

# Graphes de voisinage et apprentissage, applications à l'image

Journée Analyse de données, Statistique  
et Apprentissage pour la Fouille d'images

AS Fouille d'images, RTP 25

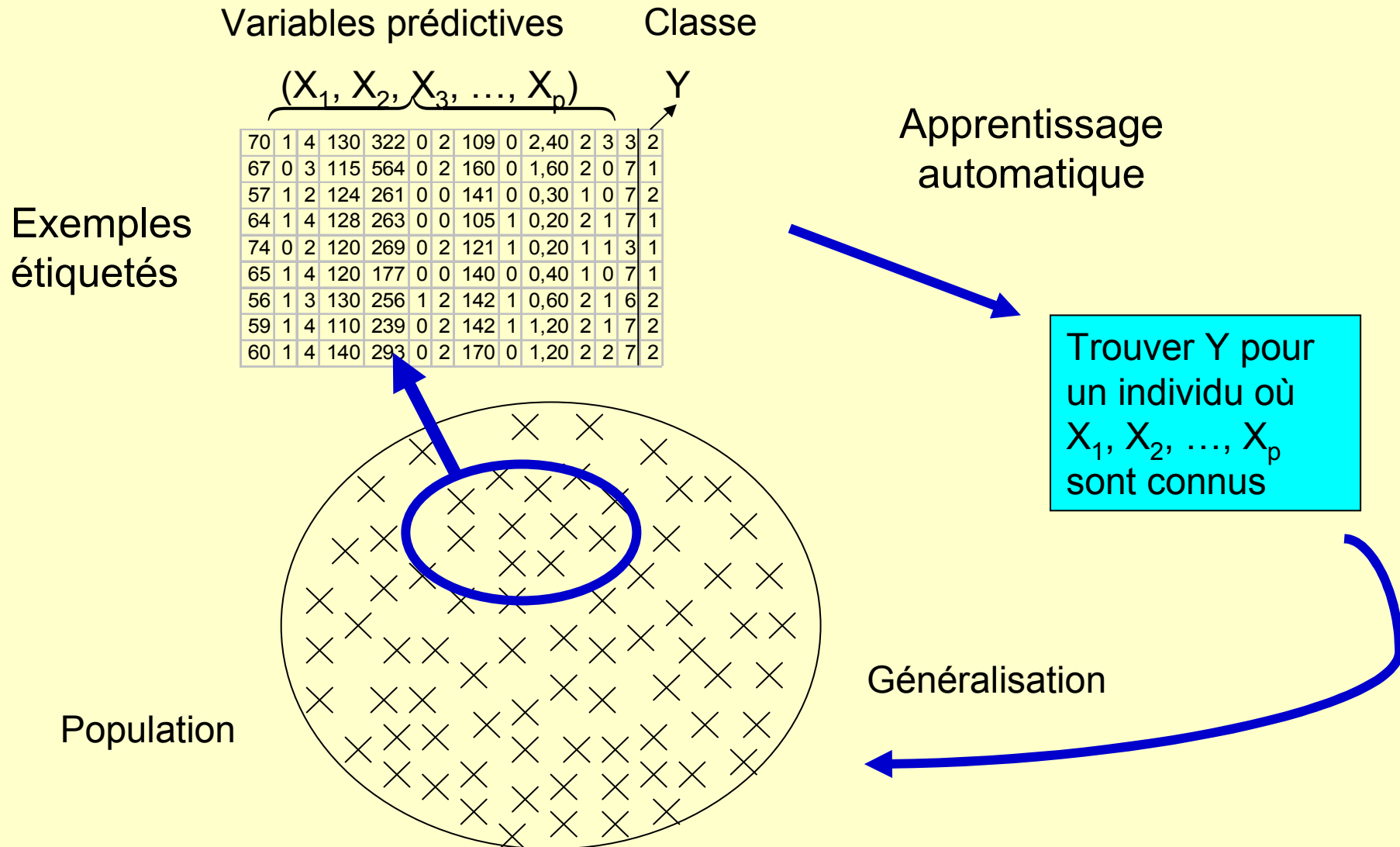
Stéphane Lallich,  
1er avril 2003

# Plan

---

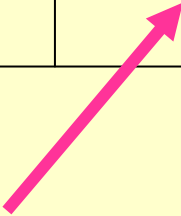
- 1. Cadre général de l'apprentissage supervisé**
- 2. Qualité de la représentation**
- 3. Graphes de voisinage**
- 4. Evaluer la qualité de la représentation**
- 5. Détecter les exemples atypiques**
- 6. Application à la segmentation d'image**
- 7. Application à la réduction d'exemples par boosting**
- 8. Perspectives**

# 1. Cadre général de l'apprentissage supervisé



## 2. Qualité de la représentation en apprentissage

Type	Principe	Validation
<b>Apprentissage fondé sur un modèle</b>	Construction d'un modèle	Adéquation du modèle
<b>Apprentissage fondé sur les voisinages</b>	Stockage des exemples	Qualité de la représentation



**Détection des  
exemples atypiques**

**Sélection de  
variables ...**

**Qualité de la  
représentation**

- Y catégorielle
- Y numérique

**Automatisation,  
Critère d'arrêt :  
segmentation d'image**

**Automatisation,  
Critère d'arrêt :  
boosting d'exemples**

# Un thème de recherche du laboratoire ERIC

## Axe de l'équipe de recherche :

- graphes de voisinage (Zighed et al.)
- réétiquetage par relaxation (Largeron, Amghar)
- nombre d'arêtes coupées (Sebban)
- thèses récentes (Muhlenbach, Clech)

## Modélisation :

- Nouveau modèle bien adapté au problème et plus riche
- Analogie avec l'analyse spatiale
- Formalisme puissant, autocorrélation de voisinage

## Produit :

- Test de séparabilité
- Test de structure

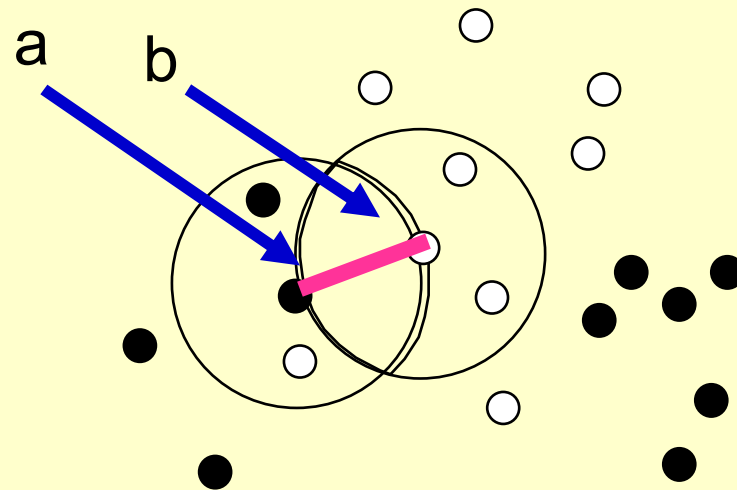
### 3. Graphe de voisinage

#### Graphe de voisinage :

- deux exemples voisins au sens des prédicteurs sont reliés par une arête (graphe connexe et symétrique)
- les sommets du graphe sont étiquetés par Y

#### Graphe de Toussaint :

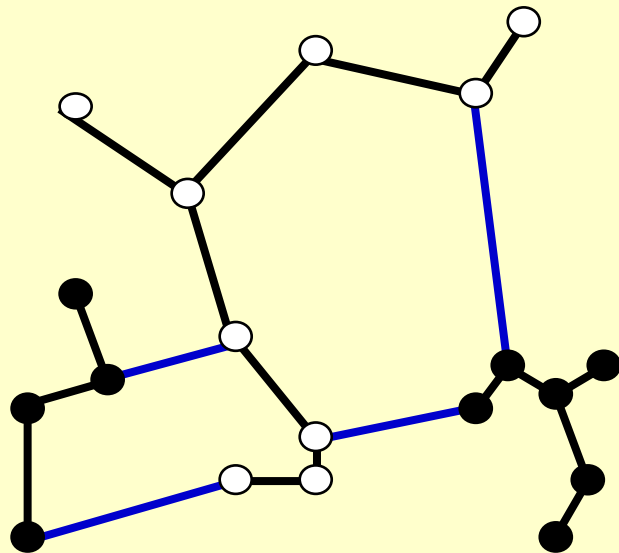
$$d(a,b) \leq \text{Max} \{d(a,c), d(b,c)\}$$
$$\forall c, c \neq a, c \neq b$$



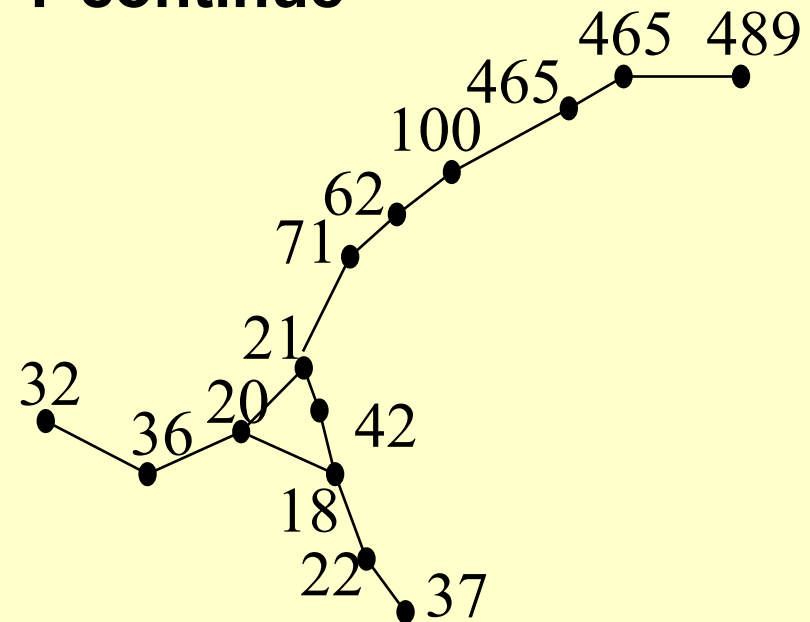
## 4. Evaluer la qualité de la représentation

### Analogie avec les graphes spatiaux

Y catégorielle



Y continue



$H_0$  : étiquettes distribuées au hasard sur les sommets du graphe



adaptation des outils d'analyse de l'autocorrélation spatiale : BW, Moran



# Statistique produit croisé et coefficients globaux

**Autocorrélation spatiale** : lien entre  $Y$  et le graphe spatial

**Définition** (Mantel 67, Hubert 76, Getis 91) :  $\Gamma = \sum_i \sum_j w_{ij} u_{ij}$

- $W$  évalue la proximité spatiale (matrice connexion, poids)
- $U$  évalue la proximité au sens de  $Y$

**Coefficients globaux usuels** : choix  $U$  et normalisation

Coefficient	$U_{ij}$	Normalisation
BW de Moran	$I(y_i \neq y_j)$	1/2
I de Moran	$(y_i - \bar{y})(y_j - \bar{y})$	$2a\text{Var}(y)$

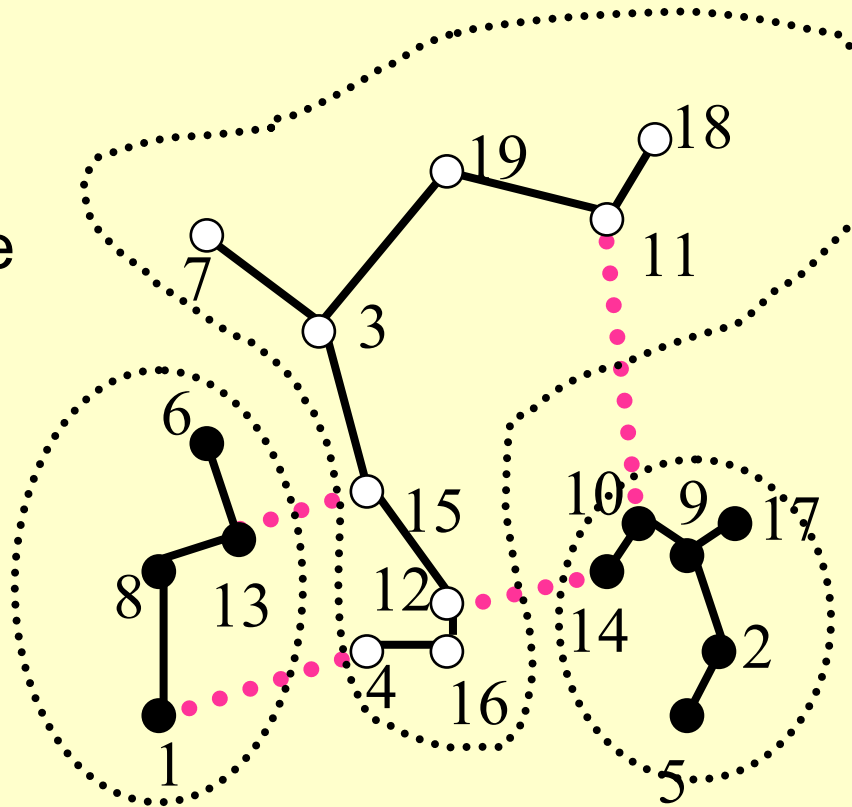
Nombre d'arêtes :  $a$  ; nombre de sommets du graphe :  $n$

**Loi sous  $H_0$**  : cf. Cliff et Ord, 81

# Y catégorielle : Coupure d'arêtes

Suppression  
des arêtes  
reliant des points de  
classes différentes

Suppression  
de 4 arêtes



Formation de 3 amas

# Y catégorielle : Cas booléen

## Formalisme des join-counts (Moran, Cliff et Ord)

- **variables indicatrices**

$X_i$  : le sommet  $i$  est de classe 1

$U_{ij}$  : les sommets  $i$  et  $j$  sont de même classe

- **$J_{12}$  : poids des arêtes coupées**

$$J_{12} = \frac{1}{2} \sum_i \sum_j w_{ij} (X_i - X_j)^2 = \frac{1}{2} \sum_i \sum_j w_{ij} U_{ij}$$

- **Hypothèse nulle  $H_0$**  : étiquettes distribuées au hasard suivant la distribution globale de  $Y$

## Loi sous $H_0$

- $X_i$  i.i.d  $B(1, \pi_1)$
- $U_{ij}$  i.d  $B(1, 2\pi_1\pi_2)$  mais non indépendants
- calcul direct des covariances du type  $\text{Cov}(U_{ij}, U_{il})$
- on en déduit  $E(J_{12})$  et  $\text{Var}(J_{12})$

## Décision

- ✓ test unilatéral ou bilatéral suivant opportunité
- ✓ p-value de  $J_{12, \text{obs}}$  par approx. normale ou par simulation

# Y catégorielle : du booléen au multiclassés

**Notations** : on considère tous les couples de classes (r, s) :

$$J = \sum_{r=1}^{k-1} \sum_{s=r+1}^k J_{rs} = \frac{1}{2} \sum_i \sum_j w_{ij} U_{ij}$$

$J_{rs}$  : poids des arêtes coupées entre sommets de classes r et s

$U_{ij}$  : les sommets i et j sont de même classe, parmi k possibles  
chaque  $J_{rs}$  correspond à une situation booléenne

## Loi de J sous $H_0$

- calcul aisé de  $E(J)$  à partir des  $E(J_{rs})$
- calcul de  $\text{Var}(J)$  nettement plus compliqué ! il faut déterminer les covariances du type  $\text{Cov}(U_{ij}, U_{il})$
- approximation normale ou simulation

# Y numérique : le coefficient global de Moran

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}$$

## Interprétation :

quotient d'une autocovariance spatiale par la variance totale

## Écriture sous forme matricielle : $I = z'Wz / z'z$

- $z$  : variable centrée associée à  $y$
- $W$  : matrice 1-normalisée

**Décomposition de la variance globale** (Lebart ; Thioulouse) :  
= autocovariance locale (avec poids) + variance locale (avec poids)  
(poids 1/n-normalisés)

# Schéma probabiliste pour global Moran

## Schéma gaussien ( $N$ ) :

tirages équiprobables et indépendants dans une population normale.

## Schéma randomisé ( $R$ ) :

tirage au hasard pur parmi les  $n!$  permutations possibles sur les sommets du graphe (schéma conditionnel aux observations).

**Test de  $H_0$**  : la réalisation de la statistique de test remet-elle en cause  $H_0$  ?

→ Choix d'un schéma et d'une statistique de test

→ Loi de la statistique de test sous  $H_0$  et calcul p-value

- Appr. normale : calcul des moments sous  $H_0$ , standardisation, p-value
- Stat. exacte ou simulation : loi exacte ou simulée, p-value

# 5. Détection des exemples atypiques

Composantes locales de la statistique produit croisé :

$$\Gamma = \sum_i \sum_j w_{ij} u_{ij} \rightarrow \Gamma_i$$

**Définition des exemples atypiques (Shekhar et al. 01).**

- Y variable à prédire
- Y' agrégation des valeurs de voisinage de Y
- $D : \mathbb{R}^2 \rightarrow \mathbb{R}$ , compare Y et Y'
- T fonction de décision,  $T : \mathbb{R} \rightarrow \{\text{Vrai}, \text{Faux}\}$

“i” GV-atypique  $(Y, Y', D, T) \Leftrightarrow T(D(Y_i; Y'_i))$  vraie.



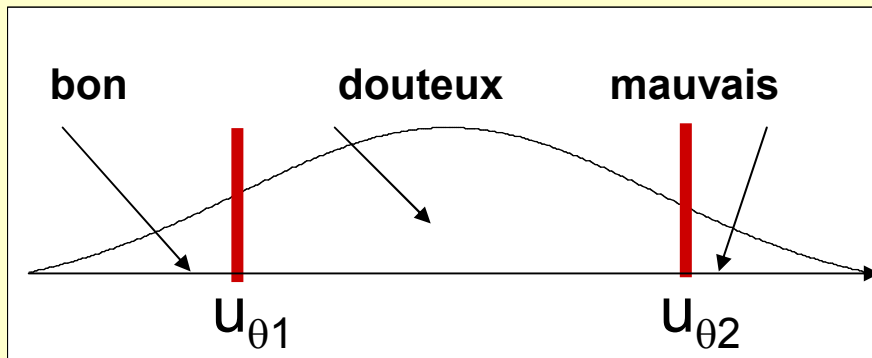
# Composantes locales suivant la nature de Y

## Y catégorielle

**Stat. locale** :  $\Gamma_i = \sum_j w_{ij} I(y_i \neq y_j)$

$H_0$  : étiquetage au voisinage de  $i$ , conforme à étiquetage global

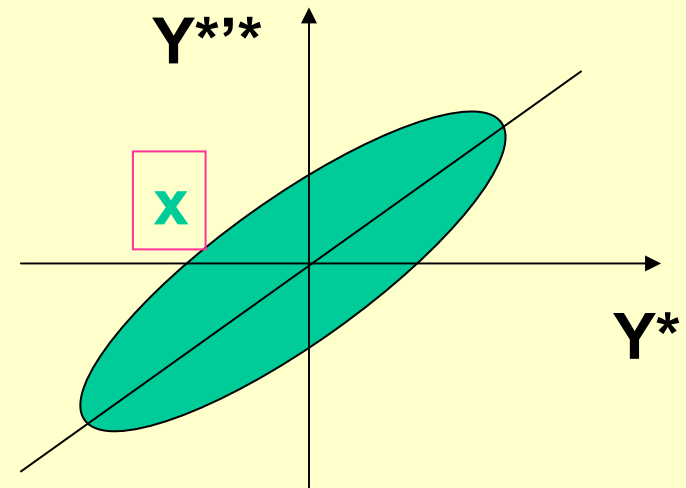
**Loi de  $\Gamma_i$  sous  $H_0$**



**Stat. locale vs Relaxation**

- cohérence locale
- similitude globale

## Y numérique

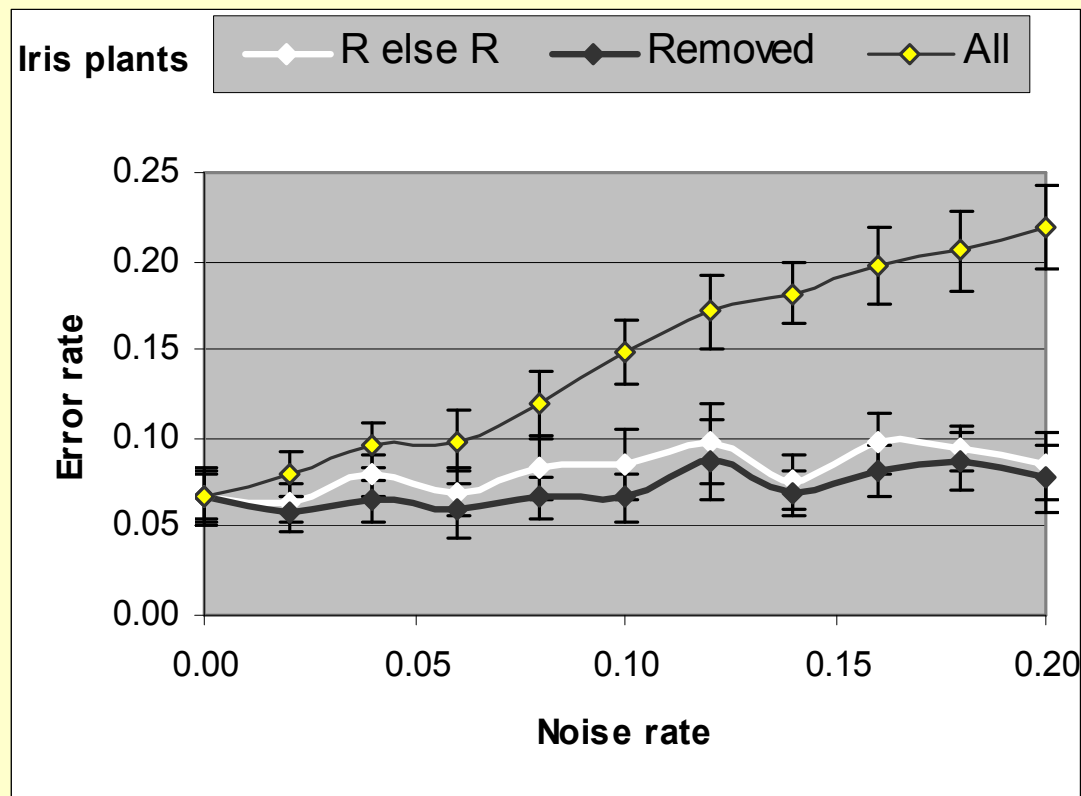


**Diagramme de Moran**

# Expériences : outliers dans le cas Y catégorielle

$\theta$  : taux d'erreur en généralisation

LS : ensemble d'apprentissage



1.  $\theta$  associé à LS non traité (*all*) **augmente linéairement** avec le niveau de bruit

2.  $\theta$  associé à LS traité (*R, R else R*) **reste à un niveau constant**

3. L'option "retrait" est **légèrement meilleure** que "réétiquetage ou retrait"

# 6. Application à la segmentation d'images

**But** : partitionner une image en régions homogènes

**Graphe des pixels** : on construit le graphe de contiguïté des pixels, avec comme étiquette le niveau de gris.

**Algorithme de décimation itérative (Meer 89)** :

- à chaque étape, entre deux voisins, un seul survit
- un sommet qui ne survit pas a au moins un voisin qui survit

**Problème**

- arrêt de la procédure itérative (global)
- comment éviter les fusions localement injustifiées (local)
  - Définition de seuils adaptés à chaque image, Jolion, Montanvert, 92

# Décimation automatique

## **Graphe de décimation**

graphe partiel où ne sont conservées que les arêtes qui relient un non survivant et un survivant ;

## **Contrôle global : Moran global**

On poursuit la décimation jusqu'à ce que l'on puisse considérer que la distribution du niveau de gris sur les sommets du graphe de décimation est aléatoire ;

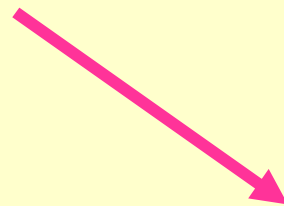
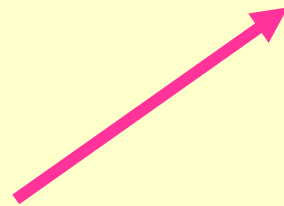
## **Contrôle local : Moran atypique**

On interdit la fusion des sommets atypiques au sens du diagramme de Moran.

**Image initiale**



**Contrôle global**

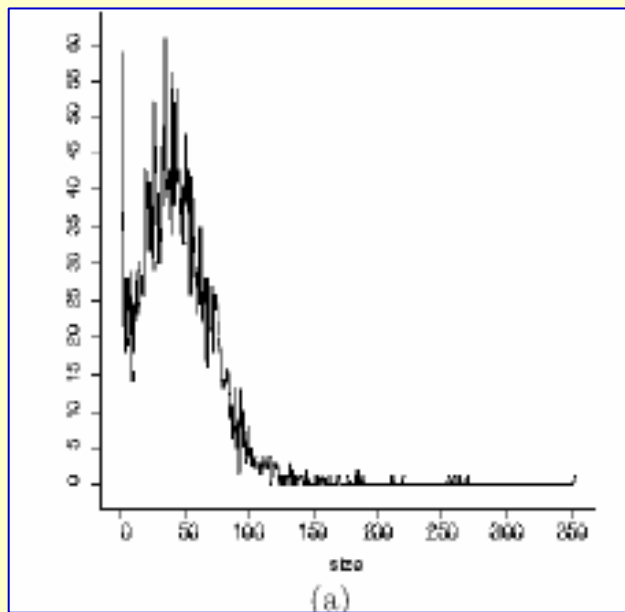


**Contrôle local**

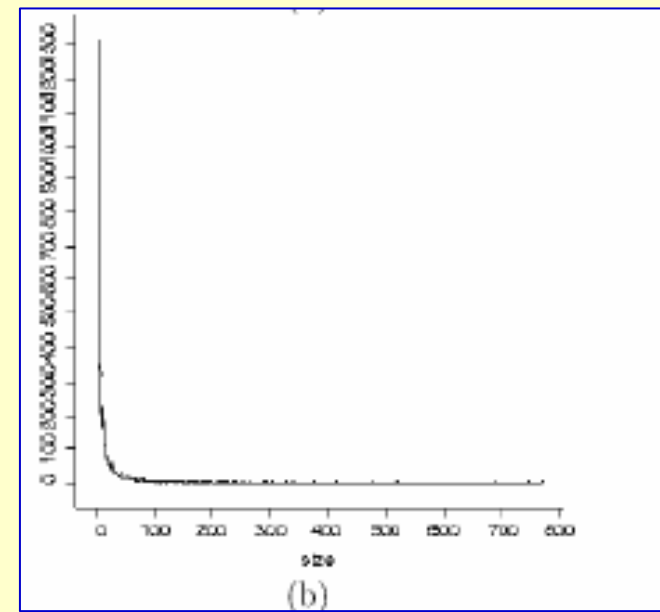


# Distribution de la taille des régions

Contrôle global



Contrôle local



## 7. Réduction d'exemples par le *boosting*

**Réduction d'exemples** : on réduit la taille de l'ensemble d'apprentissage, cf. Skalak 94, Wilson et Martinez 00.

**Boosting** : procédure itérative, repondération, agrégation de classifieurs, cf. AdaBoost, Freund et Schapire 95.

**Justification** : le boosting agit sur une borne supérieure de l'erreur en généralisation.

**Réduction d'exemples par le boosting** : tout exemple est un classifieur de ses plus proches voisins ; à chaque pas du boosting, PSBoost sélectionne l'exemple qui classe le mieux ses voisins. cf. Nock et Sebban, 01

**Problème** : Le nombre d'itérations  $T$  correspond à la taille de l'ensemble réduit. Il est fixé au préalable.

# PSBoost2 (cas booléen)

**Solution** : arrêt de la procédure lorsque le candidat ne classe pas mieux ses voisins qu'un *toss* équilibré.

## Principe

- $D_t(e)$  : poids de l'exemple  $e$  à l'itération  $t$
- on forme la statistique locale ( $R(e)$  voisinage réciproque de  $e$ )

$$W_e^+ = \sum_{e' \in R(e)} D_t(e') = \sum_{e' \in R(e)} D_t(e') I(y(e')=y(e))$$

- le formalisme de Moran permet de construire le test

## Expériences (par validation croisée, 5cv, 3NN) :

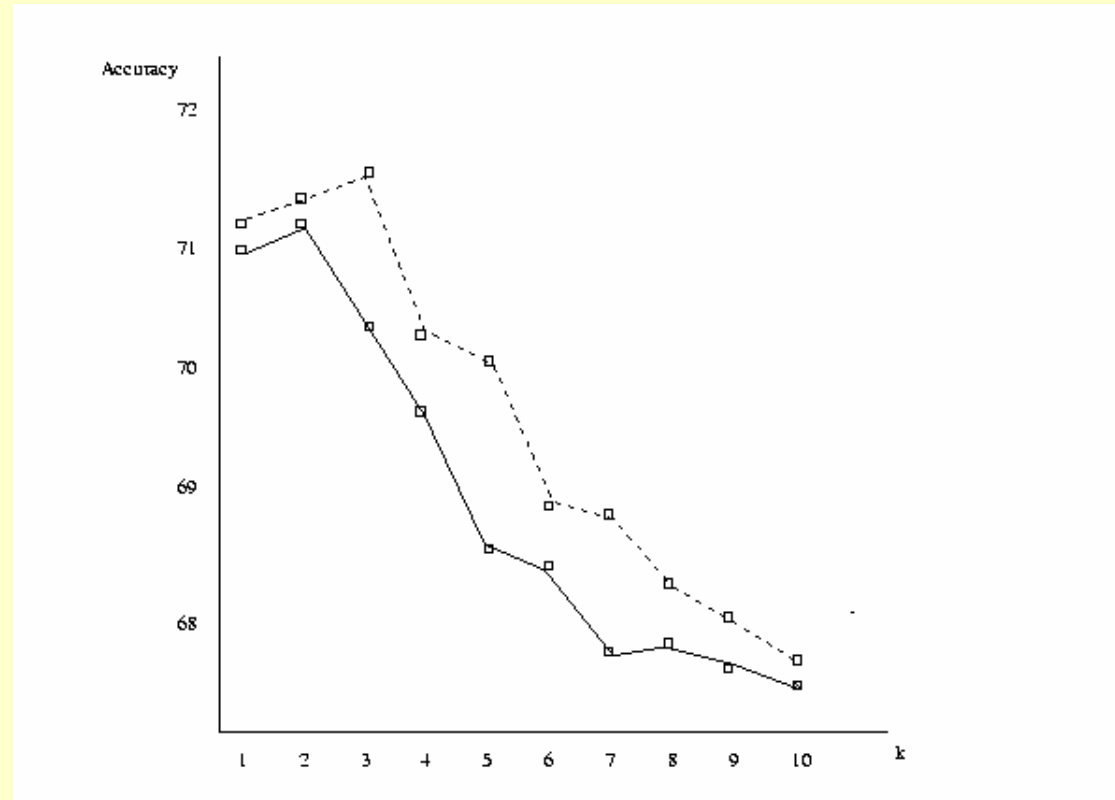
- PSBoost2 : réduction de plus de 50 %, perf. au moins égale.
- plus de 90 % des exemples bruités sont éliminés (30 % bruit)



# PSBoost\_MC

## Principe :

- Si  $Y$  a  $c$  modalités, on mène en parallèle  $c$  *boosting* du type “un contre tous” ;
- On choisit le candidat qui est en moyenne le meilleur, tout en étant en moyenne meilleur qu’un *toss* équilibré.



# 8. Perspectives

**Réduction de la complexité** : structures de voisinage, recherche des plus proches voisins

**Qualité de la représentation** : approche globale, à la fois détection des exemples atypiques et sélection de variable

**Images** : extension de la procédure de segmentation pour la couleur

**Boosting** : test pour le *boosting* de *stumps*

# Bibliographie

---

Cliff A. D., Ord J. K. (1981), *Spatial processes : models & applications*, Pion Limited, London.

Lallich S. (2002), *Mesure et validation en extraction des connaissances à partir des données*, HDR Lyon 2.

Lallich S., Muhlenbach F., Jolion J.-M. (2003), A test to control a region growing process within a hierarchical graph, à paraître *Pattern Recognition*.

Muhlenbach F., Lallich S., Zighed D. (2003), Handling mislabeled instances, accepté *Journal of Information Intelligent Systems*.

Preparata F., Shamos M. (1985), *Pattern recognition and scene analysis*, Springer-Verlag.

Sebban M., Nock R., Lallich S. (2002), Stopping criterion for boosting based data reduction techniques : from binary to multiclass problems, *Journal of Machine Learning Research*.

Toussaint G. (2002), Proximity graph for nearest neighbor decision rules : recent progress, *Proc. interface 2002, 34th Symposium on Computing and Statistics*.

Zighed D. A., Lallich S., Muhlenbach F. (2002), Separability index in supervised learning, *6th European Conference PKDD'02, Helsinki, pp. 475-487*.