

# Systemes Intelligents : Raisonnement et Reconnaissance

James L. Crowley

Deuxième Année ENSIMAG

Deuxième Semestre 2007/2008

Séance 7

9 avr 2008

## Reconnaissance Bayesienne

Notations .....	2
La classification.....	3
Partition de l'Espace des Caractéristiques.....	4
La Probabilité d'un Evénement.....	5
Définition Fréquentielle.....	5
Définition Axiomatique.....	5
La probabilité de la valeur d'une variable aléatoire.....	6
Densité de Probabilité.....	7
La Règle de Bayes.....	9
La règle de Bayes avec une ratio d'histogrammes.....	11
Exemple : Les statistiques de couleurs de la peau.....	12
Détection par ratio d'histogramme.....	14
La Classification.....	15
La Classification Bayesienne.....	16

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
$E = T_k$	L'affirmation que l'événement $E$ est $T_k$
$M_k$	Nombre d'exemples de la classe $k$ .
$M$	Nombre totale d'exemples de toutes les classes
	$M = \sum_{k=1}^K M_k$
$h(x)$	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 $h(x)$ . $Q = N^D$
$p(k) = p(E = T_k)$	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 $X$
$p(X = x)$	Probabilité q'un vecteur $X$ prendre la valeur $x$
$P(X   k)$	Densité de Probabilité pour $X$ etant donné que $k$
	$P(X) = \sum_{k=1}^K p(X   k) p(k)$

## 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.

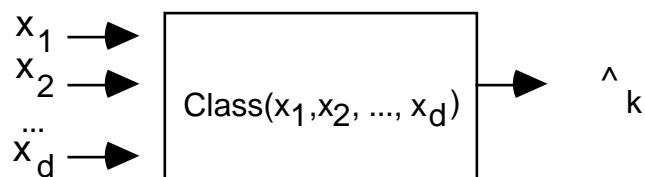
Un ensemble est défini par un test d'appartenance.

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.  $\{x_1, x_2 \dots x_D\}$ .

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).

Dans ce cas, la classification peut être fait par comparaison avec les membre de la classe. Soit  $K$  classe  $k [1, K]$ , et  $M_k$  exemplaires pour chaque classe  $k : X_m$

Soit l'observation  $Y$  :

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

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

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

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'espace 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,  $Y$ , le critère de particition est la probabilité.

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

## 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 qu'un événement  $E$  est issue de la classe  $A_k$  est

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

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

Si l'échantillonnage est "representative", nous pouvons utiliser cette probabilité pour les événements future.

### 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 :

$A_i, A_j \in S$  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 :

$$p(X_m=x) = \lim_M \left\{ \frac{1}{M} h(x) \right\}$$

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

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

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'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  $X$ .

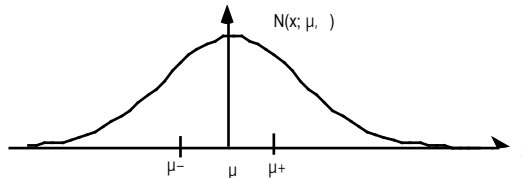
$X$  est réel entre  $[- , ]$

tel que

$$\int P(x) dx = 1$$

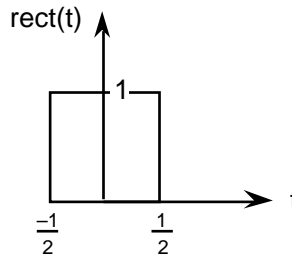
exemples :

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



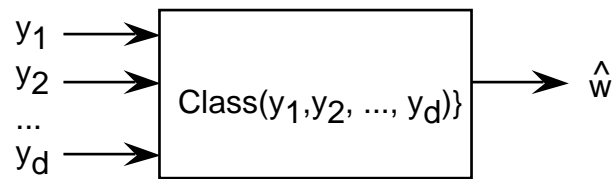
Mélange de Normales  $P(X) = \sum_{n=1}^N p_n \mathcal{N}(x; \mu_n, \sigma_n^2)$

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,  $\hat{y}$

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 un événement "E". Soit deux tribus d'événements A et B tel que certains événements sont commun à A et à B.

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

Soit deux propositions p et q.

donc  $P(p \cap q) = P(p) \cdot P(q)$  et  $P(p \cap \neg q) = P(p) \cdot P(\neg q)$ .

Par axiome 2 de la définition des systèmes de probabilités :

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

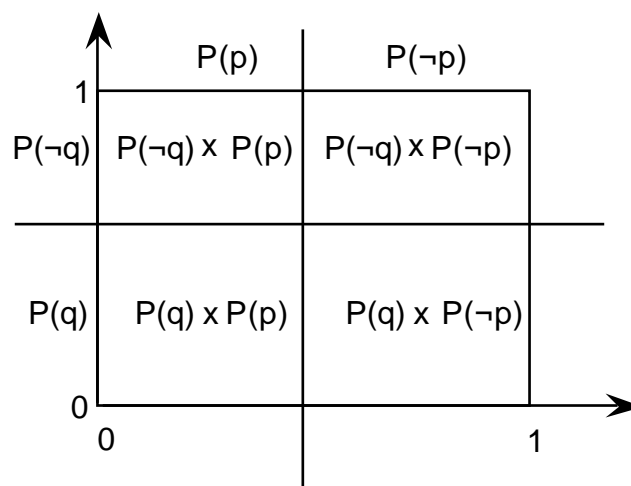
$P(p \cap q)$  est la probabilité "conjointe" de p et q.

Si p et q sont indépendentes

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

$$P(p \cap \neg q) = P(p) \cdot P(\neg q).$$

On peut voir ça d'une manière graphique :



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

Dans ce cas, les probabilités marginales sont

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

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

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

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

de la même manière :

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

Par algèbre on déduit :

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

d'où

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

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

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

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

**La règle de Bayes avec une ratio d'histogrammes.**

On peut utiliser la règle de Bayes pour calculer la probabilité d'appartenance d'un classe. Soit un vecteur de caractéristiques,  $X$  discrètes tel que  $x \in [X_{\min}, X_{\max}]$ , et une ensemble de classe  $k$ .

$$p(k | X=x) = \frac{p(X=x | k)}{P(X=x)} p(k)$$

probabilité de la classe  $k$ :  $p(k) = \frac{M_k}{M}$

probabilité conditionnelle de  $X$ :  $p(X=x | k) = \frac{1}{M_k} h_k(x)$

Probabilité à priori de  $X$ :  $p(X=x) = \frac{1}{M} h(x)$

ce qui donne :

$$p(k | X=x) = \frac{p(X=x | k)}{p(X=x)} p(k) = \frac{\frac{1}{M_k} h_k(x)}{\frac{1}{M} h(x)} \frac{M_k}{M} = \frac{h_k(x)}{h(x)}$$

Cette technique peut également marcher pour les vecteurs de caractéristiques.

La histogramme est une table a  $D$  dimensions :  $h(x)$

probabilité conditionnelle de  $X$ :  $p(X=x | k) = \frac{1}{M_k} h_k(x)$

Probabilité à priori de  $X$ :  $p(X=x) = \frac{1}{M} h(x)$

ce qui donne :

$$p(k | X=x) = \frac{p(X=x | k) p(k)}{p(x)} = \frac{\frac{M_k}{M} \frac{1}{M_k} h_k(x)}{\frac{1}{M} h(x)} = \frac{h_k(x)}{h(x)}$$

Cette technique s'avère très utile dans les cas où il y a suffisamment d'échantillons pour faire un histogramme valable. Par exemple quand on traite des images ou les signaux.

### Exemple : Les statistiques de couleurs de la peau

Une image est une table de pixels.

Chaque pixel est une observation d'une scène, et donc, une variable aléatoire.

Il y a beaucoup des pixels dans les images ( $512 \times 512 = 2^{18} = 256 \text{ K}$  pixels)

Les pixels d'une image couleur sont représenté par 3 octets R, G et B avec (8 bits par octets). Dans ce cas, chaque pixel est une vecteur aléatoire.

$$X = (R, G, B)^T$$

ou R, G, et B sont issue du  $[0, 255]$ .

Pour un vecteur de caractéristique, on peut calculer une table à 3 dimensions.

Pour un image couleur, composé de (R, G, B), avec 8 bits par pixel,  $h(X)$  contient  $256^3 = 2^{24}$  valeurs. Mais chaque image contient  $512^2 = 2^{18}$  pixels.

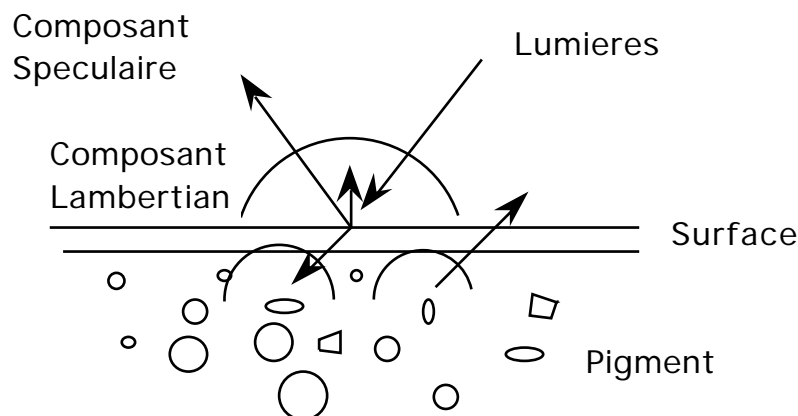
Si on suppose qu'il faut 10 exemples par cellule, Il faut  $10 \times 2^6$  images = 640 images pour une estimation valable de  $p(X) = \frac{1}{M} h(X)$ .

A 25 images per second, ceci est 25.6 seconds de vidéo.

Si on souhaite, on peut reduire le nombre de dimensions, en normalisant la luminance.

On peut transformer le vecteur  $(R, G, B)^T$  en luminance et chrominance.

La luminance, ou intensité, L, est en proportion de  $\cos(i)$  où  $i$  est l'angle entre la source et la normale de la surface. La chrominance,  $C_1, C_2$  est une signature pour la reconnaissance.



La composant "luminant" est déterminé par l'orientation de la surface.

La composant "chrominant" est déterminé par la composition de la spectre de la source et le spectre d'absorption des pigments de la surfaces. Si la spectre de la source est constante, la chrominance indique l'identité de l'objet

Par exemple :

$$L = R+G+B \quad C_1 = \frac{R}{R+G+B} \quad C_2 = \frac{G}{R+G+B}$$

R, G, B sont les entiers. Donc,  $C_1, C_2$  sont issu d'une ensemble finit de valeurs dans l'intervalle  $[0, 1]$ . On peut transformer  $C_1, C_2$  en entier entre  $[0, N-1]$ , par

$$c_1 = \text{Round} \left( N \cdot \frac{R}{R+G+B} \right). \quad c_2 = \text{Round} \left( N \cdot \frac{G}{R+G+B} \right).$$

Donc pour chaque pixel  $(i,j)$ ,  $Y = (C_1, C_2)$

On aura  $N^2$  cellules de chrominances dans l'histogramme.

Par exemple, pour  $N=32$ , on a  $32^2 = 1024$  cellules à remplir est il nous faut que  $M = 10$  K pixels d'exemples. (Une image = 256 K pixels).

Dans ce cas, pour M observations  $p(X=x) = \frac{1}{M} h(x)$

Un histogramme de couleurs,  $h(X)$ , de les M pixels dans une l'image donne une approximation de la probabilité de chaque couleur dans l'image.

$$p(X) = p(X=x) = \frac{1}{M} h(x)$$

Un histogramme des de couleurs d'un entité, A,  $h_A(X)$ , de les  $M_A$  pixels dans une région d'une image de l'objet,  $w(i, j)$ , donne une approximation de la probabilité de chaque couleur de l'objet.

$$p(X | A) = p(X= x | A) = \frac{1}{M_A} h_A(x)$$

**Détection par ratio d'histogramme**

L histogramme permet d'utiliser la règle de Bayes afin de calculer la probabilité qu'un pixel corresponde à un objet.

Pour chaque pixel de chrominance  $X(i, j)$  :  $p(\text{objet} | x) = \frac{p(x | \text{objet}) p(\text{objet})}{p(x)}$

Soit  $N$  images de  $M$  pixels. Ceci fait  $M$  Pixels.

Soit  $h(c_1, c_2)$ , l'histogramme de tous les  $M$  pixels.

Soit  $h_A(c_1, c_2)$ , l'histogramme des  $M_A$  pixels de l'objet "A".

$$p(c(i,j) = A) = \frac{M_A}{M} \quad p(x) = \frac{1}{M} h(x)$$

$$p(x | A) = \frac{1}{M_A} h_A(x)$$

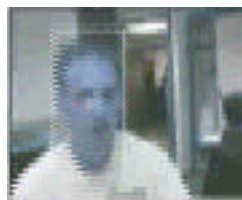
$$\text{Donc } p(A | x) = \frac{p(x | A) p(A)}{p(x)} = \frac{1}{M_A} h_A(x) \frac{\frac{M_A}{M}}{\frac{1}{M} h(x)}$$

$$p(A | Y) = \frac{h_A(x)}{h(x)}$$

Il faut assurer que  $h(x) \neq 0$  !! Pour cela il faut que  $w_A(i,j) = w(i,j)$ .

Ainsi, on peut créer une image de probabilité de l'objet.

En peut ensuite déterminer un seuil pour l'image de probabilité.

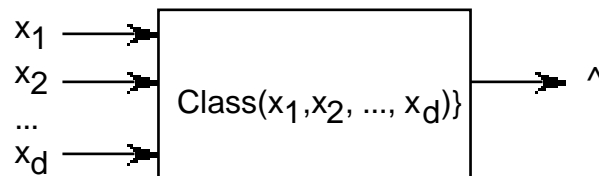


## La Classification

Soit les événements  $E$  décrit par un vecteur de caractéristiques  $X : (E, X)$ .

Soit  $K$  classes d'événements  $\{T_k\} = \{T_1, T_2, \dots, T_K\}$

La classification est un processus d'estimation de l'appartenance d'un événement à une des classes  $T_k$  fondée sur les caractéristiques de l'événement,  $X$ .



$$\hat{k} = \text{Decider}(E \quad k)$$

$\hat{k}$  est la proposition que  $(E \quad k)$ .

La fonction de classification est composée de deux parties  $d()$  et  $g_k()$ :

$$\hat{k} = d(g(X)).$$

$g(X)$  : Une fonction de discrimination :  $\mathbb{R}^D \rightarrow \mathbb{R}^K$

$d()$  : Une fonction de décision :  $\mathbb{R}^K \rightarrow \{K\}$

## La Classification Bayésienne

La technique Bayésienne de Classification repose sur une fonction de vérité probabiliste et le règle de Bayes.

Dans un système de vérité probabilisé, la valeur de vérité de la proposition une probabilité :

$$p(k) = p(E_k)$$

Le critère de décision est de minimiser le nombre d'erreur. Dans un système probabiliste, ca revient de minimiser la probabilité d'erreur. Ceci est équivalent à choisir la classe le plus probable.

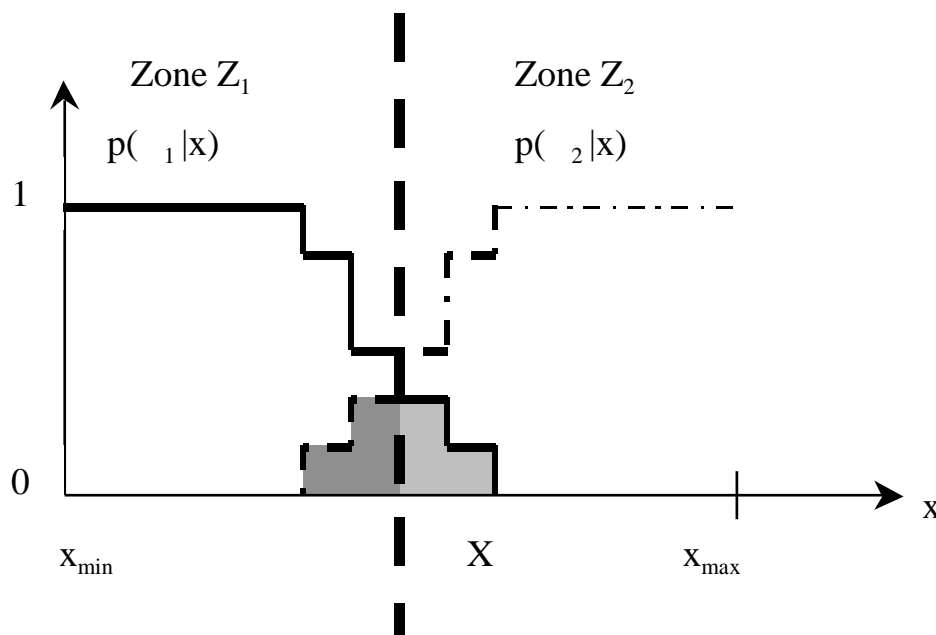
$$\hat{k} = \text{Decider}(E_k) = \arg\text{-max}_k \{p(k | X)\}$$

Pour estimer la probabilité nous utilisons les caractéristiques,  $X$ , de l'événement.

Considère le cas  $D = 1$  et  $K = 2$ . Dans ce cas, le domaine d' $X$  est un axe.

La classification est équivalente à une découpage du domaine d' $X$  en deux zones :  $Z_1$  et  $Z_2$ .

$$\hat{1} \text{ si } X \in Z_1 \text{ et } \hat{2} \text{ si } X \in Z_2$$



La probabilité d'erreur est la somme des probabilités de  $p(2)$  en  $Z_1$  et



la somme de probabilité de  $p(\cdot | X)$  en zone 2.

$$p(\text{erreur}) = \int_{Z_1} p(\cdot | X) + \int_{Z_2} p(\cdot | X)$$

La minimum est atteint quand :

$$\text{Donc } d(g_k(X)) = \arg\text{-max}_k \{p(\cdot | X)\}$$

Dans ce cas, nous avons utilisé  $\arg\text{-max}_k \{p(\cdot | X)\}$  en tant que fonction de décision

et  $g_k(X) = p(\cdot | X)$  comme la fonction de discrimination