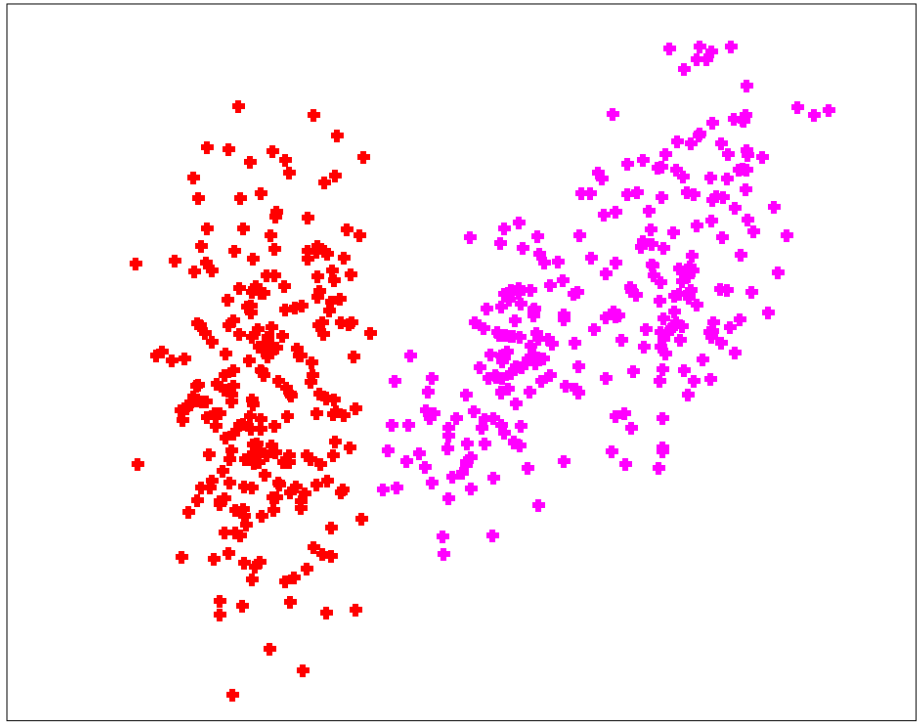
Combien de classes ?



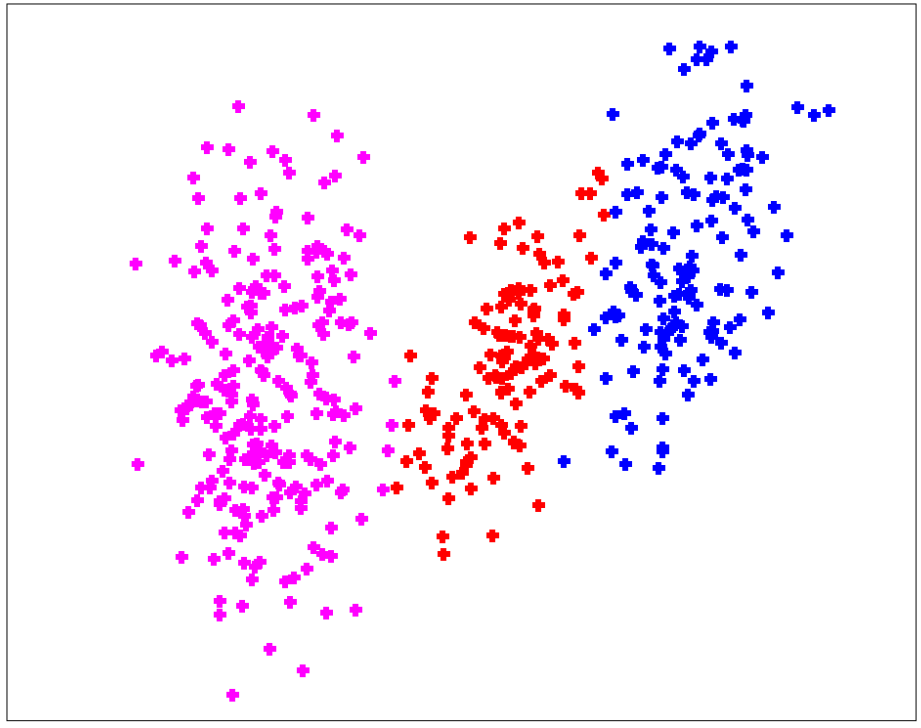
Combien de classes ?



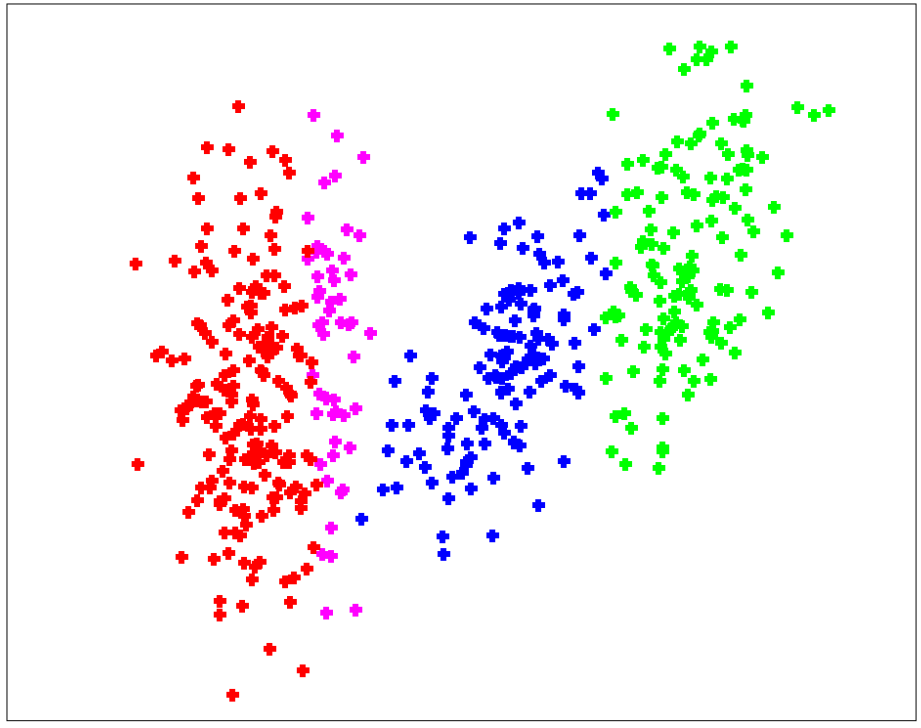
Combien de classes ?



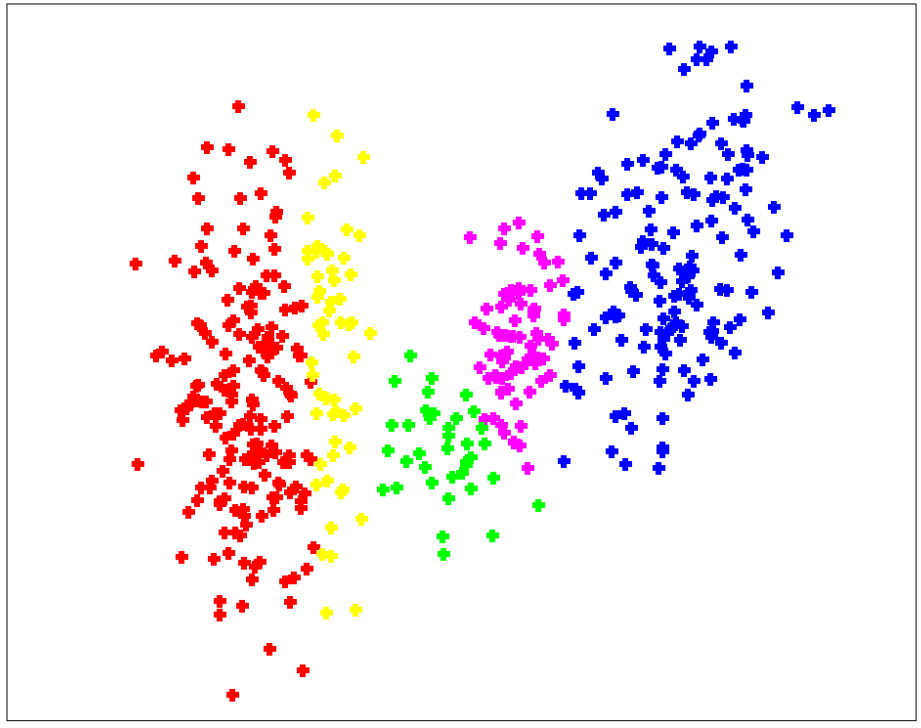
Combien de classes ?



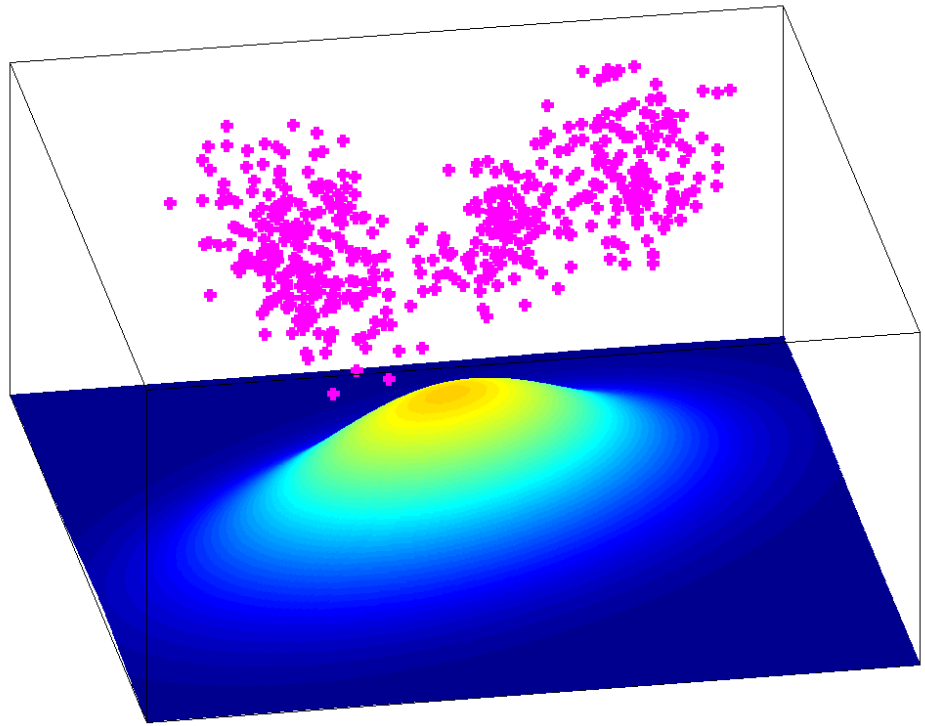
Combien de classes ?



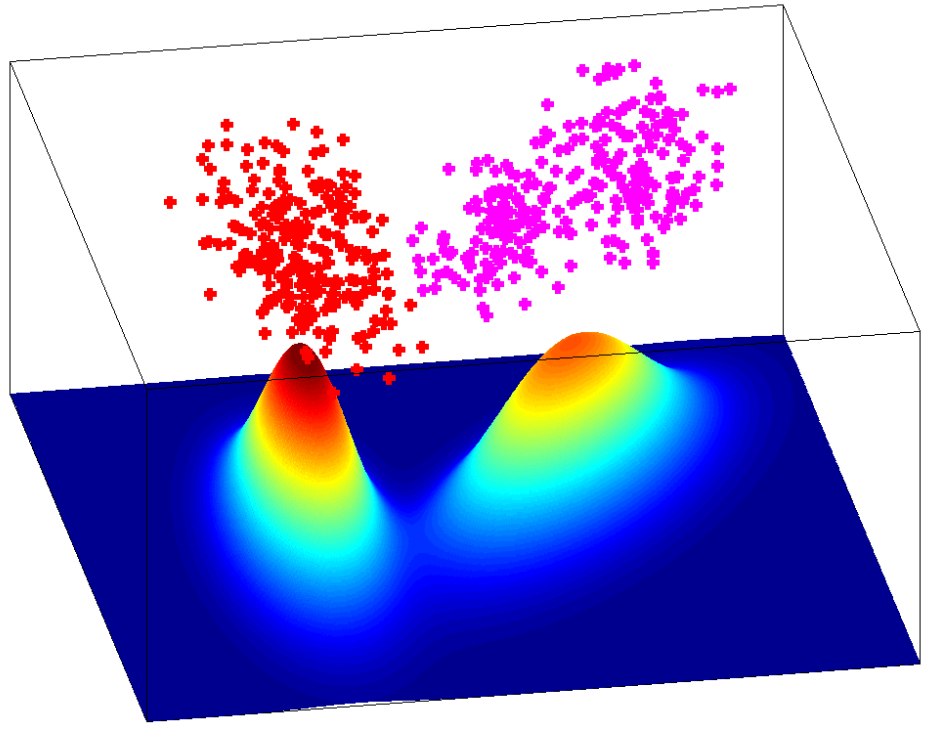
Combien de classes ?



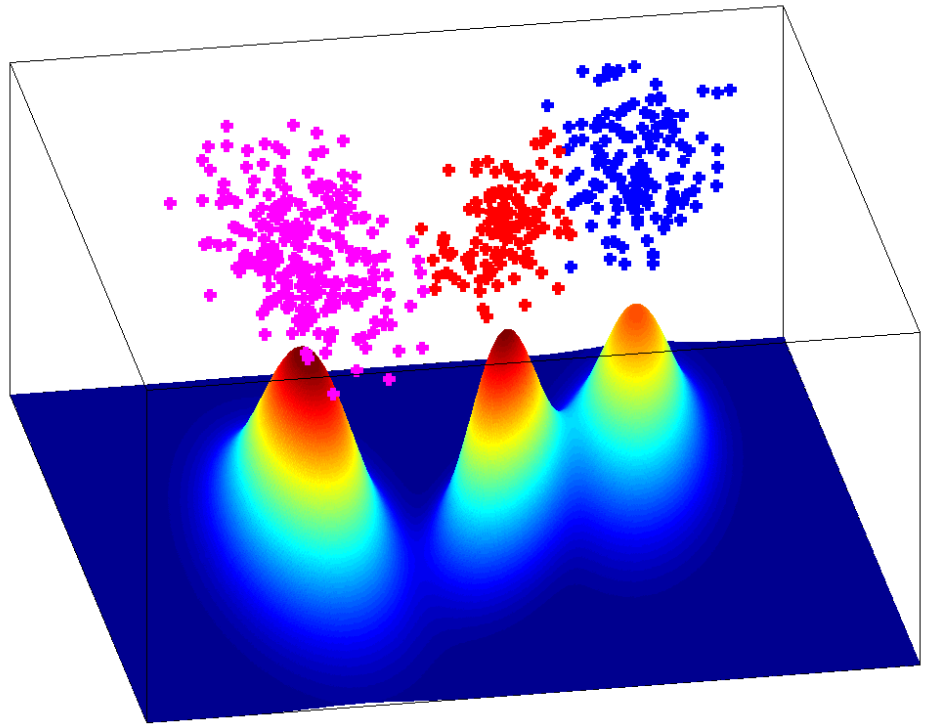
Combien de classes ?



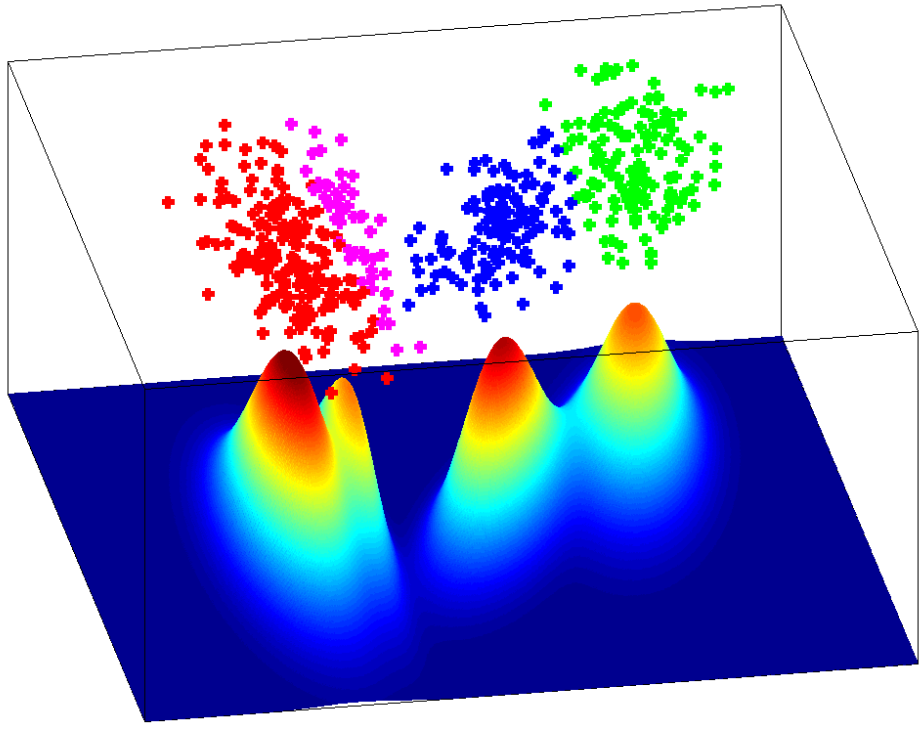
Combien de classes ?



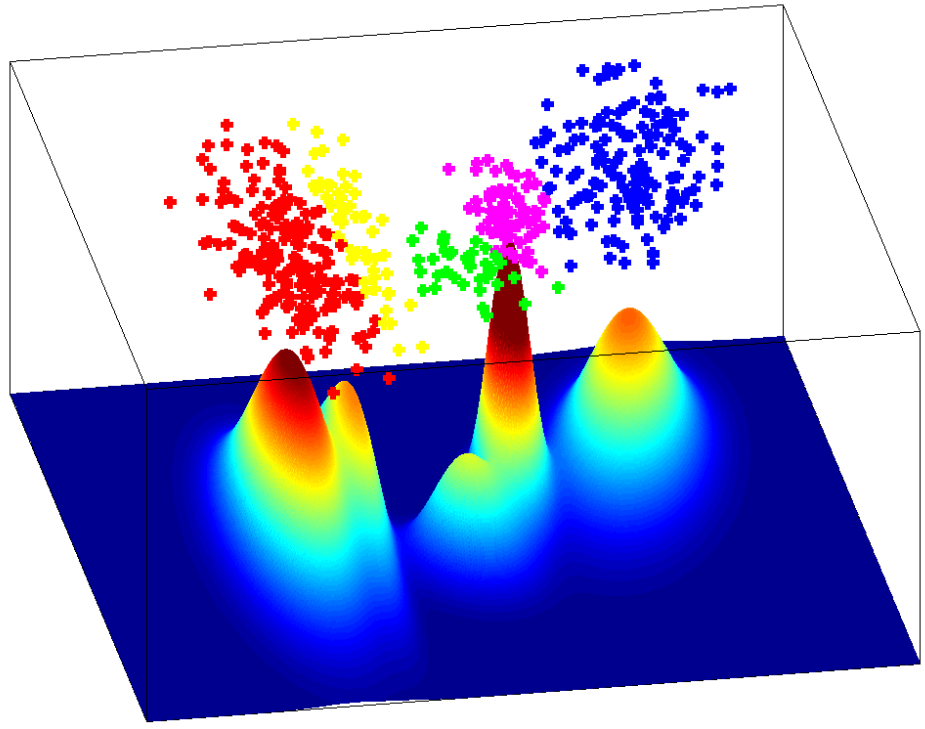
Combien de classes ?



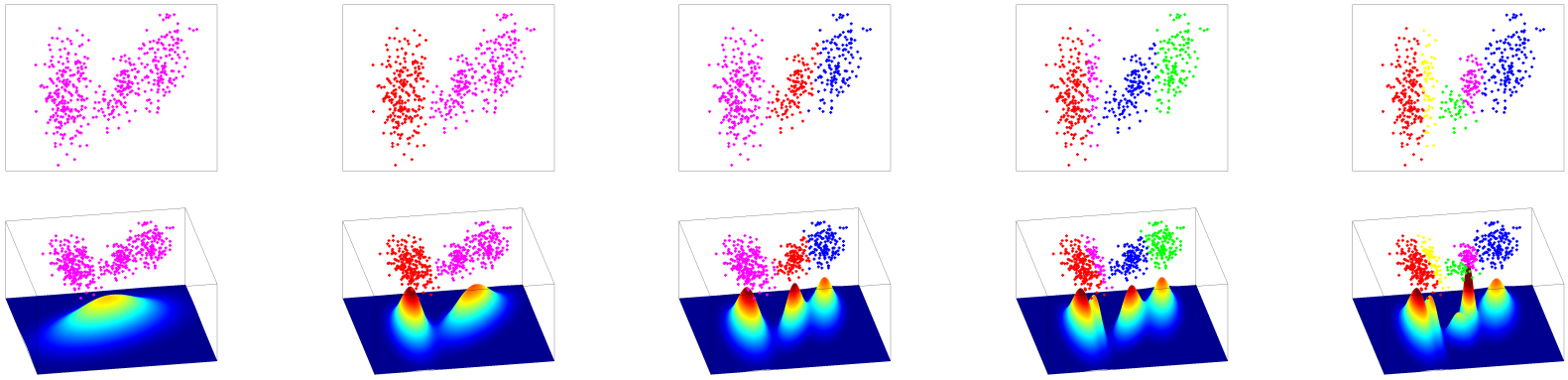
Combien de classes ?



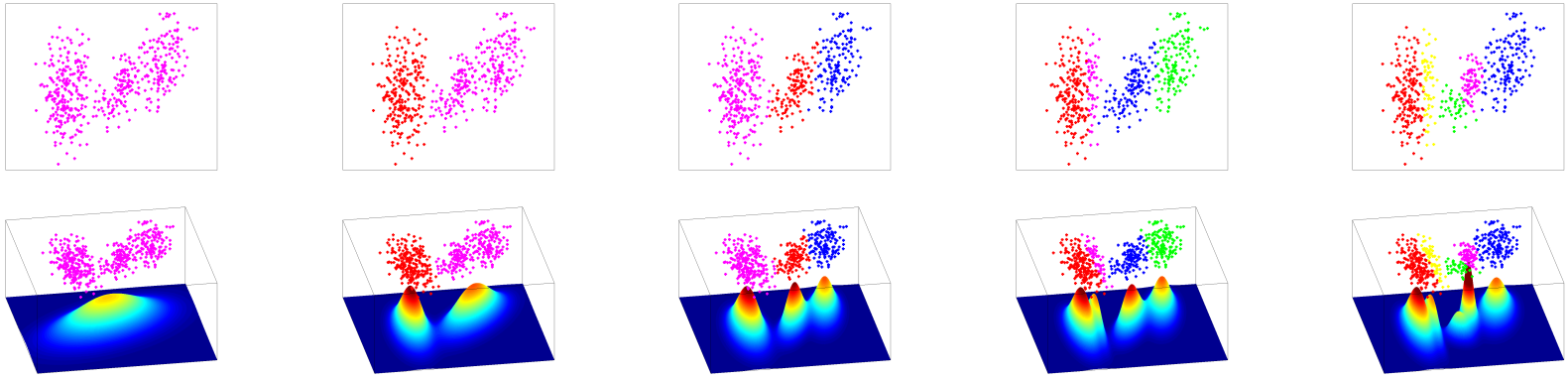
Combien de classes ?



Combien de classes ?

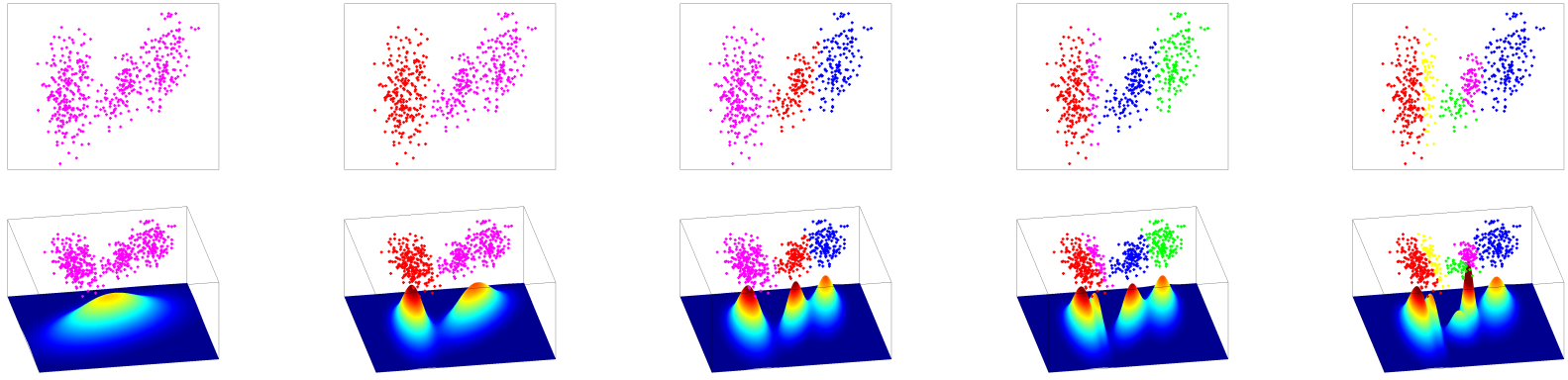


Combien de classes ?



Fidélité $\quad -- \quad + \quad +++ \quad +++\cdot \quad ++++$

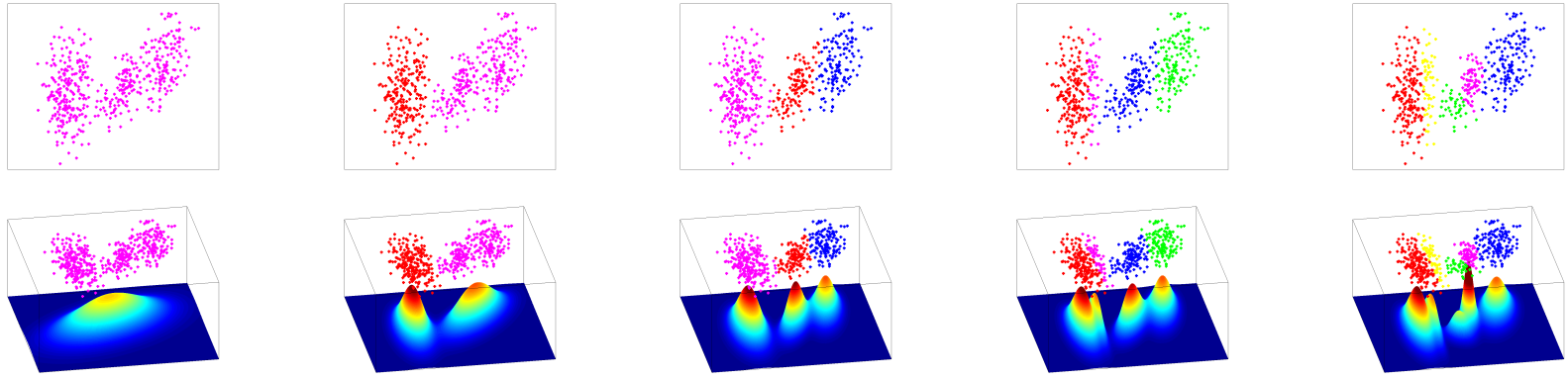
Combien de classes ?



Fidélité $\quad -- \quad + \quad +++ \quad +++\cdot \quad ++++$

- Question difficile où la vraisemblance (la fidélité) ne suffit pas !

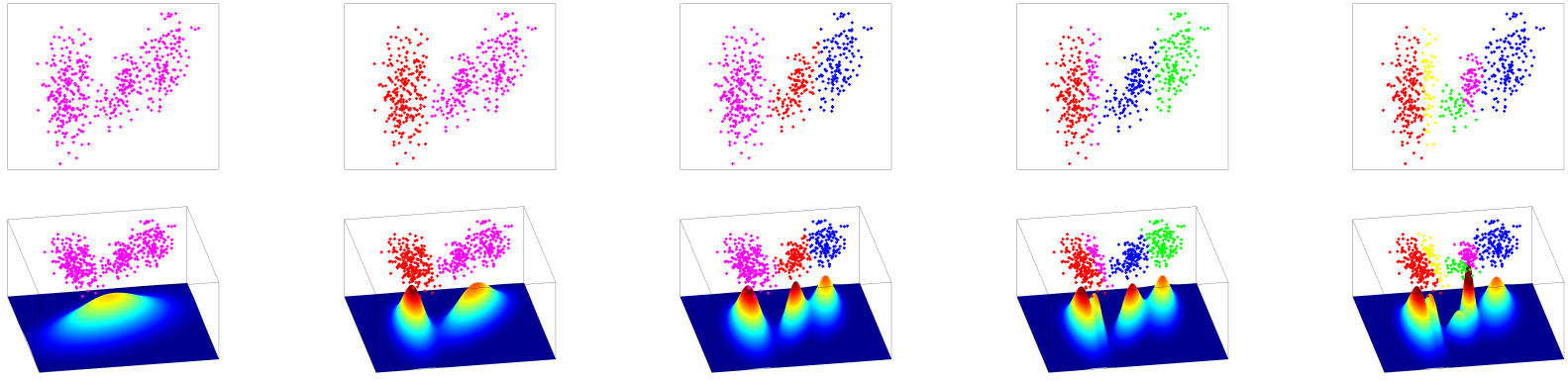
Combien de classes ?



Fidélité	$--$	$+$	$+++$	$+++\cdot$	$++++$
Simplicité	$+++++$	$+++$	$++$	$+$	$-$

- Question difficile où la vraisemblance (la fidélité) ne suffit pas !

Combien de classes ?



Fidélité	$--$	$+$	$+++$	$+++\cdot$	$++++$
Simplicité	$+++++$	$+++$	$++$	$+$	$-$

- Question difficile où la vraisemblance (la fidélité) ne suffit pas !
- Prise en compte de la complexité du modèle ?

Le rasoir d'Ockham

Le rasoir d'Ockham

Les multiples ne doivent pas être utilisés sans nécessité.
Guillaume d'Ockham ($\sim$ 1285 - 1347)

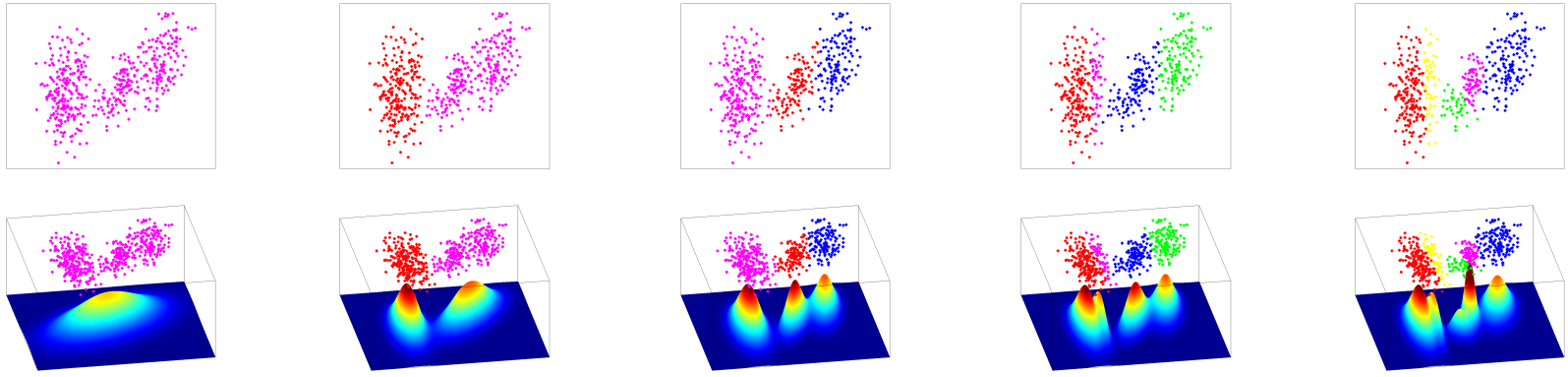
Le rasoir d'Ockham

Les multiples ne doivent pas être utilisés sans nécessité.
Guillaume d'Ockham ($\sim$ 1285 - 1347)

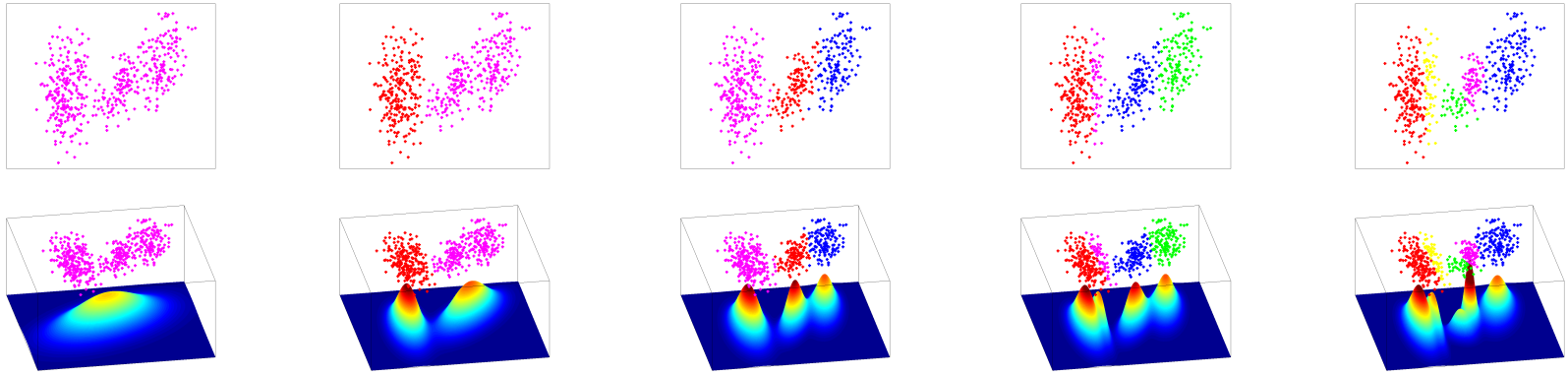
- Rasoir d'Ockham (principe de simplicité) : il ne faut pas ajouter des hypothèses, si celles utilisées suffisent déjà !
- Compromis entre pouvoir d'explication et simplicité.

Sélection par pénalisation

Sélection par pénalisation

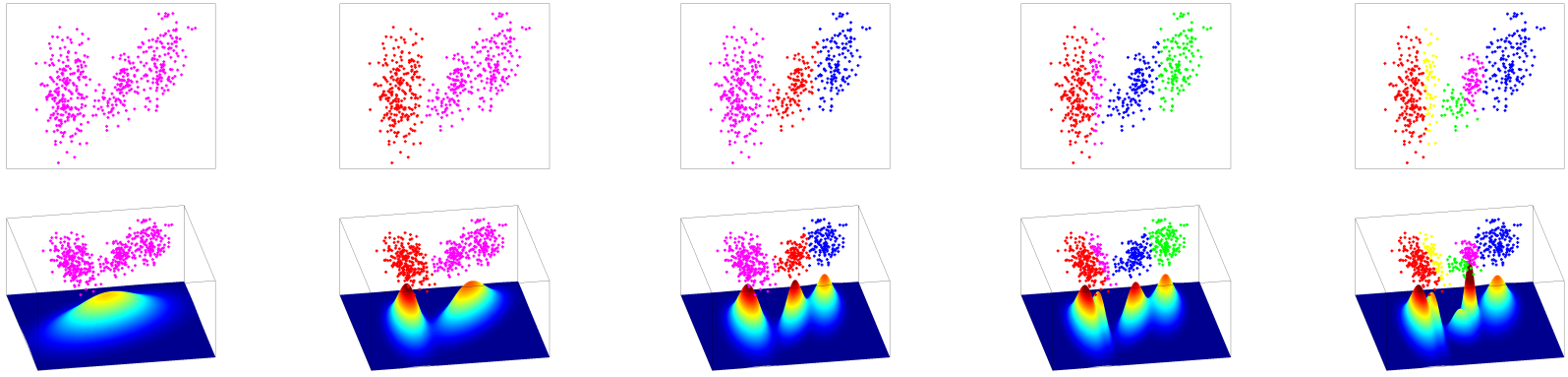


Sélection par pénalisation



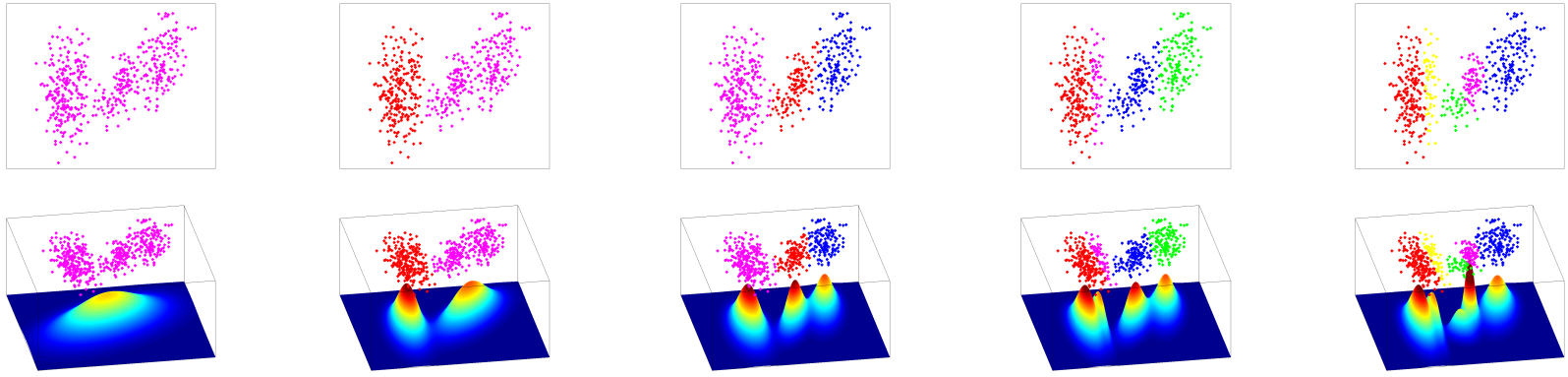
Vraisemblance	$--$	$+$	$+++$	$+++\cdot$	$++++$
Simplicité	$+++++$	$+++$	$++$	$+$	$-$

Sélection par pénalisation



Vraisemblance	$--$	$+$	$+++$	$+++\cdot$	$++++$
$+$ Simplicité	$+++++$	$+++$	$++$	$+$	$-$
$=$ Compromis	$++$	$++++$	$+++++$	$++++\cdot$	$+++$

Sélection par pénalisation

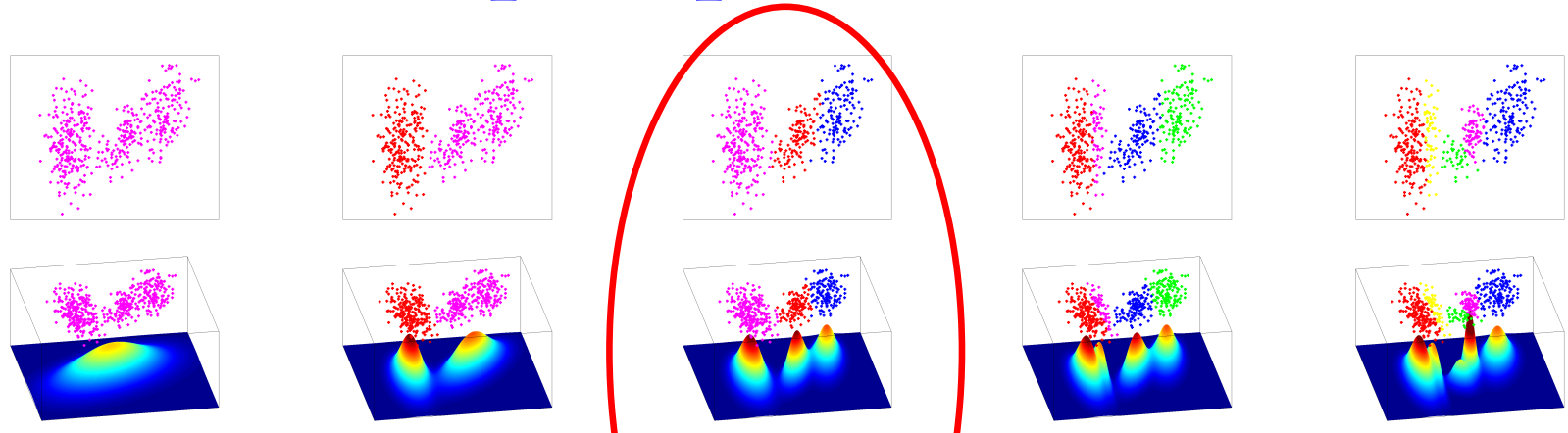


Vraisemblance	$--$	$+$	$+++$	$+++\cdot$	$++++$
$+$ Simplicité	$+++++$	$+++$	$++$	$+$	$-$
$=$ Compromis	$++$	$++++$	$+++++$	$++++\cdot$	$+++$

- Vraisemblance : $\sum_{i=1}^{n} \log \hat{s}_K(X_i)$.
- Simplicité : $-\lambda \mathrm{Dim}(S_K)$ (beaucoup de théorie derrière).
- Estimateur pénalisé :

$$\mathrm{argmax} \underbrace{\sum_{i=1}^{n} \log \hat{s}_K(X_i)}_{\text{Vraisemblance}} \underbrace{- \lambda \mathrm{Dim}(S_K)}_{\text{Pénalité}}$$

Sélection par pénalisation



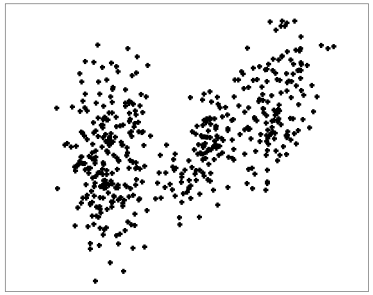
Vraisemblance	$--$	$+$	$+++$	$+++\cdot$	$++++$
$+$ Simplicité	$+++++$	$+++$	$++$	$+$	$-$
$=$ Compromis	$++$	$++++$	$+++++$	$++++\cdot$	$+++$

- Vraisemblance : $\sum_{i=1}^{n} \log \hat{s}_K(X_i)$.
- Simplicité : $-\lambda \mathrm{Dim}(S_K)$ (beaucoup de théorie derrière).
- Estimateur pénalisé :

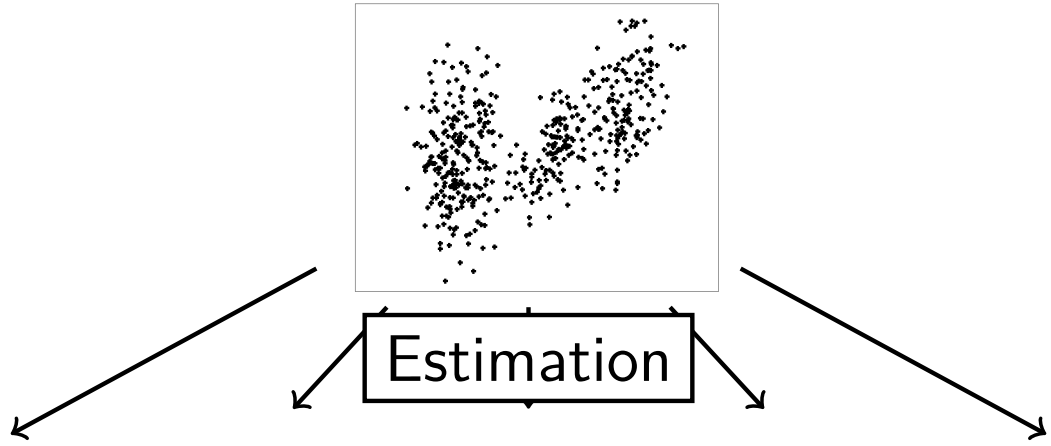
$$\mathrm{argmax} \underbrace{\sum_{i=1}^{n} \log \hat{s}_K(X_i)}_{\text{Vraisemblance}} \underbrace{- \lambda \mathrm{Dim}(S_K)}_{\text{Pénalité}}$$

Méthodologie

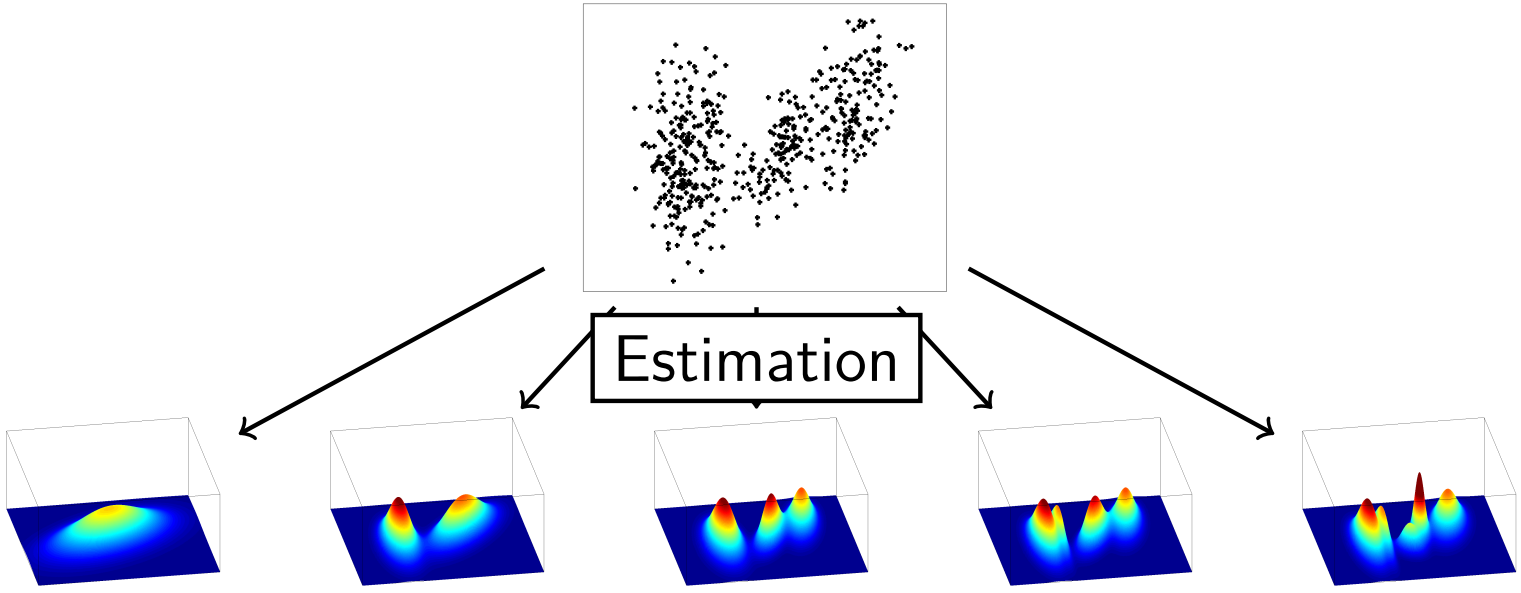
Méthodologie



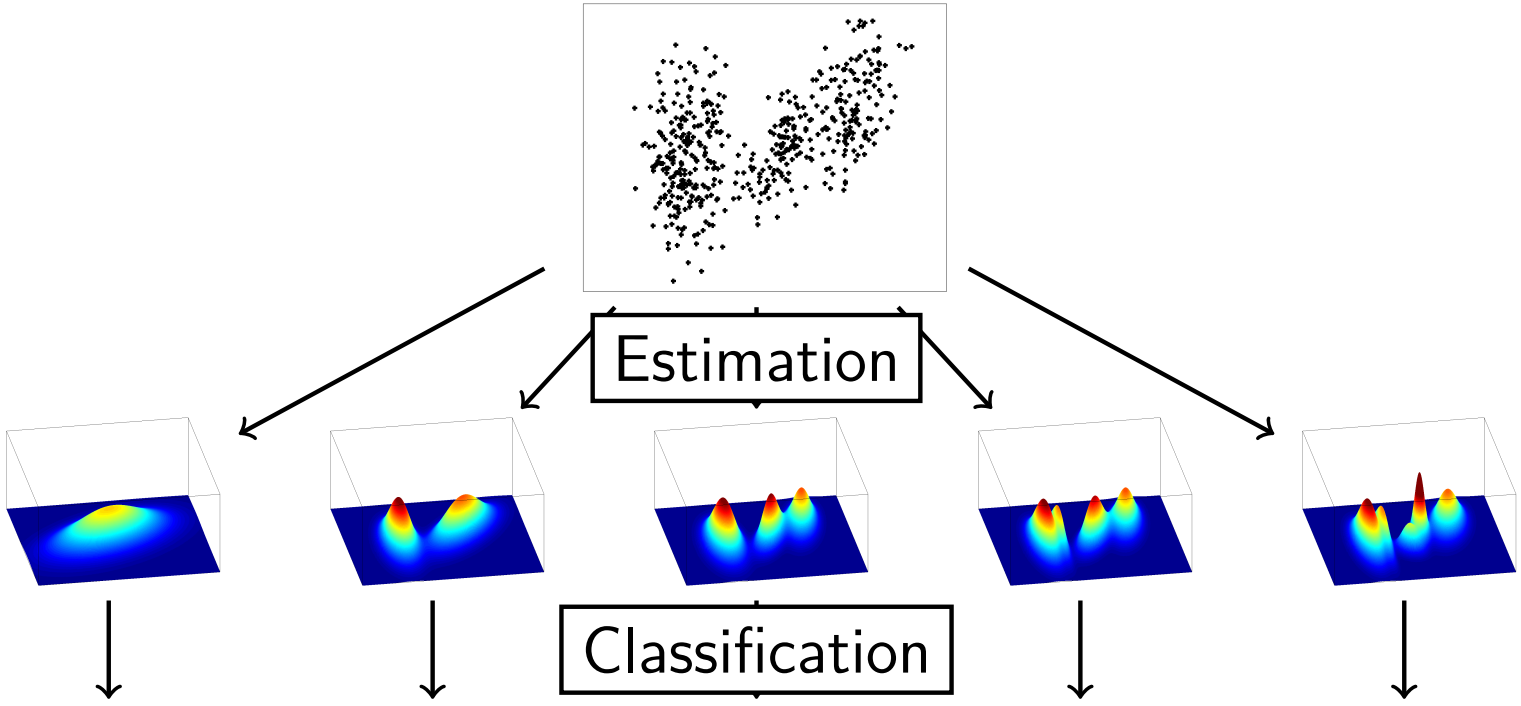
Méthodologie



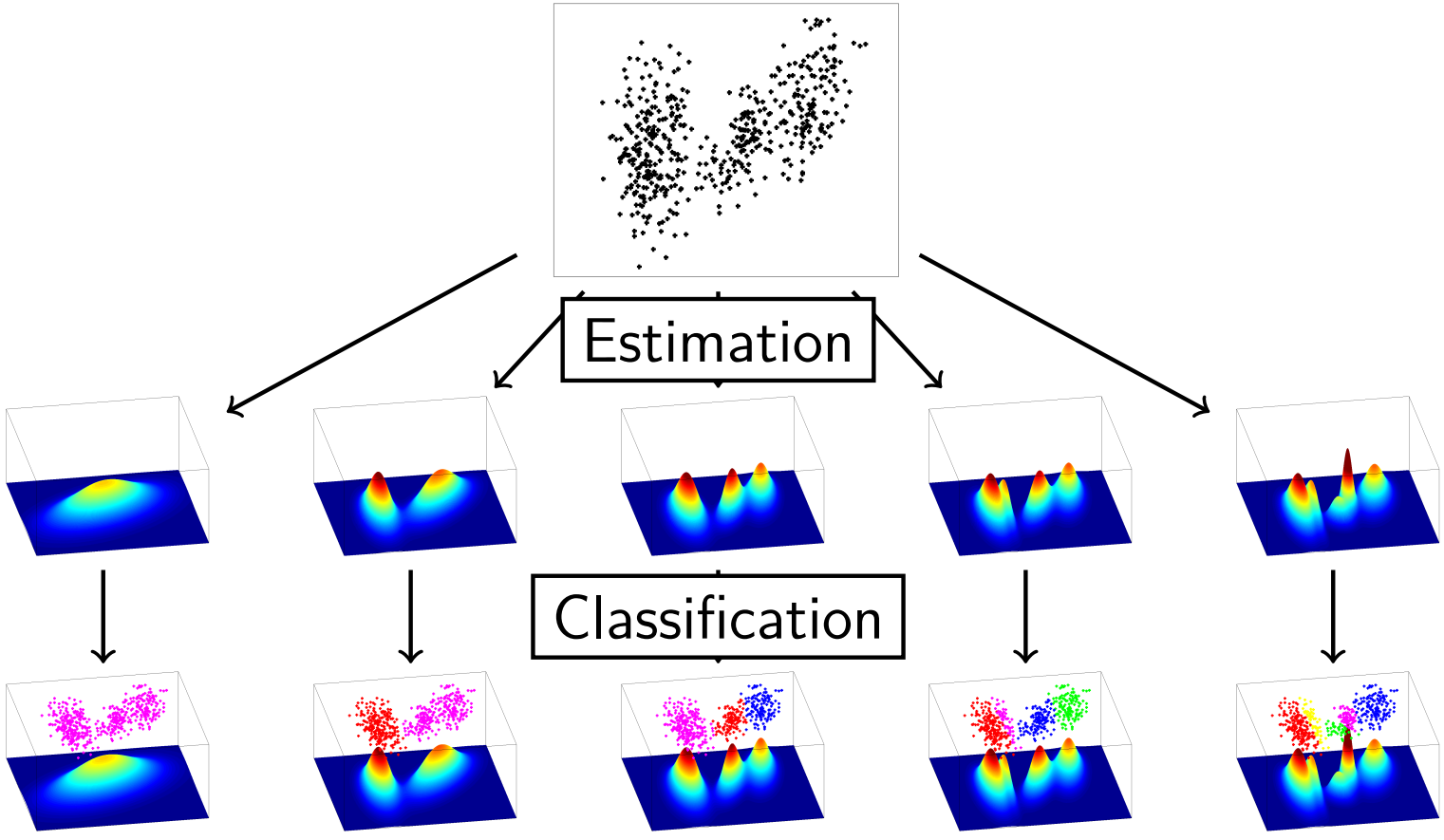
Méthodologie



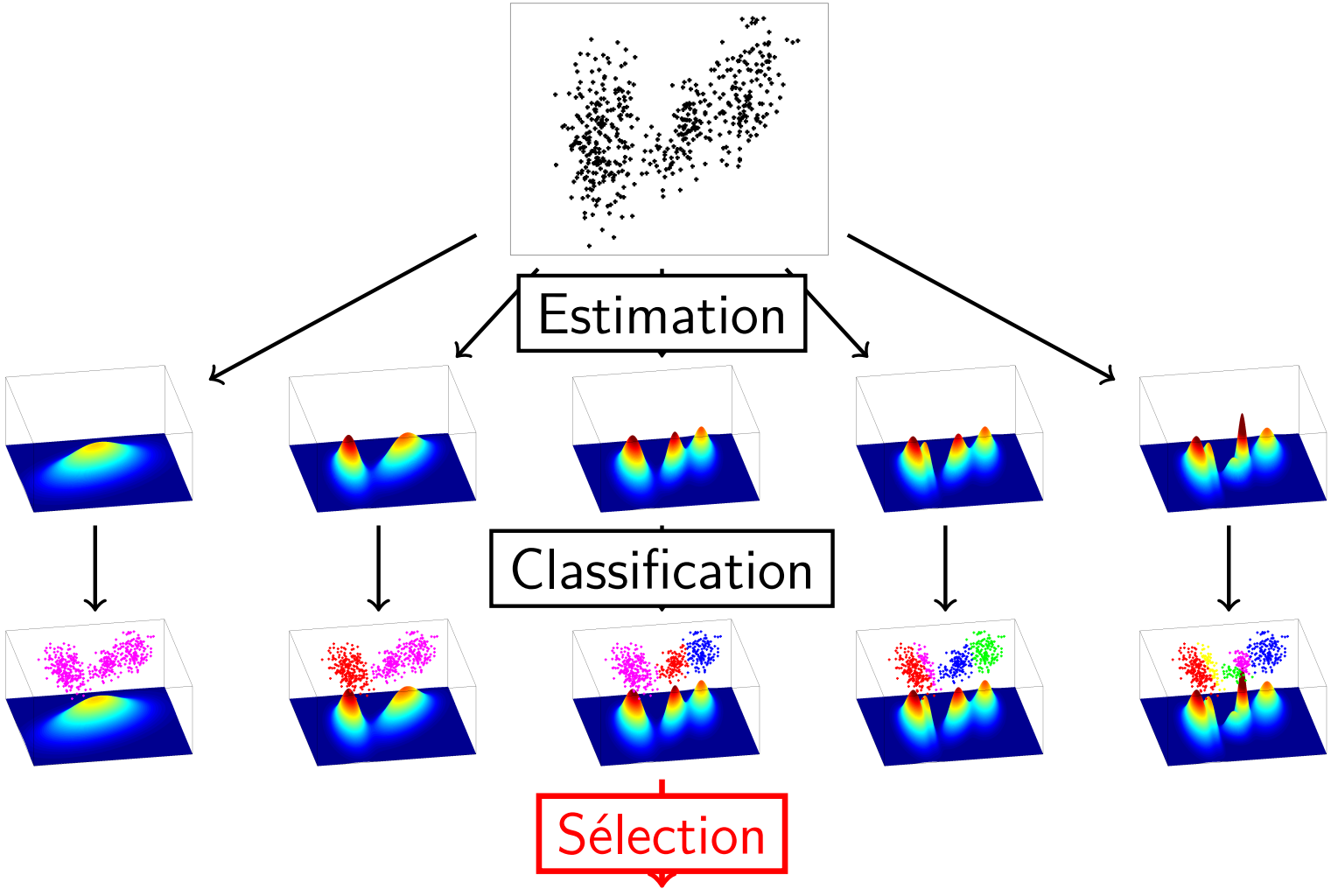
Méthodologie



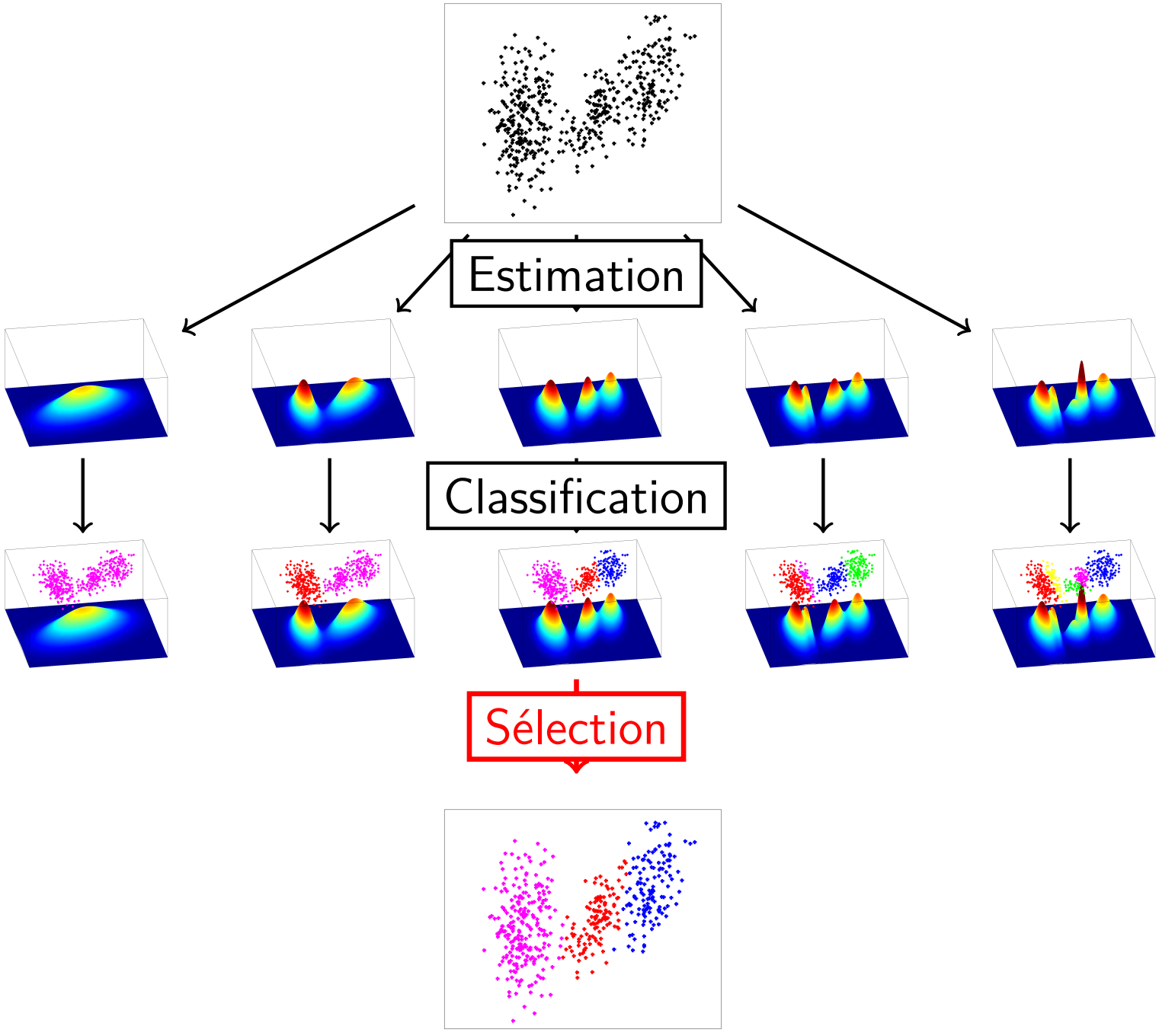
Méthodologie



Méthodologie



Méthodologie



Sélection de modèles

- Comment choisir le "modèle" S_m :
  - le nombre de classe K,
  - le modèle $[\mu\, L\, D\, A]^K$?
- Thème central du projet SELECT.
- Principe de sélection de modèles par pénalisation :
  - choix d'une collection de modèles $S_m = \{s_m\}$ avec $m \in \mathcal{M}$,
  - estimation par maximum de vraisemblance d'une densité $\hat{s}_m$ pour chaque modèle S_m,
  - sélection d'un modèle $\widehat{m}$ par

$$\widehat{m} = \operatorname{argmin} - \ln(\widehat{s}_m) + \operatorname{pen}(m).$$

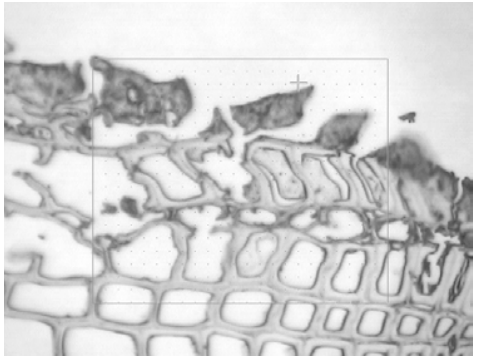
    avec $\operatorname{pen}(m) = \kappa(\ln(n))\dim(S_m)$ (dimension intrinsèque de S_m),
- Résultats (Birgé, Massart, Celeux, Maugis, Michel...) :
  - théorique d'estimation du mélange : pour κ assez grand,

$$\mathbb{E}\left[d^2(s_0, \widehat{s}_{\widehat{m}})\right] \leq C \inf_{m\in\mathcal{M}} \left(\inf_{s_m \in S_m} KL(s_0, s_m) + \frac{\operatorname{pen}(m)}{n} \right) + \frac{C'}{n}.$$

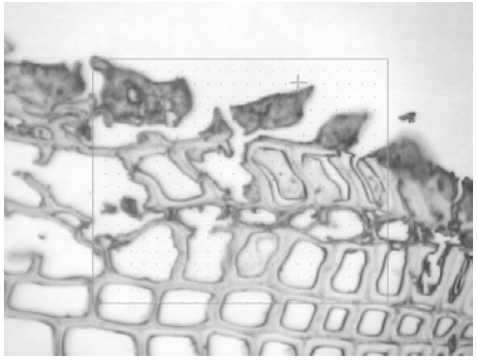
  - pratique de classification non supervisée ($\neq$ segmentation),
  - consistance de la classification si $\ln\ln(n)$ dans la pénalité...

Retour à nos violons

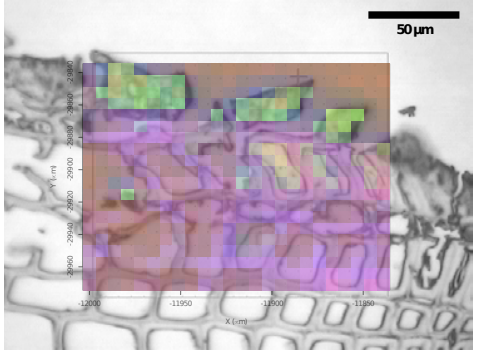
Retour à nos violons



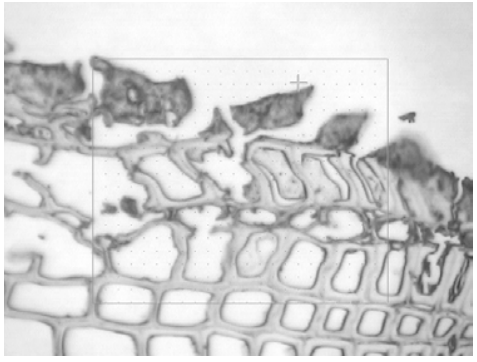
Retour à nos violons



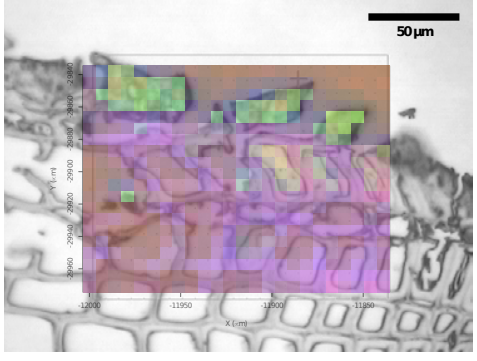
Segmentation

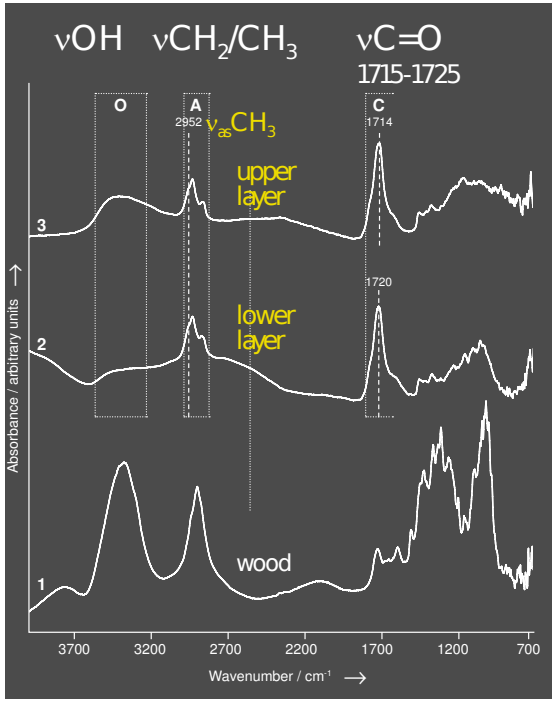


Retour à nos violons

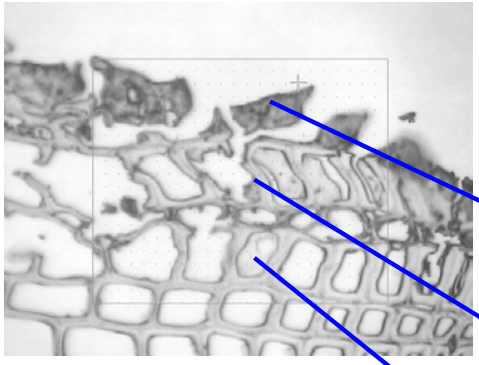


Segmentation

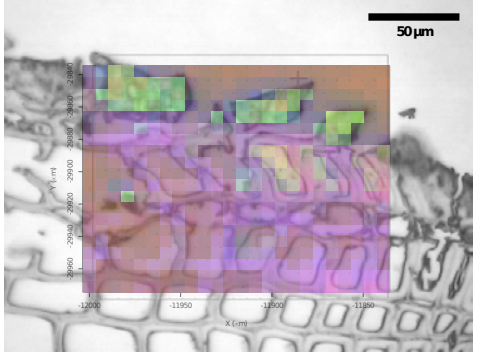


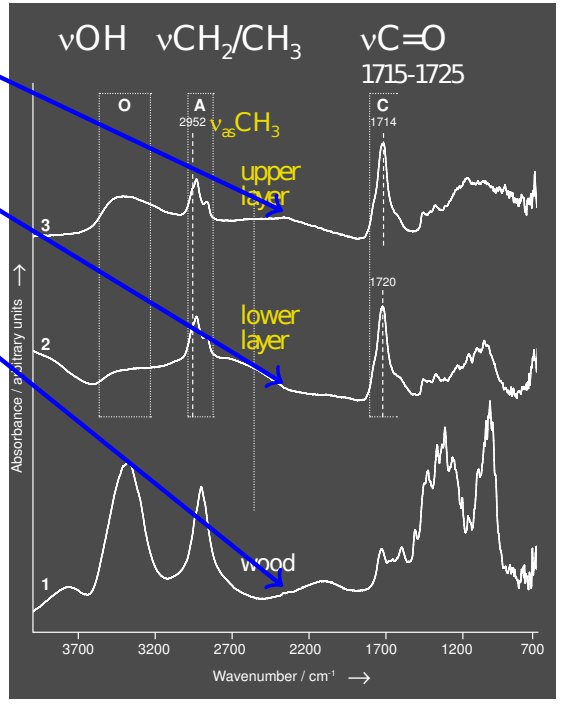


Retour à nos violons

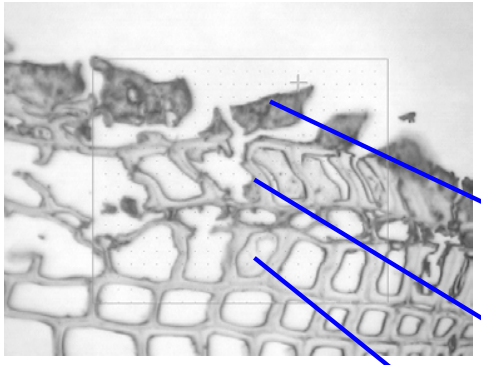


Segmentation

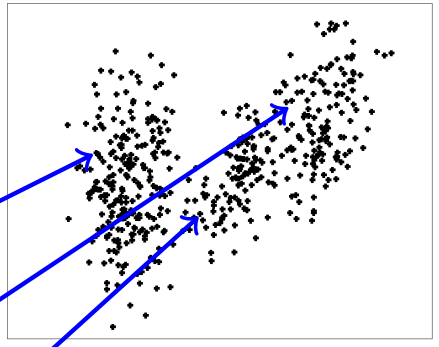


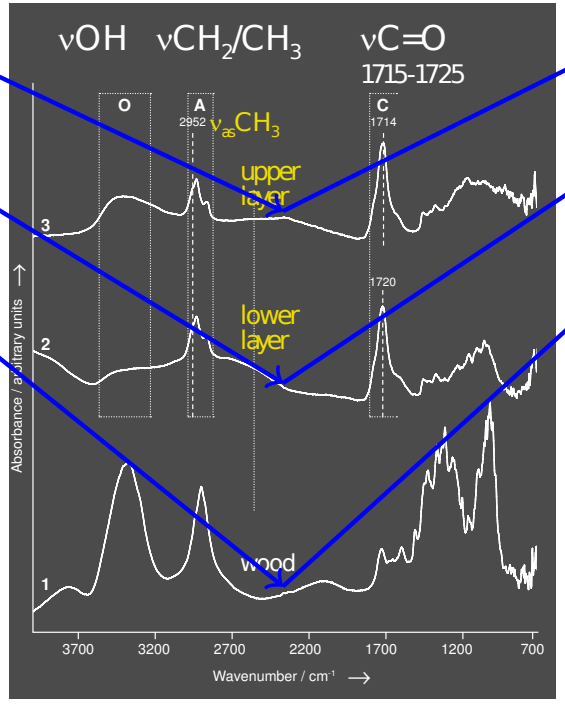


Retour à nos violons

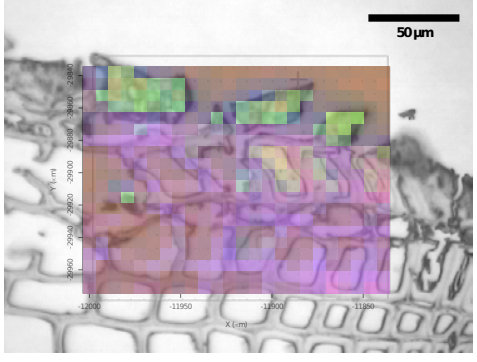


Représentation

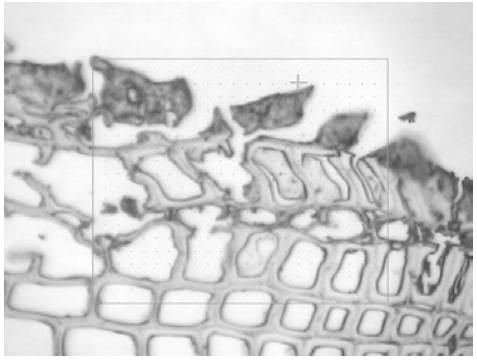




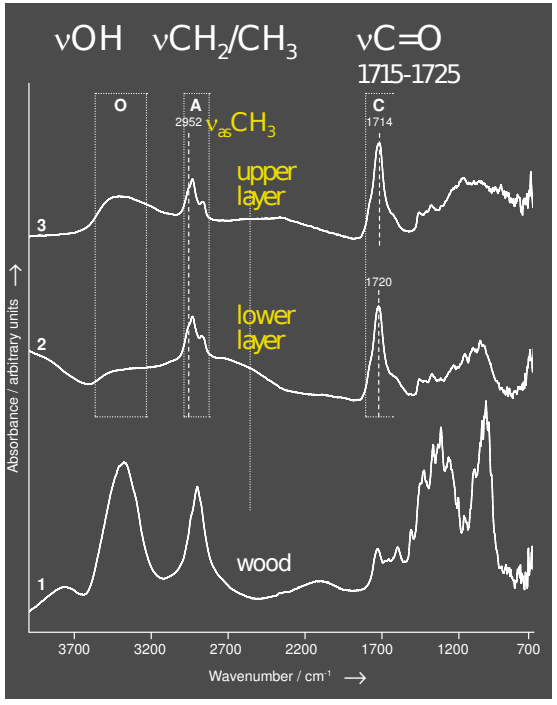
Segmentation

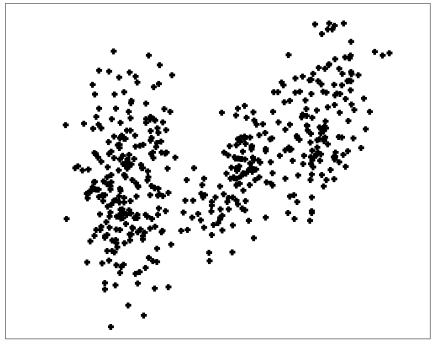


Retour à nos violons

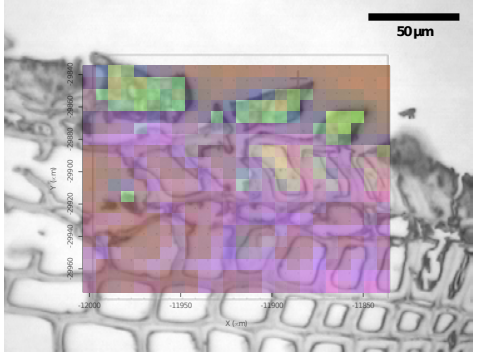


Représentation

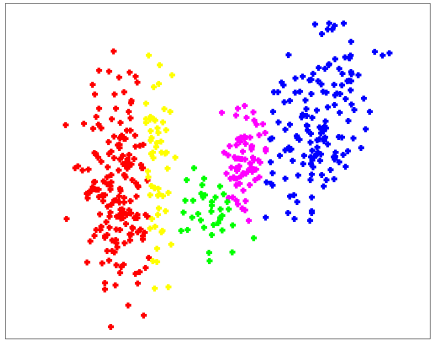




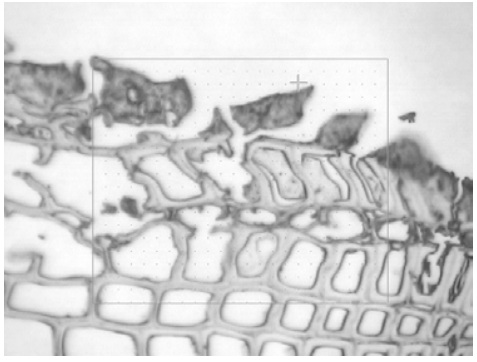
Segmentation



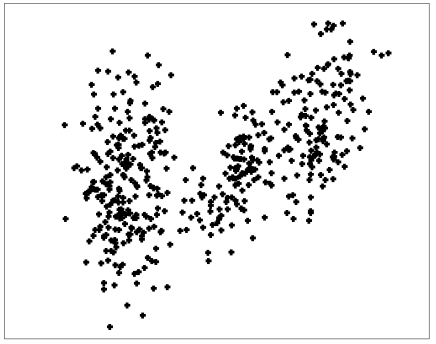
Classification

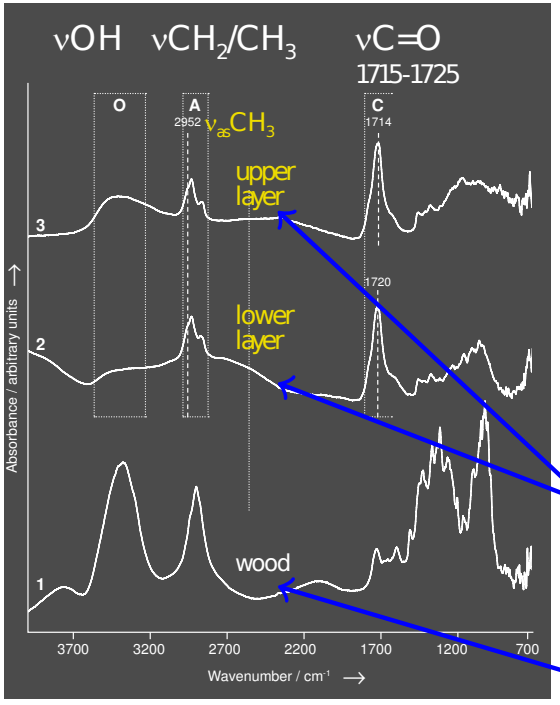


Retour à nos violons

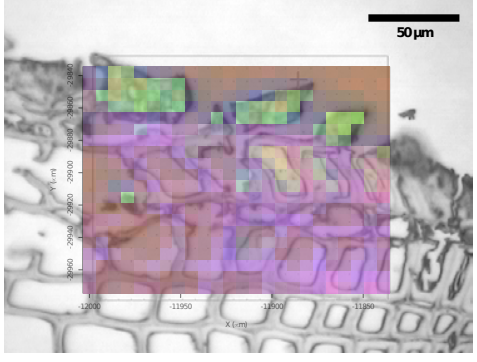


Représentation

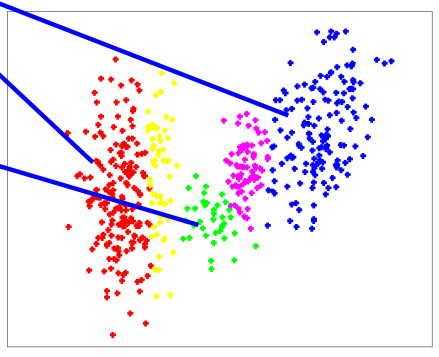




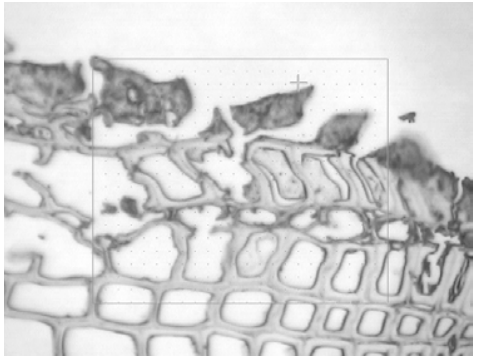
Segmentation



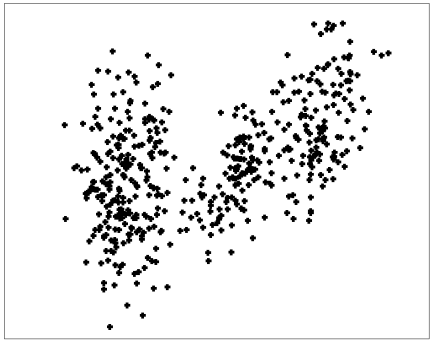
Classification

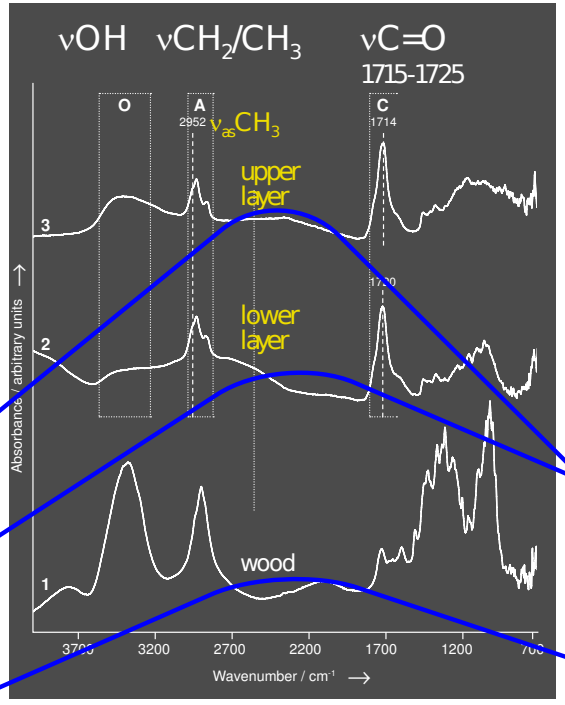


Retour à nos violons



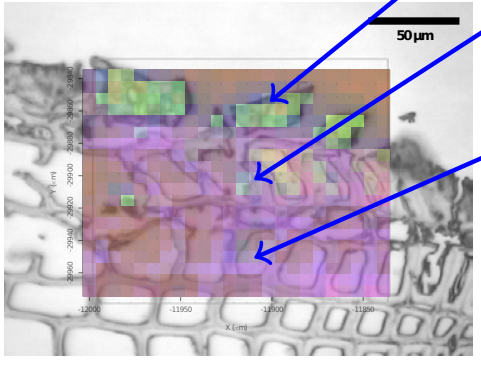
Représentation

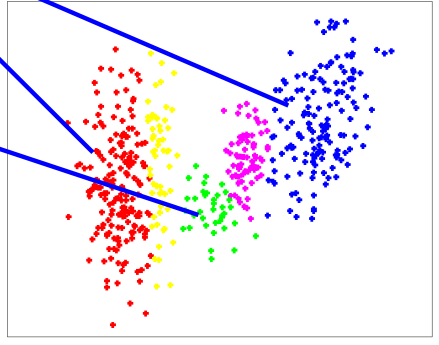




Segmentation

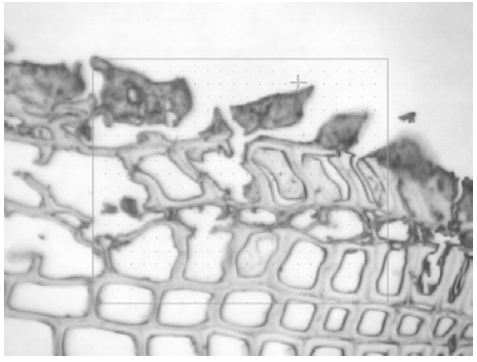
Classification



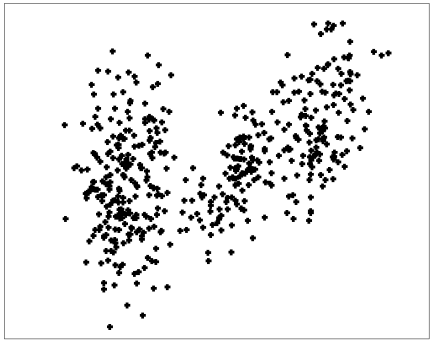


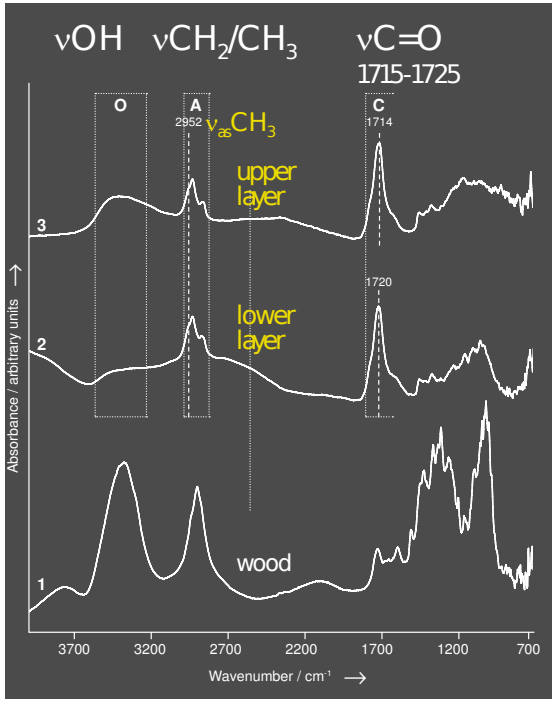
Info. Spatiale

Retour à nos violons



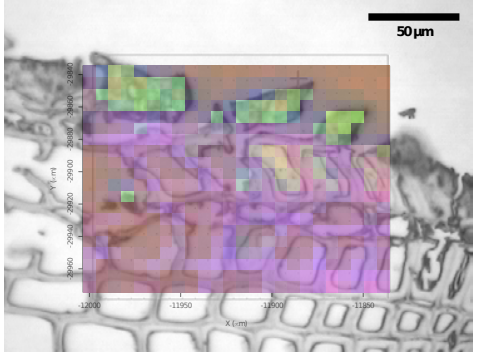
Représentation

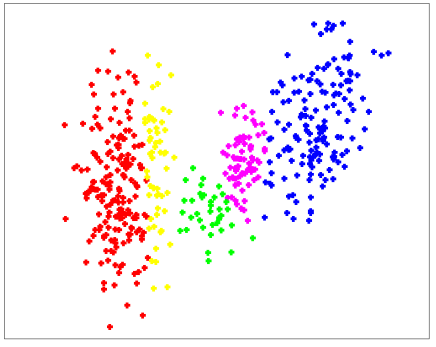




Segmentation

Classification





Info. Spatiale

Segmentation et mélange de gaussiennes

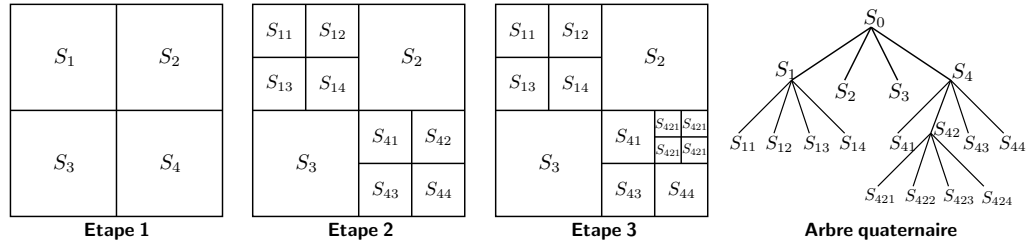
- Objectif initial : segmentation $\neq$ classification non supervisée.
- Prise en compte de la position spatiale x du spectre à travers les proportions du mélange (Kolaczyk et al) : modèle de densités conditionnelles

$$s(\mathcal{S}|x) = \sum_{k=1}^{K} \pi_k(x)\, \mathcal{N}(\mu_k, \Sigma_k)(\mathcal{S}).$$

- Modèle mélangeant paramétrique et "non-paramétrique"...
- Estimation à partir des données :
  - pour chaque classe, la moyenne μ_k et la covariance $\Sigma_k = L_k D_k A_k D_k'$,
  - de la fonction de mélange $\pi_k(x)$.
- $\pi_k(x)$ fonction : régularisation nécessaire.
- Principe de sélection de modèles...

Mélange de gaussiennes et partition hiérarchique

- Comment choisir le "modèle" S_m ? :
  - le nombre de classe K,
  - le modèle $[\mu\, L\, D\, A]^K$,
  - la structure des paramètres de mélange $\pi_k(x)$.
- Structure simple pour $\pi_k(x)$:
  - constant par morceau
    sur une partition "hiérarchique",
  - optimisation efficace possible,
  - performance d'approximation raisonnable.



- $\dim(S_m) = |\mathcal{P}|(K-1) + \dim([\mu\, L\, D\, A]^K)$.
- Pénalité $\mathrm{pen}(m) = \kappa \ln(n) \dim(S_m)$ suffisante pour
  - l'optimisation numérique (EM + programmation dynamique),
  - le contrôle théorique : pour κ assez grand,

$$\mathbb{E}\left[d^2(s_0, \widehat{s}_{\widehat{m}})\right] \leq C \inf_{m \in \mathcal{M}} \left(\inf_{s_m \in S_m} KL(s_0, s_m) + \frac{\mathrm{pen}(m)}{n} \right) + \frac{C'}{n}.$$

Densités conditionnelles

- Cadre plus général : observation de (X_i, Y_i) avec X_i indépendants et Y_i indépendants de loi de densité $s_0(y|X_i)$.
- Objectif : estimation de $s_0(y|x)$.
- Principe de sélection de modèles par pénalisation :
  - choix d'une collection de modèles $S_m = \{s_m(y|x)\}$ avec $m \in \mathcal{M}$,
  - estim. par max. de vraisemblance d'une dens. $\hat{s}_m$ pour chaque modèle S_m :

$$\hat{s}_m = \underset{s_m \in S_m}{\operatorname{argmin}} - \sum_{i=1}^{n} \ln s_m(Y_i|X_i)$$

  - avec $\mathrm{pen}(m)$ à bien choisir, sélection d'un modèle $\widehat{m}$ par

$$\widehat{m} = \underset{m \in \mathcal{M}}{\operatorname{argmin}} - \sum_{i=1}^{n} \ln \widehat{s}_m(Y_i|X_i) + \mathrm{pen}(m).$$

- Résultat d'estimation de densité du type

$$\mathbb{E}\left[d^2(s_0, \widehat{s}_{\widehat{m}})\right] \leq C \inf_{m \in \mathcal{M}} \left(\inf_{s_m \in S_m} KL(s_0, s_m) + \frac{\mathrm{pen}(m)}{n} \right) + \frac{C'}{n}.$$

- Biblio succinte : Rosenblatt, Fan et al., de Gooijer and Zerom, Efromovitch, Brunel, Comte, Lacour... / Plugin, estimation directe, perte L^2, minimax, censure...

Theorem

Under Assumption (H_{β_ϕ}) : the existence, for a given $\beta_\phi \geq 1$, of a non-decreasing function $\phi_m(\delta, \beta_\phi)$ such that $\delta \mapsto \frac{1}{\beta_\phi \delta}\phi_m(\delta, \beta_\phi)$ is non-increasing on $(0, +\infty)$ and for every $\sigma \in \mathbb{R}^+$ and every $s_m \in S_m$

$$\int_0^{\beta_\phi \sigma} \sqrt{H_{[\cdot],d^{\otimes n}}(\epsilon, S_m(s_m, \sigma))}\, d\epsilon \leq \phi_m(\sigma, \beta_\phi).$$

This assumption implies the following theorem (up to a technical measurability condition) :

Theorem

Assume we observe (X_i, Y_i) with unknown conditional s_0. Let $(S_m)_{m\in\mathcal{M}}$ a at most countable model collection.

Assume that there is a family $(x_m)_{m\in\mathcal{M}}$ of non-negative number such that $\sum_{m\in\mathcal{M}} e^{-x_m} \leq \Sigma < +\infty \qquad (K)$

and, under assumption (H), let σ_m be the unique root of $\frac{1}{\sigma}\phi_m(\sigma, \beta_\phi) = \frac{1}{\beta_\sigma}\sqrt{n}\sigma. \qquad (H_\sigma)$

and let $\widehat{s_m}$ be a ρ maximum likelihood minimizer in S_m : $\sum_{i=1}^n -\ln(\widehat{s_m}(Y_i|X_i)) \leq \inf_{s_m\in S_m}\left(\sum_{i=1}^n -\ln(s_m(Y_i|X_i))\right) + \rho$

For any $\rho \in (0,1)$, any $\beta_x > 0$ and any $C_1 > 1$, there are two absolute constants κ_0 and C_2 such as soon as for every model $m \in \mathcal{M}$

$$\mathrm{pen}(m) \geq \kappa\left(n\sigma_m^2 + \beta_x^2 x_m\right) \quad \text{with } \kappa > \kappa_0, \qquad (P)$$

the penalized likelihood estimate $\widehat{s}_{\widehat{m}}$ with $\widehat{m}$ defined by $\widehat{m} = \operatorname{argmin}_{m\in\mathcal{M}} \sum_{i=1}^n -\ln(\widehat{s_m}(Y_i|X_i)) + \mathrm{pen}(m)$

satisfies $\mathbb{E}\left[JKL_\rho^{\otimes n}(s_0, \widehat{s}_{\widehat{m}})\right] \leq C_1 \inf_{S\in\mathcal{M}}\left(\inf_{s_m\in S_m} KL^{\otimes n}(s_0, s_m) + \frac{\mathrm{pen}(m)}{n}\right) + C_2\frac{\Sigma}{n} + \frac{\rho}{n}.$

Théorème

- Inégalité oracle

$$\mathbb{E}\left[JKL_{\rho}^{\otimes n}(s_0, \widehat{s}_{\widehat{m}})\right] \leq C_1 \inf_{S \in \mathcal{M}} \left(\inf_{s_m \in S_m} KL^{\otimes n}(s_0, s_m) + \frac{\mathrm{pen}(m)}{n} \right) + C_2 \frac{\Sigma}{n} + \frac{\rho}{n}$$

dès que

$$\mathrm{pen}(m) \geq \kappa \left(n\sigma_m^2 + x_m \right) \quad \text{with } \kappa > \kappa_0,$$

où $n\sigma_m^2$ mesure la complexité du modèle S_m (entropie) et x_m le coût de codage dans la collection.

- « Distances » utilisées $KL^{\otimes n}$ et $JKL_{\rho}^{\otimes n}$: divergence de Kullback et divergence de Jensen-Kullback « tensorisées ».
- $n\sigma_m^2$ lié à l'entropie à crochet de S_m mesurée par rapport à la distance de Hellinger tensorisée $d^{2\otimes n}$.

Kullback, Hellinger et extensions

- Inégalité oracle en sélection de modèles de la forme

$$\mathbb{E}\left[d^2(s_0, \widehat{s}_{\widehat{m}})\right] \leq C\left(\inf_{m\in\mathcal{M}} \inf_{s_m\in S_m} KL(s_0, s_m) + \frac{\mathrm{pen}(m)}{n}\right) + \frac{C'}{n}.$$

- Densité : Hellinger $d^2(s, s')$ (ou affinité) (Kolaczyk, Barron, Bigot).
- Raff. avec $JKL(s, s') = 2KL(s, (s' + s)/2)$ (Massart, van de Geer).
- Jensen-Kullback : généralisation à
  $JKL_\rho(s, s') = \frac{1}{\rho} KL(s, \rho s' + (1 - \rho)s)$.
- **Prop. :** Pour toutes mesures de proba $sd\lambda$ et $td\lambda$ et tout $\rho \in (0, 1)$

$$C_\rho\, d^2_\lambda(s, t) \leq JKL_{\rho,\lambda}(s, t) \leq KL_\lambda(s, t)$$

avec $C_\rho = \frac{1}{\rho} \min(\frac{1-\rho}{\rho}, 1)\left(\ln\left(1 + \frac{\rho}{1-\rho}\right) - \rho\right)$.
De plus, si $\forall\omega \in \Omega, s(\omega) = 0 \implies t(\omega) = 0$

$$d^2_\lambda(s, t) \leq KL_\lambda(s, t) \leq \left(2 + \ln\left\|\frac{s}{t}\right\|_\infty\right) d^2_\lambda(s, t).$$

Densités conditionnelles

- Nécessité de s'adapter pour les densités conditionnelles :
  - Divergence sur la densité produit conditionnée au design (Kolaczyk, Bigot).
  - Principe de tensorisation et de passage à l'espérance sur le design :

$$KL \to KL^{\otimes n}(s, s') = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} KL\left(s(\cdot|X_i), s'(\cdot|X_i)\right)\right],$$

$$JKL_\rho \to JKL_\rho^{\otimes n} \quad \text{and} \quad d^2 \to d^{2\otimes n}.$$

- Approche similaire sauf pour Hellinger et la possibilité du passage à l'espérance sur le design dans l'inégalité oracle.
- Inégalité oracle de la forme

$$\mathbb{E}\left[JKL^{\otimes n}(s_0, \widehat{s}_{\widehat{m}})\right] \leq C \inf_{m\in\mathcal{M}}\left(\inf_{s_m\in S_m} KL^{\otimes n}(s_0, s_m) + \frac{\mathrm{pen}(m)}{n}\right) + \frac{C'}{n}.$$

- On retrouve exactement le théorème classique si $s(\cdot|X_i) = s(\cdot)$.
- Bon "scaling" de $JKL_\rho^{\otimes n}(s_{,0}\, \widehat{s}_{\widehat{m}})$ et $KL^{\otimes n}(s_0, s_m)$ avec n.
- Pb dans Bigot et al avec Hellinger : $\frac{1}{n} d^2(s_0, \widehat{s}_{\widehat{m}}) \leq \frac{2}{n}$!
- Pas ce soucis avec Bhattacharyya-Renyi de Kolaczyk et Barron mais pas de divergence "intégrée"...

Pénalité et complexité

- Pénalité liée à la complexité du modèle et de la collection.
- Complexité du modèle S_m (entropie) :
  - $H_{[\cdot],d^{\otimes n}}(\epsilon, S_m)$ entropie à crochet liée à la distance de Hellinger tensorisée $(d^{\otimes n} = \sqrt{d^{2\otimes n}} = \sqrt{\mathbb{E}\left[\frac{1}{n}\sum d^2(s(\cdot|X_i), s'(\cdot|X_i))\right]})$.
  - Hypothèse (H) : pour tout modèle S_m, il existe une fonction croissante $\phi_m(\delta)$ telle que $\delta \mapsto \frac{1}{\delta}\phi_m(\delta)$ soit décroissante sur $(0, +\infty)$ et telle que pour tout $\sigma \in \mathbb{R}^+$ et tout $s_m \in S_m$

$$\int_0^\sigma \sqrt{H_{[\cdot],d^{\otimes n}}\left(\epsilon, S_m(s_m, \sigma)\right)}\, d\epsilon \leq \phi_m(\sigma),$$

  - Complexité mesurée par $n\sigma_m^2$ avec σ_m l'unique racine de $\dfrac{1}{\sigma}\phi_m(\sigma) = \sqrt{n}\sigma$.
- Complexité de la collection (codage) :
  - complexité donnée par x_m satisfaisant Kraft $\displaystyle\sum_{m\in\mathcal{M}} e^{-x_m} \leq \Sigma < +\infty$
- Contrainte (classique) sur la pénalité

$$\mathrm{pen}(m) \geq \kappa\left(n\sigma_m^2 + x_m\right) \quad \text{avec } \kappa > \kappa_0.$$

Esquisse de preuve

- Preuve très proche de celle du théorème 7.11 du livre « Concentration Inequalities and Model Selection » de P. Massart.
- Pour toute fonction $g(x, y)$, on note $P_n^{\otimes n}(g)$ son processus empirique

$$P_n^{\otimes n}(g) = \frac{1}{n} \sum_{i=1}^{n} g(X_i, Y_i)$$

et $P^{\otimes n}(g)$ l'espérance de ce processus

$$P^{\otimes n}(g) = \mathbb{E}\left[P_n^{\otimes n}(g)\right] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} g(X_i, Y_i)\right].$$

et $\nu_n^{\otimes n}(g) = P_n^{\otimes n}(g) - P^{\otimes n}(g)$ le processus recentré.

- On note
  - $\widehat{s}_m = \operatorname{argmin}_{s_m \in S_m} P_n^{\otimes n}(-\ln s_m) = \operatorname{argmin}_{s_m \in S_m} P_n^{\otimes n}\left(-\ln \frac{s_m}{s_0}\right)$
  - $\overline{s}_m = \operatorname{argmin}_{s_m \in S_m} P^{\otimes n}\left(-\ln \frac{s_m}{s_0}\right) = \operatorname{argmin}_{s_m \in S_m} KL^{\otimes n}(s_0, s_m)$.
- On pose

$$\widehat{g}_m = -\ln\left(\frac{\widehat{s}_m}{s_0}\right) \quad \overline{g}_m = -\ln\left(\frac{\overline{s}_m}{s_0}\right) \quad \widehat{f}_m = -\frac{1}{\rho} \ln \frac{\rho \widehat{s}_m + (1-\rho) s_0}{s_0}$$

Majoration des « log-vraisemblances »

- Par convexité, $\widehat{f}_m = -\frac{1}{\rho} \ln \frac{\rho \widehat{s}_m + (1-\rho) s_0}{s_0} \leq -\ln \frac{\widehat{s}_m}{s_0} = \widehat{g}_m$
- Soit $m \in \mathcal{M}$, pour tout m' tel que

$$P_n^{\otimes n}(\widehat{g}_{m'}) + \frac{\mathrm{pen}(m')}{n} \leq P_n^{\otimes n}(\widehat{g}_m) + \frac{\mathrm{pen}(m)}{n} :$$

$$\begin{aligned} P_n^{\otimes n}(\widehat{f}_{m'}) + \frac{\mathrm{pen}(m')}{n} &\leq P_n^{\otimes n}(\widehat{g}_{m'}) + \frac{\mathrm{pen}(m')}{n} \\ &\leq P_n^{\otimes n}(\widehat{g}_m) + \frac{\mathrm{pen}(m)}{n} \\ &\leq P_n^{\otimes n}(\overline{g}_m) + \frac{\mathrm{pen}(m)}{n} \end{aligned}$$

- Soit

$$\begin{aligned} &P^{\otimes n}(\widehat{f}_{m'}) - \nu_n^{\otimes n}(\overline{g}_m) \\ &\qquad \leq P^{\otimes n}(\overline{g}_m) + \frac{\mathrm{pen}(m)}{n} - \nu_n^{\otimes n}(\widehat{f}_{m'}) - \frac{\mathrm{pen}(m')}{n} \end{aligned}$$

Inégalité oracle à déviation près

- L'inégalité précédente s'écrit

$$JKL_\rho^{\otimes n}(s_0, \hat{s}_{m'}) - \nu_n^{\otimes n}(\overline{g}_m) \leq KL^{\otimes n}(s_0, \overline{s}_m) + \frac{\mathrm{pen}(m)}{n} - \nu_n^{\otimes n}(\widehat{f}_{m'}) - \frac{\mathrm{pen}(m')}{n}$$

- On a fait apparaître
  - l'erreur intégré de l'estimateur dans le modèle m' : $JKL_\rho^{\otimes n}(s_0, \hat{s}_{m'})$
  - un processus centré et simple : $-\nu_n^{\otimes n}(\overline{g}_m)$,
  - l'oracle $KL^{\otimes n}(s_0, \overline{s}_m) + \frac{\mathrm{pen}(m)}{n}$
  - un « reste » aléatoire $-\nu_n^{\otimes n}(\widehat{f}_{m'}) - \frac{\mathrm{pen}(m')}{n}$
- On peut alors « contrôler les déviations » de $-\nu_n^{\otimes n}(\widehat{f}_{m'})$ par $\epsilon JKL_\rho^{\otimes n}(s_0, \hat{s}_{m'}) + \frac{\mathrm{pen}(m')}{n}$ dès que $\mathrm{pen}(m') \geq \kappa(n\sigma_{m'}^2 + x_m)$...

Déviation de $\nu_n^{\otimes n}(\widehat{f}_{m'})$

- Prélude : contrôle de $\nu_n^{\otimes n}(\widetilde{f}_{m'})$ avec $\widetilde{f}_{m'} = \frac{1}{\rho} \ln \frac{s_0}{\rho \widetilde{s}_{m'} + (1-\rho) s_0}$ et $\widetilde{s}_{m'} \in S_{m'}$ non aléatoire.
- $\nu_n^{\otimes n}(\widetilde{f}_{m'}) = \nu_n^{\otimes n}\left(\frac{1}{\rho} \ln \frac{s_0}{\rho \widetilde{s}_{m'} + (1-\rho) s_0} \right)$ facile à contrôler avec Bernstein ?
- Hypothèse typique pour Bernstein pour $\nu_n^{\otimes n}(\widetilde{f}_{m'})$: $\exists \sigma$ et b tels que

  pour tout entier $k \geq 2$, $\quad P^{\otimes n}(|\widetilde{f}_{m'}|^k) \leq \frac{k!}{2} \sigma^2 b^{k-2}$.

- Cas $\rho = 1$, $\widetilde{f}_{m'} = \ln \frac{s_0}{\widetilde{s}_{m'}}$: impossible à contrôler sans hypothèse sur le rapport des densités...

Jensen-Kullback et Hellinger

- **Lemme de van de Geer :** Pour toutes fonctions positives t, u et tout entier $k \geq 2$

$$P\left(\left|\ln\left(\frac{s_0+t}{s_0+u}\right)\right|^k\right) \leq \frac{k!}{2}\left(\frac{9\|\sqrt{t}-\sqrt{u}\|^2_{\lambda,2}}{8}\right)2^{k-2}.$$

- Apparition de Jensen-Kullback :

$$P^{\otimes n}\left(\left|\frac{1}{\rho}\ln\left(\frac{(1-\rho)s_0+\rho t}{(1-\rho)s_0+\rho u}\right)\right|^k\right) \leq \frac{k!}{2}\left(\frac{9d^{2\otimes n}(t,u)}{8\rho(1-\rho)}\right)\left(\frac{2}{\rho}\right)^{k-2}$$

- Soit pour $t = s_0$ et $u = \widetilde{s}_{m'}$, comme $\widetilde{f}_{m'} = \frac{1}{\rho}\ln\left(\frac{s_0}{(1-\rho)s_0+\rho\widetilde{s}_{m'}}\right)$,

$$P^{\otimes n}\left(\left|\widetilde{f}_{m'}\right|^k\right) \leq \frac{k!}{2}\left(\frac{9d^{2\otimes n}(s_0,\widetilde{s}_{m'})}{8\rho(1-\rho)}\right)\left(\frac{2}{\rho}\right)^{k-2}.$$

- Bernstein possible avec $\sigma = \sqrt{\frac{9d^{2\otimes n}(s_0,\widetilde{s}_{m'})}{8\rho(1-\rho)}}$ et $b = \frac{2}{\rho}$.

Déviation et « log-vraisemblance »

- On pose $\widetilde{s}_{m'} = \operatorname{argmin}_{s \in S_{m'}} d^{2\otimes n}(s_0, s_{m'})$.
- $-\nu_n^{\otimes n}(\widehat{f}_{m'}) = \left(-\nu_n^{\otimes n}(\widehat{f}_{m'}) + \nu_n^{\otimes n}(\widetilde{f}_{m'})\right) - \nu_n^{\otimes n}(\widetilde{f}_{m'})$
- Reste à contrôler

$$-\nu_n^{\otimes n}(\widehat{f}_{m'}) + \nu_n^{\otimes n}(\widetilde{f}_{m'}) = Z(\widehat{s}_{m'}) - Z(\widetilde{d}_{m'})$$

où $Z(s) = \nu_n^{\otimes n}\left(\frac{1}{\rho} \ln \frac{s_0}{\rho s + (1-\rho)s_0}\right)$.

- Première idée :

$$Z(\widehat{s}_{m'}) - Z(\widetilde{d}_{m'}) \leq \sup_{s_{m'} \in S_{m'}} Z(s_{m'}) - Z(\widetilde{s}_{m'}).$$

- Beaucoup trop grossier !
- Localisation pour mettre un peu de $JKL_\rho^{\otimes n}(s_0, s_{m'})$.

Épluchage et déviations locales

- En posant $S_{m'}(\widetilde{s}_{m'}, \sigma) = \{s_{m'} \in S_{m'} | d^{2\otimes n}(\widetilde{s}_{m'}, s_{m'}) \leq \sigma^2\}$, si

$$\mathbb{E}\left[\sup_{s_{m'} \in S_{m'}(\widetilde{s}_{m'}, \sigma)} Z(s_{m'}) - Z(\widetilde{s}_{m'})\right] \leq \psi(\sigma), \quad \text{pour tout } \sigma \geq \sigma_\star \geq 0,$$

avec $\psi(x)/x$ décroissante sur $\mathbb{R}^+$ alors $\forall x \geq \sigma_\star$

$$\mathbb{E}\left[\sup_{s_{m'} \in S_{m'}} \frac{Z(s_{m'}) - Z(\widetilde{s}_{m'})}{x^2 + d^{2\otimes n}(\widetilde{s}_{m'}, s_{m'}}\right] \leq 4x^{-2}\psi(x).$$

- Contrôle de

$$\mathbb{E}\left[\sup_{s_{m'} \in S_{m'}(\widetilde{s}_{indm'}, \sigma)} Z(s_{m'}) - Z(\widetilde{s}_{m'})\right] = \mathbb{E}\left[\sup_{s_{m'} \in S_{m'}(\widetilde{s}_{indm'}, \sigma)} W(s_{m'})\right]$$

avec

$$W(s_{m'}) = \nu_n^{\otimes n}\left(\frac{1}{\rho} \ln \frac{\rho \widetilde{s}_{m'} + (1-\rho)s_0}{\rho s_{m'} + (1-\rho)s_0}\right).$$

- Outil : Théorème 6.8 du livre de P. Massart...

Déviations et entropie

- **Théorème :** Soit $\mathcal{F}$ une famille dénombrable de fonctions à valeur réelles. Si ils existent des nombres positifs σ et b tels que pour tout $f \in \mathcal{F}$ et tout entier $k \geq 2$

$$P^{\otimes_n}(|f|^k) \leq \frac{k!}{2}\sigma^2 b^{k-2}$$

et que de plus pour tout nombre positif δ il existe un ensemble $B(\delta)$ de crochets recouvrant $\mathcal{F}$ tel que pour tout crochet $[g^-, g^+] \in B(\delta)$ et tout entier $k \geq 2$

$$P^{\otimes_n}(|g^+ - g^-|^k) \leq \frac{k!}{2}\sigma^2 b^{k-2}$$

Soit $e^{H(\delta)}$ le cardinal de ce recouvrement, il existe un constante absolue κ telle pour tout $\epsilon \in (0, 1]$ et tout ensemble mesurable A avec $P[A] > 0$, on a

$$\mathbb{E}^A\left[\sup_{f\in\mathcal{F}} \nu_n^{\otimes_n}(f)\right] \leq E + \frac{(1+6\epsilon)\sigma}{\sqrt{n}}\sqrt{2\ln\left(\frac{1}{\mathbb{P}\{A\}}\right)} + \frac{2b}{n}\ln\left(\frac{1}{\mathbb{P}\{A\}}\right)$$

où

$$E = \frac{\kappa}{\epsilon}\frac{1}{\sqrt{n}}\int_0^{\epsilon\sigma}\sqrt{H(u)\wedge n}\mathrm{d}u + \frac{2(b+\sigma)}{n}H(\sigma)$$

avec $\kappa \leq 27$.

Jensen-KL et entropie à crochet

- Contrôle de $\sup \nu_n^{\otimes n}(f)$ pour $f \in \mathcal{F}$ sous deux conditions
  - Hypothèse nécessaire (c.f. Bernstein) : $\exists \sigma$ et b tels que pour tout $f \in \mathcal{F}$

    et tout entier $k \geq 2$, $\quad P^{\otimes n}(|f|^k) \leq \frac{k!}{2}\sigma^2 b^{k-2}$.

  - Hypothèse d'entropie à crochet sur $\mathcal{F}$: pour tout σ, existence d'un recouvrement par des crochets $[g^-, g^+]$ de cardinal $H(\sigma)$ tel que pour

    tout entier $k \geq 2$, $\quad P^{\otimes n}(|g^+ - g^-|^k) \leq \frac{k!}{2}\sigma^2 b^{k-2}$.

- **Lemme de van de Geer :**

$$P^{\otimes n}\left(\left|\frac{1}{\rho}\ln\left(\frac{(1-\rho)s_0 + \rho t}{(1-\rho)s_0 + \rho u}\right)\right|^k\right) \leq \frac{k!}{2}\left(\frac{9d^{2\otimes n}(t,u)}{8\rho(1-\rho)}\right)\left(\frac{2}{\rho}\right)^{k-2}$$

- Importance d'utiliser la distance de Jensen-KL $JKL^{\otimes n}$.
- Condition d'entropie à crochet sur $S_{m'}(\widetilde{s}_{m'}, \sigma)$ par rapport à la distance de Hellinger tensorisée $d^{\otimes n}$!

Fin de la preuve

- Application du théorème :

$$\mathbb{E}^A\left[W_m(\sigma)\right] \leq E + \frac{(1+6\epsilon)3\sigma}{2\sqrt{2\rho(1-\rho)}\sqrt{n}}\sqrt{\ln\left(\frac{1}{\mathbb{P}\{A\}}\right)} + \frac{4}{\rho n}\ln\left(\frac{1}{\mathbb{P}\{A\}}\right)$$

avec

$$E = \frac{\kappa}{\epsilon}\frac{1}{\sqrt{n}}\int_0^{\epsilon\frac{3\sigma}{2\sqrt{2(\rho(1-\rho)}}}\sqrt{H_{[\cdot],d^{\otimes n}}\left(u, S_m(\widetilde{s}_m,\sigma)\right)\wedge n}\mathrm{d}u + \frac{2(\frac{2}{\rho}+\frac{3\sigma}{2\sqrt{2\rho(1-\rho)}})}{n}H_{[\cdot],d^{\otimes n}}\left(\sigma, S_m(\widetilde{s}_m,\sigma)\right)$$

- En utilisant les déf. de $\phi_{m'}$ et $\sigma_{m'}$, on obtient $\forall \sigma \geq \sigma_{m'}$

$$\mathbb{E}^A\left[W_m(\sigma)\right] \leq \kappa_1''\frac{\phi_m(\sigma)}{\sqrt{n}} + \frac{\kappa_2''\sigma}{\sqrt{n}}\sqrt{\ln\left(\frac{1}{\mathbb{P}\{A\}}\right)} + \frac{4}{\rho n}\ln\left(\frac{1}{\mathbb{P}\{A\}}\right).$$

- Retour à un contrôle en proba à la Berstein possible :

$$\forall A, P(A) > 0,\ \mathbb{E}^A\left[Z\right] \leq \Psi\left(\ln\left(\frac{1}{\mathbb{P}\{A\}}\right)\right) \implies \forall x > 0, \mathbb{P}\left[Z \geq \Psi(x)\right] \leq e^{-x}.$$

- Preuve sans soucis ensuite en utilisant en plus
  - $C_\rho d^{2\otimes n}(s_0, \widehat{s}_{m'}) \leq JKL_\rho^{\otimes n}(s_0, \widehat{s}_{m'})$
  - une borne d'union sur tous les modèles exploitant l'inégalité de Kraft.

Retour vers les modèles de mélanges spatiaux

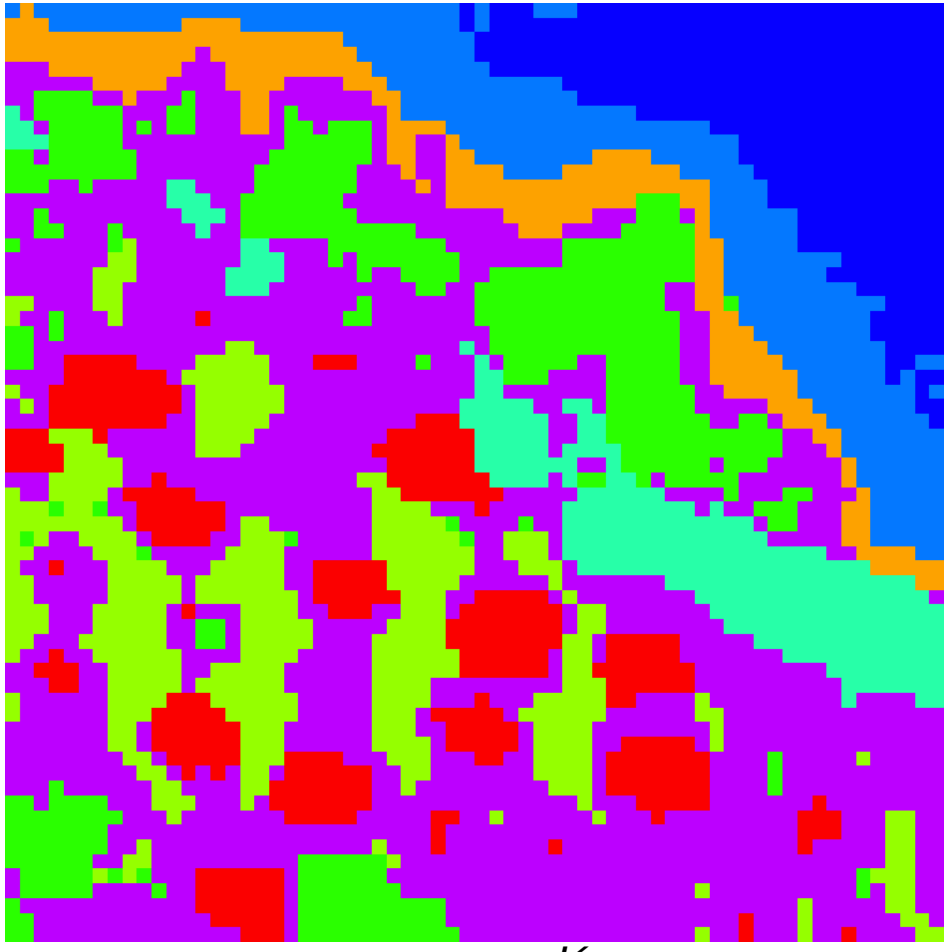
- Contrôle de $H_{[\cdot],d^{\otimes n}}(\epsilon, S_m(s_m,\sigma))$ pour les modèles de mélanges spatiaux (cf Maugis et Michel) :
  - contrôle d'un majorant de l'entropie : $H_{[\cdot],d^{\sup}}(\epsilon, S_m)$ où $d^{\sup} = \sqrt{d^{2\sup}} = \sqrt{\sup_x d^2(s(\cdot|x), s'(\cdot|x))}$,
  - résultat valide pour toutes les classes de mélanges ($[\mu\, L\, D\, A]^K$) et toutes les partitions :
    $$H_{[\cdot],d^{\sup}}(\epsilon, S_m) \leq \dim(S_m)(C + \ln\frac{1}{\epsilon})$$
    avec C presque explicite (utilisation d'un lemme de Szarek sur l'entropie de $SO(n)$ sans constante explicite...) et $\dim(S_m) = |\mathcal{P}|(K-1) + \dim([\mu\, L\, D\, A]^K)$.
  - implication : $n\sigma_m^2 \leq \kappa' \left(C' + \frac{1}{2}\left(\ln\left(\frac{n}{C'\dim(S_m)}\right)\right)_+ \right) \dim(S_m)$.
- Codage de la collection avec $x_m \leq \kappa''|\mathcal{P}| \leq \frac{\kappa''}{K-1}\dim(S_m)$.
- Condition sur la pénalité :

$$\mathrm{pen}(m) \geq \left(\kappa' \left(C' + \frac{1}{2}\left(\ln\left(\frac{n}{C'\dim(S_m)}\right)\right)_+ \right) + \frac{\kappa''}{K-1} \right) \dim(S_m).$$

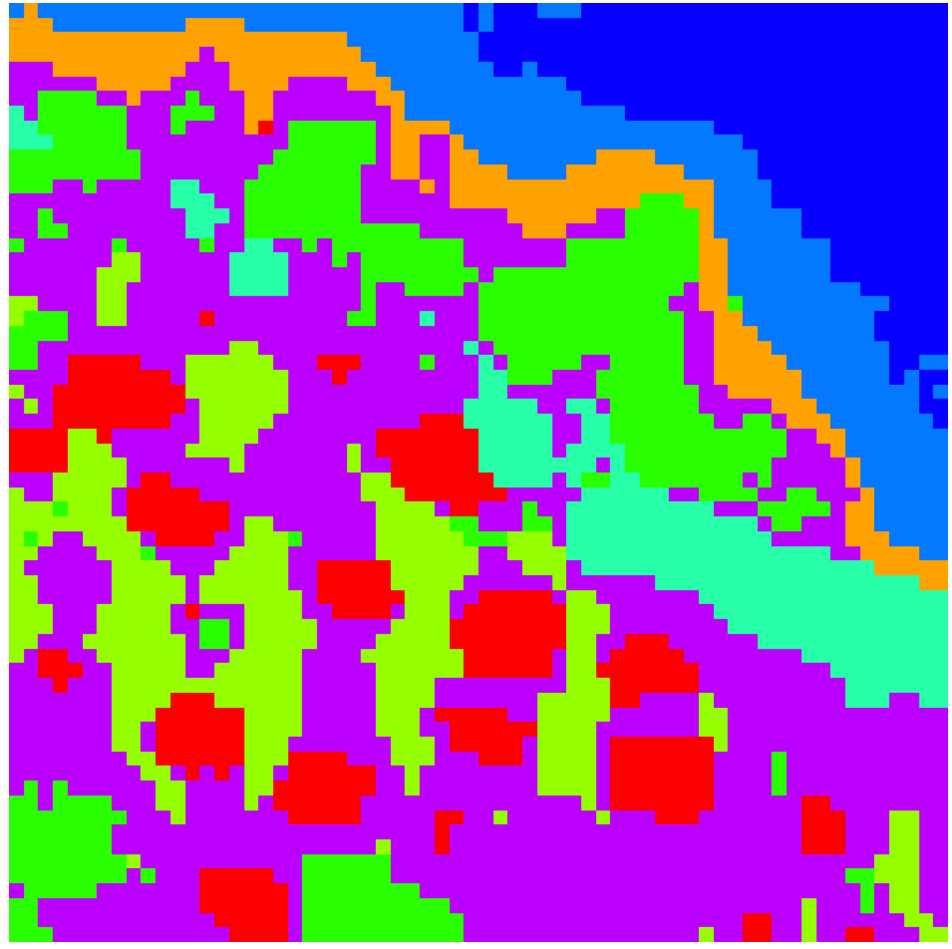
Segmentation automatique

- Résultat numérique selon la prise en compte du caractère spatial :

Sans



Avec

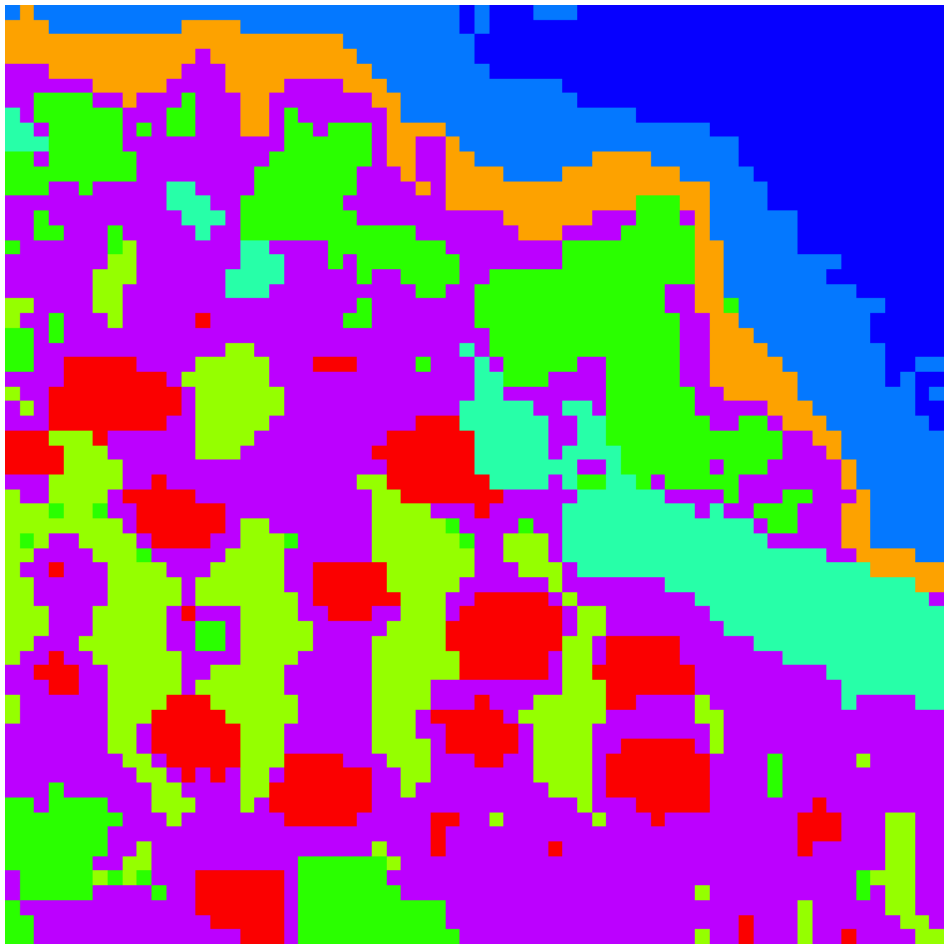


- $K = 8$, $[L_k\, D\, A_k]^K$ et partition optimale.
- Calibration de la pénalité par heuristique de pente.
- Réduction de dimension par simple ACP...

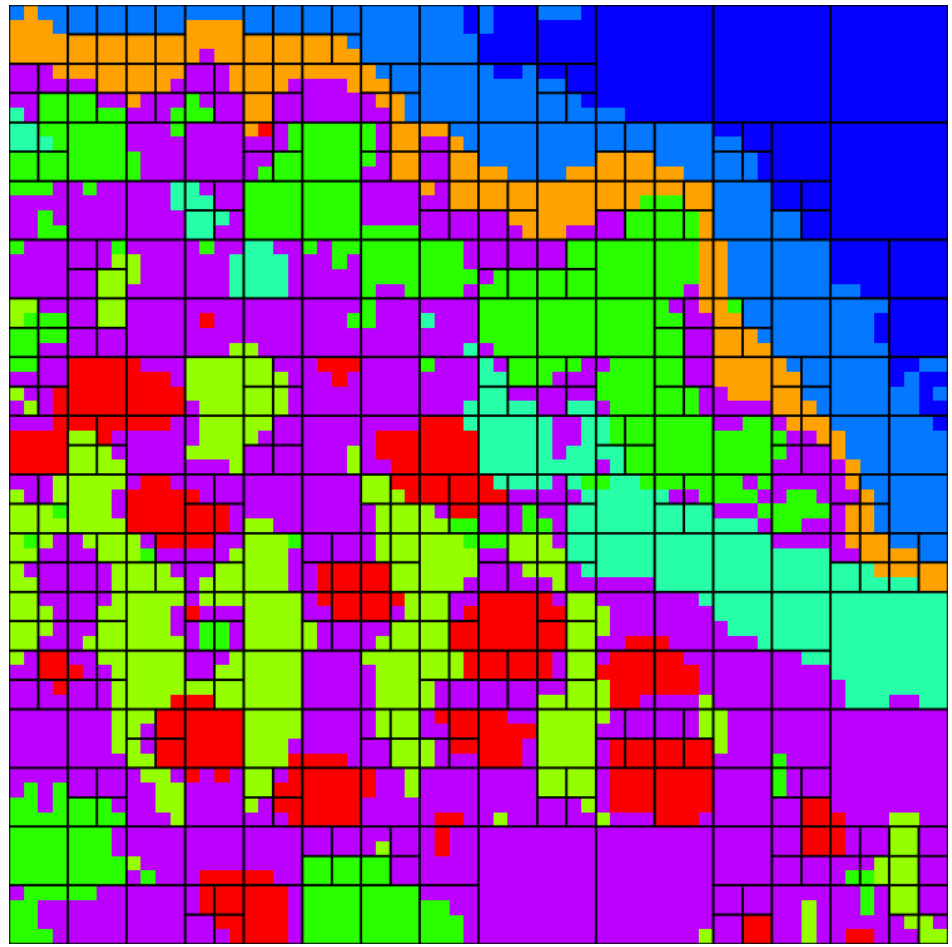
Segmentation automatique

- Résultat numérique selon la prise en compte du caractère spatial :

Sans

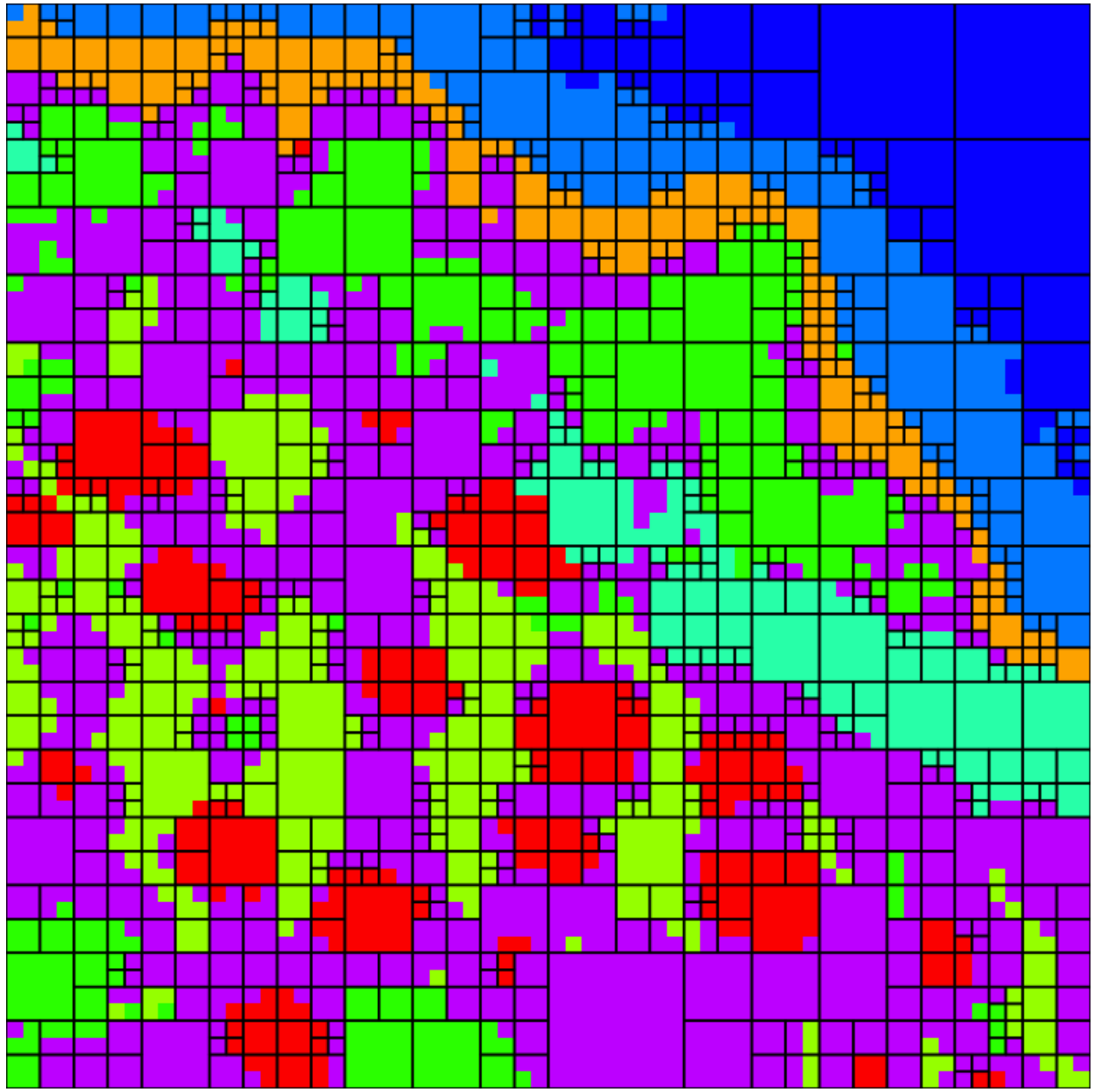


Avec

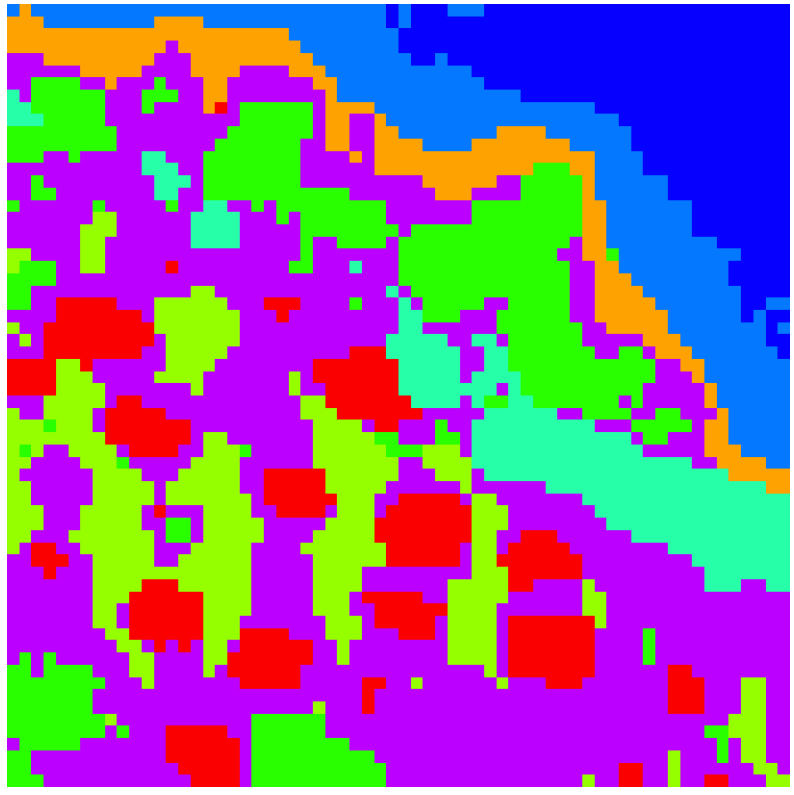


- $K = 8$, $[L_k\, D\, A_k]^K$ et partition optimale.
- Calibration de la pénalité par heuristique de pente.
- Réduction de dimension par simple ACP...

Segmentation et classification



Le secret de Stradivarius



- Deux couches fines de vernis :
  - une première couche d'huile simple, similaire à celle des peintres, pénétrant légèrement le bois,
  - une seconde d'un mélange huile, résine de pin, pigments donnant cette couleur rouge caractéristique.
- Technique classique pour l'époque.
- Le secret de Stradivarius n'est pas dans le vernis !

Conclusion

- Cadre :
  - Problème de segmentation non supervisée.
  - Estimateur de densités conditionnelles par maximum de vraisemblance et pénalisation.
- Résultats :
  - Garantie théorique pour l'estimation de densités avec des distances « tensorisées ».
  - Applicable au problème de segmentation.
  - Algorithme efficace de minimisation.
- Perspectives :
  - Applications à d'autres cas.
  - Lien entre l'estimation de densités conditionnelles et les performances de segmentation.
  - Calibration par heuristique de pente des deux problèmes.
  - Réduction de dimension adaptée à la segmentation/classification supervisée ou non.

Bibliographie

- Inspiration initiale présentant une solution complète :
  - A. Antoniadis, J. Bigot and R. von Sachs (2008). *A multiscale approach for statistical characterization of functional images.* Journal of Computational and Graphical Statistics, 18(1), 216–237
- Segmentation et sélection de modèles :
  - E.D. Kolaczyk, J. Ju and S. Gopal (2005). *Multiscale, multigranular statistical image segmentation.* Journal of the American Statistical Association, 100, 1358–1369.
  - E.D. Kolaczyk and R.D. Nowak (2004). *Multiscale likelihood analysis and complexity penalized estimation.* Annals of Statistics, 32, 500–527
- Sélection de modèles par MDL :
  - A.R. Barron, C. Huang, J. Q. Li and Xi Luo (2008). *MDL Principle, Penalized Likelihood, and Statistical Risk.* In Feschrift for Jorma Rissanen. Presented to Rissanen Nov. 2007.
- Sélection de modèles et entropie (à crochet) :
  - P. Massart (2003). *Concentration inequalities and model selection* : Ecole d'Eté de Probabilités de Saint-Flour XXXIII.
  - C. Maugis and B. Michel. (2009). *A non asymptotic penalized criterion for Gaussian mixture model selection.* Accepté à ESAIM : P&S (et l'erratum)
  - S. Cohen and E. Le Pennec (201 ?)
- Entropie de $SO(n)$:
  - J. Szarek (1998). *Metric entropy of homogeneous spaces.* Banach Center Pub., 43, 395–410
- Sélection de modèle et densité conditionnelle :
  - E. Brunel, C. Lacour et F. Comte (2007). *Adaptive estimation of the conditional density in presence of censoring.* Sankhya, 69(4), 734–763