Systèmes Intelligents : Raisonnement et Reconnaissance

James L. Crowley

Deuxième Année ENSIMAG — Deuxième Semestre 2006/2007

Séance 7 — 28 mar 2007

# Reconnaissance Bayesienne



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 |
| $\vec{x}$ | Un vecteur de D variables |
| $\vec{X}$ | Un vecteur aléatoire (non-prévisible). |
| $D$ | Nombre de dimensions de $\vec{x}$ ou $\vec{X}$ |
| $E$ | 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 $M = \sum_{k=1}^{K} M_k$ |
| $\omega_k$ | Proposition que l'evennement E la classe k |
| $p(\omega_k) = p(E \quad T_k)$ | Probabilité que E est un membre de la classe k. |
| $h(x)$ | Histogrammes des valeurs (x est entieres 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). $Q = N^D$ |
| $P(\vec{X})$ | Densité de Probabilité pour $\vec{X}$ |
| $P(\vec{X} \mid \omega_k)$ | Densité de Probabilité pour $\vec{X}$ etant donné que $\omega_k$ |

$$P(\vec{X}) = \sum_{k=1}^{K} P(\vec{X} \mid \omega_k)\, p(\omega_k)$$

# La classification

La classification est une capacité fondamentale de la vie. Pour la survie, il faut savoir reconnaître les amis, les ennemies et la nourriture.

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.

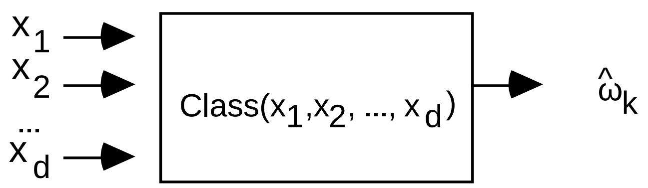
Un ensemble est défini par un test d'appartenance.

La classification est un processus d'association d'un événement à une classe. L'événement est décrit par un vecteur des caractéristiques, produit par une observation. L'affectation de l'événement à une classe est fait par un test, calculer sur le vecteur de caractéristiques.

Caractéristiques : (En anglais : Features) Signes ou ensembles de signes distinctifs.
Une ensemble de propriétés. $\{ x_1, x_2 \ldots x_D \}$.

En notation vectorielle : $\vec{X} = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_D \end{pmatrix}$

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




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 "générative" de reconnaissance.

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

$$\hat{\omega}_k = \underset{\omega_k}{\text{arg-max}}\{\ \text{Sim}(\vec{Y}\ ,\ \vec{X}_m^k\ )\}\quad \text{pour tout k, m}\quad \text{ou bien}$$

$$\hat{\omega}_k = \underset{\omega_k}{\text{arg-min}}\{\ \|\vec{Y}\ ,\ \vec{X}_m^k\|\}\quad \text{pour tout k, m}$$

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

Génératives : 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'éspace de caracteristiques.

## 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, $\vec{Y}$, le critère de particition est la probabilité.

$$\hat{\omega}_k = \underset{\omega_k}{\text{arg-max}}\{\ p(\omega_k \mid \vec{Y}\ )\ \ \}$$

# La Probabilité d'un Evénement.

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, $Pr() \quad [0, 1]$.
Le domaine de la fonction $Pr()$ est les evennements, 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

$$p(E \quad A_k) \equiv \lim_{M \to \infty} \left\{ \frac{M_k}{M} \right\}$$

Pour le cas pratique ou M est fini, $p(E \quad A_k) \approx \frac{M_k}{M}$

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 : $A_k \quad S : p(E \quad A_k) \geq 0$
Postulat 2 : $p(E \quad S) = 1$
Postulat 3 :
$A_i, A_j \quad S$ tel que $A_i \cap A_j = \quad : p(E \quad A_i \quad A_j) = p(E \quad A_i) + p(E \quad A_j)$

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

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

Si les valeurs de x sont entières, tel que x   $[x_{min}, x_{max}]$
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   $[X_{min}, X_{max}]$ est la fréquence d'occurence 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   $[x_{min}, x_{max}]$.

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

$$p(X_m=x) \equiv \lim_{M \to \infty} \left\{\frac{1}{M} h(x)\right\}$$

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

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

La validité de l'approximation depend du nombre 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 ration de le nombre de cellule de h(x), N, sur le nombre d'echantillions, M.

$$\sigma_{MSE} \approx \frac{N}{M}$$

Que faire si la masse d'exemple est insuffisante : M << N?
Que faire si x n'est pas entier ? Il faut une fonction paramétrique pour p(X).

**Densité de Probabilité.**

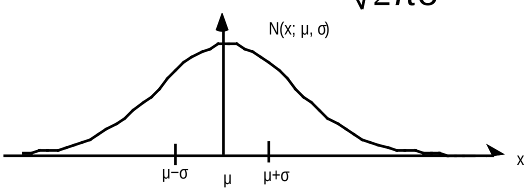
Une fonction de densité de probabilité P(X), est une fonction tel que

P(X) est Real et positive pour tout X.
X est réal entre [–∞, ∞]
tel que

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

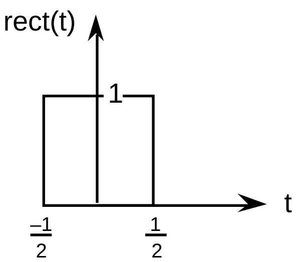
exemples :

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



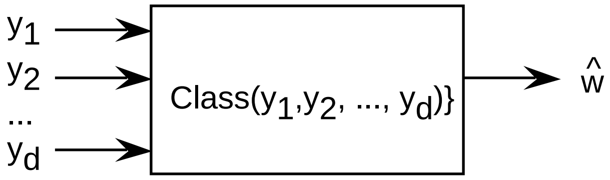

Mélange de Normales $P(X) = \sum_{n=1}^{N} \alpha_n \mathcal{N}(x; \mu_n, \sigma_n)$

rect : P(X) = rect (X).



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

$$\hat{\omega}_k \; ?= \text{Class}\{ \vec{Y} \}$$




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

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 caractèristique 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. $\hat{\omega}_k = \underset{\omega_k}{\text{arg-max}}\{ p(\omega_k \mid \vec{X} ) \}$

Cette probabilité est fournie par la règle de Bayes.

$$p(\omega_k \mid \vec{X} ) = \frac{p(\vec{X} \mid \omega_k ) \; p(\omega_k)}{p(\vec{X})}$$

# La Règle de Bayes

Soit un evennement "E". Soit deux tribus d'evennements A et B tel que certains evennements sont common à A et à B.
E peut appartenir à $A \cap B$ ou a $\neg A \cap B$ ou a $A \cap \neg B$ ou a $\neg A \cap \neg B$

Soit deux propositions p et q.

p ≡ E   A et q ≡ E   B
donc P(p) ≡ Pr{E   A} et P(q) ≡ Pr{E   B}.

Par axiome 2 de la définition des systèmes de probabilités :

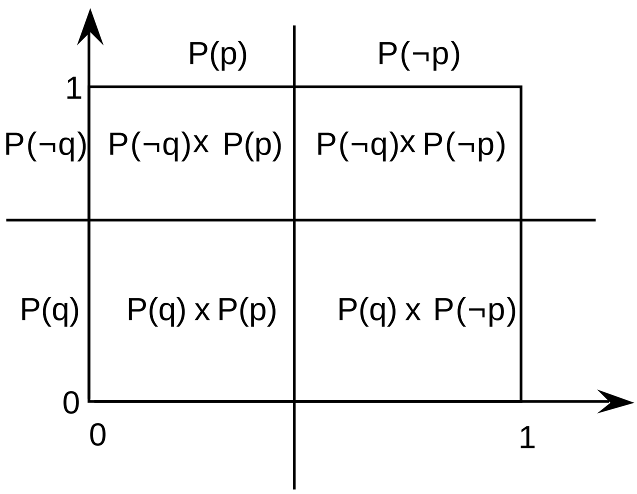
$P(q) + P(\neg q) = 1.$

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

Si p et q sont independentes

P(p   q) = P(p) · P(q),
P(p   q) = P(p) + P(q).

On peut voir ça d'une manière graphique :




P(p   q) + P(p   ¬q) + P(¬p   q) + P(¬p   ¬q) = 1

Dans ce cas, les probabilités marginales sont

$$P(p) = P(p \quad q) + P(p \quad \neg q)$$
$$P(q) = P(p \quad q) + P(\neg p \quad q)$$

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

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

de la même manière :

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

Par algèbre on déduire :

$$P(q \mid p)\, P(p) = P(p \quad q) = P(p \mid q)\, P(q)$$

d'ou

$$P(q \mid p)\, P(p) = P(p \mid q)\, P(q)$$

Ceci est une forme de <u>règle de Bayes</u>. On peut écrire :

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

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