Méthodes stochastiques en analyse d'images

- Champs de Markov (MRF), méthodes paramètriques
- **Analyse multi-résolution**, ondelettes
- **Méthodes semi et non paramètriques**, réduction de dimension

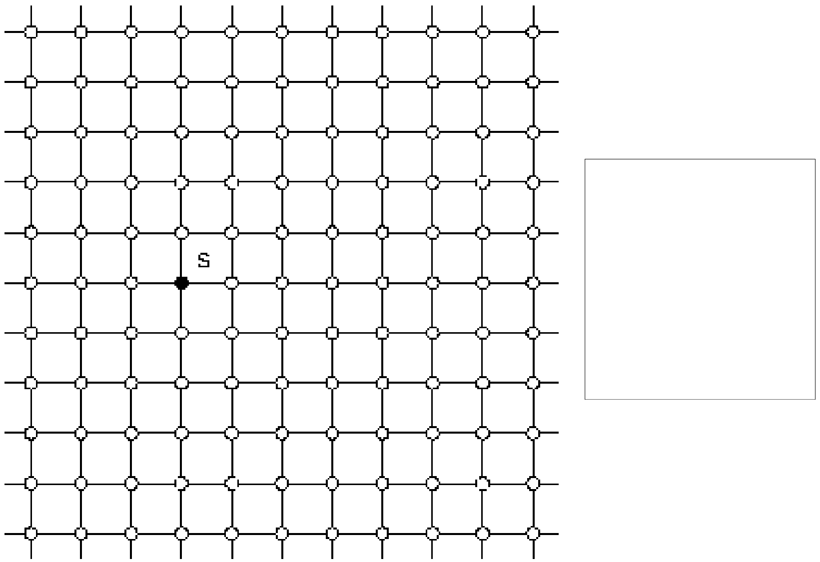
- **Couplage**: ex. Arbres de Markov sur coefficient d'ondelettes, etc.

Champs de Markov en analyse d'images

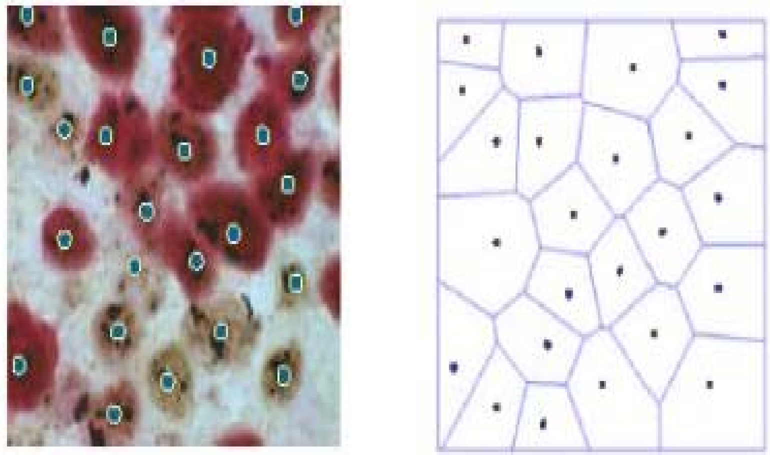
$$\forall z, \quad p(z_i \mid z_{S\setminus\{i\}}) = p(z_i \mid z_j \,, i \in N(i))$$

$$N(i) \text{ : voisins de } i$$

- Modéliser le contexte (spatial) et les interactions



Non indépendance des pixels

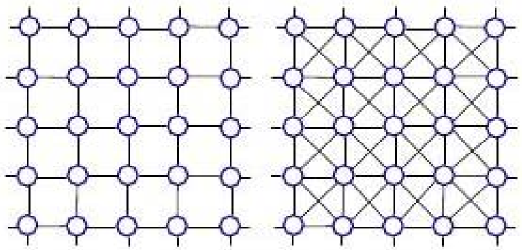


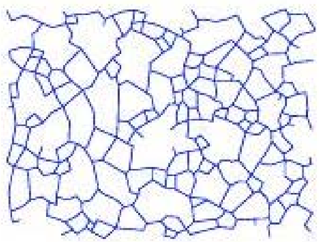
non indépendance des objets

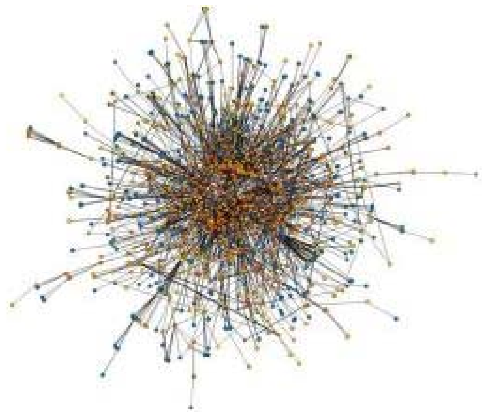
Classification de données structurées

Objectif: Faire des groupes d'individus à partir d'observations

Hypothèses: Dépendances particulières entre les individus, définies à l'aide d'un graphe d'interactions



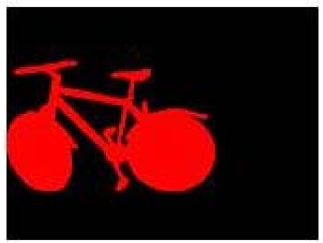




Noeuds = individus/sites

Aretes = interactions directes

Exemple: segmentation d'images



Classification supervisée : observations étiquetées (base d'apprentissage)

Classification non supervisée

Exemples d'applications

Images médicales

- Segmentation d'IRM du cerveau
- Détection des régions d'activation en IRMf
- Détection de tumeurs du sein (angio IRM)

Autres images

- Segmentation d'images satellitaires
- Reconnaissance de textures
- Reconnaissance d'objets

- **Utilisation bas niveau:**

restauration, segmentation, détection de contours

→ **MRF sur grilles régulieres (pixels)**

- **Utilisation + haut niveau:**

reconnaissance d’objets, tracking, matching

→ **MRF sur graphes (points d’intéret et descripteurs d’images)**

Modèles de mélanges pour la classification

Modèle génératif, décrit la formation des classes

$\mathbf{Z} = (Z_1, ..., Z_N)$ les classes (1,..K) les observations $\mathbf{Y} = (Y_1, ..., Y_N)$

Lois des classes a priori $P(z_i = k) = \pi_k$

Les individus (i=1,…N) sont **indépendants**

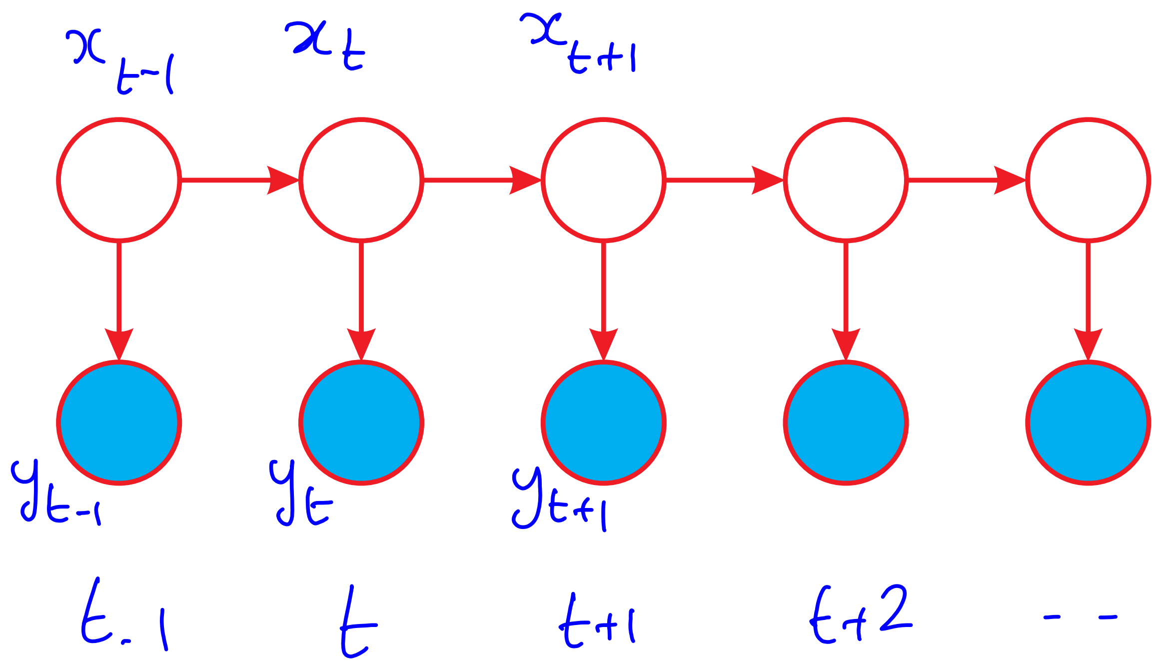
Les lois (multivariées) P(yi | zi=k) dépendent de paramètres

ex. Mélange de gaussiennes $P(y_i | z_i = k) = \mathcal{N}(y_i | \mu_k, \sigma_k^2)$

Example 2: State Space Models

- Hidden Markov chain
- Kalman filter

$$\ldots\, p(x_t|x_{t-1})\, p(y_t|x_t)\, p(x_{t+1}|x_t) \ldots$$



- Frequently wish to solve the problem of computing

$$P(x_t | y_1, \ldots, y_n)$$

Modèle de Champ de Markov caché, interaction de paires

Observations : $\mathbf{Y} = (Y_1,\dots,Y_N)$ Etiquettes manquantes : $\mathbf{Z} = (Z_1,\dots,Z_N)$

Z champ de Markov : $\forall z,\ P(z_i \mid \mathbf{Z}_{S\setminus i} = \mathbf{z}_{S\setminus i}) = P(z_i \mid \mathbf{Z}_{N(i)} = \mathbf{z}_{N(i)})$

➔**Dépendance spatiale** entre les étiquettes qui permet une régularisation de la segmentation.

Un seul voxel gris au milieu d'un bloc blanc : peu de chance.

+ Hypothese standard d'independance conditionnelle: $P(Y|Z) = \prod_i P(Y_i \mid Z_i)$

- Segmenter l'image : Maximiser la probabilité :

$$p(\mathbf{z}|\mathbf{y},\psi) = W^{-1}\exp(-H(\mathbf{z}|\mathbf{y},\psi)) \quad avec \quad W = \sum_{z}\exp(-H(\mathbf{z}|\mathbf{y},\psi))$$

$$avec \quad H(\mathbf{z}|\mathbf{y},\psi) = \sum_{i=1}^{N}\left({}^t z_i \alpha_i + \frac{\beta}{2}\sum_{j\in N(i)} {}^t z_i z_j\right) + \sum_{i\in S}\log p(y_i|z_i,\theta_y)$$

Champ externe

Régularisation spatiale

Terme d'attache aux données (Modèles d'intensités gaussiens)

$\alpha_i = (\alpha_i^1, \dots, \alpha_i^K)$

Idée intuitive : α_i pondère l'importance relative des différentes étiquettes au site i

$$avec \quad \psi = (\theta_y, \theta_z),\quad \theta_z = (\alpha_1,\dots,\alpha_N,\beta),\quad \alpha_i \in R^K$$

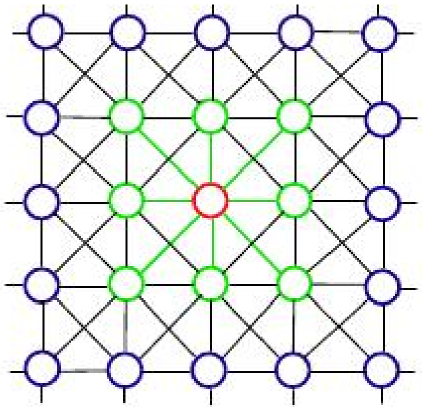
Modèle de Potts et généralisations

Définition : $\mathbf{Z}$ est un champ de Markov ssi

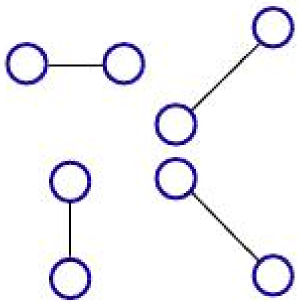
$$\forall \mathbf{z},\ \forall i,\ P(z_i|\mathbf{z}_{\mathcal{I}\setminus i}) = P(z_i|\mathbf{z}_{N_i})$$

$\Leftrightarrow$ $\mathbf{Z}$ suit une distribution de Gibbs :

$$P_G(\mathbf{z}) = W^{-1}\exp(\sum_{i\sim j} V(z_i, z_j))$$



en supposant les potentiels homogènes et isotropes.



modèle de Potts:

$V(k,k') = \beta\ \delta_{k=k'}$

Paramètres = matrice symétrique $\mathbb{V} = (V(k,k'))_{k,k'\in\mathcal{K}}$

$V(k,k') \approx$ coefficient de compatibilité entre les classe k et k'

W ne peut être calculé.

Des approximations sont nécessaires.

Un problème de vision (incertitude, contraintes spatiales)

→ **un problème d'optimisation**

3 piliers (critiques) d'une approche par MRF

- **Modélisation (choix de modèles:** **modele de bruit, distributions des classes, nombre de classes, dimensions Intrinseques…)**
- **Estimation des paramètres:** **en presence de donnees manquantes**
- **Méthodes d'optimisation:** **algorithmes d'inference, approximations…**

Classification non supervisée (paramètres inconnus)

Approche *image*: objectif classification

Méthodes de classification (segmentation) supervisée adaptées au cas des paramètres inconnus → *Classifications dures*

mises à jour itérativement, alternées avec une étape d'estimation des paramètres

Ex: recuit simulé, ICM, etc…

Approche *mélange*: objectif estimation des paramètres

Classification vue comme un problème à données manquantes

→ *Classifications floues* (probabilités d'etre dans chaque classe)

Algorithme EM et variantes

Richesses et points positifs du modèle de champ de Markov caché

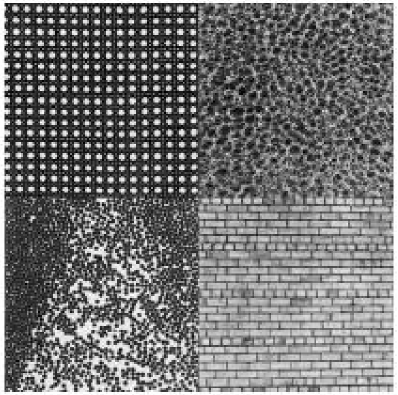
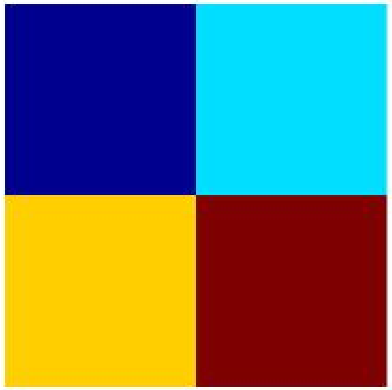
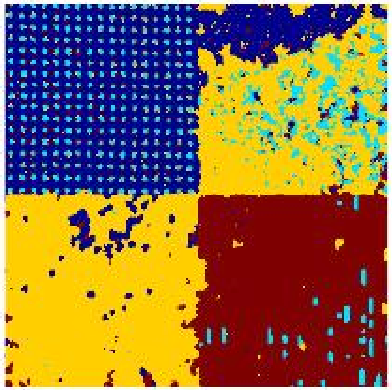
- Champ de Markov
  - Intégration de connaissances a priori : eg. champ externe
  - Régularisation spatiale: interaction de paires (fusion d'information)
  - Techniques d'inférences: algorithmes divers et variés
- Mélanges
  - Modèles gaussiens pour la grande dimension
  - Détection de outliers, classes supplémentaires
  - S'adapte au cas d'observations manquantes
  - Classification *probabiliste*, Interprétation simple
- **Limites principales**
  - Indépendance conditionnelle: bruit trop simple
  - Distributions gaussiennes unimodales
  - Hypothèse de classes non chevauchantes

Modèles de Markov Triplets

Le modèle de champ de Markov caché à bruit indépendant suppose :

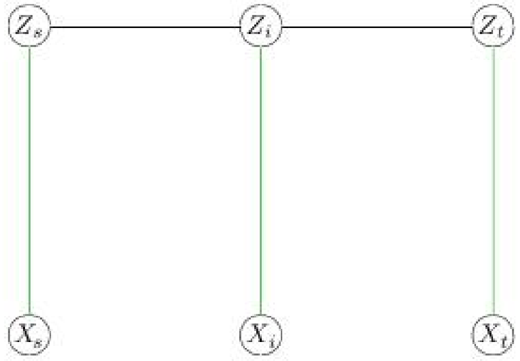
(A) $\mathbf{Z}$ est un champ de Markov

(B) Le bruit est indépendant : $P(\mathbf{x}|\mathbf{z}) = \prod_{i \in \mathcal{I}} P(x_i|z_i)$

 $\rightarrow$  $\neq$ 

On va chercher à relâcher l'hypothèse de bruit indépendant (B).

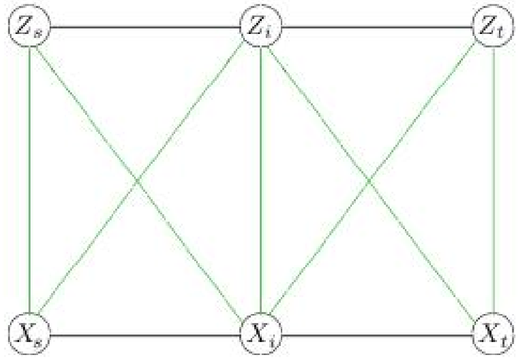
Remarque : En pratique dans les algorithmes, l'hypothèse (A) n'est pas nécessaire puisque seule est utilisée la markovianité de $\mathbf{Z}|\mathbf{X} = \mathbf{x}$ (et non directement celle de $\mathbf{Z}$).



$(\mathbf{X},\mathbf{Z})$ est un champ de Markov caché à bruit indépendant

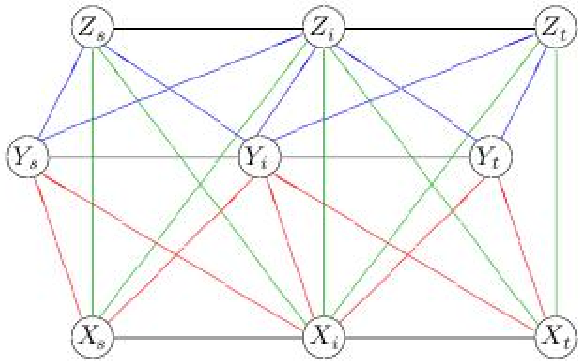
(A) $\mathbf{Z}$ est markovien

(B) $P(\mathbf{x}|\mathbf{z}) = \prod_{i\in\mathcal{I}} P(x_i|z_i)$



$(\mathbf{X},\mathbf{Z})$ est un champ de Markov couple si sa distribution est markovienne :

$$P_G(\mathbf{x},\mathbf{z}) \propto \exp(\sum_{c\in\mathcal{C}} V_c(\mathbf{x}_c,\mathbf{z}_c))$$



Ajout d'un champ auxilliaire $\mathbf{Y} = (Y_i)_{i\in\mathcal{I}}$ discret, $Y_i \in \mathcal{L}$

$(\mathbf{X},\mathbf{Y},\mathbf{Z})$ est un champ de Markov triplet si sa distribution est markovienne :

$$P_G(\mathbf{x},\mathbf{y},\mathbf{z}) \propto \exp(\sum_{c\in\mathcal{C}} V_c(\mathbf{x}_c,\mathbf{y}_c,\mathbf{z}_c))$$

Classes chevauchantes

Un individu dans plusieurs classes à la fois

→ **Pertinence dans les applications**

- Problème de résolution (eg. Volume Partiel en IRM)
- Multi-fonctionnalité/genre (eg. Classification de gènes, films)

→ **Modèles de mélanges plus généraux**

– Astuce de modélisation

– Différents des méthodes de seuillage de probabilités

→ Voir les fuzzy MRFs….

Réduction de dimension

Nouveaux modèles gaussiens de données de grande dimension pour la classification

- Discrimination: ref Bouveyron &al HDDA
- Clustering: ref Bouveyron &al HDDC

- Réduction de dimension et interactions spatiales [Blanchet &al]

- Autres modèles (non gaussiens): ACP généralisée [ref SG]

Détection de données aberrantes

Contexte mélanges:

- Ajout d'une ou plusieurs classes (eg. Uniforme)
- Mélange de lois de student (queues lourdes)

M-estimateurs

Méthodes de Trimming…

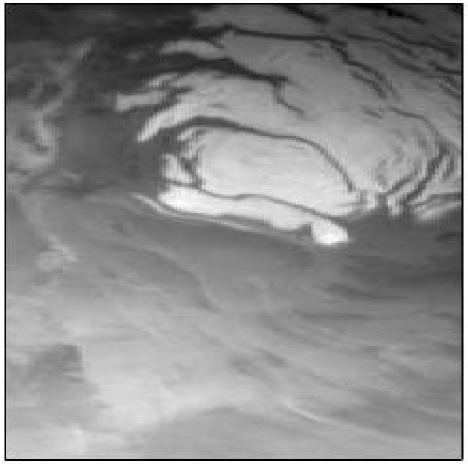
Nombreuses applications: IRM, télédétection, génomique, vision…

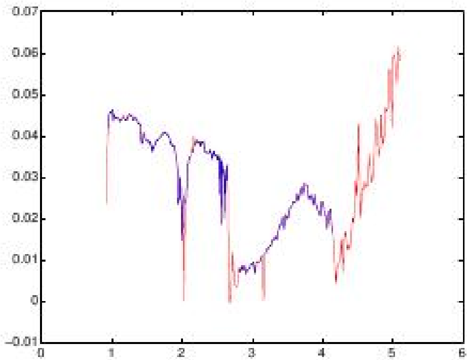
Etat de l'art à faire….

Cas d'observations manquantes

Certaines observations sont manquantes :

- elles n'ont pas pu être observées.
  Ex : sondé ne répondant pas à une question
- elles sont incohérentes.
  Ex : spectres de Mars Express (Labo. de Planétologie de Grenoble)





Objectif : classer les individus à partir d'observations incomplètes et du graphe d'interaction.

Se ramener à des observations complètes...

- suppression des vecteurs d'observations incomplets
- imputation : remplacer les observations manquantes par une valeur "raisonnable" (zéro, moyenne, k-plus proches voisins...)

... puis classer les individus à partir des observations complétées.

→ Ne tient compte ni de l'incertitude sur les valeurs imputées, ni du graphe, ni de l'objectif de classification.

→ Modifie les relations entre les observations (sous-estime la variabilité).

Approche probabiliste ne nécessitant pas
le remplacement préalable
des valeurs manquantes

Conclusion: Pistes de recherche

- Sources de complexité des données:
  - Grande dimension
  - Modèles d'interactions et de bruit
  - Observations manquantes

- Nature des classifications
  - Probabilistes: mesure de confiance, combinaison de modèles et d'estimateurs
  - Classes chevauchantes
  - Classe de données atypiques

- Choix de Modèles:
  - quel modèle de classification? modèle de bruit, distributions des classes, nombre de classes, dimensions intrinsèques…

- Mélange indépendant :

$$P(\mathbf{x},\mathbf{z}) \quad = \quad \prod_{i\in\mathcal{I}} \pi_{z_i} f(x_i|\theta_{z_i})$$

- Champ de Markov caché :

$$P(\mathbf{x},\mathbf{z}) \quad \propto \quad \text{exp}(c\sum_{i\sim j} 1_{z_i=z_j}) \times \prod_{i\in\mathcal{I}} f(x_i|\theta_{z_i}))$$

- Mélange de mélange indépendant :

$$P(\mathbf{x},\mathbf{y},\mathbf{z}) \quad = \quad \prod_{i\in\mathcal{I}} \lambda_{y_i z_i} \pi_{z_i} f(x_i|\theta_{y_i z_i})$$

- Champ de Markov triplet :

$$P_G(\mathbf{x},\mathbf{y},\mathbf{z}) \propto \text{exp}(\sum_{i\sim j}[B_{z_i z_j}(y_i,y_j)+c]\, 1_{z_i=z_j}) \times \prod_{i\in\mathcal{I}} f(x_i|\theta_{y_i z_i})$$

où f est gaussienne de matrice de covariance diagonale ou paramétrée pour la grande dimension [Bouv07]

Approche et spécificités

- **Originalité:**

Stratégie de modélisation statistique qui repose sur la notion de structure dans

- **les systèmes/phénomènes complexes**: modélisation, interprétation
- **les modèles**: estimation, sélection, techniques d'approximation
- **les données**: réduction de dimension, analyse de données non linéaire

Etude théorique et pratique des méthodes:

- **qualité** des approximations,
- **comportement asymptotique** des estimateurs,
- **convergence** des algorithmes

- **Pertinence: Couplage modèles/données, au-dela des modèles statistiques traditionnels**