

Classification de Zones Urbaines en Télédétection Satellitaire Haute Résolution

Mathieu Fauvel

LIS INPG, Iceland University

December 7, 2005

Plan de la Présentation

- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 Expériences
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion

Plan de la Présentation

- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 Expériences
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion

Plan de la Présentation

- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 Expériences
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion

Plan de la Présentation

- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 Expériences
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion

- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 Expériences
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion

Images Satellites et Aériennes

Une image est caractérisée par sa résolution:

- spatiale: 1m/pixel.
- spectrale: plus de 200 bandes.

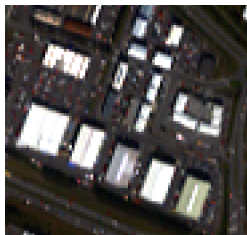
L'objectif est de classifier l'image: affecter à chaque pixel une classe prédéfinie.

La fine résolution spatiale permet une plus grande précision dans la définition/délimitation des structures.

La séparabilité des classes augmente avec le nombre de bandes disponibles.



Panchromatique



Multi Spectral

Nouvelles Problématiques

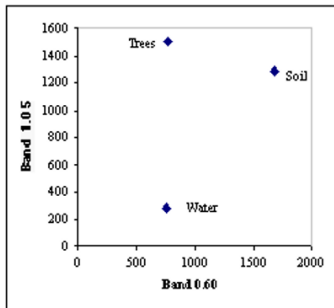
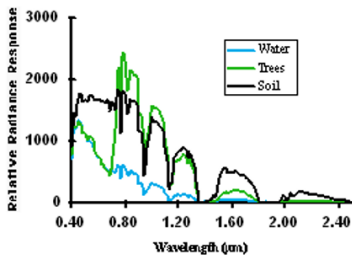
Ces progrès techniques impliquent que:

- Les classes doivent être aussi définies en fonction de leur taille.
- Le volume des données à traiter devient de plus en plus important.
- L'estimation de paramètres statistiques devient très délicate.
- Phénomène de Hughes.

Mises en œuvre de nouvelles méthodologies:

- Approche morphologique (Voir Séminaire Jocelyn Chanussot du 02/12/2004).
- Caractérisation des classes par des critères géométriques.
- Utilisation de Méthodes résistantes au phénomène de Hughes.

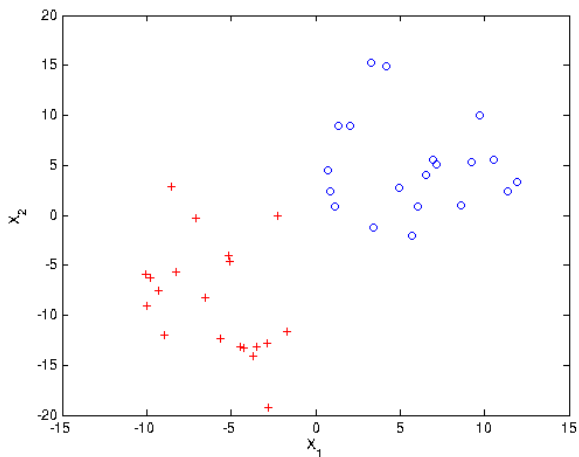
Représentation des données



- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 Expériences
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion

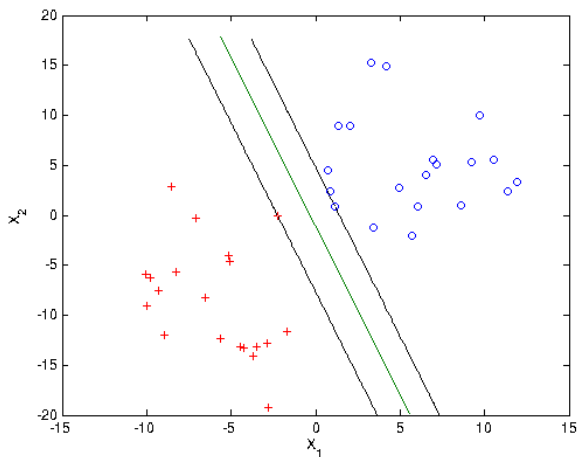
Exemple

$$(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N) \in \mathbb{R}^n \times \{\pm 1\}$$



Exemple

$$(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N) \in \mathbb{R}^n \times \{\pm 1\}$$



- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 Expériences
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion

Définitions

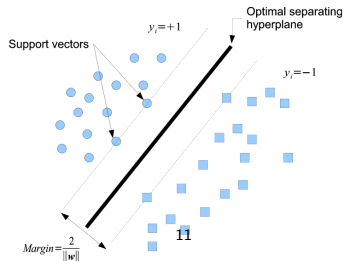
Hyperplan H_p :

- \mathbf{w} : vecteur normal à l'hyperplan, b : biais.
- $\mathbf{w} \cdot \mathbf{x} + b = 0, \forall \mathbf{x} \in H_p$.
- Distance vecteur-hyperplan: $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b, \forall \mathbf{x} \notin H_p$.
- $DIM(H_p) = n - 1$.

Décision: $y = \text{sgn}(f(\mathbf{x}))$

Entraînement: Trouver H_p tels que: $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \forall i \in [1, M]$

Hyperplan optimal: maximise les distances des échantillons d'entraînement les plus proches de l'hyperplan.



Cas Séparable

$$\min \left[\frac{\mathbf{w} \cdot \mathbf{w}}{2} \right] \text{ tel que } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \forall i \in [1, N]$$

Problème d'optimisation sous contraintes: Lagrange.

$$\left\{ \begin{array}{l} L(\mathbf{w}, b, \alpha_i) = \frac{\mathbf{w} \cdot \mathbf{w}}{2} + \sum_{i=1}^N \alpha_i (1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)) \\ \alpha_i \geq 0 \\ y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \end{array} \right.$$

$$\left\{ \begin{array}{l} \frac{\partial L(\mathbf{w}, b, \alpha_i)}{\partial b} = 0 \Rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \\ \frac{\partial L(\mathbf{w}, b, \alpha_i)}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \end{array} \right.$$

Cas Séparable

Formulation duale (Duale de Wolf):

$$W(\alpha) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

tel que $\alpha_i \geq 0$ et $\sum_{i=1}^M \alpha_i y_i = 0$

Au point selle, d'après les conditions de Karush-Kuhn-Tucker:

$$L(\bar{\mathbf{w}}, \bar{b}, \bar{\alpha}_i) = \frac{\bar{\mathbf{w}} \cdot \bar{\mathbf{w}}}{2} \Rightarrow \alpha_i (1 - y_i (\bar{\mathbf{w}} \cdot \mathbf{x}_i + \bar{b})) = 0, \quad \forall i \in [1, M]$$

$\alpha_i \neq 0$, \mathbf{x}_i est un **Vecteur Support**.

En pratique, lors de la phase d'entraînement, seulement les $M (\leq N)$ échantillons nécessaires pour séparer les classes sont retenus.

$$y = \operatorname{sgn}\left(\sum_{i=1}^M y_i \alpha_i (\mathbf{x} \cdot \mathbf{x}_i) + b\right)$$

Cas Non Séparable

Cas non séparable: on introduit des variables d'erreurs:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

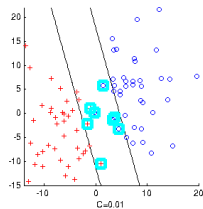
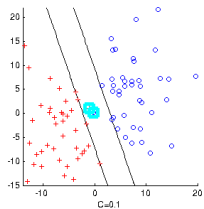
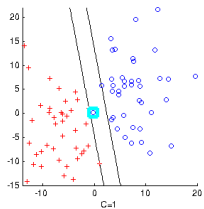
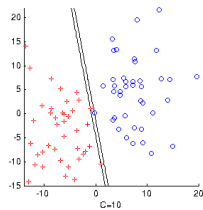
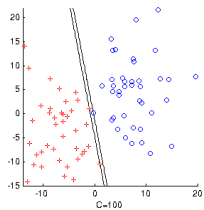
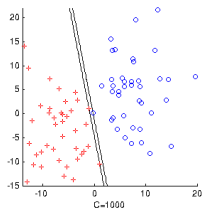
La phase d'entraînement devient alors:

$$\min \left[\frac{\mathbf{w} \cdot \mathbf{w}}{2}, C \sum_{i=1}^N \xi_i \right] \text{ tel que } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \forall i \in [1, N]$$

De la même manière, on obtient le problème de maximisation:

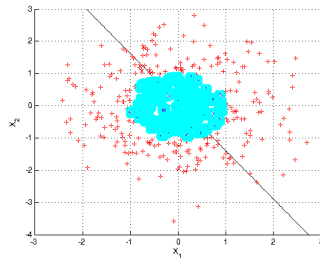
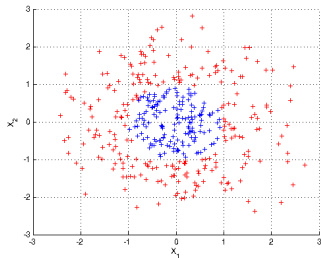
$$W(\alpha) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{tel que } 0 \leq \alpha_i \leq C \text{ et } \sum_{i=1}^M \alpha_i y_i = 0$$

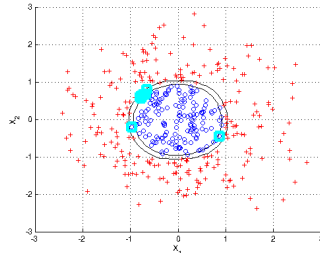
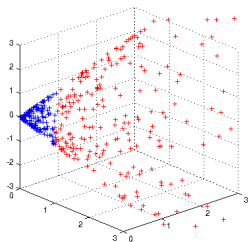
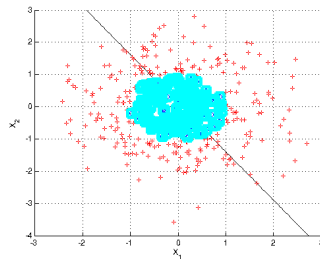
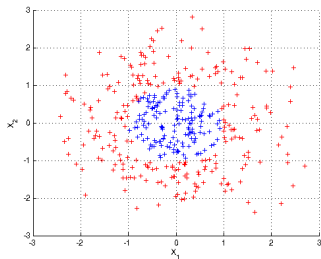
Influence de C 

- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 Expériences
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion

Exemple: cas non linéairement séparable



Exemple: cas non linéairement séparable



Coup du Noyau: exemple

Idée: projeter dans un espace de plus grande dimension où les données seront linéairement séparables.

$$\begin{aligned}\Phi : \mathbb{R}^n &\rightarrow \mathbb{H} \\ \mathbf{x} &\rightarrow \Phi(\mathbf{x}) \\ \mathbf{x}_i \cdot \mathbf{x}_j &\rightarrow \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)\end{aligned}$$

$$y = \text{sgn}\left(\sum_{i=1}^M y_i \alpha_i (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}_i)) + b\right)$$

Problème: les calculs deviennent très lourds!!!

Dans l'exemple précédent:

$$\begin{aligned}\Phi : \mathbb{R}^2 &\rightarrow \mathbb{R}^3 \\ \mathbf{x} = (x_1, x_2) &\rightarrow (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)\end{aligned}$$

Coup du Noyau: Généralisation

On peut réécrire le produit scalaire:

$$\begin{aligned}\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y}) &= (x_1^2, \sqrt{2}x_1x_2, x_2^2) \cdot (y_1^2, \sqrt{2}y_1y_2, y_2^2) \\ &= x_1^2y_1^2 + 2x_1y_1x_2y_2 + x_2^2y_2^2 \\ &= (x_1y_1 + x_2y_2)^2 \\ &= (\mathbf{x} \cdot \mathbf{y})^2 \\ &= K(\mathbf{x}, \mathbf{y})\end{aligned}$$

Pour qu'elles fonctions $K(\mathbf{x}, \mathbf{y})$ peut-on associé un couple $\{H, \Phi\}$?
Si $K(\mathbf{x}, \mathbf{y})$ vérifie les conditions de Mercer, alors on peut *toujours* y associé un couple.

Exemples de noyaux

Noyaux globaux:

$$\textit{Polynomial} : K(\mathbf{x}_i, \mathbf{x}_j) = ((\mathbf{x}_i \cdot \mathbf{x}_j) + 1)^d$$

$$\textit{Sigmoid} : K(\mathbf{x}_i, \mathbf{x}_j) = \tanh((\mathbf{x}_i \cdot \mathbf{x}_j) + 1)$$

Noyaux locaux:

$$\textit{Gaussien} : K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$$

$$\textit{Inverse multiquadratique} : K(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{\sqrt{(\|\mathbf{x}_i - \mathbf{x}_j\|^2 + 1)}}$$

Noyaux adaptés aux images hyperspectrales:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \alpha(\mathbf{x}_i, \mathbf{x}_j)^2\right)$$

$$\text{où } \alpha(\mathbf{x}_i, \mathbf{x}_j) = \arccos\left(\frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\| \cdot \|\mathbf{x}_j\|}\right)$$

- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 Expériences
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion

Problème à m classes

La solution consiste en une combinaison de classifieurs binaires:

- *Un Contre Tous*: m classifieurs sont appliqués. Chaque pixel est assigné à la classe qui obtient la distance à H_p la plus grande.
- *Un Contre Un*: $\frac{m(m-1)}{2}$ classifieurs sont appliqués. Un pour chaque paire de classes. Pour chaque classe on compte le nombre de votes *pour*. La classe obtenant le plus grand nombre de votes l'emporte.
- *Arbre*: Chaque noeud de l'arbre est un classifieur binaire.

Conclusions

Les SVMs sont intéressantes car:

- Approche géométrique du problème.
- Solution optimale.
- Gère les espaces de grandes dimensions.
- Grande capacité de généralisation.
- Fonctions noyaux: *personnalisation* du problème.

Mais:

- En pratique?
- Problème à plusieurs classes?
- Sélection des paramètres?

- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 **Expériences**
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion

- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 Expériences
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion



Image Hyperspectrale: 102 bandes.

Table: Ensemble d'entraînement et de test.

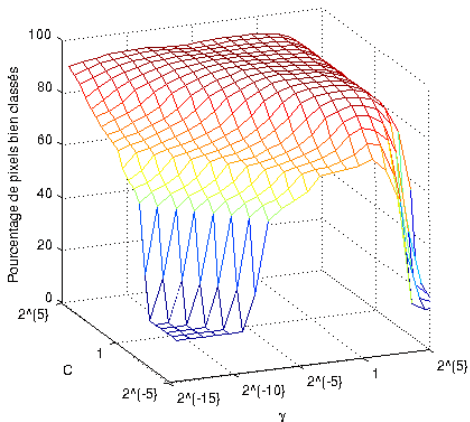
Class		Samples	
No	Name	Train	Test
1	Water	745	65278
2	Trees	785	6508
3	Meadow	797	2900
4	Brick	485	2140
5	Soil	820	6549
6	Asphalt	816	7555
7	Bitumen	808	6479
8	Tile	223	3122
9	Shadow	195	2165
Total		5536	103504

Des ensembles d'entraînement de faibles tailles sont construits aléatoirement.

Noyaux Gaussien: $K_{GAUSS}(\mathbf{x}_i, \mathbf{x}_j) = \exp \left[-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2 \right]$.

Noyaux prenant en compte les variations d'intensité dues à l'acquisition: $K_{SAM}(\mathbf{x}_i, \mathbf{x}_j) = \exp \left[-\gamma \alpha(\mathbf{x}_i, \mathbf{x}_j)^2 \right]$.

Sélection des paramètres (C, γ) par *grid search* et *cross-validation*.



Ici, les paramètres optimaux sont $\gamma = 0.125$ et $C = 128$.



Table: Classification Accuracies for the Gaussian and the SAM RBF kernel.

Training Set Size	OA%	
	Gaussian	SAM
10	93,85	93,32
20	94,51	93,87
40	99,51	93,79
60	94,71	94,23
80	95,36	94,40
100	95,29	94,54
All	96,45	95,56

Eau- Arbre- Prairie- Brique- Terre-
Asphalte- Bitume- Tuile- Ombre

- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 Expériences
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion

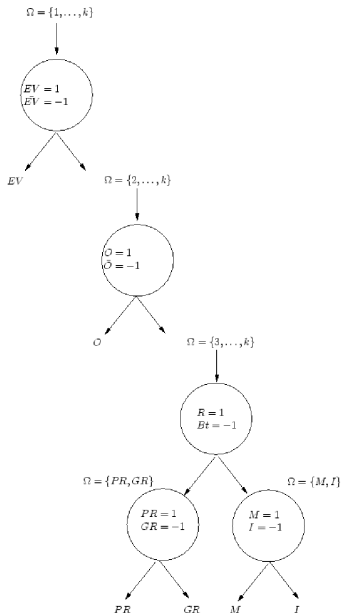
Données IKONOS (\approx QUICKBIRD, SPOT ...):

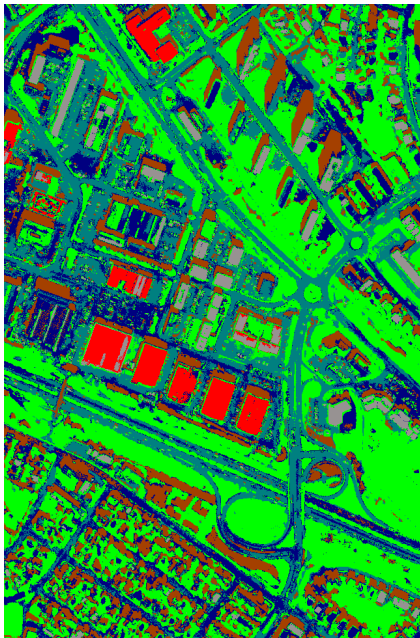
image panchromatique haute résolution spatiale +
image multispectrale (4 bandes) de résolution spatiale deux fois inférieure.

Discrimination des classes sur des critères physiques (MS) puis sur des critères morphologiques (Panchro).

Noyaux: Gaussien.

Ensemble d'entraînement: 100 pix/cl
5 Machines.





- 1 Présentation du Problème
- 2 Machines à Support Vecteur
 - Formulation Mathématique
 - Kernel Trick
 - Classification non binaire
- 3 Expériences
 - Ensemble d'entraînement de faible taille
 - Classification par arbre
- 4 Conclusion

Quelques références:

- *A Tutorial on Support Vector Machines for Pattern Recognition*, Christopher J.C. Burges.
- *An Introduction to Kernel-Based Learning Algorithms*, Muller *et al.*, IEEE Transaction on Neural Network, Vol. 12, No. 2, March 2001.
- *Kernel methods for pattern analysis*, John. Shawe-Taylor.
- <http://www.kernel-machines.org/>