

# Systemes Intelligents : Raisonnement et Reconnaissance

James L. Crowley

Deuxième Année ENSIMAG

Deuxième Semestre 2005/2006

Séance 7

22 mar 2006

## Reconnaissance Bayesienne

Notations .....	2
Introduction.....	3
La classification.....	4
Partition de l'Espace des Caractéristiques.....	5
La Probabilité d'un Evénement.....	6
Définition Fréquentielle.....	6
Définition Axiomatique.....	6
La probabilité de la valeur d'une variable aléatoire.....	7
Densité de Probabilité.....	8
La Règle de Bayes.....	10

Sources Bibliographiques :

"Neural Networks for Pattern Recognition", C. M. Bishop, Oxford Univ. Press, 1995.

"Pattern Recognition and Scene Analysis", R. E. Duda and P. E. Hart, Wiley, 1973.

**Notations**

$x$	Une variable
$X$	Une valeur aléatoire (non-prévisible).
$N$	Le nombre de valeurs possible pour $x$ ou $X$
$x$	Un vecteur de $D$ variables
$X$	Un vecteur aléatoire (non-prévisible).
$D$	Nombre de dimensions de $x$ ou $X$
$E$	Une événement.
$A, B$	des classes d'événements.
$T_k$	La classe $k$
$k$	Indice d'une classe
$K$	Nombre de classes
$M_k$	Nombre d'exemples de la classe $k$ .
$M$	Nombre totale d'exemples de toutes les classes
	$M = \sum_{k=1}^K M_k$
$k$	L'affirmation que l'événement $E = T_k$
$h(x)$	Histogrammes des valeurs ( $x$ est entières avec range limité)
$h_k(x)$	Histogramme des valeurs pour la class $k$ .
	$h(x) = \sum_{k=1}^K h_k(x)$
$Q$	Nombre de Cellules dans $h(x)$ . $Q = N^D = 2^{BD}$
$p_k$	Proposition que l'événement $E = T_k$ la classe $k$
$p(k) = p(E = T_k)$	Probabilité que $E$ est un membre de la classe $k$ .
$Y$	La valeur d'une observation (un vecteur aléatoire).
$P(X)$	Densité de Probabilité pour $X$
$P(X   k)$	Densité de Probabilité pour $X$ etant donné que $k$
	$P(X) = \sum_{k=1}^K P(X   k) p(k)$

## Introduction

En Séances 5 et 6 on a posé la question :

Qu'est que c'est à comprendre?

Association entre choses perçues et choses connus.

On a proposé que la compréhension est la capacité de trouver la correspondance entre les entités perçues et les éléments d'un "schéma".

Les schémas fournissent une représentation déclarative pour la connaissance.

On a défini les Schémas comme une structure déclarative pour représenter des contextes. Ils sont composés de rôles et de relations.

Les concepts fondamentaux sont :

Observation : mesure produit par un capteur.(numérique ou symbolique)  
ou calculé à partir des mesures

Entité : Une association d'Observations.  
L'association est typiquement par corrélation de position, mais peut être par d'autres propriétés.

Rôle Un prédicat sur les entités dans un domaine  
Cette définition est fondée sur les actions rendus possible par l'entité.

Comme déterminer si une entité peut assumer un "rôle"?

Par la classification.

## La classification

La classification est une capacité fondamentale de l'intelligence.

Reconnaissance : Le fait de reconnaître, d'identifier un objet, un être comme tel.

Identifier : Reconnaître un individu

Classer : Reconnaître un membre d'une catégorie, ou d'une classe.

Un ensemble est défini par un test d'appartenance.

La classification est un processus d'association d'une observation à une classe par un teste d'appartenance. Le test d'appartenance s'applique à un vecteur de caractéristiques. Ce vecteur des caractéristiques décrit un "événement", par exemple l'observation d'une entité.

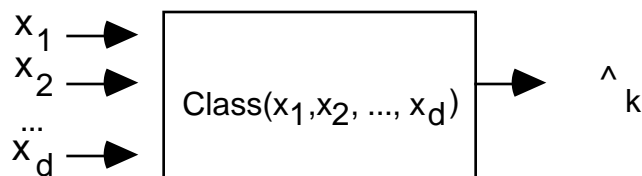
Caractéristiques : (En anglais : Features) Signes ou ensembles de signes distinctifs.

Une ensemble de propriétés.  $\{ x_1, x_2 \dots x_D \}$ .

En notation vectorielle :

$$X = \begin{matrix} x_1 \\ x_2 \\ \dots \\ x_D \end{matrix}$$

Pour un vecteur de caractéristique,  $X$ , un processus de classification propose une estimation de la classe,  $\hat{k}$



Les techniques de reconnaissance de formes, statistiques fournissent une méthode pour induire des tests d'appartenance à partir d'un ensemble d'échantillons.

Les classes peuvent êtres définis par

extension : une liste complète des membres

intention : une conjonction de prédicats.

par extension : Une comparaison d'une observation avec des membres connus de la classe (des prototypes). Ceci correspond (grosso modo) à des méthode dite "généralisante" de reconnaissance.

Dans ce cas, la classification peut être fait par comparaison avec les membre de la classe.

$$\hat{k} = \arg\text{-max}_k \{ \text{Sim}(Y, X_m^k) \} \quad \text{pour tout } k, m \quad \text{ou bien}$$

$$\hat{k} = \arg\text{-min}_k \{ \|Y - X_m^k\| \} \quad \text{pour tout } k, m$$

par intention : Conjonction de prédicats définis sur les propriétés observées. Ceci correspond (grosso modo) à des méthode dite "discriminative" de reconnaissance.

Généralisantes : Les techniques fondaient sur un modèle.

Discriminatives : Les techniques fondaient sur des tests quelconques.

La teste d'appartenance est une forme de partition de l'espace de caractéristiques.

### Partition de l'Espace des Caractéristiques

La classification se résume à une division de l'espace de caractéristique en partition disjoint. Cette division peut-être fait par estimation de fonctions paramétrique ou par une liste exhaustives des frontières.

Les caractéristiques définit une espace.

La classification se résume à une division de cet espace en partition disjoint de région selon la probabilité d'appartenance dans les classes.

Etant une observation,  $Y$ , le critère de partition est la probabilité.

$$\hat{k} = \arg\text{-max}_k \{ p(k | Y) \}$$

## La Probabilité d'un Événement.

La sémantique (ou "sens") de la probabilité d'un événement peut être fourni par sa fréquence d'occurrence ou par un système d'axiomes. L'approche fréquentielle a l'avantage d'être facile à comprendre. Par contre, elle peut entraîner les difficultés dans l'analyse. La définition axiomatique favorise les analyses mathématiques.

Dans le deux cas, la probabilité est une fonction numérique,  $\Pr() \in [0, 1]$ .

Le domaine de la fonction  $\Pr()$  est les événements,  $E$ .

### Définition Fréquentielle.

Une définition "Fréquentielle" de la probabilité sera suffisante pour la plupart des techniques vues dans ce cours.

Soit  $M$  observations des événement aléatoire dont  $M_k$  appartiennent à la classe  $A_k$ .

La Probabilité d'observer un événement  $E$  de la classe  $A_k$  est

$$p(E = A_k) = \lim_M \left\{ \frac{M_k}{M} \right\}$$

Pour le cas pratique où  $M$  est fini,  $p(E = A_k) = \frac{M_k}{M}$

La validité (ou précision) de l'approximation dépend du nombre d'échantillons  $M$ .

### Définition Axiomatique.

Une définition axiomatique permet d'appliquer certaines techniques d'analyse de systèmes probabilistes. Trois postulats sont suffisants :

Postulat 1 :  $A_k \in S : p(E = A_k) \geq 0$

Postulat 2 :  $p(E \in S) = 1$

Postulat 3 :

$$A_i, A_j \in S \text{ tel que } A_i \cap A_j = \emptyset : p(E \in A_i \cup A_j) = p(E \in A_i) + p(E \in A_j)$$

## La probabilité de la valeur d'une variable aléatoire

Pour  $x$  entier, tel que  $x \in [x_{\min}, x_{\max}]$ , on peut traiter chacun des valeurs possibles comme une classe d'événement.

Si les valeurs de  $x$  sont entières, tel que  $x \in [x_{\min}, x_{\max}]$  on peut estimer la probabilité a partir de  $M$  observations de la valeur,  $\{X_m\}$ .

Pour estimer la probabilité d'une valeur on peut compter le nombre d'observation de chaque valeur,  $x$ , dans une table,  $h(x)$ .

L'existence des ordinateurs avec des centaines de megabytes rendre des tables de fréquence très pratique pour la mise en œuvre en temps réel des algorithmes de reconnaissance. Dans certains domaines, comme l'analyse d'images, par abus de langage, un tel table s'appelle une histogramme. Proprement dit, l'histogramme est une représentation graphique de  $h(x)$

Ainsi la probabilité d'une valeur de  $X \in [X_{\min}, X_{\max}]$  est la fréquence d'occurrence de la valeur. Avec  $M$  observations de la valeur,  $X$ , on peut faire une table,  $h(x)$ , de fréquence pour chacun des valeurs possibles. On observe  $M$  exemples de  $X$ ,  $\{X_m\}$ .

Pour chaque observation on ajoute "1" à son entré dans la table.

$$m=1, M : h(X_m) := h(X_m) + 1; M := M+1;$$

$h(x)$  est une table de fréquence pour chaque  $x \in [x_{\min}, x_{\max}]$ .

Ainsi, on peut définir la probabilité d'une valeur  $x$  par sa fréquence :

$$p(X_m=x) = \frac{1}{M} \sum_{m=1}^M h(x)$$

Quand  $M$  est fini, on peut faire appel à l'approximation.

$$P(X=x) \approx \frac{1}{M} \sum h(x)$$

La validité de l'approximation depend dunombre de valeurs possible et de  $M$ . En règle générale, on dit qu'il faut 10 exemples par cellule de l'histogramme.

La précision de cette estimation depend de la vrai densité  $p(X)$ .

La pire de cas et une densité uniforme. Dans ce cas, on peut demontrer que l'ecart type moyenne de l'erreur est en proportion avec la ratiion de le nombre de cellule de  $h(x)$ ,  $N$ , sur le nombre d'echantillons,  $M$ .

MSE —

Que faire si la masse d'exemple est insuffisante :  $M \ll N$ ?

Que faire si  $x$  n'est pas entier ? Il faut une fonction paramétrique pour  $p(X)$ .

**Densité de Probabilité.**

Une fonction de densité de probabilité  $P(X)$ , est une fonction tel que

$P(X)$  est Real et positive pour tout  $X$ .

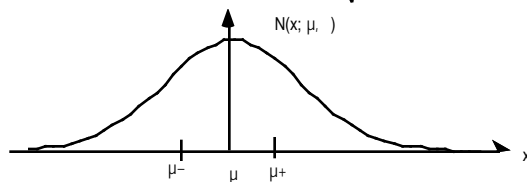
$X$  est réal entre  $[- , ]$

tel que

$$\int_{-} P(x) dx = 1$$

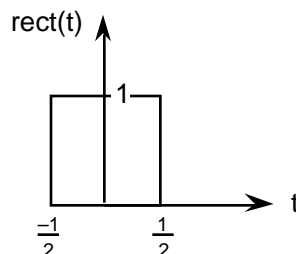
exemples :

Loi Normale  $P(X) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



Mélange de Normales  $P(X) = \sum_{n=1}^N p_n \mathcal{N}(x; \mu_n, \sigma_n^2)$

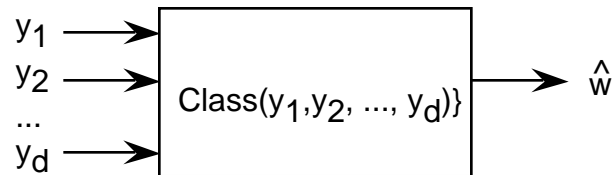
rect :  $P(X) = \text{rect}(X)$ .





La classification est un processus d'association d'une observation à une classe par un teste d'appartenance.

$$\hat{y}_k = \text{Class}\{ Y \}$$



Pour un vecteur de caractéristique il sort une estimation de la classe,  $\hat{w}$

Les techniques de reconnaissance de formes statistiques fournissent une méthode pour induire des tests d'appartenance à partir d'un ensemble d'échantillons.

La classification se résume à une division de l'espace de caractéristique en partition disjoint. Cette division peut-être fait par estimation de fonctions paramétrique ou par une liste exhaustives des frontières.

Le critère est la probabilité.  $\hat{y}_k = \arg\text{-max}_k \{ p(y_k | X) \}$

Cette probabilité est fournie par la règle de Bayes.

$$p(y_k | X) = \frac{p(X | y_k) p(y_k)}{p(X)}$$

## La Règle de Bayes

Soit deux tribus d'événements A et B tel que certains événements sont communs à A et à B.  $A \cap B$  et que A et B sont indépendants.

Soit un événement "E". E peut appartenir à A ou  $\neg A$  à B ou  $\neg B$ .

E peut appartenir à  $A \cap B$  ou à  $\neg A \cap B$  ou à  $A \cap \neg B$  ou à  $\neg A \cap \neg B$

L'appartenance de E à A et B est indépendante de l'appartenance de E à  $\neg A$  et  $\neg B$ .

Soit deux propositions p et q.

donc  $P(p \cap q) = P(p) \cdot P(q)$  et  $P(p \cup q) = P(p) + P(q) - P(p \cap q)$ .

Par axiome 2 de la définition des systèmes de probabilités :

$$P(q) + P(\neg q) = 1.$$

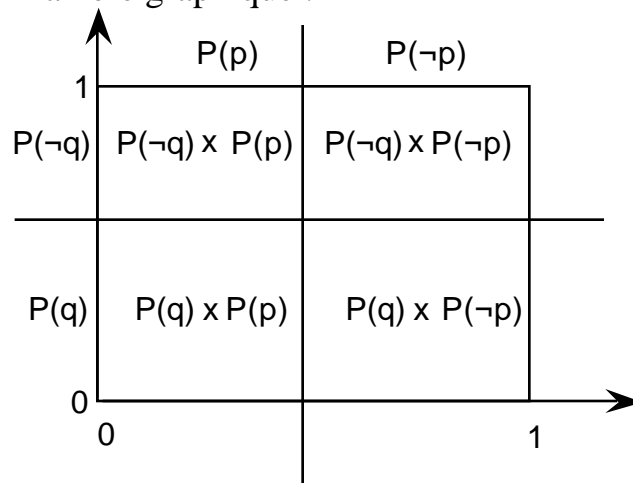
$P(p \cap q)$  est la probabilité "conjointe" de p et q.

Si p et q sont indépendantes

$$P(p \cap q) = P(p) \cdot P(q),$$

$$P(p \cup q) = P(p) + P(q) - P(p \cap q).$$

On peut voir ça d'une manière graphique :



$$P(p \cap q) + P(p \cap \neg q) + P(\neg p \cap q) + P(\neg p \cap \neg q) = 1$$

soit :  $P(p \cup q) = P(p) + P(q) - P(p \cap q)$

$$\begin{aligned} &P(p \quad \neg q) \\ &P(\neg p \quad q) \\ &P(\neg p \quad \neg q) \end{aligned}$$

Dans ce cas, les probabilités marginales sont

$$\begin{aligned} P(p) &= P(p \quad q) + P(p \quad \neg q) \\ P(q) &= P(p \quad q) + P(\neg p \quad q) \end{aligned}$$

La probabilité conditionnelle de q étant donnée p s'écrit  $P(q | p)$

$$P(q | p) = \frac{P(p \quad q)}{P(p)} = \frac{P(p \quad q)}{P(p \quad q) + P(p \quad \neg q)}$$

de la même manière :

$$P(p | q) = \frac{P(p \quad q)}{P(q)} = \frac{P(p \quad q)}{P(p \quad q) + P(\neg p \quad q)}$$

Par algèbre on déduit :

$$P(q | p) P(p) = P(p \quad q) = P(p | q) P(q)$$

d'où

$$P(q | p) P(p) = P(p | q) P(q)$$

Ceci est une forme de règle de Bayes. On peut écrire :

$$P(q | p) = \frac{P(p | q) P(q)}{P(p)}$$

$P(q | p)$  est la probabilité "conditionnelle" ou "postérieur"