Formation et Analyse d'Images

James L. Crowley

ENSIMAG 3 Premier Sémestre 2007/2008

Séance 8 23 novembre 2007

Analyse et Réconnaissance d'Images

Plan de la Séance :

Densités de Probabilité.

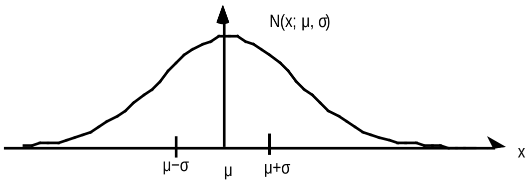
Une fonction de densité de probabilité P(X), est une fonction tel que

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

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

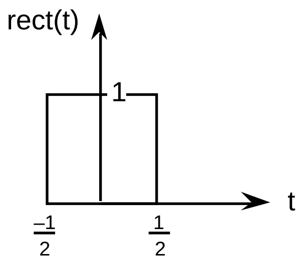
exemples :

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



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

rect : $\quad$ 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.

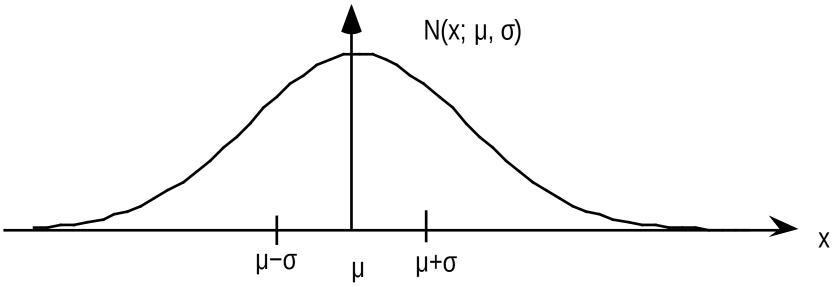
La loi Normale décrit une population d'exemples $\{X_m\}$.

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.

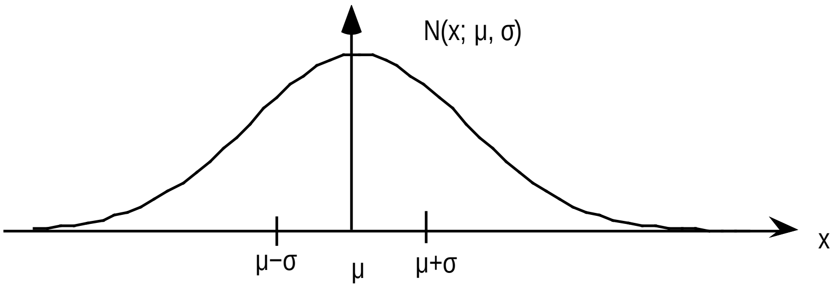
La différence normalisée par la variance est connue sous le nom de "Distance de Mahalanobis". La Distance de Mahalanobis est un test naturel de similarité

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

Dans tous les cas :

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

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

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

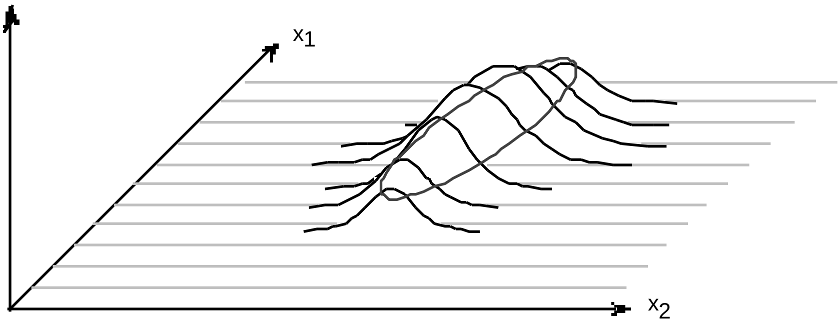
Ces coefficients composent une matrice de covariance. C

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

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

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.)

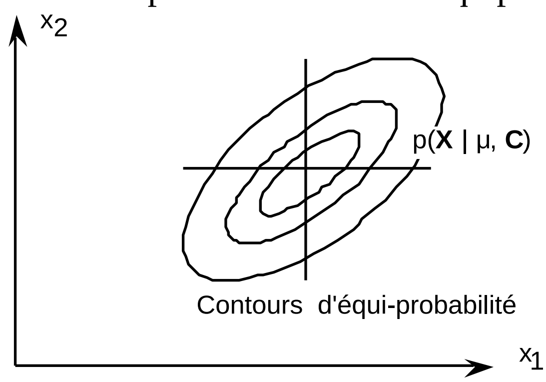
Cette valeur est connue comme la "distance de Mahalanobis".

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

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 :

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

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

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

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 $\{\vec{\mathbf{X}}_m\}$

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. $\mathbf{C}_x$

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

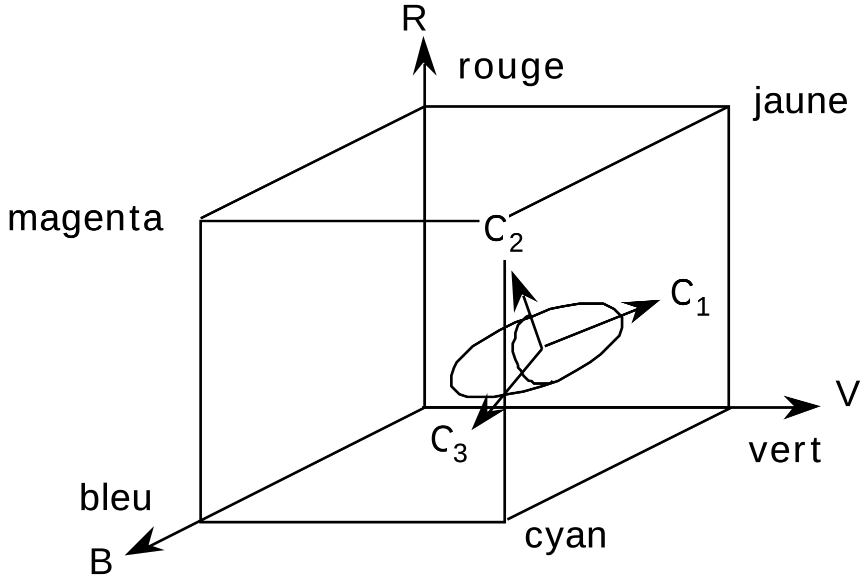
ou $\sigma_{ij}^2 = \frac{1}{M}\sum_{m=1}^{M} (X_{im} - \mu_i)(X_{jm} - \mu_j)$)

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 $p(\vec{Y})$

$$w_k = \underset{k}{\mathrm{Max}}\{\, p(w_k \mid \vec{Y})\, \} = \underset{k}{\mathrm{Max}}\left\{ p(\vec{Y} \mid w_k)\, p(w_k) \right\}$$

$$= \underset{k}{\mathrm{Max}}\left\{ \mathrm{Log}\{\, p(\vec{Y} \mid w_k)\, p(w_k)\, \} \right\}$$

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

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

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

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

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

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 :

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

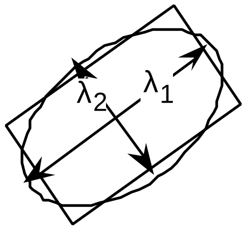
$$\sigma_j^2 = \frac{1}{S}\sum_{i=1}^{M}\sum_{j=1}^{N} w(i, j)\cdot(j-\mu_j)^2$$

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

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 <u>une analyse en composantes principales</u> 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.

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

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

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

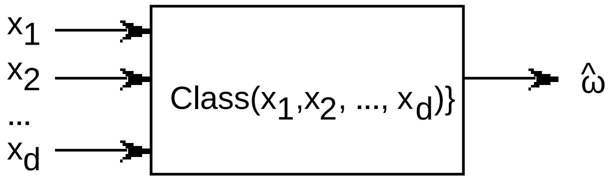
Les lignes du Φ sont des vecteurs propres du **C.**
La longueur des axes majeurs et mineur est les valeurs propres de la matrice **C**.

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



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

$\hat{\omega}_k$ 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 : $R^K \rightarrow \{\omega_K\}$

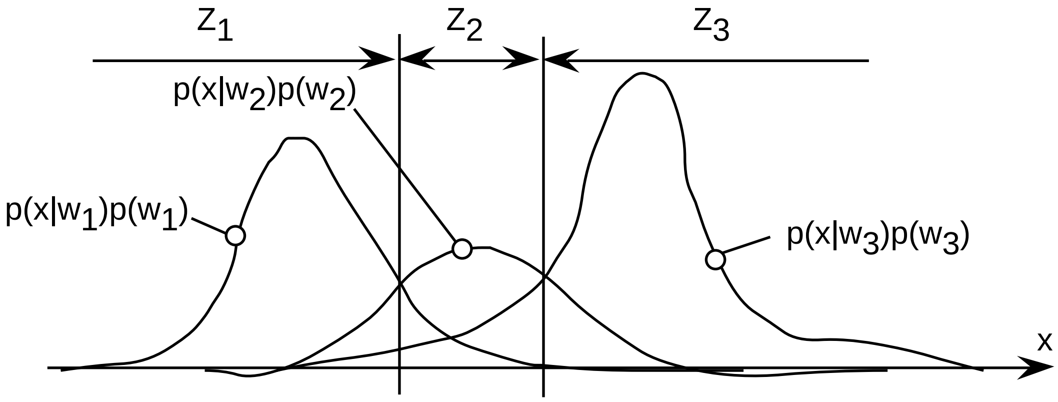
en génerale $g(\vec{X}) = \begin{pmatrix} g_1(\vec{X}) \\ g_2(\vec{X}) \\ \ldots \\ g_K(\vec{X}) \end{pmatrix}$ est une vecteur de K fonctions $g_k(\vec{X})$

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

$$k = \underset{k}{\text{arg-max}}\{g_k(X)\} \qquad \text{avec } g_k(X) = p(x \mid \omega_k)\, p(\omega_k)$$

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.

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

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 :

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

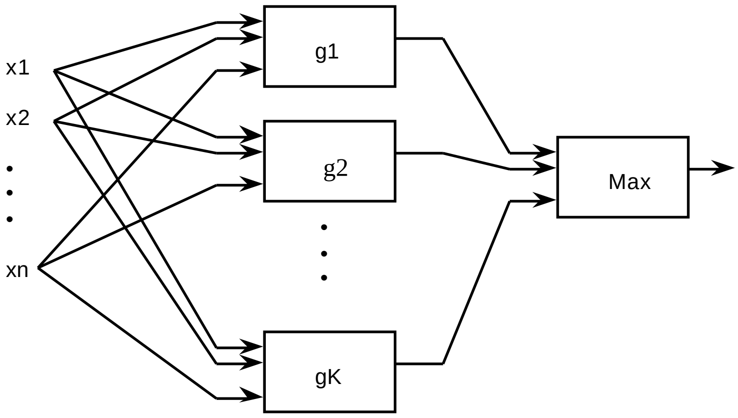
Donc notre fonction de discrimination devient :

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

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

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 :

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

Cecit peut être traduit dans une forme cannonique :

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

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