Analyse et Reconnaissance d'Images

James L. Crowley

M2R IVR Premier Semestre 2007/2008

Séance 6 26 novembre 2007

Reconnaissance Probabiliste

Plan de la Séance :

La Reconnaissance.....2
La Probabilité d'un Evénement.....4
Définition Fréquentielle.....4
Définition Axiomatique.....4
La probabilité de la valeur d'une variable aléatoire.....5
La Règle de Bayes.....6
Classification des Pixels par Ratio d'Histogramme 8
Histogrammes.....8
Example : Détection d'object par ratio d'histogramme de couleur.8
Histogrammes de Champs Réceptifs.....10
Densités de Probabilité.....12
La Loi Normale.....13
La Loi Normale pour D = 1.....14
La Loi Normale pour D > 1.....15
Exemple : Analyse des Images Terrestre.....19
Exemple : Caractérisation d'un région par moments.....22
Composantes principales.....23
Fonctions de Discrimination 24

La Reconnaissance

La reconnaissance est une capacité fondamentale de l'intelligence et même 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 entité comme un individu

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

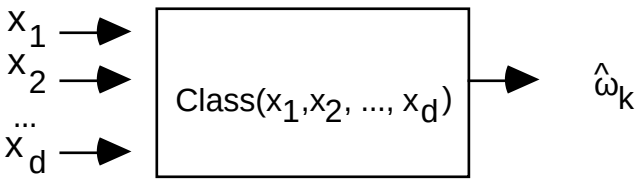
Un ensemble est défini par un test d'appartenance.

La classification est un processus d'association un entité (un événement) à une classe. L'entité est décrit par un vecteur des caractéristiques, produit par une observation. L'affectation de l'entité 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.

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

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

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.

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 conditionelle d'appartenance.

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.

Ayant une observation, $\vec{X}$, le critère de partition est la probabilité.

$$p(\omega_k \mid \vec{X}) = \text{Pr}(E \quad T_k \text{ etant donnée l'observation de } \vec{X})$$

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

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 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() [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

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

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'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 $x \in [x_{min}, x_{max}]$, 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'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 $x \in [x_{min}, x_{max}]$.

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 du nombre de valeurs possible et de M.
En règle générale, on dit qu'il faut 10 exemples par cellule de l'histogramme.

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 ∩ B ou a ¬A ∩ B ou a A ∩ ¬B ou a ¬A ∩ ¬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 :

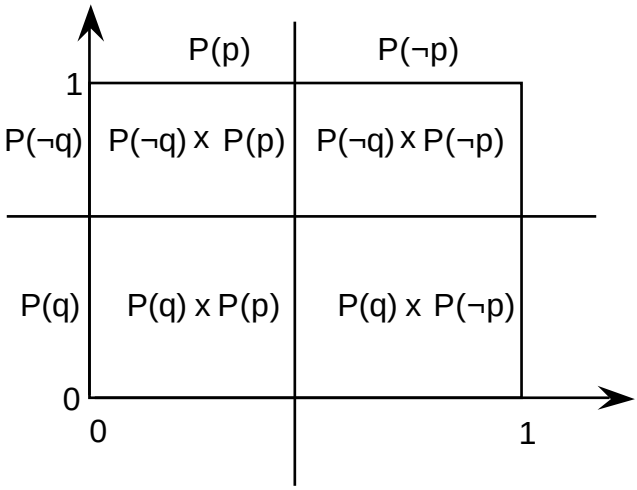
P(q) + P(¬q) = 1.

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 q) + P(p ¬q)
P(q) = P(p q) + P(¬p 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 règle de Bayes. 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"

Classification des Pixels par Ratio d'Histogramme

Histogrammes

La probabilité d'une valeur x est sa fréquence d'occurence.

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

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 du ratio entre le nombre de Cellules, $Q = N^d$, de h(x) et le nombre d'echantillions, M.

L'erreur moyenne entre $\frac{1}{M} h(\vec{C})$ et $P(\vec{C})$ est $E_{ms} \sim O(\frac{Q}{M})$

Pour que l'estimation soit "raisonable", il faut assuré que $M >> Q = N^d$
En règle générale, on dit qu'il faut 10 exemples par cellule de l'histogramme.

Example : Détection d'object par ratio d'histogramme de couleur

On peut utiliser les histogrammes avec la regle de Bayes pour detecter les objects.

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

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

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

Supposone qu'on code c_1 et c_2 avec les entiers entre 0 et N – 1

$$c_1 = \text{Round}\left((N-1) \cdot \frac{R}{R+G+B}\right) \quad c_2 = \text{Round}\left((N-1) \cdot \frac{G}{R+G+B}\right)$$

On alloue un tableau 2D, $h(c_1, c_2)$., de taille N x N cellulles.
(exemple Q = 32 x 32 = 1024 cellules)

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

l'histogramme qui correspond à $\vec{C}$: $h(\vec{C}) := h(\vec{C}) + 1$

c'a-dire $h(c_1, c_2) := h(c_1, c_2) + 1$

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

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

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,

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

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

Pour chaque pixel $\vec{C}(i, j)$ $p(\text{objet} | C) = p(C | \text{objet}) \frac{p(\text{objet})}{p(C)}$

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".

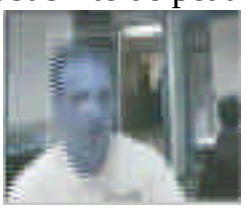
$$p(\text{objet}) = \frac{M_o}{M}$$

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

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

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

Voici une image de la probabilité de peau fait par ratio d'histogramme de r, v



Histogrammes de Champs Réceptifs.

Cette méthode peut être généralisé en remplacant la chrominance par une vecteur de champs réceptifs. Mais il faut bien gerer la ration Q/M !

Soit une image p(i,j) et une vecteur de "d" champs réceptifs $\vec{G}$

$\vec{V}(i,j) = < \vec{G}, p(i,j) >$ est un vecteur de caractéristiques de d dimensions.

$h(\vec{V})$ aura $Q = N^d$

N \ d	1	2	3	4	5	6
2	2^1	2^2	2^3	2^4	2^5	2^6
4	2^2	2^4	2^6	2^8	2^{10} =1 Kilo	2^{12} =2 Kil
8	2^3	2^6	2^9	2^{12}	2^{15}	2^{18}
16	2^4	2^8	2^{12}	2^{16}	2^{20} = 1 Meg	2^{24} = 4 M
32	2^5	2^{10} =1 Kilo	2^{15}	2^{20} = 1 Meg	2^{25}	2^{30} = 1 Gi
64	2^6	2^{12}	2^{18}	2^{24}	2^{30} = 1 Gig	2^{36}
128	2^7	2^{14}	2^{21} = 2 Meg	2^{28}	2^{35}	2^{42} =2 Te
256	2^8	2^{16}	2^{24}	2^{32} = 2 Gig	2^{40} = 1 Tera	2^{48}

Soit les champs receptifs achromatique

$\vec{G}_\sigma = (G_x, G_y, G_{xx}, G_{xy}, G_{yy})$

d = 5

ou chromatique avec normalisation de l'orientation et echelle :

$\vec{G}_{\sigma,\theta} = (G^L_x\ G^{C_1}_x, G^{C_2}_x, G^L_{xx}\ G^L_{xy}\ G^{C_1}_{xx}, G^{C_2}_{xx})$

d = 7.

On peut faire

$$p(\text{objet}(i,j) \mid \vec{V}(i,j)) = \frac{p(\vec{V}(i,j) \mid \text{objet}(i,j))\, p(\text{objet}(i,j))}{p(\vec{V}(i,j))} \approx \frac{h_o(\vec{V}(i,j))}{h_{tot}(\vec{V}(i,j))}$$

sur condition de gerer M et Q.

Rappel :

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

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.

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).

Densités de Probabilité.

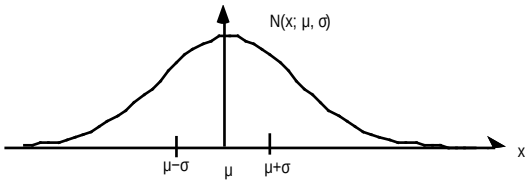
Une fonction de densité de probabilité P(X), est une fonction tel que

P(X) est Real et positive pour tout X.
X est réal entre $[-\infty, \infty]$
tel que

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

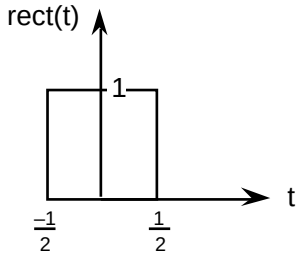
exemples :

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 Loi Normale

La fonction paramétrique la plus utilisée est la loi Normale.

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.

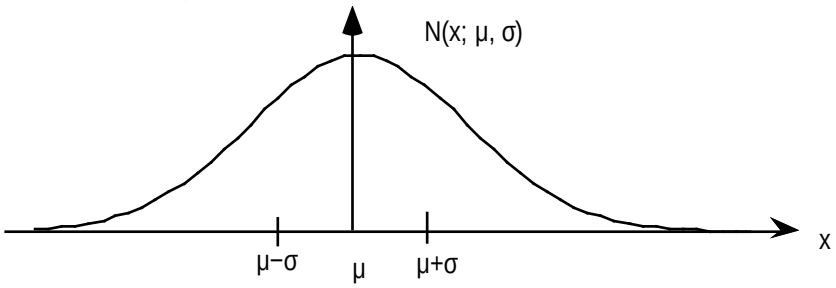
La loi Normale décrit une population d'exemples $\{X_m\}$.

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)
On peut même estimer les moments quand il n'existe pas les bornes (X_{max}–X_{min})
ou quand X est une variable continue.

Dans ce cas, p( ) est une "densité" et il faut une fonction paramétrique pour p().

Dans la plupart des cas, on peut utiliser $\mathcal{N}(\mu, \sigma)$ comme une fonction de densité pour p(x).

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



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.

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

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.

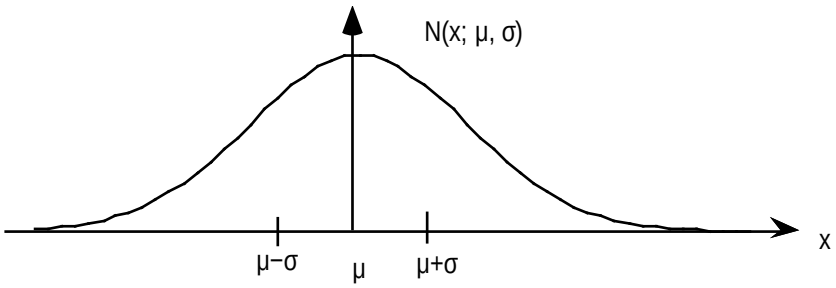
La différence normalisée par la variance est connue sous le nom de "Distance de Mahalanobis". La Distance de Mahalanobis est un test naturel de similarité

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 :



La moyenne est le premier moment de la densité p(x).

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

La variance σ^2 est le deuxième moment de p(x).

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

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

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

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

Dans tous les cas :

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

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

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

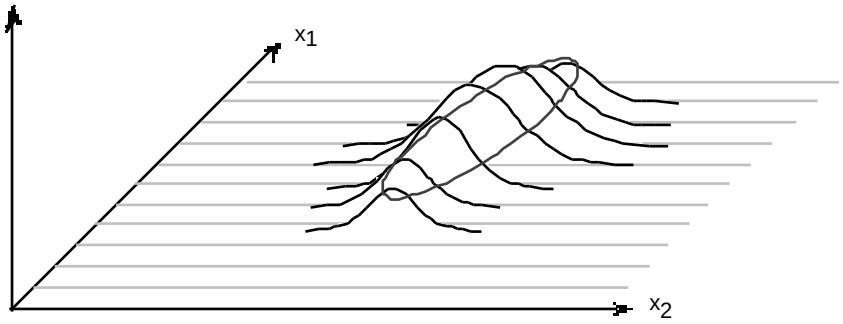
Ces coefficients composent une matrice de covariance. C

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

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

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

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



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$^{2\cdot}$)

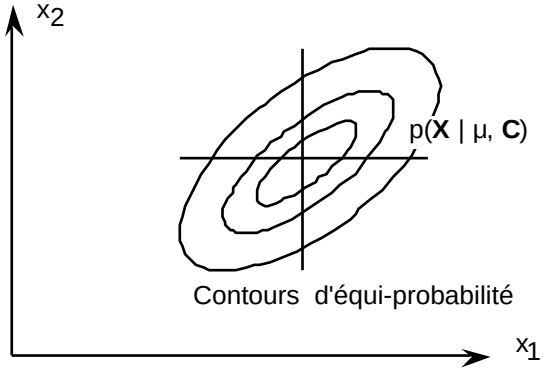
Cette valeur est connue comme la "distance de Mahalanobis".

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

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

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

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

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

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

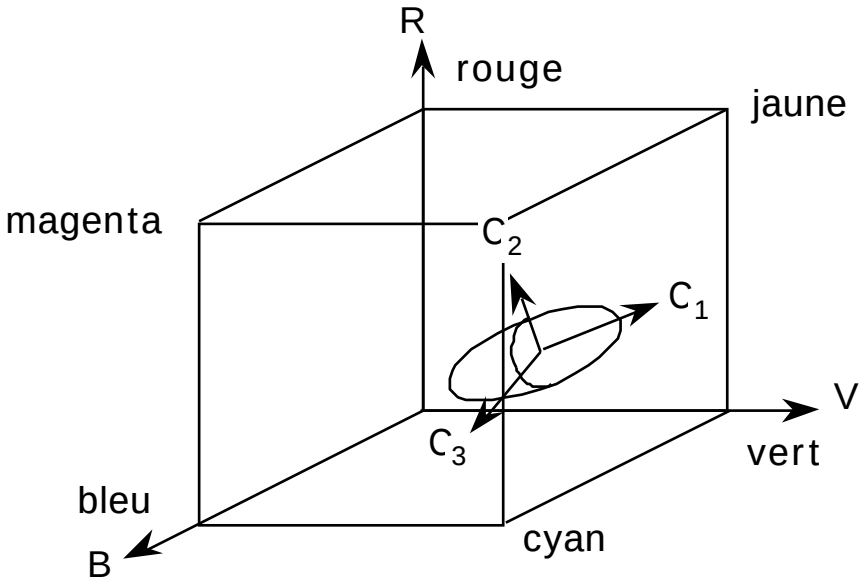
$$\mathbf{C}_x = \begin{pmatrix}\sigma_{11}^2 & \sigma_{12}^2 & \dots & \sigma_{1D}^2\\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2D}^2\\ \dots & \dots & & \dots\\ \sigma_{D1}^2 & \sigma_{D2}^2 & \dots & \sigma_{DD}^2\end{pmatrix}$$

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

Exemple : Analyse des Images Terrestre

Les statistiques de la couleur sont utilisées couramment afin d'analyser les images pris par les satellites.

On peut utiliser la moyenne et covariance afin de caractériser la couleur d'une région. Soit un échantillon R d'une couleur dans une scène, nous pouvons calculer la distribution normale qui représente la couleur de R par la Moyenne,
$\hat{X}_R = [r, v, b]^T$ et Covariance, $\mathbf{C_x}$.



La somme de l'histogramme : $S = \sum_{r,v,b=0}^{255} h(r,v,b)$

Moyenne : $\hat{X} = \begin{bmatrix} \mu_r \\ \mu_v \\ \mu_b \end{bmatrix}$ où $\mu_r = \frac{1}{S} \sum_{r,v,b=0}^{255} h(r,v,b)r$

$\mu_v = \frac{1}{S} \sum_{r,v,b=0}^{255} h(r,v,b)v$ $\mu_b = \frac{1}{S} \sum_{r,v,b=0}^{255} h(r,v,b)b$

Covariance : $\mathbf{C}_x \triangleq \begin{bmatrix} \sigma_r^2 & \sigma_{rv} & \sigma_{rb} \\ \sigma_{br} & \sigma_b^2 & \sigma_{bv} \\ \sigma_{vr} & \sigma_{vb} & \sigma_b^2 \end{bmatrix}$ où

$\sigma_r^2 \equiv \frac{1}{S} \sum_{r,v,b=0}^{255} h(r,v,b)(r - \mu_r)^2$ $\sigma_v^2 + \frac{1}{S} \sum_{r,v,b=0}^{255} h(r,v,b)(v - \mu_v)^2$

$\sigma_b^2 \equiv \frac{1}{S} \sum_{r,v,b=0}^{255} h(r,v,b)(b - \mu_b)^2$

et

$$\sigma_{rv} \equiv \sigma_{vr} \equiv \frac{1}{S} \sum_{r,v,b=0}^{255} h(r,v,b)(r - \mu_r)(v - \mu_v)$$

$$\sigma_{rb} \equiv \sigma_{vb} \equiv \frac{1}{S} \sum_{r,v,b=0}^{255} h(r,v,b)(r - \mu_r)(b - \mu_b)$$

$$\sigma_{vb} \equiv \sigma_{vb} \equiv \frac{1}{S} \sum_{r,v,b=0}^{255} h(r,v,b)(v - \mu_v)(b - \mu_b)$$

Avec ces statistiques, on peut estimer la probabilité d'avoir observé une instance d'une classe, étant donné l'observation d'un pixel Y= (r, v, b).

Soit K "classes", $\{w_k\}$, chacun décrit par $\hat{X}_k, C_k^{-1}$

Pour un observation $\vec{Y}$

$$P(\vec{Y} \mid \hat{X}_k) = \frac{1}{(2\pi)^{\frac{3}{2}} \det(C_k)^{\frac{1}{2}}} \quad e^{-\frac{1}{2}(\vec{Y} - \hat{X}_k)^T C_k^{-1} (\vec{Y} - \hat{X}_k)}$$

La probabilité d'obtenir w_k ayant observé $\vec{Y}$

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

La classe le plus "probable" est indépendant de $p(\vec{Y})$

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

$$= \max_k \{ \mathrm{Log}\{ p(\vec{Y} \mid w_k)\, p(w_k) \} \}$$

$$= \max_k \left\{ -\frac{1}{2} (\vec{\mathbf{Y}} - \vec{\mu_k})^T \mathbf{C}_k^{-1} (\vec{\mathbf{Y}} - \vec{\mu_k})) - \mathrm{Log}\left(\sqrt[n]{2\pi}\, \det(\mathbf{C}_k)^{\frac{1}{2}} \right) + \mathrm{Log}\{p(Cw_k)\} \right\}$$

Si la bruit est indépendant de la classe, $\forall j, k : \det(\mathbf{C}_j) = \det(\mathbf{C}_k)$

$$\underset{k}{\text{Max}}\{\, p(w_k \mid \vec{Y}) \,\} = \underset{k}{\text{Max}} \left\{ -\frac{1}{2}(\vec{\mathbf{Y}} - \vec{\mu_k})^T \mathbf{C}_k^{-1}(\vec{\mathbf{Y}} - \vec{\mu_k})) + \text{Log}\{p(w_k)\} \right\}$$

Si les classes sont equi-probable, $\forall j, k : \text{Log}\{p(w_j)\} = \text{Log}\{p(w_k)\}$

$$\underset{k}{\text{Max}}\{\, p(w_k \mid \vec{Y}) \,\} = \underset{k}{\text{Max}} \left\{ -\frac{1}{2}(\vec{\mathbf{Y}} - \vec{\mu_k})^T \mathbf{C}_k^{-1}(\vec{\mathbf{Y}} - \vec{\mu_k})) \right\}$$

Pour trouver les pixels "blé" dans les images de satellite.

Déterminer la couleur "caractéristique" du blé $\hat{X}_b = (R_b, V_b, B_b)$
et sa covariance. $\mathbf{C}_b$.

Pour un vecteur de propriétés de 3 Dimensions :

$$P(\vec{Y} \mid \hat{X}_b) = \frac{1}{(2\pi)^{\frac{3}{2}} \det(C_b)^{\frac{1}{2}}} \; e^{-\frac{1}{2}(\vec{Y} - \hat{X}_b)^T C_x^{-1} (\vec{Y} - \hat{X}_b)}$$

Pour chaque pixel de l'image, $\vec{Y}(i,j)$, on peut calculer la distance normalisée par la covariance.

$$D_b^2(i,j) = \frac{1}{2} \{ (\vec{Y}(i,j) - \hat{X}_b)^T \mathbf{C}_b^{-1} (\vec{Y}(i,j) - \hat{X}_b) \}$$

Une image de la "confiance" ou CF du blé :

$$p(i, j) = e^{-\frac{1}{2}(\vec{Y}(i,j) - \hat{X}_k)^T C_k^{-1} (\vec{Y}(i,j) - \hat{X}_k)}$$

Les pixels avec une couleur proche de celle de l'échantillon apparaissent intenses.

Exemple : Caractérisation d'un région 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 fênetre (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 :

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

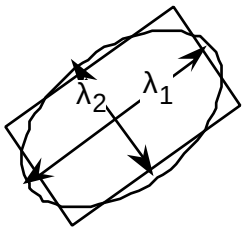
$$\sigma_j^2 = \frac{1}{S} \sum_{i=1}^{M} \sum_{j=1}^{N} w(i, j) \cdot (j - \mu_j)^2$$

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

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.

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

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

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

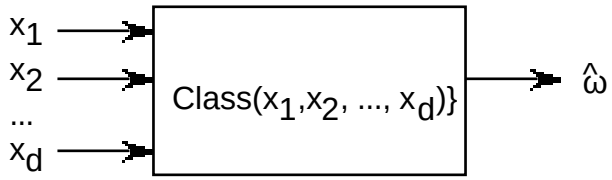
Les lignes du Φ sont des vecteurs propres du **C.**
La longueur des axes majeurs et mineur est les valeurs propres de la matrice **C**.

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.

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 : $R^K \rightarrow \{\omega_K\}$

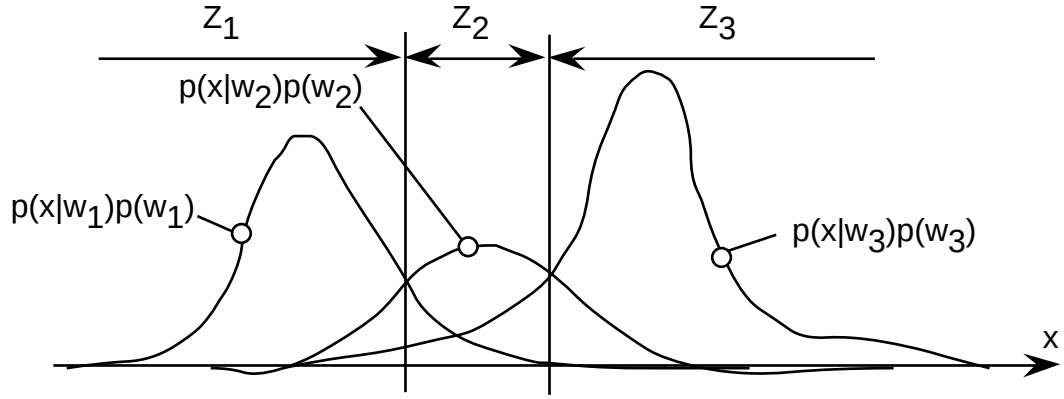
en génerale $g(\vec{X}) = \begin{pmatrix} g_1(\vec{X}) \\ g_2(\vec{X}) \\ \dots \\ g_K(\vec{X}) \end{pmatrix}$ est une vecteur de K fonctions $g_k(\vec{X})$

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

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

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

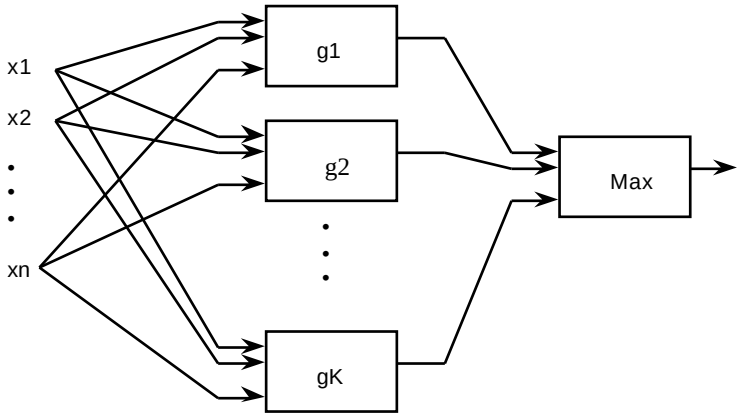
Donc notre fonction de discrimination devient :

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

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

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

La discriminateur 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 $k = \arg\max_{k}\{g_k(X)\} = \arg\max_{\omega_k}\{\mathrm{Log}\{g_k(X)\}\}$
parce que Log{} est une fonction monotone.

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

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

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

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

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