
Bases du traitement des images



Classification d'image



Dufrenoix Franck
LISIC

Université du littoral

Classification d'images

Introduction

- **Objectif:** partitionner l'image en groupe de pixels ayant des caractéristiques communes:
 - Couleurs
 - Textures
 - Formes
- **Un cadre mathématique basé sur:** mab
Lois de probabilité - la distance - la densité
- **Avons-nous des informations au préalable?**
 - Classification supervisée
 - Classification non supervisée

Classification Bayesienne

Ingrédients?

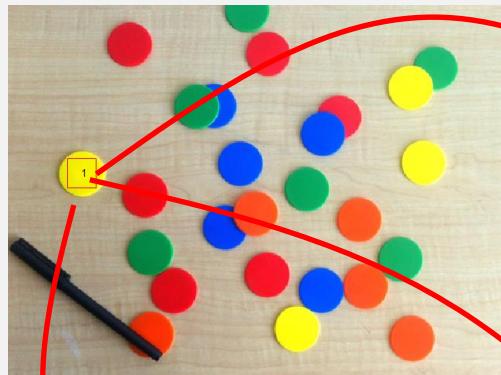


- L'image: I
- Un pixel: $x \in \{0,255\}^3$
- 7 Classes: F, S, R,O,V,J,B

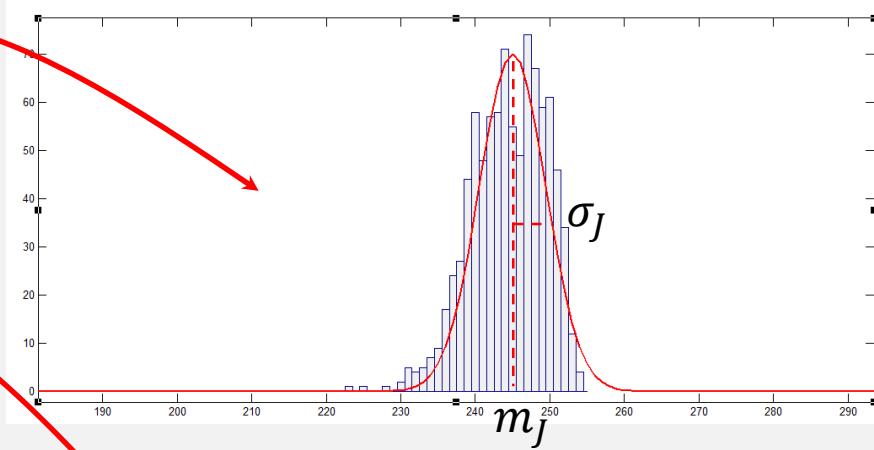
Question: Quel est la probabilité que le pixel x appartienne à la classe 'F', 'S', 'R',....?

Classification Bayesienne

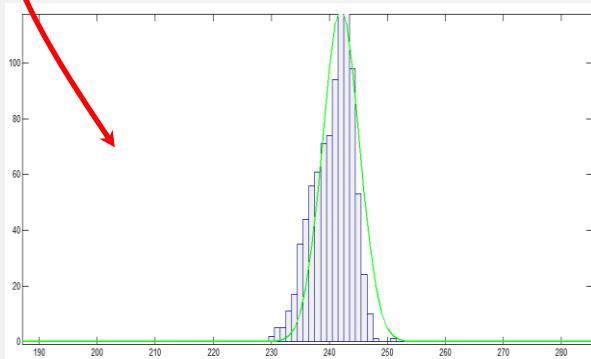
Connaissance à priori?



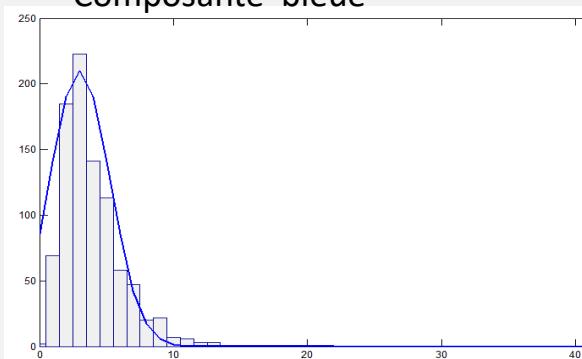
Composante rouge



Composante verte



Composante bleue



Classification Bayesienne

Connaissance à priori?

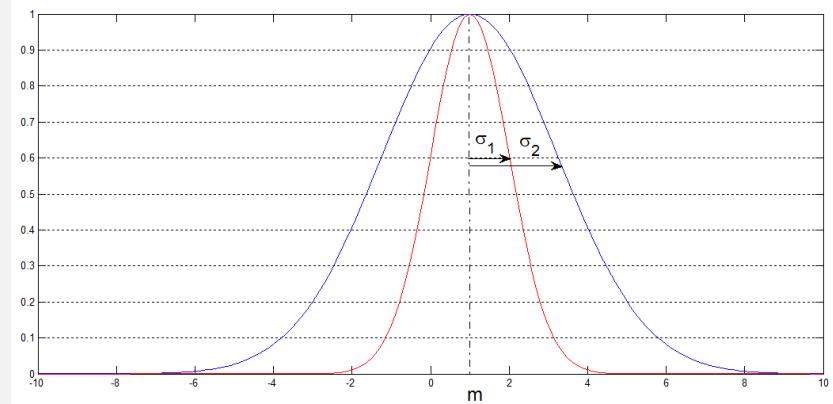
Loi de probabilité à priori: On peut définir une probabilité à priori que x appartienne à la classe $C=\{F,S,R,O,J,V,B\}=\{C_i\}_{i=1..7}$

$$P(x/C_i)$$

En pratique on suppose que la loi de probabilité qui régit x est une loi normale ou gaussienne: 2 paramètres (m :moyenne, σ : ecart type)

A une dimension:

$$P(x/C_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-m_i)^2}{2\sigma^2}}$$



Classification Bayesienne

Connaissance à priori?

Loi Normale à plusieurs dimensions

- C'est le cas pour une image couleur, on a 3 composantes (R,G,B) : la dimension est 3
- Généralisation de la loi gaussienne
- $x = (r, g, b)$,
- La classe Ci est définie par:
 - Sa moyenne: $\mathbf{m}_i = (\mathbf{m}_i^r, \mathbf{m}_i^g, \mathbf{m}_i^b)$
 - Sa matrice de variance-covariance: $\mathbf{E}_i = \begin{bmatrix} (\sigma_i^r)^2 & \sigma_i^r \sigma_i^g & \sigma_i^r \sigma_i^b \\ \sigma_i^g \sigma_i^r & (\sigma_i^g)^2 & \sigma_i^g \sigma_i^b \\ \sigma_i^b \sigma_i^r & \sigma_i^b \sigma_i^g & (\sigma_i^b)^2 \end{bmatrix}$

Classification Bayesienne

Connaissance à priori?

Loi Normale à plusieurs dimensions:

$$P(x/c_i) = \frac{1}{(2\pi)^{3/2} |E_i|^{1/2}} \exp \left[-\frac{1}{2} (x - m_i) E_i^{-1} (x - m_i)^T \right]$$

$|E_i|$: déterminant de la matrice de variance-covariance

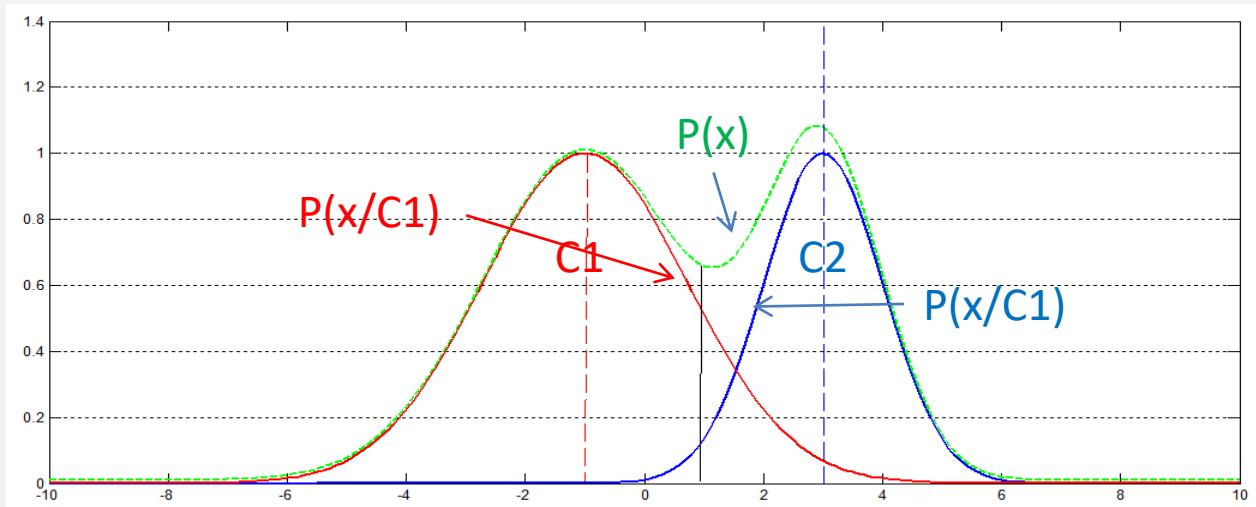
Classification Bayesienne

Règle de Bayes

- **Règle de Bayes:** $P\left(\frac{C_i}{x}\right)P(x) = P\left(\frac{x}{C_i}\right)P(Ci)$
 - $P(Ci)$: Probabilité à priori de la classe Ci
 - $P\left(\frac{x}{Ci}\right)$: Probabilité à priori: sachant les caractéristiques de la classe Ci, quelle est la probabilité que x appartient à Ci
 - $P\left(\frac{Ci}{x}\right)$: Probabilité à posteriori: sachant x, quelle est la probabilité qu'il appartient à Ci → **c'est ce qu'on recherche**
 - $P(x)$: probabilité de x:

Classification Bayesienne

Règle de décision: cas à 2 classes



La décision qu'un pixel x appartienne à la classe C_1 ou C_2 , va dépendre des valeurs prises par les probabilités à posteriori $P(\frac{C_1}{x})$ et $P(\frac{C_2}{x})$. La loi de décision sera basée sur le maximum à posteriori (MAP):

Règle de décision:

$$\frac{P(\frac{C_1}{x})}{P(\frac{C_2}{x})} > \begin{cases} \frac{P(C_2)}{P(C_1)} & x \in C_1 \\ \frac{P(C_1)}{P(C_2)} & x \in C_2 \end{cases}$$

Classification Bayesienne

Règle de décision

- **Règle de décision pour K classes**

$$\frac{P(\frac{C_i}{x})}{P(\frac{C_j}{x})} > \frac{P(C_j)}{P(C_i)} \quad \forall j = \{1 \dots k\} \setminus \{i\}$$

- **Règle de Bayes:** $P\left(\frac{C_i}{x}\right) P(x) = P\left(\frac{x}{C_i}\right) P(C_i)$

- **Simplification:**

On suppose que les classes sont équiprobables $P(C_1)=P(C_2)=\dots=P(C_k)$, et en s'appuyant sur le théoreme de Bayes, on obtient:

$$P\left(\frac{x}{C_i}\right) > P\left(\frac{x}{C_j}\right) \quad \forall j = \{1 \dots k\} \setminus \{i\}$$

Classification Bayesienne

Règle de décision

$$\frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-m_i)^2}{2\sigma^2}} \stackrel{x \in C_i}{>} \stackrel{x \in C_j}{<} \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(x-m_j)^2}{2\sigma^2}}$$

On prend le logarithme:

$$-\log(\sqrt{2\pi}\sigma_i) - \frac{(x-m_i)^2}{2\sigma^2} \stackrel{x \in C_i}{>} \stackrel{x \in C_j}{<} -\log(\sqrt{2\pi}\sigma_j) - \frac{(x-m_j)^2}{2\sigma^2}$$

On pose $D_i(x) = \log(\sqrt{2\pi}\sigma_i) + \frac{(x-m_i)^2}{2\sigma^2}$

Classification Bayesienne

Règle de décision

- **Règle de décision:** on décide que x appartient à la classe I si

$$l = \operatorname{argmin}_{i=1 \dots k} (D_i(x))$$

Avec $D_i(x) = \log(\sqrt{2\pi}\sigma_i) + \frac{(x - m_i)^2}{2\sigma^2}$

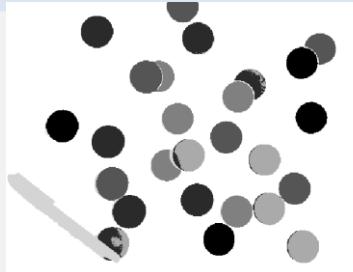
- **Algorithme**

Pour chaque pixel x de l'image, calculer: $D_1(x), D_2(x), \dots, D_K(x)$

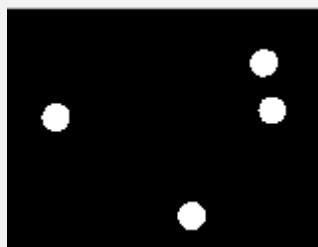
Le minimum de ces valeurs fournira la classe prédictive de x

Classification Bayesienne

Résultats



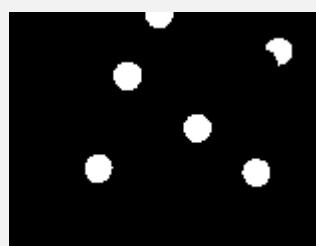
Etiquetage des pixels: 0,1,...,6



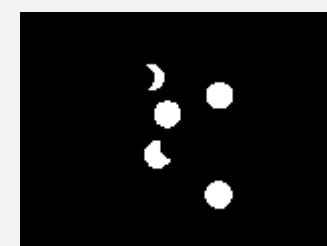
$R==0$



$R==1$



$R==2$



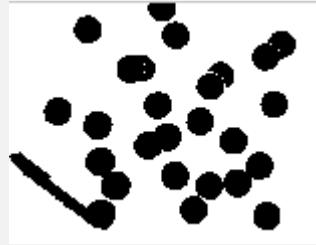
$R==3$



$R==4$



$R==5$



$R==6$

Classification Bayesienne

Frontière de décision

- La règle de décision bayésienne permet de déterminer la surface de séparation entre 2 classes, c'est-à-dire le lieu des points où les probabilités d'appartenir aux deux classes sont égales

$$P(x/c_1) = \frac{1}{(2\pi)^{d/2}|E_1|^{1/2}} \exp \left[-\frac{1}{2}(x - m_1)E_1^{-1}(x - m_1)^T \right]$$

$$P(x/c_2) = \frac{1}{(2\pi)^{d/2}|E_2|^{1/2}} \exp \left[-\frac{1}{2}(x - m_2)E_2^{-1}(x - m_2)^T \right]$$

$$P(x/c_2) = P(x/c_1)$$



$$x^T A x + x^T b + c = 0$$

A: matrice
b: vecteur
c: constante

Classification Bayesienne

Frontière de décision

- Calcul de la moyenne m_i et la matrice de covariance E_i de la classe Ci (ni observations)

$$m_i = \frac{1}{n_i} \sum_{k=1}^{n_i} x_k \quad E_i = \frac{1}{n_i} \sum_{k=1}^{n_i} (x_k - m_i)(x_k - m_i)^T$$

- Calcul de l'inverse de matrice de covariance E_i de la classe Ci (ni observations)

$$A^{-1} = \frac{1}{\det A} {}^t \text{com } A. \quad (\text{com } A)_{i,j} := \det \left(A'_{i,j} \right) = (-1)^{i+j} \det (A_{i,j}),$$

- $A'_{i,j}$ est la sous-matrice carrée de taille $n - 1$ déduite de A en supprimant la i -ème ligne et la j -ème colonne

- Exemple matrice 2*2

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \text{com } A = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}, \quad {}^t \text{com } A = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

$$A^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Classification Bayesienne

Exercice

Considérons l'ensemble d'apprentissage suivant:

$$S = \left\{ \left(\begin{pmatrix} 0 \\ 4 \end{pmatrix}, c_1 \right), \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, c_1 \right), \left(\begin{pmatrix} 3 \\ 3 \end{pmatrix}, c_1 \right), \left(\begin{pmatrix} 4 \\ 0 \end{pmatrix}, c_1 \right), \left(\begin{pmatrix} 4 \\ 0 \end{pmatrix}, c_2 \right), \left(\begin{pmatrix} 7 \\ 1 \end{pmatrix}, c_2 \right), \left(\begin{pmatrix} 8 \\ 4 \end{pmatrix}, c_2 \right), \left(\begin{pmatrix} 5 \\ 3 \end{pmatrix}, c_2 \right) \right\}$$

- 1) Représenter ces classes sur un graphique
- 2) Calculer les moyennes
- 3) Calculer les matrices de variances covariance
- 4) Déterminer la surface séparatrice entre les deux classes
- 5) A quelle classe appartient le point (3,0) et le point (5,-1)?

Classification Bayesienne

Exercice

$$\widehat{\boldsymbol{\mu}}_1 = \frac{1}{4} \sum_{j=1}^4 \mathbf{x}_{1j} = \frac{1}{4} \begin{pmatrix} 0+1+3+4 \\ 4+1+3+0 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$\widehat{Q}_1 = \frac{1}{4} \sum_{j=1}^4 (\mathbf{x}_{1j} - \widehat{\boldsymbol{\mu}}_1)(\mathbf{x}_{1j} - \widehat{\boldsymbol{\mu}}_1)^T$$

On trouve au total :

$$\widehat{Q}_1 = \frac{1}{4} \left(\begin{pmatrix} 4 & -4 \\ -4 & 4 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 4 & -4 \\ -4 & 4 \end{pmatrix} \right) = \begin{pmatrix} 5/2 & -3/2 \\ -3/2 & 5/2 \end{pmatrix}$$

$$\text{D'où : } |\widehat{Q}_1| = 4 \text{ et } \widehat{Q}_1^{-1} = \begin{pmatrix} 5/8 & 3/8 \\ 3/8 & 5/8 \end{pmatrix}$$

De même, on trouve :

$$\widehat{\boldsymbol{\mu}}_2 = \begin{pmatrix} 6 \\ 2 \end{pmatrix}, \widehat{Q}_2 = \begin{pmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{pmatrix}, |\widehat{Q}_2| = 4 \text{ et } \widehat{Q}_2^{-1} = \begin{pmatrix} 5/8 & -3/8 \\ -3/8 & 5/8 \end{pmatrix}$$

Les deux classes ont le même déterminant, donc la surface est définie par

$$(\mathbf{x} - \widehat{\boldsymbol{\mu}}_1)^T \widehat{Q}_1^{-1} (\mathbf{x} - \widehat{\boldsymbol{\mu}}_1) = (\mathbf{x} - \widehat{\boldsymbol{\mu}}_2)^T \widehat{Q}_2^{-1} (\mathbf{x} - \widehat{\boldsymbol{\mu}}_2)$$

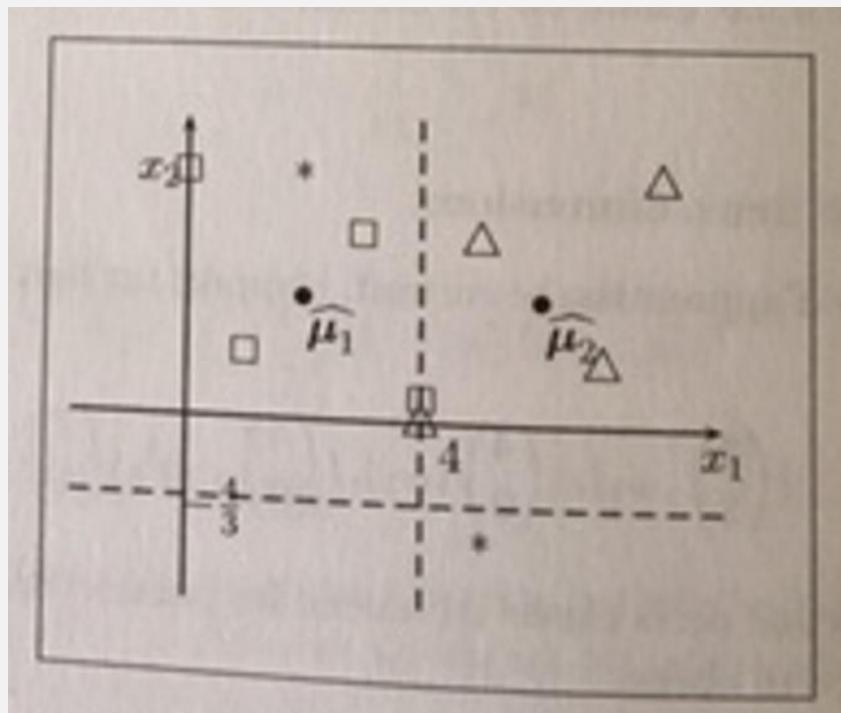
$$(x_1 - 2x_2 - 2) \begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} x_1 - 2 \\ x_2 - 2 \end{pmatrix} = (x_1 - 6x_2 - 2) \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix} \begin{pmatrix} x_1 - 6 \\ x_2 - 2 \end{pmatrix}$$

$$(x_1 - 4)(x_2 + 4/3) = 0$$

Classification Bayesienne

Exercice

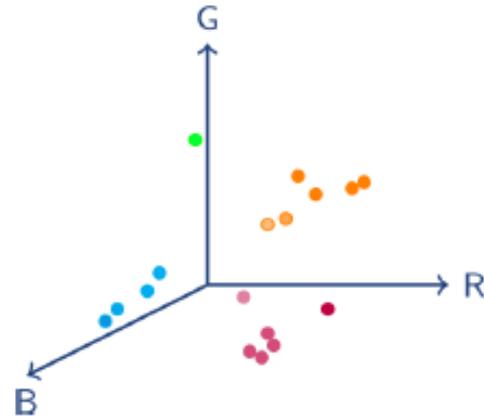
La surface séparatrice est une hyperbole dégénérée en 2 droites et partageant le plan en 4 zones: 2 zones pour C1 et 2 zones pour C2



Clustering: k-moyenne

Introduction

Un pixel est maintenant représenté par un vecteur à valeurs dans $\mathbb{R}^B \Rightarrow$ travailler directement dans l'espace \mathbb{R}^B !



Le principe des méthodes de classification (ou plus exactement de coalescence) (*clustering*) est de regrouper les pixels en groupes homogènes.

Les k-moyennes

Illustration par l'image

Algorithme assez simple pour détecter plusieurs classes (≥ 2) dans une image
Considérons par exemple une image couleur contenant 3 classes , chaque plan contient n pixels (x_1, \dots, x_n), chaque pixel x_i contient d = 3 valeurs (RGB).



K-moyennes

Projections dans le plan RGB

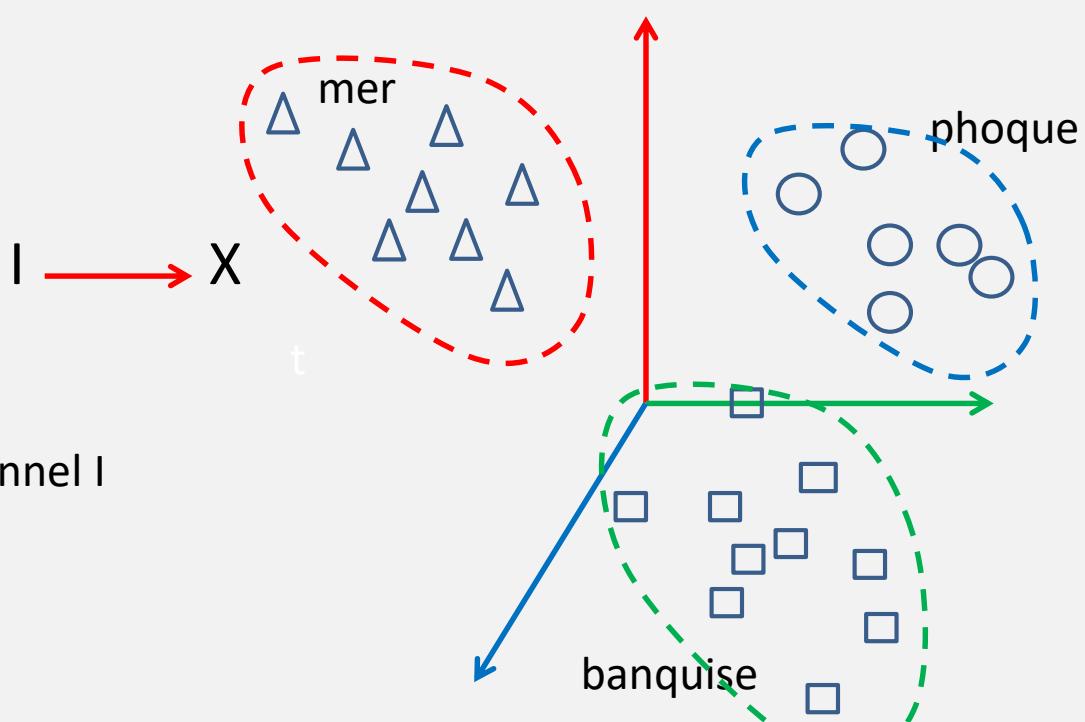


Image: tableau tridimensionnel I

$R=I(:,:,1);$

$G=I(:,:,2);$

$B=I(:,:,3);$

$X=[R(:),G(:),B(:)];$

K-moyennes

principes

Initialisation:

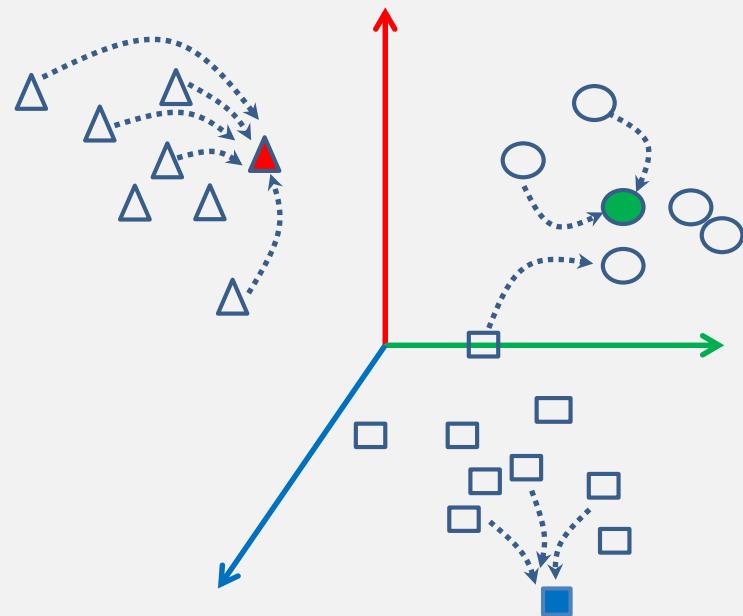
On définit le nombre de classes à détecter: **K**

On tire aléatoirement K données dans le tableau:
exemple des données en couleur (K=3).

Ces points deviennent le centre des classes.

Répétitions:

Tous les autres points sont affectés au centre le plus proche

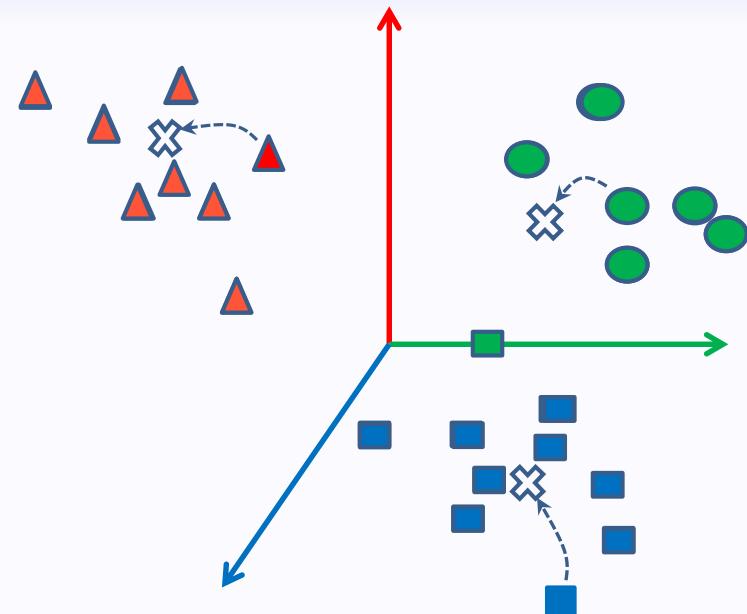


K-moyennes

principes

Répétitions:

- Tous les autres points sont affectés au centre le plus proche
- On rajuste les centres des classes en calculant le centre de chaque regroupement.
- On répète l'opération jusqu'à stabilisation des centres.



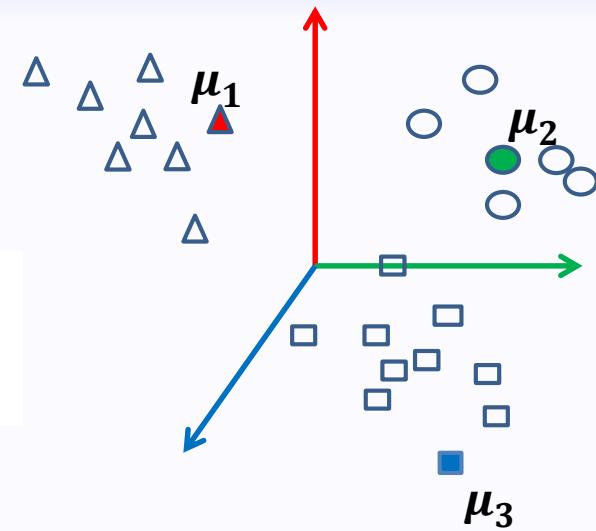
K-moyennes

Formellement

Soit un ensemble de données (x_1, \dots, x_n),
K-means minimise le critère d'erreur
(distorsion) suivant:

$$\mathcal{J}(\mu_1, \dots, \mu_K, z) = \sum_{k=1}^K \sum_{i=1}^n z_{ik} \|x_i - \mu_k\|^2$$

$$z_{ik} = \begin{cases} 1 & \text{si } i \in \text{classe } k \\ 0 & \text{sinon} \end{cases}$$



$$\|x_i - \mu_k\| = d(x_i, \mu_k) = \sqrt{\sum_{j=1}^d (x_{ij} - \mu_{kj})^2}$$

K-moyennes

Les étapes clés

Initialisation : On initialise les centres des classes $(\mu_1^{(0)}, \dots, \mu_K^{(0)})$ (à votre choix) pour donner le pas de départ de l'algorithme (par exemple on choisissant aléatoirement des centres "virtuels", ou K données parmi les données à traiter). Il s'agit donc de démarrer à l'itération $t = 0$ avec des valeurs initiales pour les paramètres du modèle $(\mu_1^{(0)}, \dots, \mu_K^{(0)})$.

Etape d'affectation (classification) : Chaque donnée est assignée à la classe du centre dont elle est la plus proche : $\forall i = 1, \dots, n$

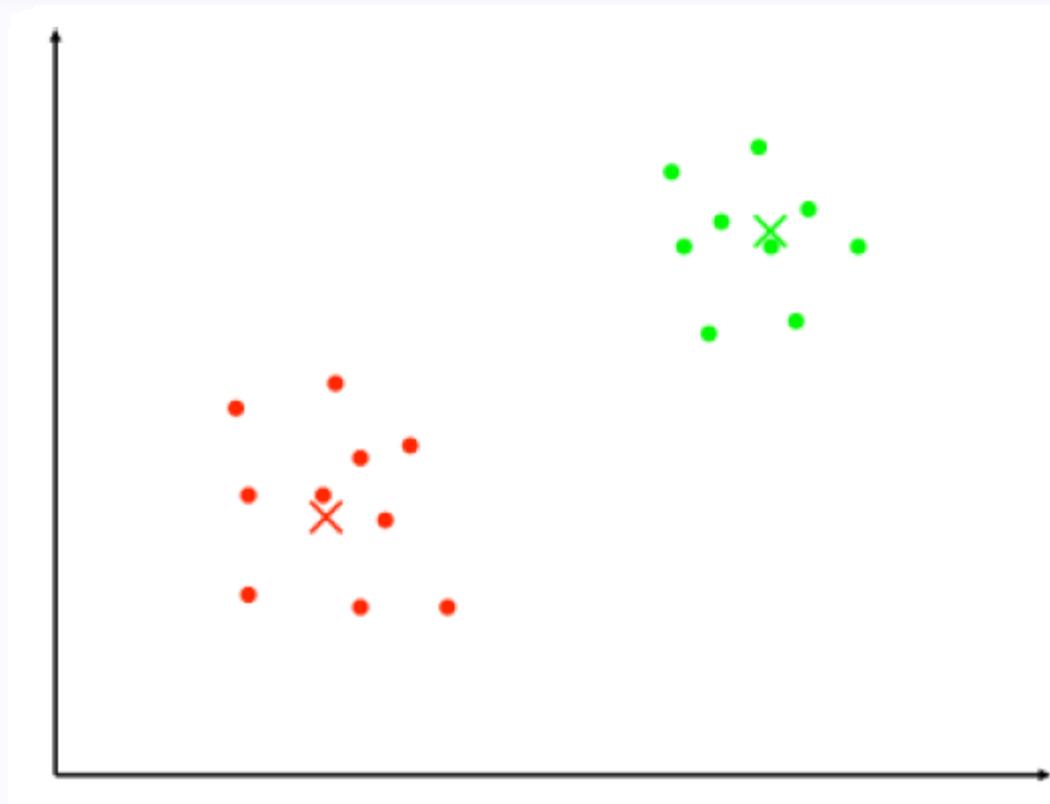
$$z_{ik}^{(t)} = \begin{cases} 1 & \text{si } k = \arg \min_{z \in \{1, \dots, K\}} \| \mathbf{x}_i - \mu_z \|^2 \\ 0 & \text{sinon.} \end{cases} \quad (3)$$

Etape de recalage des centres : le centre μ de chaque classe k est recalculé comme étant la moyenne arithmétique de toutes les données appartenant à cette classe (suite à l'étape d'affectation précédente) : $\forall k = 1, \dots, K$

$$\mu_k^{(t+1)} = \frac{\sum_{i=1}^n z_{ik}^{(t)} \mathbf{x}_i}{\sum_{i=1}^n z_{ik}^{(t)}}, \quad (4)$$

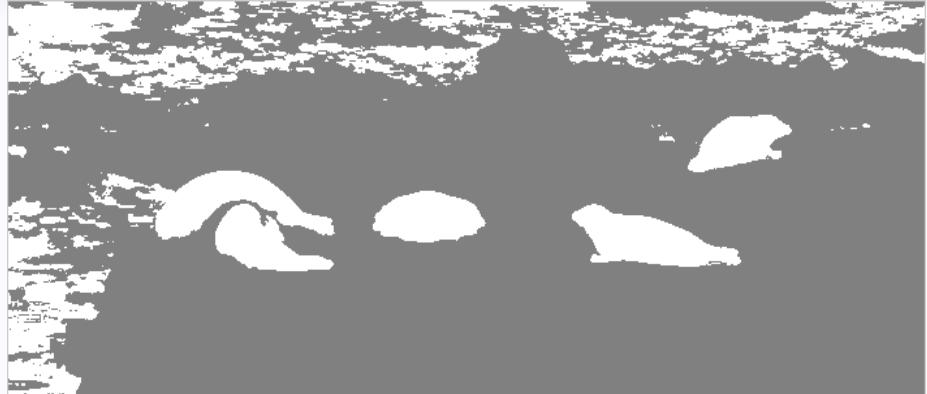
K-moyennes

Illustration en 2D

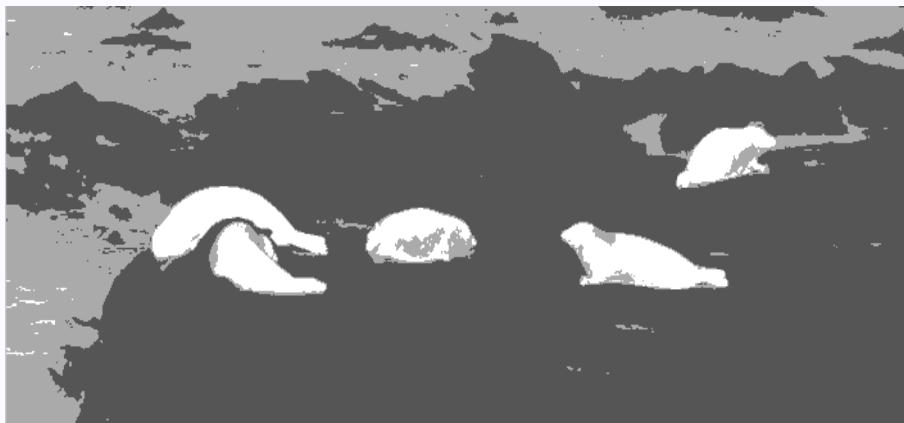


K-moyennes

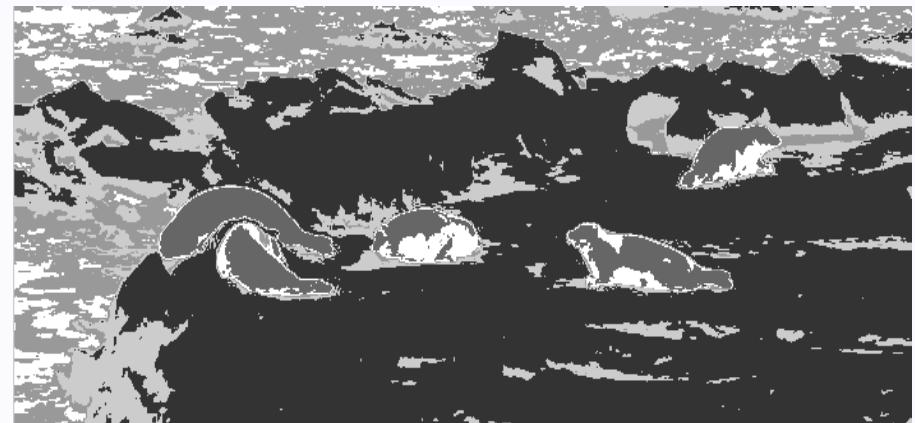
Application :segmentation d'images



K=2



K=3



K=5

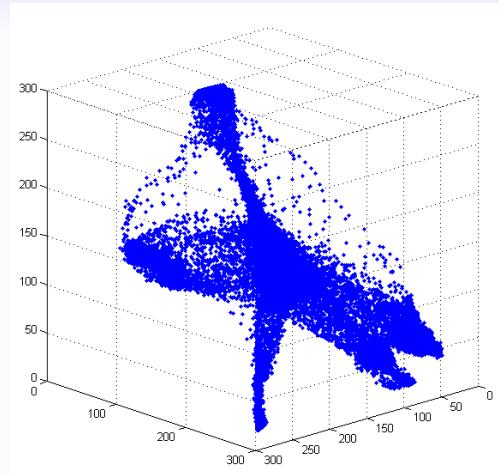
K-moyennes

Application: Les pieces

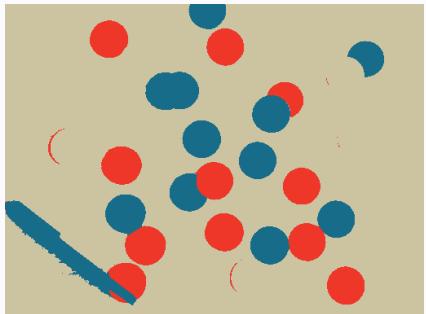
I



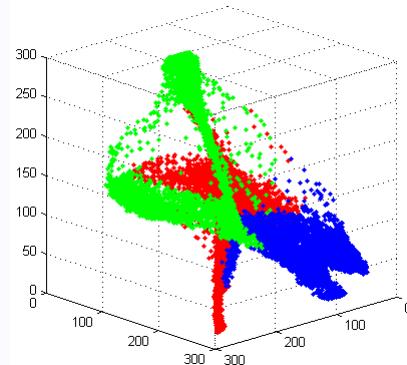
X



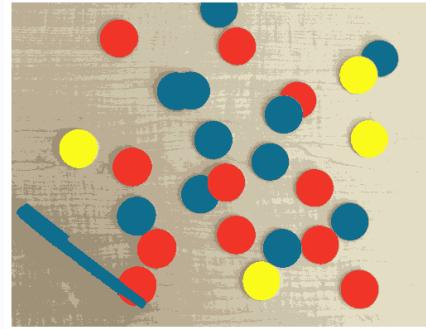
K=3



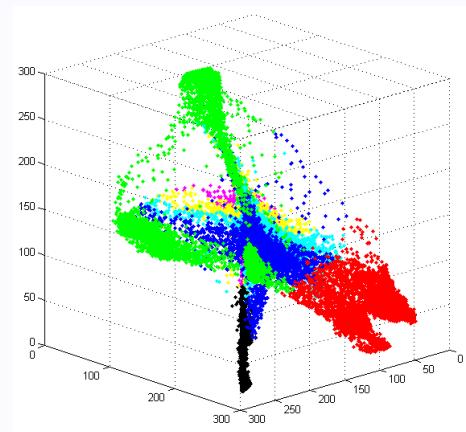
Etiquetage
des pixels



K=7



Etiquetage des
points RGB



K-moyennes

Avantages-Inconvénients

Avantages

- Méthode simple
- Implémentation facile
- Méthode généralement rapide
- Classes de variance conditionnelle minimale
- Fonctionne correctement lorsque les clusters sont sphériques



K-moyennes

Avantages-Inconvénients

Inconvénients

- Nécessite de connaître le nombre de classes
- Sensible aux minima locaux, donc à l'initialisation
- Peut être lent en grande dimension
- Échoue pour des structures non sphériques



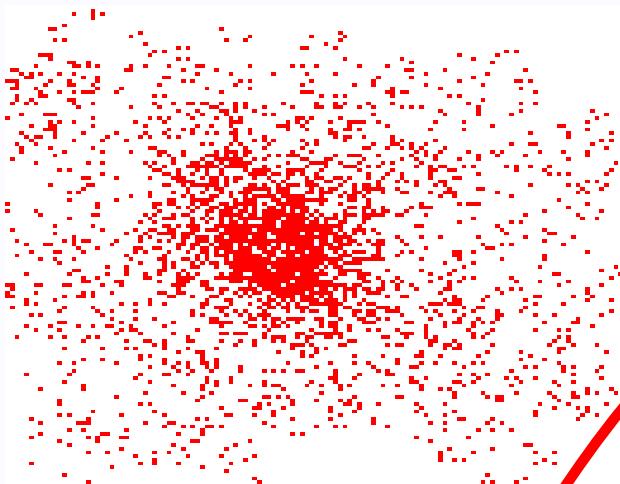
- Sensible aux valeurs aberrantes



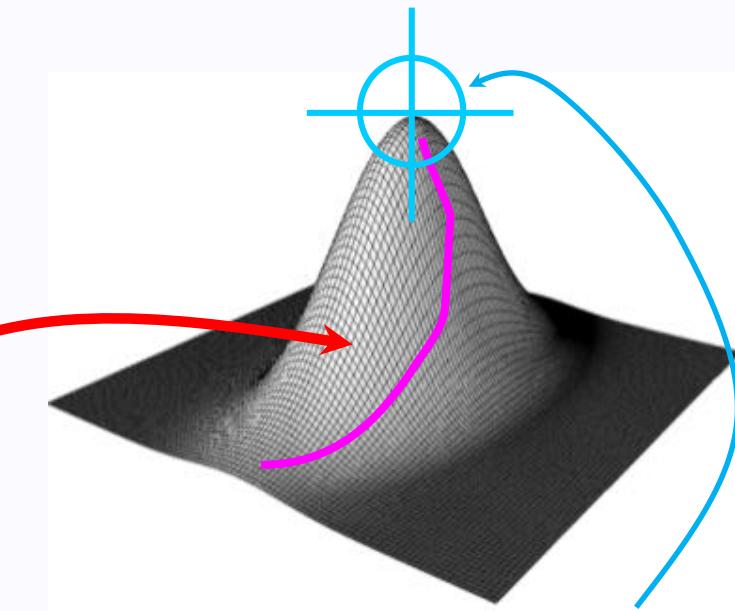
Mean Shift

Principe

- Le décalage moyen est une procédure permettant de localiser les maxima - les modes - d'une fonction de densité à partir de données discrètes échantillonnées à partir de cette fonction.



Points échantillonnés représentent une fonction de densité



Objectif: déterminer le maxima de cette fonction par une méthode du gradient

Mean Shift

Mathématiquement parlant....

- Il s'agit d'une méthode itérative, et nous commençons par une estimation initiale x . On considère une fonction noyau $K(x_i - x)$. Cette fonction détermine les poids du points x_i par rapport au centre x et permet d'évaluer la moyenne des points x_i ayant une influence (les points les plus proches).
- Plusieurs noyaux possibles: noyau uniforme – noyau gaussien - ...

- Déplacement moyen:

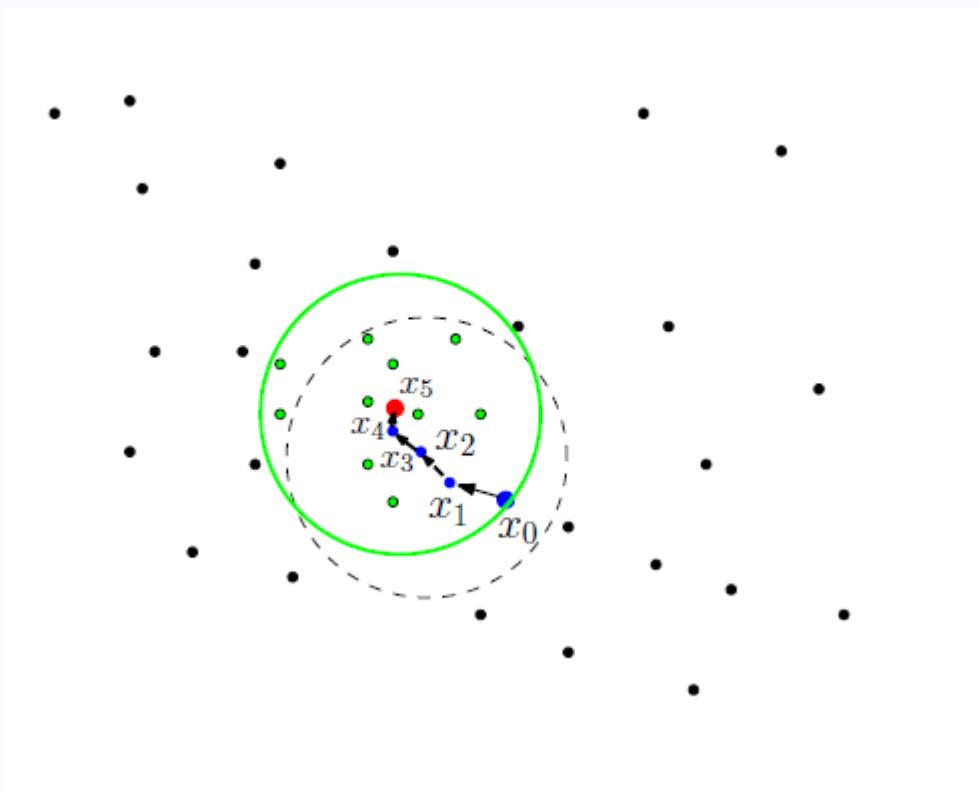
$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x)x_i}{\sum_{x_i \in N(x)} K(x_i - x)}$$

ou $N(x)$ est un voisinage de x tel que $K(x_i - x) \neq 0$.

- **le décalage $m(x)-x$ est appelé le vecteur mean shift.**
- **Procédure itérative: $x_0=x$, $x_1=m(x_0)$, $x_2=m(x_1)$,..., jusqu'à stabilisation**

Mean Shift

Illustration du mean shift



Initialisation

x_0 : donnée initiale

R : rayon de la fenêtre

Iter 1

x_1 : centre des points dans le cercle
Déplacement du cercle de centre x_1

Iter 2

x_2 : centre des points dans le cercle
Déplacement du cercle de centre x_2

⋮

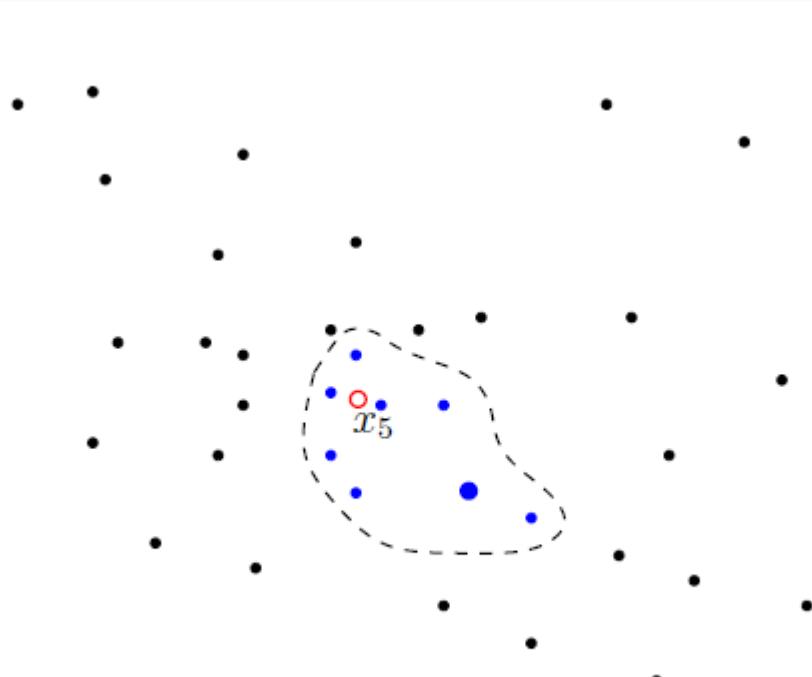
Iter 5

x_5 : point stationnaire

Convergence : $d(x_i, x_{i+1}) < \varepsilon$

Mean Shift

Création d'un cluster (groupe)



Étape 1

Un cluster associé au centre x_5 est créé :
Il est initialisé avec le premier point x_0

$$C_5 = \{x_0\}$$

Étape 2

Définition d'un rayon d'influence R_i :
Les points contenus dans le cercle
appartiennent au cluster:

$$C_5 = \{x_0, p_0\}$$

Étape 3

Mise à jour du cluster par déplacement du
cercle sur x_1 :

$$C_5 = \{x_0, p_0, p_1, p_2, p_3\}$$

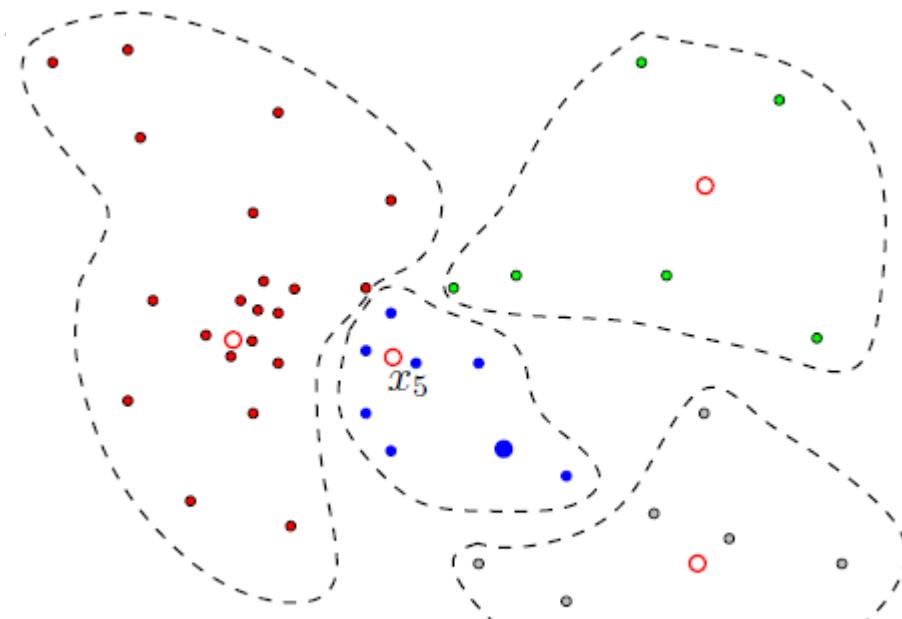
Étape n

Cluster détecté:

$$C_5 = \{x_0, p_0, p_1, p_2, p_3, p_4, p_5, p_6\}$$

Mean Shift

Création de plusieurs clusters



Étape 1

Répétition sur les autres points:
création d'autres clusters

$$C_5 = \{x_0, p_0, p_1, p_2, p_3, p_4, p_5, p_6\}$$

$$C_{10} = \{p_7, p_8, \dots\}$$

$$C_{\dots} = \{\dots\}$$

Étape 2

Fusion de clusters proches:
Définition d'un rayon de fusion: R_m

x_p, x_q points stationnaires du Mean-Shift

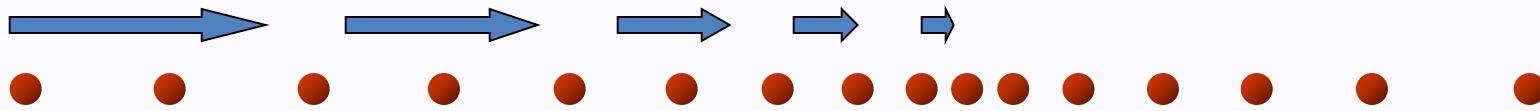
$$d(x_p, x_q) \leq R_m$$



fusionner C_p et C_q

Mean Shift

propriétés



- Vitesse de convergence automatique – La longueur du vecteur de déplacement dépend de l'amplitude du gradient.
- A proximité des maxima, les déplacements sont de plus en plus petits
- Pour le noyau uniforme, la convergence est atteinte en un nombre fini de pas.
- L'utilisation du noyau normal fournit une trajectoire lissée, mais sa convergence est plus faible que celle obtenue à partir du noyau uniforme.

Mean Shift

Avantages - Inconvénients

Avantages :

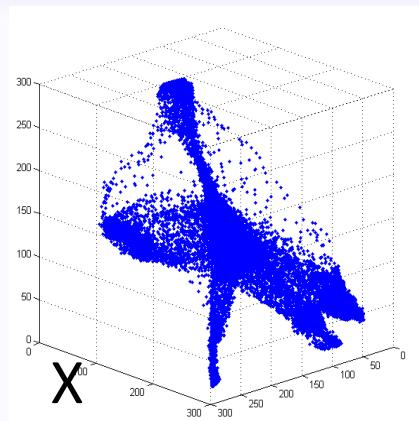
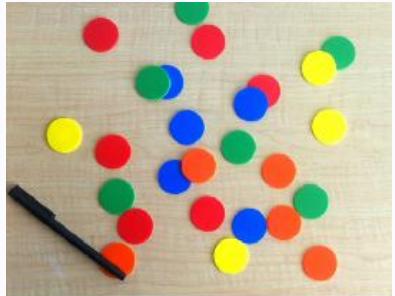
- Un outil indépendant de l'application
- Aucune hypothèse à priori sur la forme des classes
- L'espace des caractéristiques peut être arbitraire.
- Un SEUL paramètre à fixer : R

Désavantages :

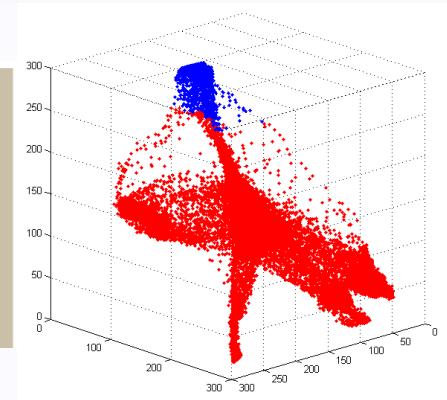
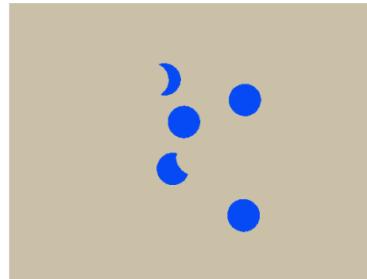
- La taille de la fenêtre n'est pas trivial
- Une taille inappropriée peut entraîner l'agglomération ou la fusion de plusieurs modes distincts. Elle peut générer également des modes "superficiels". → Utiliser une fenêtre adaptative

Mean Shift

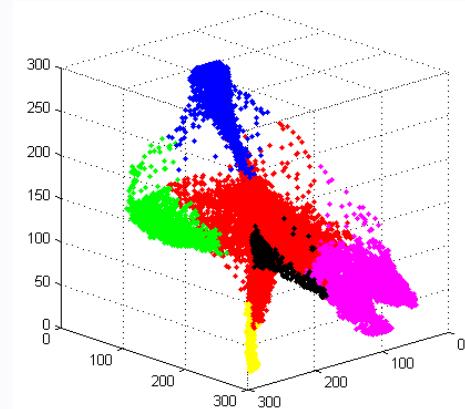
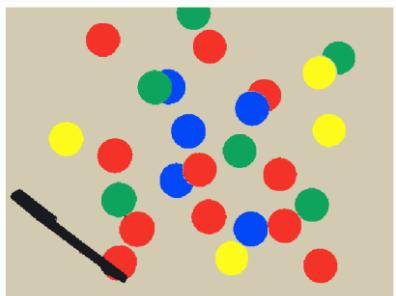
Exemple



R=0.5, K=2



R=0.2, K=6



R=0.1, K=58



Mean Shift

Avantages - Inconvénients

Avantages :

- Un outil indépendant de l'application
- Aucune hypothèse à priori sur la forme des classes
- L'espace des caractéristiques peut être arbitraire.
- Un SEUL paramètre à fixer : R

Désavantages :

- La taille de la fenêtre n'est pas trivial
- Une taille inappropriée peut entraîner l'agglomération ou la fusion de plusieurs modes distincts. Elle peut générer également des modes "superficiels". → Utiliser une fenêtre adaptative