POLITECNICO DI MILANO

Earth Observation

MSc Geoinformatics Engineering

School of Civil Environmental and Land Planning Engineering
ICAT

Ludovico Biagi

Department of Civil and Environmental Engineering - Geodesy and Geomatics area

GEOlab – Geomatics and Earth Observation Laboratory

Supervised Classification

The known pixels in each one of the predecided classes $\omega_1, \omega_2, \ldots, \omega_K$, form corresponding **"sample sets"** $S_1, S_2, \ldots, S_K$ with $n_1, n_2, \ldots, n_K$ number of pixels respectively.

Estimates from each sample set S_i, $(i = 1, 2, \ldots, K)$:

Class mean vectors:

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in S_i} \mathbf{x}$$

Class covariance matrices:

$$\mathbf{C}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in S_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$$

Supervised classification methods:

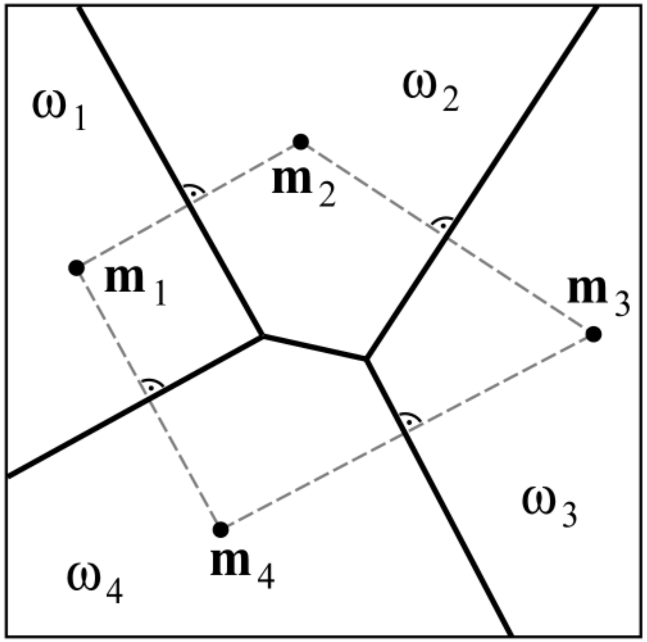
- Parallelepiped
- Euclidean distance (minimization)
- Mahalanobis distance (minimization)
- Maximum likelihood
- Bayesian (maximum a posteriori probability density)

Classification with Euclidean distance

$$d_E(\mathbf{x},\mathbf{x}') = \| \mathbf{x} - \mathbf{x}' \| = \sqrt{(x^1 - x'^1)^2 + (x^2 - x'^2)^2 + \ldots + (x^B - x'^B)^2}$$

(a) Simple

$$\| \mathbf{x} - \mathbf{m}_i \| = \min_k \| \mathbf{x} - \mathbf{m}_k \| \quad \Rightarrow \quad \mathbf{x} \in \omega_i$$



Assign each pixel to the class
of the closest center (class mean)

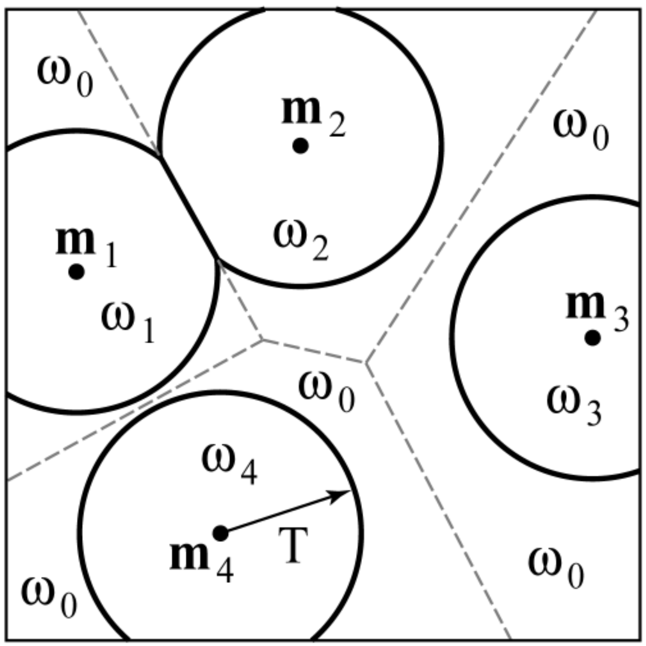
Boundaries between class regions =

= perpendicular at middle of segment
joining the class centers

Classification with Euclidean distance

$$d_E(\mathbf{x}, \mathbf{x}') = \| \mathbf{x} - \mathbf{x}' \| = \sqrt{(x^1 - x'^1)^2 + (x^2 - x'^2)^2 + \ldots + (x^B - x'^B)^2}$$

(b) with threshold T



$$\left. \begin{array}{c} \| \mathbf{x} - \mathbf{m}_i \| = \min\limits_k \| \mathbf{x} - \mathbf{m}_k \| \\ \| \mathbf{x} - \mathbf{m}_i \| \le T \end{array} \right\} \Rightarrow \quad \mathbf{x} \in \omega_i$$

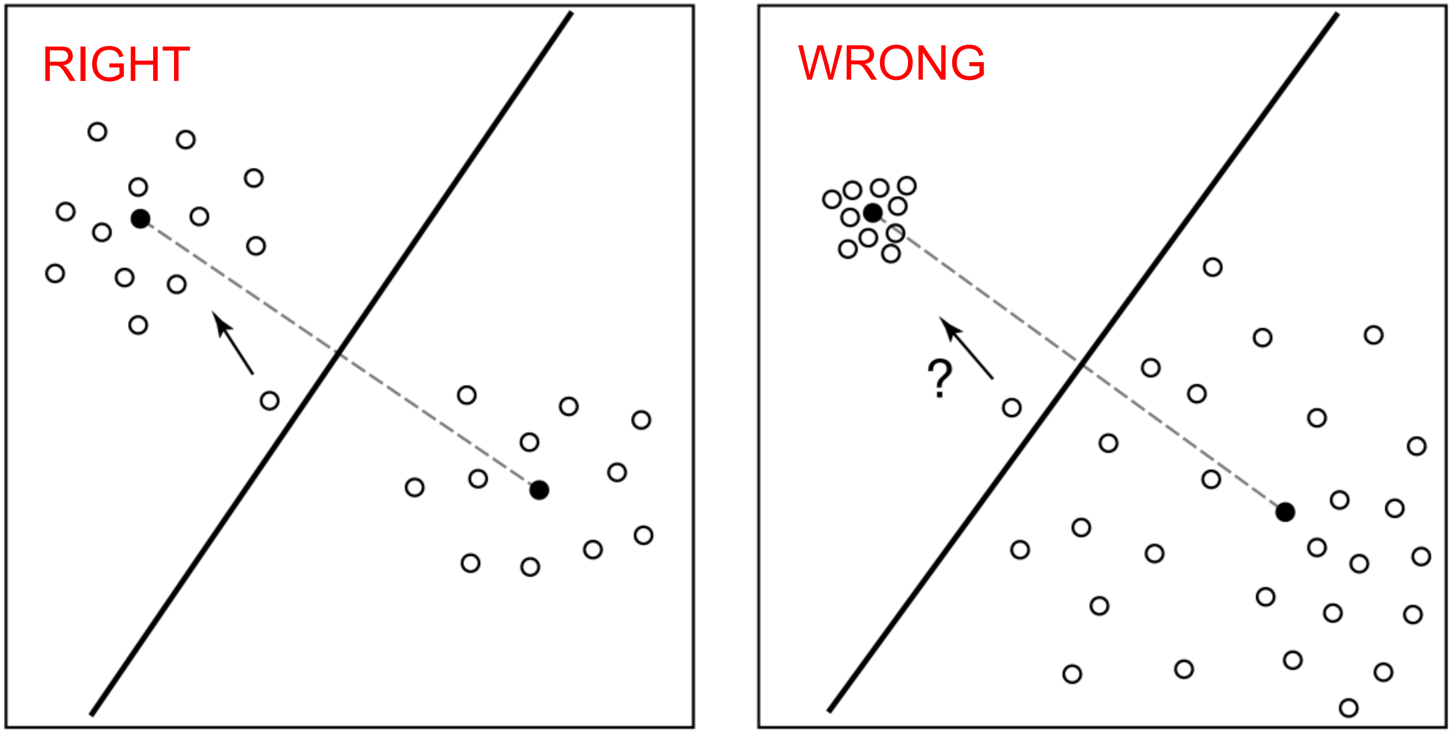
Assign each pixel to the class
of the closest center (class mean)
if
distance < threshold

$$\| \mathbf{x} - \mathbf{m}_i \| > \mathrm{T}, \ \forall i \quad \Rightarrow \quad \mathbf{x} \in \omega_0$$

Leave pixel unclassified (class ω_0)
if
all class centers are at distances
larger than threshold

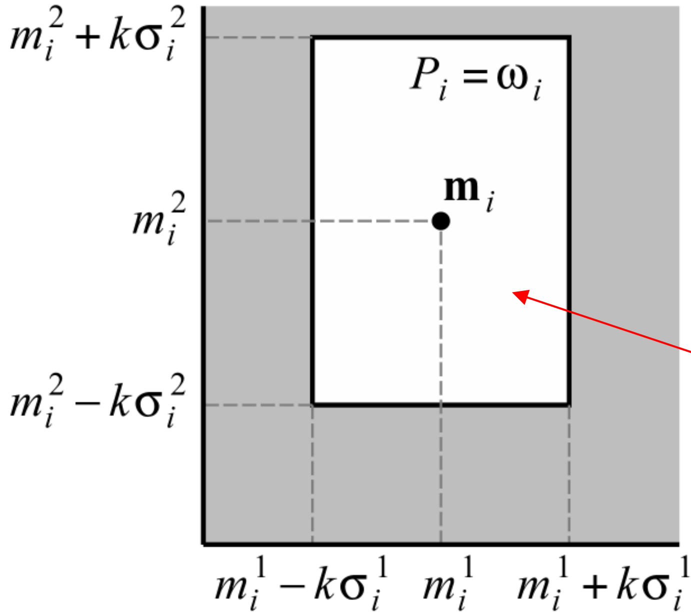
Classification with Euclidean distance

$$d_E(\mathbf{x}, \mathbf{x}') = \| \mathbf{x} - \mathbf{x}' \| = \sqrt{(x^1 - x'^1)^2 + (x^2 - x'^2)^2 + \ldots + (x^B - x'^B)^2}$$



The role of statistics (dispersion) in classification

Classification with the parallelepiped method



standard deviations for each band

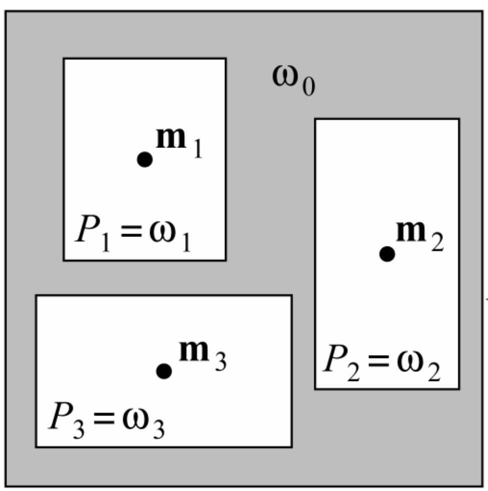
$$\sigma_i^j = \sqrt{(\mathbf{C}_i)_{jj}} \qquad j = 1, 2, \ldots, B$$

parallelepipeds P_i

$$\mathbf{x} = [x^1 \ \ldots \ x^j \ \ldots \ x^B]^T \in P_i$$

$$\Leftrightarrow \quad m_i^j - k\ \sigma_i^j \le x^j \le m_i^j + k\ \sigma_i^j$$

$$j = 1, 2, \ldots, B$$



Classification:

$$\mathbf{x} \in P_i \quad \Rightarrow \quad \mathbf{x} \in \omega_i$$

$$\mathbf{x} \notin \boxed{?}_i P_i \quad \Rightarrow \quad \mathbf{x} \in \omega_0$$

Classification with the Mahalanobis distance

Mahalanobis distance: $d_M(\mathbf{x},\mathbf{x}') = \sqrt{(\mathbf{x}-\mathbf{x}')^T\,\mathbf{C}^{-1}\,(\mathbf{x}-\mathbf{x}')}$

$$\mathbf{C} = \frac{1}{N}\sum_i \sum_{\mathbf{x}\in S_i} (\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^T = \frac{1}{N}\sum_i n_i \mathbf{C}_i$$

(total covariance matrix)

Classification (simple): $d_M(\mathbf{x},\mathbf{m}_i) < d_M(\mathbf{x},m_k),\quad \forall k \neq i \quad \Rightarrow \quad \mathbf{x}\in\omega_i$

Classification with threshold:

$$\left.\begin{array}{l} d_M(\mathbf{x},\mathbf{m}_i) < d_M(\mathbf{x},m_k),\ \forall k\neq i \\ d_M(\mathbf{x},\mathbf{m}_i) \leq \mathrm{T}, \end{array}\right\} \Rightarrow \mathbf{x}\in\omega_i$$

$$d_M(\mathbf{x},\mathbf{m}_i) > T,\ \forall i \quad \Rightarrow \quad \mathbf{x}\in\omega_0$$

Classification with the maximum likelihood method

Probability distribution density function or likelihood function of class ω_i:

$$l_i(\mathbf{x}) = \frac{1}{(2\pi)^{B/2}\,|\,\mathbf{C}_i\,|^{1/2}} \;\exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m}_i)^T\,\mathbf{C}_i^{-1}\,(\mathbf{x}-\mathbf{m}_i)\,\right]$$

Classification: $\quad l_i(\mathbf{x}) > l_k(\mathbf{x}) \quad \forall k \neq i \quad \Leftrightarrow \quad \mathbf{x} \in \omega_i$

Equivalent use of decision function:

$$d_i(\mathbf{x}) = 2\,\ln[l_i(\mathbf{x})] + B\,\ln(2\pi) = -\ln|\,\mathbf{C}_i\,| - (\mathbf{x}-\mathbf{m}_i)^T\,\mathbf{C}_i^{-1}\,(\mathbf{x}-\mathbf{m}_i)$$

$$d_i(\mathbf{x}) > d_k(\mathbf{x}) \quad \forall k \neq i \quad \Leftrightarrow \quad \mathbf{x} \in \omega_i$$

Classification using the Bayesian approach

N : total number of pixels in the image (i.e. in each band)

B : number of bands,

$\omega_1, \omega_2, \ldots, \omega_K$: the K classes present in the image

N_i : number of image pixels belonging to the class ω_i ($i = 1,2, \ldots, K$)

n_x : number of pixels with value $\mathbf{x}$ (= vector of values in all bands)

n_{xi} : number of pixels with value $\mathbf{x}$ which also belong to the class ω_i

Classification using the Bayesian approach

N : total number of pixels in the image (i.e. in each band)

B : number of bands,

$\omega_1, \omega_2, \ldots, \omega_K$: the K classes present in the image

N_i : number of image pixels belonging to the class ω_i $(i = 1,2, \ldots, K)$

n_x : number of pixels with value $\mathbf{x}$ (= vector of values in all bands)

n_{xi} : number of pixels with value $\mathbf{x}$ which also belong to the class ω_i

$$\sum_{\mathbf{x}} n_{\mathbf{x}} = N \qquad \sum_{i} N_i = N \qquad \sum_{\mathbf{x}} n_{\mathbf{x}i} = N_i \qquad \sum_{i} n_{\mathbf{x}i} = n_{\mathbf{x}}$$

Classification using the Bayesian approach

N : total number of pixels in the image (i.e. in each band)

B : number of bands,

$\omega_1, \omega_2, \ldots, \omega_K$: the K classes present in the image

N_i : number of image pixels belonging to the class ω_i $(i = 1,2, \ldots, K)$

n_x : number of pixels with value $\mathbf{x}$ (= vector of values in all bands)

n_{xi} : number of pixels with value $\mathbf{x}$ which also belong to the class ω_i

$$\sum_{\mathbf{x}} n_{\mathbf{x}} = N \qquad \sum_{i} N_i = N \qquad \sum_{\mathbf{x}} n_{\mathbf{x}i} = N_i \qquad \sum_{i} n_{\mathbf{x}i} = n_{\mathbf{x}}$$

Basic identity:

$$\frac{n_{xi}}{n_x} = \frac{\dfrac{n_{xi}}{N}}{\dfrac{n_x}{N}}$$

Classification using the Bayesian approach

$N:$ total number of pixels in the image (i.e. in each band)

$B:$ number of bands,

$\omega_1, \omega_2, \ldots, \omega_K:$ the K classes present in the image

$N_i:$ number of image pixels belonging to the class ω_i $(i = 1,2, \ldots, K)$

$n_x:$ number of pixels with value $\mathbf{x}$ (= vector of values in all bands)

$n_{xi}:$ number of pixels with value $\mathbf{x}$ which also belong to the class ω_i

$$\sum_{\mathbf{x}} n_{\mathbf{x}} = N \qquad \sum_{i} N_i = N \qquad \sum_{\mathbf{x}} n_{\mathbf{x}i} = N_i \qquad \sum_{i} n_{\mathbf{x}i} = n_{\mathbf{x}}$$

Basic identity:

$$\frac{n_{xi}}{n_x} = \frac{\dfrac{n_{xi}}{N}}{\dfrac{n_x}{N}} = \frac{\dfrac{n_{xi}}{N_i}\dfrac{N_i}{N}}{\dfrac{n_x}{N}}$$

Classification using the Bayesian approach

N : total number of pixels in the image (i.e. in each band)

B : number of bands,

$\omega_1, \omega_2, \ldots, \omega_K$: the K classes present in the image

N_i : number of image pixels belonging to the class ω_i ($i = 1,2, \ldots, K$)

n_x : number of pixels with value $\mathbf{x}$ (= vector of values in all bands)

n_{xi} : number of pixels with value $\mathbf{x}$ which also belong to the class ω_i

$$\sum_{\mathbf{x}} n_{\mathbf{x}} = N \qquad \sum_{i} N_i = N \qquad \sum_{\mathbf{x}} n_{\mathbf{x}i} = N_i \qquad \sum_{i} n_{\mathbf{x}i} = n_{\mathbf{x}}$$

Basic identity:

$$\frac{n_{xi}}{n_x} = \frac{\dfrac{n_{xi}}{N}}{\dfrac{n_x}{N}} = \frac{\dfrac{n_{xi}}{N_i}\dfrac{N_i}{N}}{\dfrac{n_x}{N}} \qquad \left(\frac{n_{xi}}{n_x}\right) = \frac{\left(\dfrac{n_{xi}}{N_i}\right)\left(\dfrac{N_i}{N}\right)}{\left(\dfrac{n_x}{N}\right)}$$

$$p(\omega_i) = \frac{N_i}{N}$$

probability of a pixel to belong to the class ω_i

$$p(\mathbf{x}) = \frac{n_{\mathbf{x}}}{N}$$

probability of a pixel to have the value **x**

$$p(\mathbf{x} \mid \omega_i) = \frac{n_{\mathbf{x}i}}{N_i}$$

probability of a pixel belonging to the class ω_i to have value **x** (conditional probability)

$$p(\omega_i \mid \mathbf{x}) = \frac{n_{\mathbf{x}i}}{n_{\mathbf{x}}}$$

probability of a pixel having value **x** to belong to the class ω_i (conditional probability)

$$p(\mathbf{x}, \omega_i) = \frac{n_{\mathbf{x}i}}{N}$$

probability of a pixel to have the value **x** and to simultaneously belong to ω_i (joint probability)

$$p(\omega_i) = \frac{N_i}{N}$$ probability of a pixel to belong to the class ω_i

$$p(\mathbf{x}) = \frac{n_{\mathbf{x}}}{N}$$ probability of a pixel to have the value **x**

$$p(\mathbf{x} \mid \omega_i) = \frac{n_{\mathbf{x}i}}{N_i}$$ probability of a pixel belonging to the class ω_i to have value **x** (conditional probability)

$$p(\omega_i \mid \mathbf{x}) = \frac{n_{\mathbf{x}i}}{n_{\mathbf{x}}}$$ probability of a pixel having value **x** to belong to the class ω_i (conditional probability)

$$p(\mathbf{x}, \omega_i) = \frac{n_{\mathbf{x}i}}{N}$$ probability of a pixel to have the value **x** and to simultaneously belong to ω_i (joint probability)

$$\left(\frac{n_{xi}}{n_x}\right) = \frac{\left(\frac{n_{xi}}{N_i}\right)\left(\frac{N_i}{N}\right)}{\left(\frac{n_x}{N}\right)}$$

$$p(\omega_i) = \frac{N_i}{N}$$ probability of a pixel to belong to the class ω_i

$$p(\mathbf{x}) = \frac{n_{\mathbf{x}}}{N}$$ probability of a pixel to have the value **x**

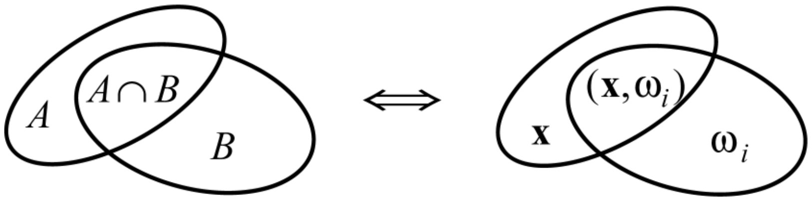
$$p(\mathbf{x} \,|\, \omega_i) = \frac{n_{\mathbf{x}i}}{N_i}$$ probability of a pixel belonging to the class ω_i to have value **x** (conditional probability)

$$p(\omega_i \,|\, \mathbf{x}) = \frac{n_{\mathbf{x}i}}{n_{\mathbf{x}}}$$ probability of a pixel having value **x** to belong to the class ω_i (conditional probability)

$$p(\mathbf{x}, \omega_i) = \frac{n_{\mathbf{x}i}}{N}$$ probability of a pixel to have the value **x** and to simultaneously belong to ω_i (joint probability)

$$\left(\frac{n_{xi}}{n_x}\right) = \frac{\left(\frac{n_{xi}}{N_i}\right)\left(\frac{N_i}{N}\right)}{\left(\frac{n_x}{N}\right)} \quad \text{fi} \quad p(\omega_i \,|\, \mathbf{x}) = \frac{p(\mathbf{x} \,|\, \omega_i)\, p(\omega_i)}{p(\mathbf{x})}$$

formula of Bayes



The Bayes theorem:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

$$\Pr(A|B)\Pr(B) = \Pr(A \cap B) = \Pr(B|A)\Pr(A)$$

$$\Pr(B|A) = \frac{\Pr(A|B)\Pr(B)}{\Pr(A)}$$

event A = occurrence of the value $\mathbf{x}$

event B = occurence of the class ω_i

$$p(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)\, p(\omega_i)}{p(\mathbf{x})}$$

Classification: $p(\omega_i|\mathbf{x}) > p(\omega_k|\mathbf{x}) \quad \forall k \neq i \quad \Rightarrow \quad \mathbf{x} \in \omega_i$

$p(\mathbf{x})$ = not necessary (common constant factor)

Classification: $p(\mathbf{x}|\omega_i)\, p(\omega_i) > p(\mathbf{x}|\omega_k)\, p(\omega_k) \quad \forall k \neq i \quad \Rightarrow \quad \mathbf{x} \in \omega_i$

Classification: $$p(\mathbf{x}|\omega_i)p(\omega_i) = \max_k \,[p(\mathbf{x}|\omega_k)p(\omega_k) \quad \Rightarrow \quad \mathbf{x} \in \omega_i$$

for Gaussian distribution:

$$p(\mathbf{x}\,|\,\omega_i) = l_i(\mathbf{x}) = \frac{1}{(2\pi)^{B/2}\,|\,\mathbf{C}_i\,|^{1/2}} \,\exp\left\{-\frac{1}{2}\,(\mathbf{x}-\mathbf{m}_i)^T\,\mathbf{C}_i^{-1}\,(\mathbf{x}-\mathbf{m}_i)\,\right\}$$

Instead of $$p(\mathbf{x}\,|\,\omega_i)\ \mathrm{p}(\omega_i) = \max$$

equivalent $$\ln[p(\mathbf{x}\,|\,\omega_i)\ \mathrm{p}(\omega_i)] = \ln[\mathrm{p}(\mathbf{x}\,|\,\omega_i) + \ln[p(\omega_i) = \max$$

$$-\frac{1}{2}\,(\mathbf{x}-\mathbf{m}_i)^T\,\mathbf{C}_i^{-1}\,(\mathbf{x}-\mathbf{m}_i) - \frac{1}{2}\,\ln[\,|\,\mathbf{C}_i\,| + \ln[p(\omega_i)] = \max$$

or finally:

$$(\mathbf{x}-\mathbf{m}_i)^T\,\mathbf{C}_i^{-1}\,(\mathbf{x}-\mathbf{m}_i) + \ln[\,|\,\mathbf{C}_i\,| + \ln[p(\omega_i)] = \min$$

Bayesian Classification for Gaussian distribution :

$$\boxed{(\mathbf{x}-\mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{x}-\mathbf{m}_i) + \ln[\,|\mathbf{C}_i| + \ln[p(\omega_i)] = \min}$$

SPECIAL CASES:

$$p(\omega_1) = p(\omega_2) = \ldots = p(\omega_K) \quad \Rightarrow \quad \boxed{(\mathbf{x}-\mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{x}-\mathbf{m}_i) + \ln[\,|\mathbf{C}_i| = \min}$$

Maximum Likelihood !

$$\left.\begin{array}{c} p(\omega_1) = p(\omega_2) = \ldots = p(\omega_K) \\ \mathbf{C}_1 = \mathbf{C}_2 = \ldots = \mathbf{C}_K = \mathbf{C} \end{array}\right\} \Rightarrow \quad \boxed{(\mathbf{x}-\mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{x}-\mathbf{m}_i) = \min}$$

Mahalanobis distance !

$$\left.\begin{array}{c} p(\omega_1) = p(\omega_2) = \ldots = p(\omega_K) \\ \mathbf{C}_1 = \mathbf{C}_2 = \ldots = \mathbf{C}_K = \mathbf{I} \end{array}\right\} \Rightarrow \quad \boxed{(\mathbf{x}-\mathbf{m}_i)^T (\mathbf{x}-\mathbf{m}_i) = \min}$$

Euclidean distance !

Evaluation of results

Two possibilities for the training samples:

1) The training samples are used to set the chracteristics of the classes, then are by definition attributed to their class.

2) The training samples are used to set the chracteristics of the classes, then they are classified with the others.

In the second case, the classification of the training can be investigated to check the accuracy.

The assessment is not independent because the same pixels are used both to train and to check.

More rigorous

A percentage of the sample pixels are used to train the algorithm.

The remaining part can be used to check the accuracy of results

Evaluation of results: the confusion matrix

The confusion matrix:
in the rows: the true classes of check pixels
in the columns: the predicted class of check pixels

The perfect classification

Classified \ True class (%)	Forest	Grass	…	Urban
Forest	100	0		0
Grass	0	100		0
…			…	
Urban	0			100

The perfect classification is impossible:
confusion between Forest and Grass is acceptable…
but, for example, confusion between Forest and Urban is not acceptable!