Vision par Ordinateur

James L. Crowley

DEA IVR Premier Bimestre 2006/2007

Séance 4 24 octobre 2006

Reconnaissance Probabilistique

Plan de la Séance :

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

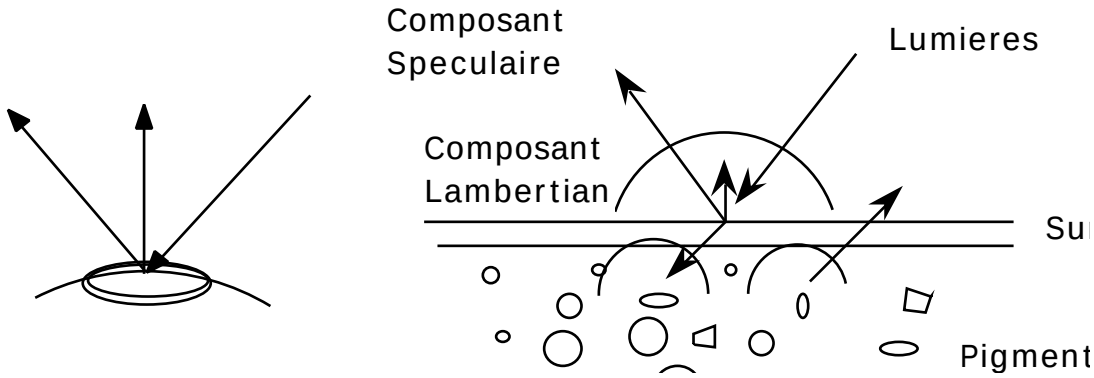
M_k	Nombre d'exemples de la classe k.
M	Nombre totale d'exemples de toutes les classes $M = \sum_{k=1}^{K} M_k$
ω_k	L'affirmation que Evennement E T_k

$h(x)$	Histogrammes des valeurs (x est entieres avec range limité)
$h_k(x)$	Histogramme des valeurs pour la class k. $h(x) = \sum_{k=1}^{K} h_k(x)$
ω_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.
$\vec{Y}$	Une observation (un vecteur aléatoire).
$P(\vec{X})$	Densité de Probabilité pour $\vec{X}$
$p(\vec{X} = \vec{x})$	Valeur de probabilité q'un vecteur $\vec{X}$ prendre la valeur $\vec{x}$
$P(\vec{X} \mid \omega_k)$	Densité de Probabilité pour $\vec{X}$ etant donné que ω_k

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

Analyse Probabiliste au niveau Pixel

Invariants au niveau pixels



Rappelons que l'albédo d'un objet non-métalique peut être approximé par la composition d'une réflexion "spéculaire" et d'une réflexion "lambertienne".

$$R(i, e, g, \lambda) = \rho_S R_S(i, e, g, \lambda) + \rho_l R_l(i, e, g, \lambda)$$

Hors reflets speculaire, pour une reflection lambertian :

$$R_l(i, e, g, \lambda) = p(\lambda) \cos(i)$$

Ceci peut est vue comme une produit de luminance terme cos(i) qui module le luminance et $p(\lambda)$ modulant la chromiance.

Le composant "chrominance" $p(\lambda)$ est déterminé par la composition du spectre de la source et le spectre d'absorption des pigments de la surface. Si le spectre de la source est constant, la chrominance indique l'identité de l'objet.

Ceci donne une indice pour la detection d'objet au niveau pixel.

Description de la chrominance

L'axe luminance, L, peut etre défini par

$$L = R + V + B$$

La chrominance $C = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ est une signature pour l'object.

La chrominance peut être définit par plusieur codages.

Par example, pour la détection du peau, il est fréquent de voir
Normalisation par la luminance laisse deux axes chromatiques : r, v

$$c_1 = r = \frac{R}{R+V+B} \qquad c_2 = v = \frac{V}{R+V+B}$$

Une autre codage fréquente est une codage en "couleur opposée", par exemple :

$$L = \frac{R+G+B}{3}, \qquad c_1 = \frac{R-G}{2} \quad \text{et} \quad c_2 = \frac{R+G}{2} - B.$$

$$\begin{pmatrix} L \\ C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 0.33 & 0.33 & 0.33 \\ 0.5 & -0.5 & 0 \\ 0.5 & 0.5 & -1 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

En tout cas, on peut apprendre et detecter la signature de le chrominance avec d'une maniere statistique.

Histogrammes

Un histogramme est une table de fréquence. Il peut fournir une estimation d'une densité de probabilités.

On peut utiliser les histogrammes pour calculer la probabilité qu'un pixel est une projection d'une objet.

Par example, construisons une histogramme pour le vecteur de chrominance (r,v).

Chaque pixel est un vecteur (r, v) : $\vec{C}(i, j) = \begin{pmatrix} r \\ v \end{pmatrix}(i, j)$

On alloue un tableau 2D de taille N_h (exemple 32 x 32 = 1024 cellules : h(r, v).

Pour chaque pixel $\vec{C}$ = C(i, j) dans l'image, on incrémente la cellule de l'histogramme qui correspond à $\vec{C}$ = (r, v)

h(r, v) := h(r, v) +1 c'a-dire $h(\vec{C}) := h(\vec{C}) + 1$

Soit M Pixels dans l'image. Un histogramme des chrominance, h(**C**), des M pixels dans une l'image donne leurs fréquences d'occurence.

$$P(\vec{C}) \approx \frac{1}{M} h(\vec{C})$$

Pour que l'estimation soit "raisonable", il faut assuré que $M >> N_h$

Considére une région W de M_o pixels du même image correspondance à l'objet O.

$$(i,j) \quad W \; : \; h_o(\vec{C}(i,j)) := h_o(\vec{C}(i,j)) + 1$$

Ensuite: pour tout pixel $\vec{C}(i, j) = \begin{pmatrix} r \\ v \end{pmatrix}(i, j)$: $p(\vec{C}| \text{ objet }) \approx \frac{1}{M_o} \; h_o(\vec{C})$

Parce que W est dans l'image, la probabilité de rencontrer un pixel de W,

$$P(W) = \frac{M_o}{M}$$

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 $\vec{C}(i, j)$ $\quad p(\text{objet} \mid C) = p(C \mid \text{objet}) \frac{p(\text{objet})}{p(C)}$

Soit M images de I x J pixels. Ceci fait N = I x J x M Pixels.
Soit h(r, v), l'histogramme de tous les N pixels.
Soit $h_o(r, v)$, l'histogramme des N_o pixels de l'objet "o".

$$p(\text{objet}) = \frac{M_o}{M}$$

$$p(\vec{C}) = \frac{1}{M} h(\vec{C})$$

$$p(\vec{C} \mid \text{objet}) = \frac{1}{M_o} h_o(\vec{C})$$

Donc $\quad p(\text{objet} \mid \vec{C}) = p(\vec{C} \mid \text{objet}) \frac{p(\text{objet})}{p((\vec{C}))} = \frac{1}{M_o} h_o(\vec{C}) \frac{\frac{M_o}{M}}{\frac{1}{M} h((\vec{C}))}$

$$p(\text{objet} \mid \vec{C}) = \frac{h_o(\vec{C})}{h(\vec{C})}$$

Par exemple, voici une image de la probabilité de peau fait par ratio d'histogramme de r, v

Caractérisation par moments

Les ensemble connexes de pixels s'appelles les "blobs".

On peut décrire une blob par une vecteur de caractéristiques "invariantes" à l'orientation grâce aux "moments"

Les moments sont invariants aux transformations affines.

Pour une febetre (imagette) w(i, j) de taille N x M

Somme des Pixels : $$S = \sum_{i=1}^{M} \sum_{j=1}^{N} w(i, j)$$

Premiers moments :

$$\mu_i = \frac{1}{S} \sum_{i=1}^{M} \sum_{j=1}^{N} w(i, j) \cdot i \qquad \mu_j = \frac{1}{S} \sum_{i=1}^{M} \sum_{j=1}^{N} w(i, j) \cdot j$$

Le premier moment est le centre de gravité de la forme :

Deuxième moment :

$$\sigma_i^2 = \frac{1}{S} \sum_{i=1}^{M} \sum_{j=1}^{N} (w(i, j)) \cdot (i - \mu_i)^2$$

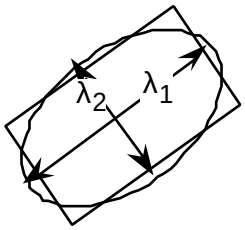
$$\sigma_j^2 = \frac{1}{S} \sum_{i=1}^{M} \sum_{j=1}^{N} w(i, j) \cdot (j - \mu_j)^2$$

$$\sigma_{ji}^2 = \frac{1}{S} \sum_{i=1}^{M} \sum_{j=1}^{N} w(i, j) \cdot (i - \mu_i)(j - \mu_j)$$

Ceci permet de définir les "axes", majeur, λ_1 et mineur, λ_2, de la forme par analyse des composantes principales de la deuxième moment

$$\mathbf{C}_o \triangleq \begin{bmatrix} \sigma_i^2 & \sigma_{ij}^2 \\ \sigma_{ij}^2 & \sigma_j^2 \end{bmatrix}$$

Composantes principales



Les deuxièmes moments sont "invariants" à l'orientation

Les axes sont calculés par une analyse en composantes principales de la matrice C. Il s'agit de trouver une rotation, Φ, dans l'espace de caractéristiques $\Phi \mathbf{C}_P \Phi^T = \Lambda$ telles que Λ soit diagonale.

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \quad \text{tel que } \lambda_1 > \lambda_2 \qquad \Phi = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

$$\Phi \mathbf{C}_P \Phi^T = \Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \qquad \Phi^T \Phi = \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\Phi \mathbf{C}_P \Phi^T \Phi = \Phi \mathbf{C}_P = \Lambda \Phi = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

Les lignes du Φ sont des vecteurs propres du **C.**
La longueur des axes majeurs et mineur est les valeurs propres de la matrice **C**.

θ est l'orientation de l'axe "majeur" et λ_1 / λ_2 est le rapport entre la longueur et la largeur.

λ_1 / λ_2 est une caractéristique invariante de la taille et de l'orientation.

La Classification des Formes

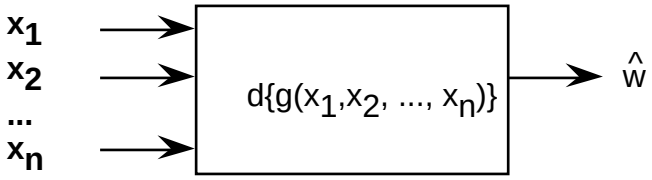
La classification est une capacité fondamentale de l'intelligence.

Classer : Reconnaître un membre d'une catégorie, ou d'une classe.
On peut distinguer "reconnaissance" et "identification".

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

Identifier : Reconnaître un individu

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



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 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 d'appartenance.
Cette probabilité est fournie par la règle de Bayes.

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

Les Caractéristiques

Forme n. f. : A. Apparence, aspect visible. 1) ... 2) apparence extérieure donnant à un objet ou à un être sa spécificité.

Les méthodes statistique de la reconnaissance de forme traite les observations sous forme de vecteur de caractéristiques.

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

En notation vectorielle :

$$\vec{X} = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_D \end{pmatrix}$$

Par exemple, $\vec{X} = \begin{pmatrix} \mu_j \\ \lambda_1 \\ \lambda_2 \\ \theta \end{pmatrix}$ est une vecteur de caractéristique pour les "blobs".

La formation des vrais objets physiques est sujette aux influences aléatoires. Pour les objets d'une classe, w_k, les propriétés des objets individuels sont, les valeurs aléatoires. On peut resume ceci par une somme d'une forme "intrinsèque" $\vec{x}$ plus ces influences aléatoires individuelles, $\vec{B}_i$.

$$\vec{X} = \vec{x} + \vec{B}_i$$

Les techniques probabiliste de RF suppose un bruit additif.

En notation vectorielle :

$$\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \ldots \\ X_n \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix} + \begin{pmatrix} B_1 \\ B_2 \\ \ldots \\ B_n \end{pmatrix}$$

Les Observations

Une observation : une constatation attentive des phénomènes.

Pour des machines, des observations sont fournies par les capteurs.

Ceci donne une observation (un phénomène) sous forme d'une ensemble de caractéristiques : $\{Y_1, Y_2 \ldots Y_D\}$.

$$\vec{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{pmatrix}$$

Les observations sont corrompues par un bruit, $\vec{B}_o$.

$$\vec{Y} = \vec{y} + \vec{B}_o$$

Le bruit est, par définition, imprévisible. Il est aléatoire.
Donc les caractéristiques observées sont des vecteurs aléatoires.
La corruption des observations par un bruit aléatoire est fondamentale aux capteurs physiques.

Pour chaque classe k, la probability d'observé $\vec{Y}$ est fournie par la règle de Bayes.

$$p(\omega_k \mid \vec{Y}) = \frac{p(\vec{Y} \mid w_k)\, p(w_k)}{p(\vec{Y})}$$

Si $\vec{Y} = \vec{X} + \vec{B}_o$ est issu de la classe w_k ayant characteristique $\vec{X} = \vec{x} + \vec{B}_i$

Par exemple :

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

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

$$C_1 = \text{Round}\left(N \cdot \frac{R}{R+G+B} \right). \qquad C_2 = \text{Round}\left(N \cdot \frac{G}{R+G+B} \right).$$

On aura N^2 cellulles 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 $p(\vec{X}) \approx \frac{1}{M} h(\vec{X})$

La probabilité à posteriori peut être calculé par la règle de Bayes.

soit ω_k, la proposition que l'observation $\vec{X}$ la classe k

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

Dans le cas des valeurs de $\vec{X}$ discrètes tel que x $[X_{min}, X_{max}]$, on a

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

probabilité conditionnelle de $\vec{X}$: $p(\vec{X} = \vec{x} \mid \omega_k) = \frac{1}{M_k} h_k(\vec{x})$

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

ce qui donne :

$$p(\omega_k \mid \vec{X}=\vec{x}) = \frac{p(\vec{X}= \vec{x} \mid \omega_k)\ p(\omega_k)}{p(\vec{x})} \approx \frac{\frac{M_k}{M}\ \frac{1}{M_k}\ h_k(\vec{x})}{\frac{1}{M}\ h(\vec{x})} = \frac{h_k(\vec{x})}{h(\vec{x})}$$

La Probabilité

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

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

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'éviter certaine difficulté dans l'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 $x \quad [x_{min}, x_{max}]$, on peut traiter chacun des valeurs possibles comme une classe d'événement.

Si les valeurs de x sont entières, tel que $x \quad [x_{min}, x_{max}]$
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.

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

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 :

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

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

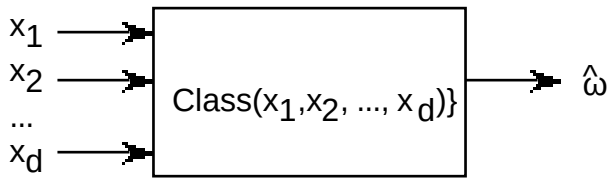
$$P(X=x) \approx \frac{1}{M} h(x)$$

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

Que faire si la masse d'exemple est insuffisante : $M < 10 (X_{max}–X_{min})$?
Que faire si x n'est pas entier ?
Dans ces cas, on peut faire appel à une fonction paramétrique pour p(X).

Fonctions de Discrimination

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



$\hat{\omega}_k$ = Classer(E) = Decider(E A_k)

$\hat{\omega}_k$ est la proposition que (E ω_k).

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

$\hat{\omega}_k = d(g(\vec{X}))$.

$g(\vec{X})$: Une fonction de discrimination : $R^D \rightarrow R^K$
d() : Une fonction de décision : $R^K \rightarrow \{\omega_K\}$

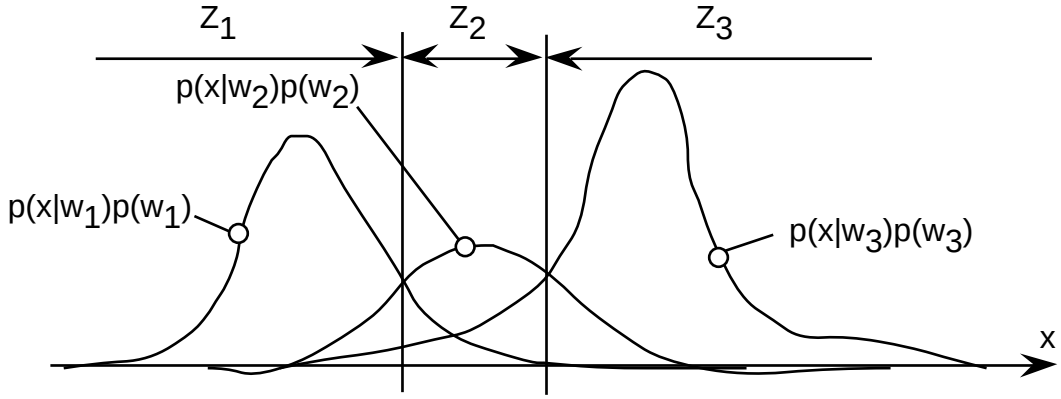
en génerale $g(\vec{X}) = \begin{pmatrix} g_1(\vec{X}) \\ g_2(\vec{X}) \\ \dots \\ g_K(\vec{X}) \end{pmatrix}$ est une vecteur de K fonctions $g_k(\vec{X})$

Dans le cas général, K > 2, la nombre minimum d'erreur sont fait si ω_k est choisi tel que :

$k = \arg\max_k \{g_k(X)\}$ avec $g_k(X) = p(x \mid \omega_k)\, p(\omega_k)$

Les frontières entre régions i et j sont les valeurs pour lesquelles

$g_i(X) = g_j(X)$



Une fonction de discrimination partition l'espace de caractéristique en régions disjointes Z_1, ..., Z_k pour chaque classe.

$$k = \underset{k}{\text{arg-max}}\{g_k(X)\}$$

Mais comment calculer $g_k(X)$?

Les caractéristiques X de l'événement E sont aléatoires avec une dispersion due aux variations naturelles de sa classe.

Ceci est modélisé par une variable aléatoire B_k autours d'une valeur "type" x_k. La valeur type est spécifique à la classe.

$$X = x_k + B_k$$

Si D=1, les membres de la classe ω_k auront les caractéristiques X tel que :

$$p(X=x \mid \omega_k) = \mathcal{N}(x; \mu_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}$$

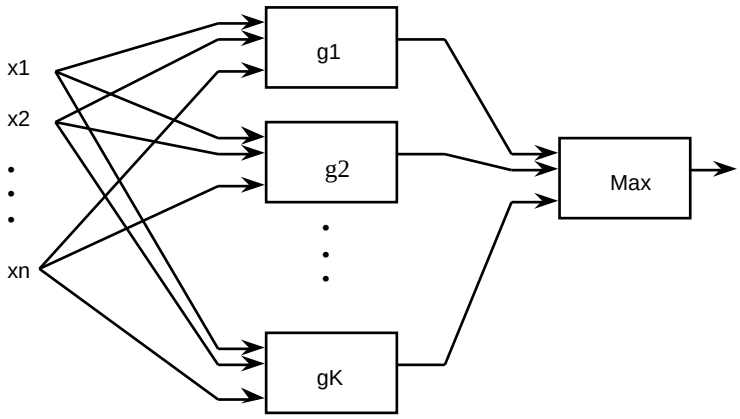
Donc notre fonction de discrimination devient :

$$g_k(X) = p(\omega_k) \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}$$

pour D > 1 il faut la fonction normales multi-variate

$$p(\vec{X}) = \mathcal{N}(\vec{X}; \vec{\mu}, \mathbf{C}_x) = \frac{1}{(2\pi)^{\frac{D}{2}} \det(\mathbf{C_x})^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{X} - \vec{\mu})^T C_x^{-1} (\vec{X} - \vec{\mu})}$$

La classificateur est une machine qui calcule K fonctions $g_k(\vec{x})$ suivie d'une sélection du maximum.



Soit D = 1. (une seul caractéristique).

On peut noter que $k = \underset{k}{\text{arg-max}}\{g_k(X)\} = \underset{\omega_k}{\text{arg-max}}\{\text{Log}\{g_k(X)\}\}$
parce que Log{} est une fonction monotone.

$$k = \underset{k}{\text{arg-max}}\{\text{Log}\{ \frac{1}{\sqrt{2\pi}\sigma_k} \; e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}} \} + \text{Log}\{p(\omega_k)\} \}$$

ou $$k = \underset{k}{\text{arg-max}}\{-\text{Log}\{\sigma_k\} - \frac{(x-\mu_k)^2}{2\sigma_k^2} + \text{Log}\{p(\omega_k)\} \}$$

Dans le cas générale D > 1 La fonction de discrimination devient :

$$g_k(\vec{x}) = -\frac{1}{2} \text{Log}\{\det(C_k)\} - \frac{1}{2}(\vec{X} - \vec{\mu}_k)^T C_k^{-1}(\vec{X} - \vec{\mu}_k) + \text{Log}\{p(\omega_k)\}$$

Cecit peut être traduit dans une forme cannonique :

$$g_k(\vec{X}) = \vec{X}^T (D_k) \vec{X} + \vec{d}_k^{\,T} \vec{X} + d_{ko.}$$

avec une terme 2ieme ordre : $D_k = \frac{1}{2} C_k^{-1}$

une terme 1iere ordre : $\vec{d}_k = C_k^{-1}\vec{\mu}_k$

Constant : $d_{ko} = -\frac{1}{2}(\vec{\mu}_k^{\,T} C_k^{-1}\vec{\mu}_k) - \frac{1}{2} \text{Log}\{\det(C_k)\} + \text{Log}\{p(\omega_k)\}$

La Loi Normale :

Représentation Paramètrique de la Probabilité.

Une "Dénsité" de probabilité (écrit "pdf" pour "Probability Density Function") est une fonction, P(X) represetnant la probabilité pour une variables aléateoire, X
X [–∞, ∞] tel que

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

Quand les variables aléatoires sont issues d'une séquence d'événements aléatoires, leur densité de probabilité prend la forme de la loi normale, $\mathcal{N}(\mu, \sigma)$. Ceci est démontré par le théorème de la limite centrale. Il est un cas fréquent en nature.

Soit M exemple d'observation d'une Evennement E_m : X_m
La loi Normale décrit une population d'exemples $\{X_m\}$.

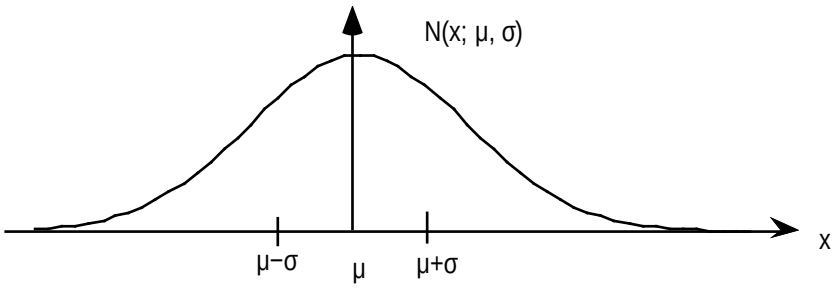
Les paramètres de $\mathcal{N}(\mu, \sigma)$ sont les premiers et deuxième moments de la population.
On peut estimer les moments pour n'importe quel nombre d'exemples (M > 0)

La Loi Normale pour D = 1

La cas le plus simple concerne une seule caractéristique.
Avec μ et σ^2, on peut estimer la densité p(x) par $\mathcal{N}(x; \mu, \sigma)$

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

$\mathcal{N}(x; \mu, \sigma)$ a la forme :



Le base "e" est : e = 2.718281828....Il s'agit du fonction tel que $\int e^x dx = e^x$

Le terme $\frac{1}{\sqrt{2\pi}\sigma}$ sert à normaliser la fonction en sorte que sa surface est 1.

$$\int_{-\infty}^{0} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sqrt{2\pi}\ \sigma.$$

Le terme $d^2(x)= \frac{(x-\mu)^2}{\sigma^2}$ est la difference entre x et μ normalisée par la variance.
La différence $(x - \mu)^2$ est la "distance" entre une caractéristique et la moyenne de l'ensemble $\{X_m\}$. La variance, σ^2, sert à "normaliser" cette distance.

La différence normalisée par la variance est connue sous le nom de "Distance de Mahalanobis". La Distance de Mahalanobis est un test naturel de similarité

Les parametres du $\mathcal{N}(x; \mu, \sigma)$ sont le premier moment μ et le deuxieme moment σ

Estimations des moments d'une densité

Le premier moment : La Moyenne

Soit M observations d'un variable aléatoire, $\{X_m\} : \{ X_1, X_2, ..., X_M\}$
La moyenne est l'espérance de $\{X_m\}$. .

$$\mu \equiv E\{X\} = \frac{1}{M} \sum_{m=1}^{M} X_m$$

Il s'agit d'une somme sur M (le nombre exemples).
On note que dans le cas ou il existe une histogramme pour X, on peut aussi estimer la moyenne par la table de fréquence. La masse d'un histogramme, h(x) est le nombre d'échantillons qui composent l'histogramme, M.

$$M = \sum_{x=x_{min}}^{x_{max}} h(x)$$

Pour X entier, tel que X $[x_{min}, x_{max}]$ on peut démontrer que

$$\mu \equiv E\{X\} = \frac{1}{M} \sum_{x=x_{min}}^{x_{max}} h(x) \cdot x$$

donc : $\mu \equiv E\{X\} = \frac{1}{M} \sum_{m=1}^{M} X_m = \frac{1}{M} \sum_{x=x_{min}}^{x_{max}} h(x)$

Pour X continue la moyenne peut être calculé par le 1e moment du pdf.

$$\mu \equiv E\{X\} = \int p(x) \cdot x \, dx$$

Le deuxième moment (La variance)

La variance σ^2 est le deuxième moment de la densité de probabilité.
Pour un ensemble de M observations $\{X_m\}$

$$\sigma^2 \equiv E\{(X_m-\mu)^2\} = \frac{1}{M}\sum_{m=1}^{M}(X_m - \mu)^2$$

Mais l'usage d'μ estimé avec le même ensemble, introduit un biais dans σ^2.
Pour l'éviter, on peut utiliser une estimation sans biais.

$$\sigma^2 = \frac{1}{M-1}\sum_{m=1}^{M}(X_m - \mu)^2$$

Lequel est correct ? (les deux !) Ils ont les usages différents.

Pour X entier, tel que X $[X_{min}, X_{max}]$ on peut démontrer que

$$\sigma^2 = E\{(X_m-\mu)^2\} = \frac{1}{M}\sum_{x=x_{min}}^{Xmax} h(x)(x-\mu)^2$$

Ceci est vrai par ce que la table h(x) est faite de $\{X_m\}$.

Donc :

$$\sigma^2 = \frac{1}{M}\sum_{m=1}^{M}(X_m - \mu)^2 = \frac{1}{M}\sum_{x=x_{min}}^{xmax} h(x)(x-\mu)^2$$

Pour X réel, la variance est la deuxième moment de la pdf.

$$\sigma^2 \equiv E\{(X_m-\mu)^2\} = \int p(x)\cdot(x-\mu)^2\, dx$$

La Loi Normale pour D > 1

Soit les événements E décrit par une vecteur de D caractéristiques $\vec{X}$

Soit une ensemble de M événements, $\{E_m\}$ avec leurs caractéristiques. $\{\vec{X}_m\}$
Cet ensemble est dit l'ensemble d'entrainement (training set)

$$\mu_d \equiv E\{x_d\} = \frac{1}{M}\sum_{m=1}^{M} X_{dm}$$

Pour le vecteur de D caractéristiques :

$$\vec{\mu} \equiv E\{\vec{X}\} = \frac{1}{M}\sum_{m=1}^{M} \vec{X}_m = \begin{pmatrix}\mu_1\\ \mu_2\\ \dots\\ \mu_D\end{pmatrix} = \begin{pmatrix}E\{X_1\}\\ E\{X_2\}\\ \dots\\ E\{X_D\}\end{pmatrix}$$

Pour M observations $\{\vec{\mathbf{X}}_m\}$, la covariance entre les variables x_i et x_j est

ou $\sigma_{ij}^2 \equiv E\{ (X_i - E\{X_i\})(X_j - E\{X_j\})\} = \frac{1}{M}\sum_{m=1}^{M} (X_{im} - \mu_i)(X_{jm} - \mu_j))$

Ces coefficients composent une matrice de covariance. $\mathbf{C}_x$

$$\mathbf{C}_x = \begin{pmatrix}\sigma_{11}^2 & \sigma_{12}^2 & \dots & \sigma_{1D}^2\\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2D}^2\\ \dots & \dots & & \dots\\ \sigma_{D1}^2 & \sigma_{D2}^2 & \dots & \sigma_{DD}^2\end{pmatrix}$$

En matrice en écrit :

Soit $\vec{V} = \vec{X} - E\{\vec{X}\} = \vec{X} - \vec{\mu}$

$$\mathbf{C}_x \equiv E\{ \vec{V}\ \vec{V}^T \} \equiv E\{[\vec{\mathbf{X}} - \vec{\mu})(\vec{\mathbf{X}} - \vec{\mu})^T\}$$

Pour $\vec{X}$ entier, tel que pour chaque d [1, D], X_d $[x_{dmin}, x_{dmax}]$ on peut démontrer que

$$\mu_d \equiv E\{x_d\} = \frac{1}{M} \sum_{x_1=x_{1min}}^{x_{1max}} \ldots \sum_{x_D=x_{Dmin}}^{x_{Dmax}} h(\vec{x})\ x_d$$

Pour x réel, $\mu_d \equiv E\{x_d\} = \int \int \ldots \int p(\vec{x}) \cdot x_d \, d\vec{X}$

Dans tous les cas :

$$\vec{\mu} = E\{\vec{\mathbf{X}}\} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \ldots \\ \mu_n \end{pmatrix} = \begin{pmatrix} E\{x_1\} \\ E\{x_2\} \\ \ldots \\ E\{x_n\} \end{pmatrix}$$

Pour D dimensions, la covariance entre les variables x_i et x_j est estimée à partir de M observations $\{\vec{\mathbf{X}}_m\}$

$$\sigma_{ij}^2 \equiv E\{ (X_i - E\{X_i\})(X_j - E\{X_j\})\} = \frac{1}{M} \sum_{m=1}^{M} (X_{im} - \mu_i)(X_{jm} - \mu_j))$$

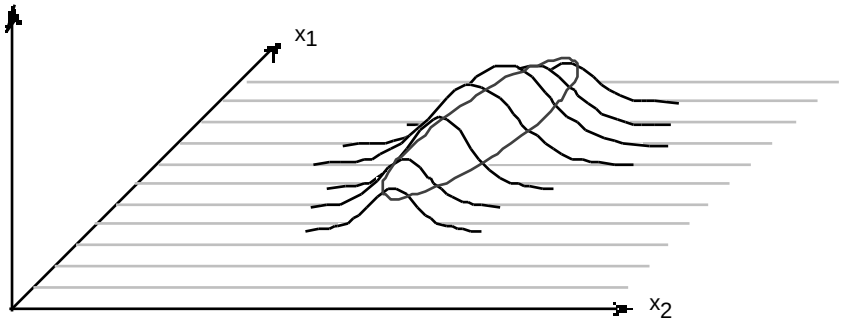
Ces coefficients composent une matrice de covariance. C

$$\mathbf{C}_x \equiv E\{[\vec{\mathbf{X}} - \vec{\mu})(\vec{\mathbf{X}} - \vec{\mu})^T\} = E([\vec{\mathbf{X}} - E\{\vec{\mathbf{X}}\})(\vec{\mathbf{X}} - E\{\vec{\mathbf{X}}\})^T\}$$

$$\mathbf{C}_x \equiv \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \ldots & \sigma_{1D}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \ldots & \sigma_{2D}^2 \\ \ldots & \ldots & & \ldots \\ \sigma_{D1}^2 & \sigma_{D2}^2 & \ldots & \sigma_{DD}^2 \end{pmatrix}$$

Dans le cas d'un vecteur de propriétés, $\vec{X}$, la loi normale prend la forme :

$$p(\vec{X}) = \mathcal{N}(\vec{X}; \vec{\mu}, \mathbf{C}_x) = \frac{1}{(2\pi)^{\frac{D}{2}} \det(\mathbf{C_x})^{\frac{1}{2}}} \quad e^{-\frac{1}{2}(\vec{X} - \vec{\mu})^T C_x^{-1} (\vec{X} - \vec{\mu})}$$



Le terme $(2\pi)^{\frac{D}{2}} \det(\mathbf{C}_x)^{\frac{1}{2}}$ est un facteur de normalisation.

$$\int\int \ldots \int e^{-\frac{1}{2}(\vec{X}-\vec{\mu})^T C_x^{-1}(\vec{X}-\vec{\mu})} dX_1\, dX_2 \ldots dX_D = (2\pi)^{\frac{D}{2}} \det(\mathbf{C})^{\frac{1}{2}}$$

La déterminante, det(C) est une opération qui donne la "énergie" de C.

Pour D=2 $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\cdot b - c\cdot d$

Pour D=3

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a\cdot\det \begin{pmatrix} e & f \\ h & i \end{pmatrix} + b\cdot\det \begin{pmatrix} f & d \\ i & g \end{pmatrix} + c\cdot \det \begin{pmatrix} d & e \\ g & h \end{pmatrix}$$

$$= a(ei-fh)+ b(fg-id)+ c(dh-eg)$$

pour D > 3 on continue récursivement.

L'exposant est une valeur positive et quadrique.
(si X est en mètre, $\frac{1}{2}(\vec{\mathbf{X}}-\vec{\mu})^T \mathbf{C}_x^{-1}(\vec{\mathbf{X}}-\vec{\mu})$ est en mètre$^{2\cdot}$)

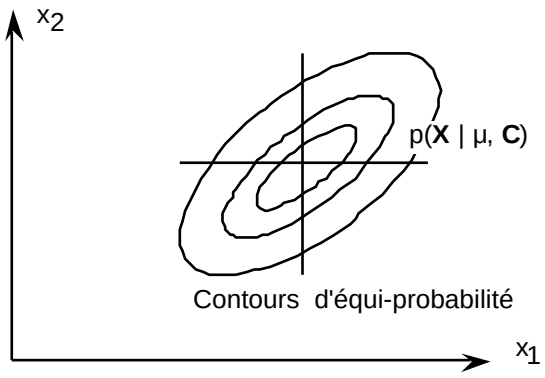
Cette valeur est connue comme la "distance de Mahalanobis".

$$d^2(\vec{X}) = \frac{1}{2}(\vec{X}-\vec{\mu})^T C_x^{-1}(\vec{X}-\vec{\mu})$$

Il s'agit d'une distance euclidienne, normalisé par la covariance C_x.

Cette distance est bien définie, même si les composants de $\vec{X}$ n'ont pas les mêmes unités. (Ceci est souvent le cas).

La loi Normale peut être visualisé par ses contours d'"équiprobabilité"



Ces contours sont les contours de constant $d^2(\vec{X})$

La matrice C est positif et semi-definite. Nous allons nous limiter au cas ou C est positif et definite (C.-à-d. $\det(C) = |C| > 0$

si x_i et x_j sont statistiquement indépendants, $\sigma_{ij}^2 = 0$.

Soit les événements E décrit par une vecteur de caractéristiques $\vec{X}$: $(E,\vec{X})$.
Soit une ensemble aléatoire de M événements avec leurs caractéristiques.

Cet ensemble est dit l'ensemble d'entrainement (training set) $\{\vec{X}_m\}$

Pour un vecteur de D caractéristiques :

$$\vec{\mu} \equiv E\{\vec{X}\} = \frac{1}{M}\sum_{m=1}^{M} \vec{X}_m = \begin{pmatrix}\mu_1\\ \mu_2\\ \dots\\ \mu_D\end{pmatrix} = \begin{pmatrix}E\{X_1\}\\ E\{X_2\}\\ \dots\\ E\{X_D\}\end{pmatrix}$$

Pour $\vec{X}$ entier, tel que pour chaque d [1, D], X_d $[x_{dmin}, x_{dmax}]$ on peut démontrer que

$$\mu_d \equiv E\{x_d\} = \frac{1}{M}\sum_{x_1=x_{1min}}^{x_{1max}} \dots \sum_{x_D=x_{Dmin}}^{x_{Dmax}} h(\vec{x})\, x_d$$

Pour x réel, $\mu_d \equiv E\{x_d\} = \int\int \dots \int p(\vec{x}) \cdot x_d \, d\vec{X}$

Dans tous les cas : $\vec{\mu} = E\{\vec{\mathbf{X}}\} = \begin{pmatrix}\mu_1\\ \mu_2\\ \dots\\ \mu_n\end{pmatrix} = \begin{pmatrix}E\{x_1\}\\ E\{x_2\}\\ \dots\\ E\{x_n\}\end{pmatrix}$

Pour D dimensions, la covariance entre les variables x_i et x_j est estimée à partir de M observations $\{\vec{\mathbf{X}}_m\}$

Soit $\vec{V} = \vec{X} - E\{\vec{X}\} = \vec{X} - \vec{\mu}$

$$\mathbf{C}_X \equiv E\{\vec{V}\,\vec{V}^T\} \equiv E\{[\vec{\mathbf{X}} - \vec{\mu})(\vec{\mathbf{X}} - \vec{\mu})^T\}$$

Ces coefficients composent une matrice de covariance. $\mathbf{C}_x$

$$\mathbf{C}_x = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \dots & \sigma_{1D}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2D}^2 \\ \dots & \dots & & \dots \\ \sigma_{D1}^2 & \sigma_{D2}^2 & \dots & \sigma_{DD}^2 \end{pmatrix}$$

ou $$\sigma_{ij}^2 = \frac{1}{M} \sum_{m=1}^{M} (X_{im} - \mu_i)(X_{jm} - \mu_j)\ )$$

Forme en Algebre Linéaire

Une expression en algebre de Matrice est souvent utile.

Soit une ensemble aléatoire de M événements avec leurs caractéristiques. $\{\vec{X}_m\}$

$$\vec{\mu} \equiv E\{\vec{X}\} = \frac{1}{M} \sum_{m=1}^{M} \vec{X}_m$$

Soit $\vec{V}_m = \vec{X}_m - \vec{\mu}$

On peut faire une matrix V composé de M colones $\{\vec{V}_m\}$

$$V = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1M} \\ v_{21} & v_{22} & \cdots & v_{2M} \\ \cdots & \cdots & & \cdots \\ v_{D1} & v_{D2} & \cdots & v_{DM} \end{pmatrix}$$

$$\mathbf{C}_X \equiv V\ V^T = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

$$= \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

Transformations Linéaire

La transformation linéaire d'une loi normale et une loi normale. Les moments d'une transformation linéaire sont les transformations linéaires des moments.

Soit un vecteur unitaire $\vec{R} = \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \dots \\ \Delta x_n \end{pmatrix} = \begin{pmatrix} \cos(\alpha_1) \\ \cos(\alpha_2) \\ \dots \\ \cos(\alpha_D) \end{pmatrix}$ tel que $\| \vec{R} \| = 1$.

La projection (transformation linéaire) de $\vec{X}$ sur y est

$$y = \vec{R}^T \vec{X}.$$

Pour la covariance :

$$\begin{aligned} \sigma_y^2 \equiv\ & E\{(\vec{R}^T \vec{V})(\vec{R}^T \vec{V})^T\} \\ =\ & E\{(\vec{R}^T \vec{V})(\vec{V}^T \vec{R})\} && \text{car } (\vec{R}^T \vec{V})^T = (\vec{V}^T \vec{R}) \\ =\ & E\{\vec{R}^T (\vec{V} \vec{V}^T) \vec{R}\} \\ =\ & \vec{R}^T E\{\vec{V} \vec{V}^T\} \vec{R} = \vec{R}^T \mathbf{C}_x \vec{R} \end{aligned}$$

La projection de la covariance est la covariance de la projection.
La projection de la moyenne et la covariance sur un axe, $\vec{R}^T$ donne une moyenne μ_y et variance, σ_y^2 dans la direction $\vec{R}$.

$$\mu_y = \vec{R}^T \vec{\mu}_x, \qquad \sigma_y^2 = \vec{R}^T \mathbf{C}_x \vec{R}$$

$$p(y) = \mathcal{N}(y;\ \vec{R}^T \vec{\mu}_x, \vec{R}^T \mathbf{C}_x \vec{R}^T) = \mathcal{N}(y;\ \mu_y, \sigma_y^2)$$

Les moments d'une projection sont les projections des moments.

$$\mu_y = E\{p(y)\} = \vec{R}^T \vec{\mu}_x \qquad \sigma_y^2 = E\{(p(y)-\mu_y)(p(y)-\mu_y)\} = \vec{R}^T \mathbf{C}_x \vec{R}$$

Bruit d'observation

Chaque observation d'un événement corrompu par une bruit d'observation.

$$Y = x_k + B_k + B_o$$

B_o est souvent Normale, avec moyenne 0 et variance σ_o^2.
On dit que la variance σ_o^2 est la précision de la capteur.
Cette bruit ne depend pas de la classe.

Dans ce cas, on observe Y avec

$$p(Y=y \mid \omega_k) = \mathcal{N}(y; \mu_k, \sigma_k^2 + \sigma_o^2) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2}\frac{(y-\mu_k)^2}{(\sigma_k^2+\sigma_o^2)}}$$

et

$$k = \arg\max_k \{-\mathrm{Log}\{\sqrt{\sigma_k^2+\sigma_o^2}\} - \frac{(y-\mu_k)^2}{2(\sigma_k^2+\sigma_o^2)} + \mathrm{Log}\{p(\omega_k)\}\}$$

$$= \arg\max_k \{-\mathrm{Log}\{\sigma_k\} - \frac{(y-\mu_k)^2}{2(\sigma_k^2+\sigma_o^2)} + \mathrm{Log}\{p(\omega_k)\}\}$$

Classification pour K > 2 et D > 1.

Dans le cas général, il y a D caractéristique.

$$g_k(\vec{X}) = p(\omega_k \mid \vec{X})\, p(\omega_k)$$

Et le règle de décision est :

$$\hat{\omega}_i \ : \text{si} \quad j \neq i \quad g_i(\vec{X}) > g_j(\vec{X})$$

Dans cette forme le classificateur est une machine qui calcule K fonctions $g_k(\vec{x})$ suivie d'une sélection du maximum.

La fonction de discrimination est : $g_k(\vec{X}) = p(\vec{X} \mid \omega_k)\, p(\omega_k)$

On sélection la classe ω_k pour laquelle $\arg\max_k\{g_k(\vec{X})\}$

par règle de Bayes :

$$\arg\max_k\{p(\omega_k \mid \vec{X})\} = k = \arg\max_k\{p(\vec{X} \mid \omega_k)\, p(\omega_k)\}$$

$$= \arg\max_k\{\mathrm{Log}\{p(\vec{X} \mid \omega_k)\} + \mathrm{Log}\{p(\omega_k)\}$$

Si les caractéristiques suivent une densité Normale :

$$p(\vec{X} \mid w_k) = \mathcal{N}(\vec{X}, \vec{\mu}_k, C_k)$$

$$\mathrm{Log}\{p(\vec{X} \mid \omega_k)\} = \mathrm{Log}\left\{ \frac{1}{(2\pi)^{\frac{D}{2}} \det(C_k)^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{X} - \vec{\mu}_k)^T C_k^{-1} (\vec{X} - \vec{\mu}_k)} \right\}$$

$$\mathrm{Log}\{p(\vec{X} \mid \omega_k)\} = -\frac{D}{2} \mathrm{Log}\{2\pi\} - \frac{1}{2} \mathrm{Log}\{\det(C_k)\} - \frac{1}{2}(\vec{X} - \vec{\mu}_k)^T C_k^{-1} (\vec{X} - \vec{\mu}_k)$$

One note que $-\frac{D}{2} \mathrm{Log}\{2\pi\}$ peut être éliminé parce qu'il est constant pour tout k.

La fonction de discrimination devient :

$$g_k(\vec{x}) = -\frac{1}{2}\ \mathrm{Log}\{\det(C_k)\} - \frac{1}{2}(\vec{X} - \vec{\mu}_k)^T C_k^{-1}(\vec{X} - \vec{\mu}_k) + \mathrm{Log}\{p(\omega_k)\}$$

Les classifiers Bayesiennes sont définies par les variations de cette formule.

Forme Canonique de la fonction de discrimination

La décision ω_k est celle qui donne un maximum pour

$$g_k(\vec{X}) = -\frac{1}{2}(\vec{X} - \vec{\mu}_k)^T C_k^{-1}(\vec{X} - \vec{\mu}_k) + \frac{1}{2}\ \mathrm{Log}\{\det(C_k)\} + \mathrm{Log}\{p(\omega_k)\}$$

On peut réécrire $(\vec{X} - \vec{\mu}_k)^T C_k^{-1}(\vec{X} - \vec{\mu}_k)$ comme

$$\vec{X}\, C_k^{-1}\vec{X} - \vec{X}\ C_k^{-1}\vec{\mu}_k - \vec{\mu}_k^T C_k^{-1}\vec{X} + \vec{\mu}_k^T C_k^{-1}\,\vec{\mu}_k$$

On note que C_k^{-1} est symétrique, et donc $\vec{X}\ C_k^{-1}\vec{\mu}_k = \vec{\mu}_k^T C_k^{-1} X$

Donc $-\ \vec{X}\ C_k^{-1}\vec{\mu}_k - \vec{\mu}_k^T C_k^{-1}\vec{X} = 2\,(\vec{\mu}_k^T C_k^{-1})^T \vec{X} = 2\,(C_k^{-1}\vec{\mu}_k)^T \vec{X}$

On peut réécrire $g_k(\vec{X})$ comme

$$g_k(\vec{X}) = -\vec{X}^T(\tfrac{1}{2}C_k^{-1})\,\vec{X} + (C_k^{-1}\vec{\mu}_k)^T\vec{X} - \frac{1}{2}(\vec{\mu}_k^T C_k^{-1}\vec{\mu}_k) - \frac{1}{2}\mathrm{Log}\{\det(C_k)\} + \mathrm{Log}\{p(w_k)\})$$

ou bien

$$g_k(\vec{X}) = \vec{X}^T(D_k)\,\vec{X} + \vec{d}_k^{\,T}\vec{X} + d_{ko}.$$

avec $D_k = \frac{1}{2}\ C_k^{-1}$

$$\vec{d}_k = C_k^{-1}\vec{\mu}_k$$

$$d_{ko} = -\frac{1}{2}(\vec{\mu}_k^T C_k^{-1}\vec{\mu}_k) - \frac{1}{2}\ \mathrm{Log}\{\det(C_k)\} + \mathrm{Log}\{p(\omega_k)\}$$

Cette fonction est composée de trois termes :

une terme quadratique $\vec{X}^T(D_k)\,\vec{X}$,

une terme linéaire : $\vec{d}_k^{\,T}\vec{X}$

et une terme constant : d_{ko}

Bruit et Choix de la Fonction de Discrimination

La conception d'un classifier dépend de la choix de caractéristiques, $\vec{x}$ et du bruit observé sur ses caractéristiques :

Rappel qu'une observation $\vec{Y} = \vec{x}_k + \vec{B}_k + \vec{B}_o$

où

$\vec{x}_k$ Est la forme type (moyenne) de la classe w_k

$\vec{B}_k$: Les variations aléatoires intra-classe.
Elle est spécifiques à chaque classe et chaque individus.
Elles n'est change pas entre observations.

$\vec{B}_o$: Les variations aléatoires des observations.
Elles est indépendantes de la classe et de l'individu.
Elles changent avec les observations.

Selon la nature de $\vec{B}_k$, $\vec{B}_o$ et de $p(\omega_k)$ on peut faire certaines simplifications.

par exemple :

Si $\vec{B}_o >> \vec{B}_k$ on peut trouver que $w_k : C_k \approx C$: On a une classifier linéaire.

Si $i,j \quad \sigma_{ij}^2 = \sigma^2 \quad C$: On a la détecteur optimale utilisé en communication hz.

Exemples de caractéristiques : $\vec{x}$

1) Les échantillons d'un signal : $x(n)$ pour $n \in [1, N]$
2) Les caractéristiques d'un individu : [hauteur, poids, yeux, cheveux etc.]
3) Les caractéristiques géométriques d'un objet : Hauteur, largeur, nombre de faces, etc.