



Máxima Verossimilhança Gaussiana

O problema pode também ser resolvido com uma abordagem probabilística...

Qual a probabilidade de uma determinada classe ocorrer em cada posição da imagem?

Máxima Verossimilhança

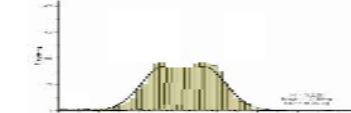
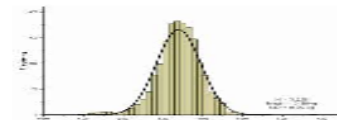
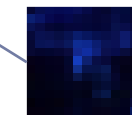
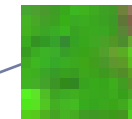
Objetivo: Dado um pixel “x” (ou seja, lemos um pixel na imagem e conhecemos seus valores digitais) definir a classe mais provável w_i , dentre um conjunto de classes possíveis, com base nas estatísticas das amostras:

$$[W = w_1, w_2, \dots, w_n].$$

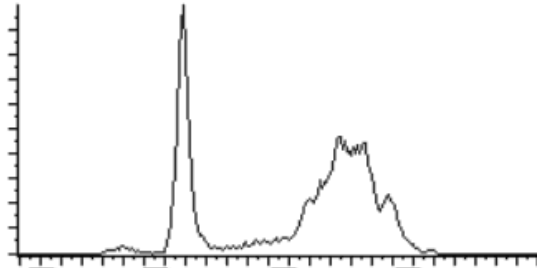
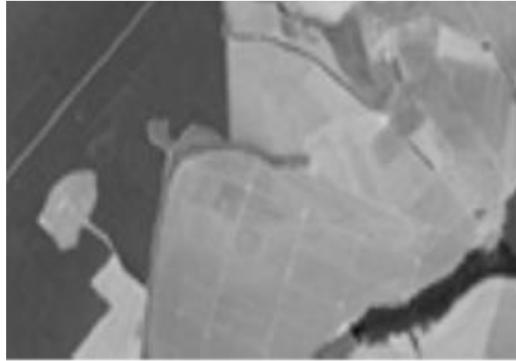
estatísticas



$$x = \begin{bmatrix} B \\ G \\ R \\ NIR \\ MIR \end{bmatrix}$$

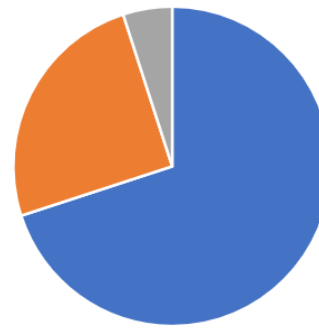
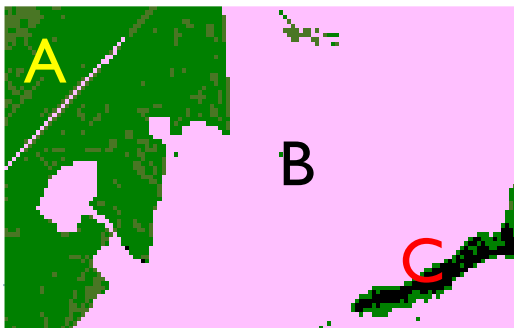


Claro que...



Na uma imagem digital

- ▶ todos os valores digitais podem ocorrer, porém com diferentes probabilidades $p(x)$
- ▶ As classes não ocorrem com a mesma frequência, logo existe diferente probabilidade de cada classe ocorrer $p(w)$

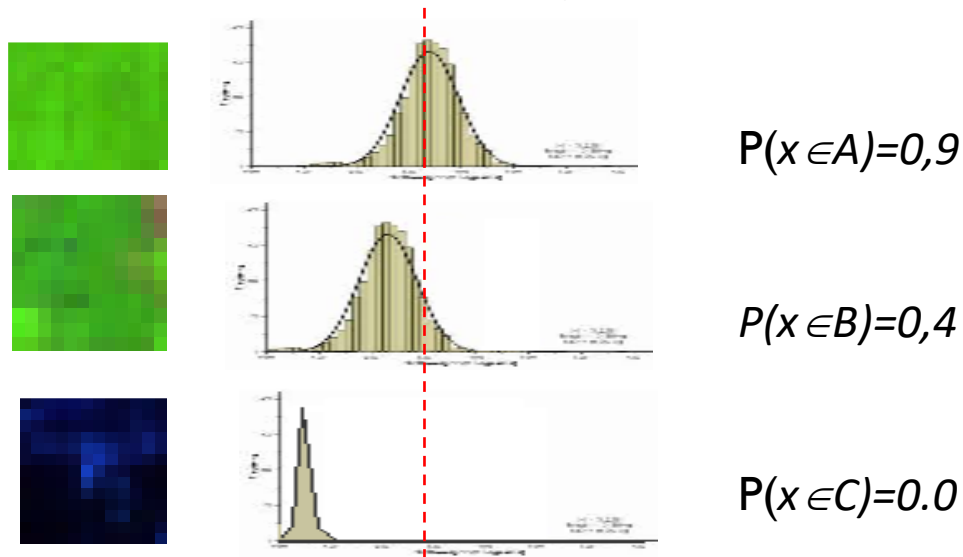


■ vegetação ■ solo ■ água

Máxima Verossimilhança

Com base nas amostras podemos estimar a probabilidade com que cada valor ocorre em cada classe. Se um valor ocorre muitas vezes em uma classe, então tem alta probabilidade.

Se conhecemos todas as probabilidades $P(x \in w_i)$, então podemos escolher a classe associada à maior probabilidade.



Ou seja, atribuir o pixel “ x ” à classe w_i , se:

$$P(x \in w_i) > P(x \in w_k)$$

Máxima Verossimilhança

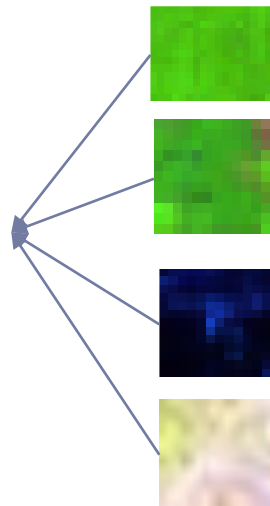
Então, queremos saber a probabilidade de uma determinada classe ocorrer, dado que o valor digital “x” ocorre $P(w/x)$.

Se conhecemos todas as probabilidades $P(w_k/x)$, então escolhemos a classe associada à maior probabilidade.

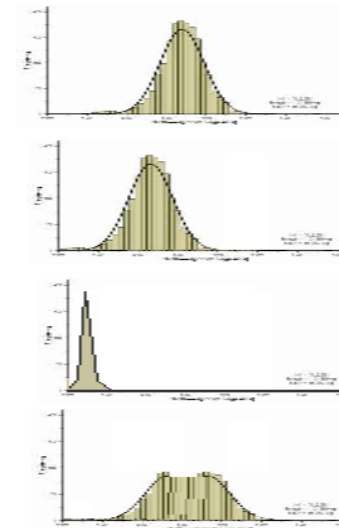


DADO

$$x = \begin{bmatrix} B \\ G \\ R \\ NIR \\ MIR \end{bmatrix}$$



estatísticas



$P(x, \text{dado } A)$

$P(x, \text{dado } B)$

$P(x, \text{dado } C)$

...

Dois eventos, duas possibilidades

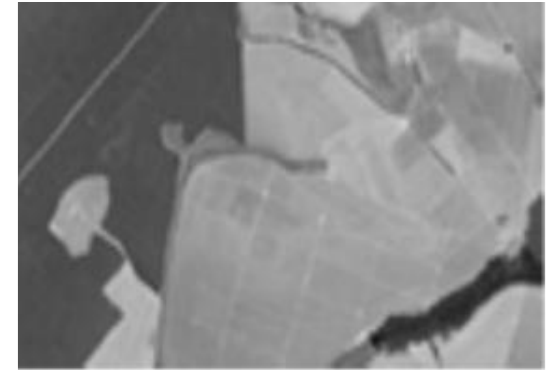
O pixel tem valor x ;
A classe “ k ” ocorre nesse local.

Exemplo:

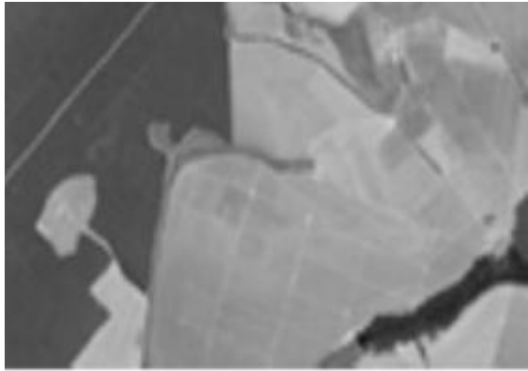
- ▶ O pixel tem valor 21 no IVP (na escala 1-255);
- ▶ A classe “água” ocorre nesse local.

- ▶ **Se o pixel tem valor= x , qual a probabilidade dele ser da classe “ k ”?**
- ▶ *Se o pixel tem valor=21, qual a probabilidade dele ser da classe água?*

- ▶ **Se a classe é “ k ” qual a probabilidade dele ter o valor x ?**
- ▶ *Se sabemos que o pixel é da classe água, qual a probabilidade dele ter valor=21?*



Os dois eventos ocorrem ...



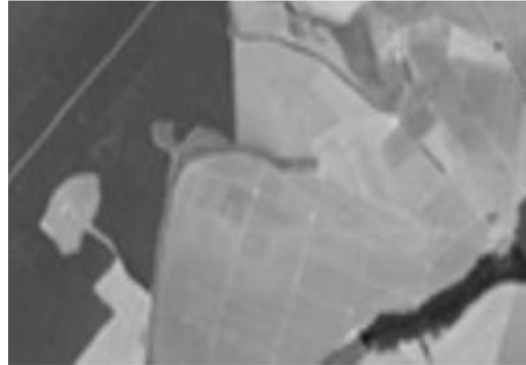
O pixel tem valor "x"



A classe é "k"

Qual é probabilidade de uma classe " w_k " ocorrer ao mesmo tempo que o pixel com valor "x" ocorre ... $P(\text{pixel}=x \text{ e classe}=w_k)$

Se conhecemos o valor x



Se conhecemos o valor do pixel...

A probabilidade de uma classe “ w_k ” ocorrer ao mesmo tempo que o pixel com valor “ x ” ocorre ... $P(\text{pixel}=x \text{ e classe}=w_k)$ pode ser descrita pela probabilidade condicional da classe “ w_k ” dado que o pixel “ x ” ocorre

$$P(x, w_k) = p(w_k | x) * p(x)$$

- ▶ Se o pixel tem valor= x , qual a probabilidade dele ser da classe “ k ”?
-

Se soubermos qual é a classe...

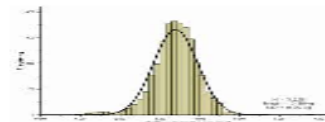
Por outro lado, se sabemos que o pixel é da classe “k”, e como temos as estatísticas das amostras, podemos calcular a mesma probabilidade conjunta como o evento:

O valor “x” ocorre, dado que a classe w_k ocorreu

$$P(x, w_k) = p(x | w_k) * p(w)$$



Se o pixel é de solo, qual a probabilidade dele ter valor 10?



$P(x, \text{ dado SOLO})$



$P(x, \text{ dado VEG})$



$P(x, \text{ dado ÁGUA})$



► Igualando:

$$p(w_A|x) * p(x) = p(x|w_A) * p(w_A)$$

Podemos calcular a probabilidade da classe “ w_A ” ocorrer, dado que lemos um pixel com valor digital “ x ” da imagem:

$$p(w_A|x) = \frac{p(x|w_A) * p(w_A)}{p(x)}$$



A probabilidade é calculada como:

$$P(w|x) = \frac{p(x|w) * p(w)}{p(x)}$$

$P(x)$: Prob. do pixel “ x ” ocorrer na imagem.

$P(x/w)$: função densidade de probabilidade da classe “ w ”.

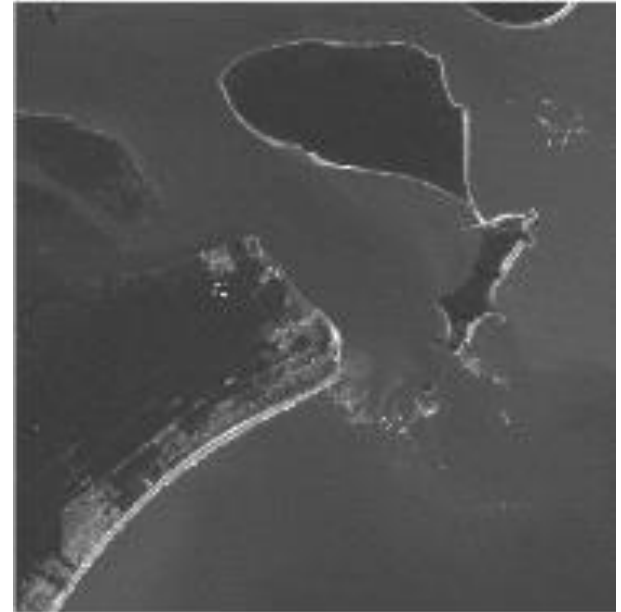
$P(w)$: Probabilidade “a priori” da classe.

- ▶ Explicados a seguir...
-

$P(x)$: Prob. do pixel “x” ocorrer na imagem.

$$P(w|x) = \frac{p(x|w) * p(w)}{p(x)}$$

Os valores digitais na imagem não ocorrem com a mesma frequência. Qual a probabilidade de encontrar um valor 255 nesta imagem?



Como as probabilidades serão comparadas, este valor é igual para todas as classes e se torna redundante, e pode ser desconsiderado

$$P(x \in W_A) > P(x \in W_B)$$
$$\frac{p(x|W_A) * p(W_A)}{p(x)} \text{ vs. } \frac{p(x|W_B) * p(W_B)}{p(x)}$$



$P(x)$: Prob. do pixel “x” ocorrer na imagem.

Então calcula-se

$$G(w|x) = p(x|w) * p(w)$$

Uma função discriminante que é proporcional à probabilidade, mas que não é o verdadeiro valor da probabilidade.

Porém, esta função serve para encontrar a classe “mais provável”

Com isto, a regra passa ser:

atribuir o pixel “x” à classe A, se:

$$G(W_A|x) > G(W_B|x)$$

$$p(x|W_A) * p(W_A) > p(x|W_B) * p(W_B)$$

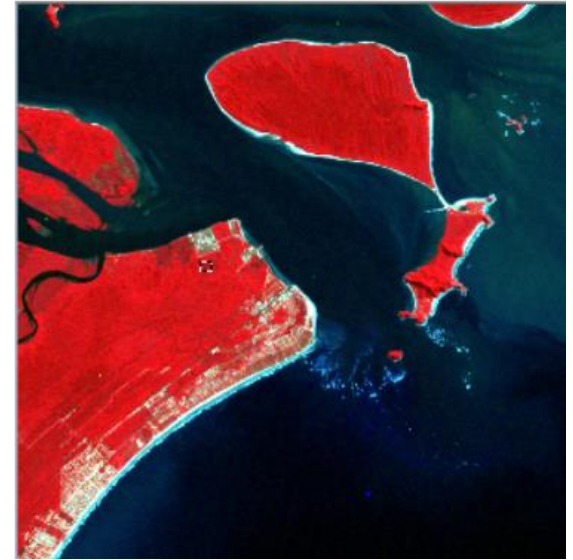
$P(w)$: Probabilidade “a priori”

$$P(w|x) = \frac{p(x|w) * p(w)}{p(x)}$$

$P(w)$ descreve a frequência com que a classe ocorre na imagem.

Qual a probabilidade de encontrar um pixel de areia nesta imagem?

O que é mais provável, areia ou água?



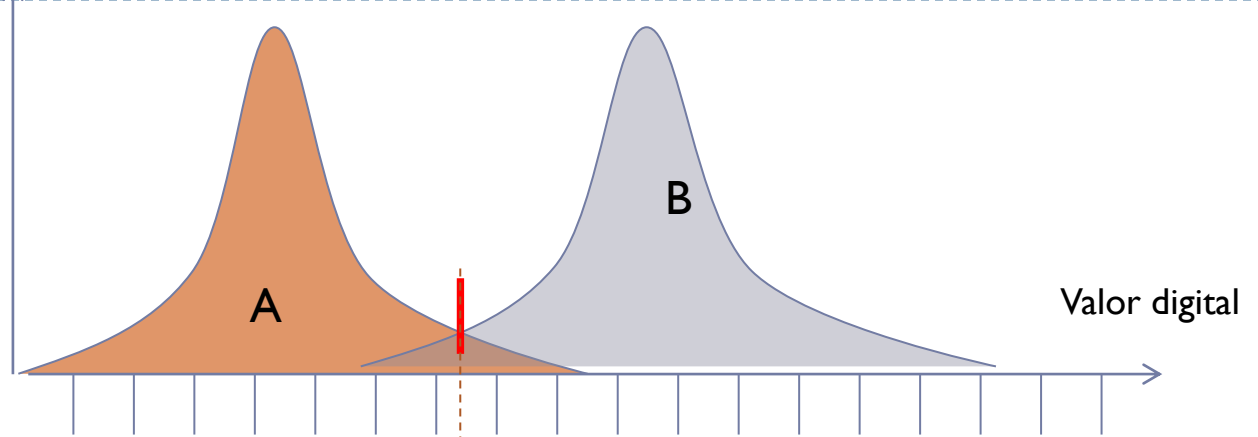
Para saber isto, deveríamos conhecer quantos pixels de cada classe tem na imagem, mas a verdadeira classificação dos pixels é desconhecida (é o que queremos)

Para não favorecer nenhuma classe injustamente, o valor de $p(x)$ é igual para todas as classes.

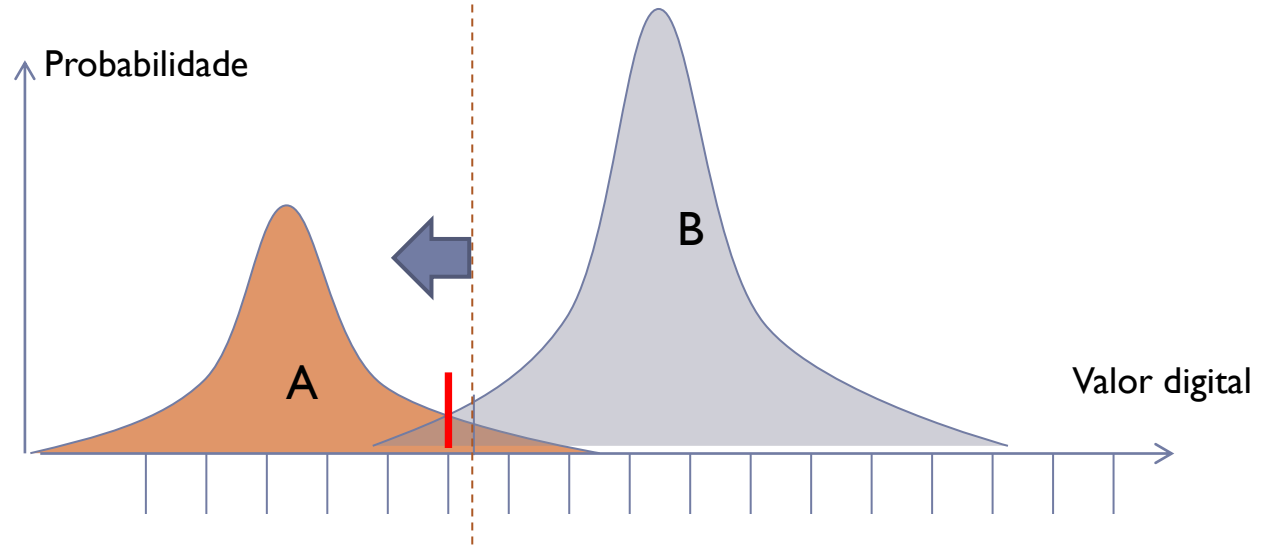
Efeito da Probabilidade a priori

Probabilidade

$$P(A)=P(B)$$



$$P(A)<P(B)$$



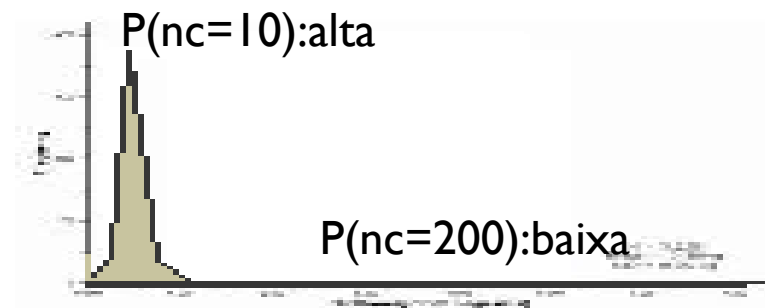
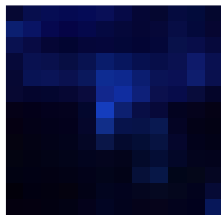
$P(x/w)$: função densidade de probabilidade

$$P(w|x) = \frac{p(x|w) * p(w)}{p(x)}$$

$P(x/w)$: A função densidade de probabilidade da classe descreve a variação dos valores digitais dentro de cada classe.

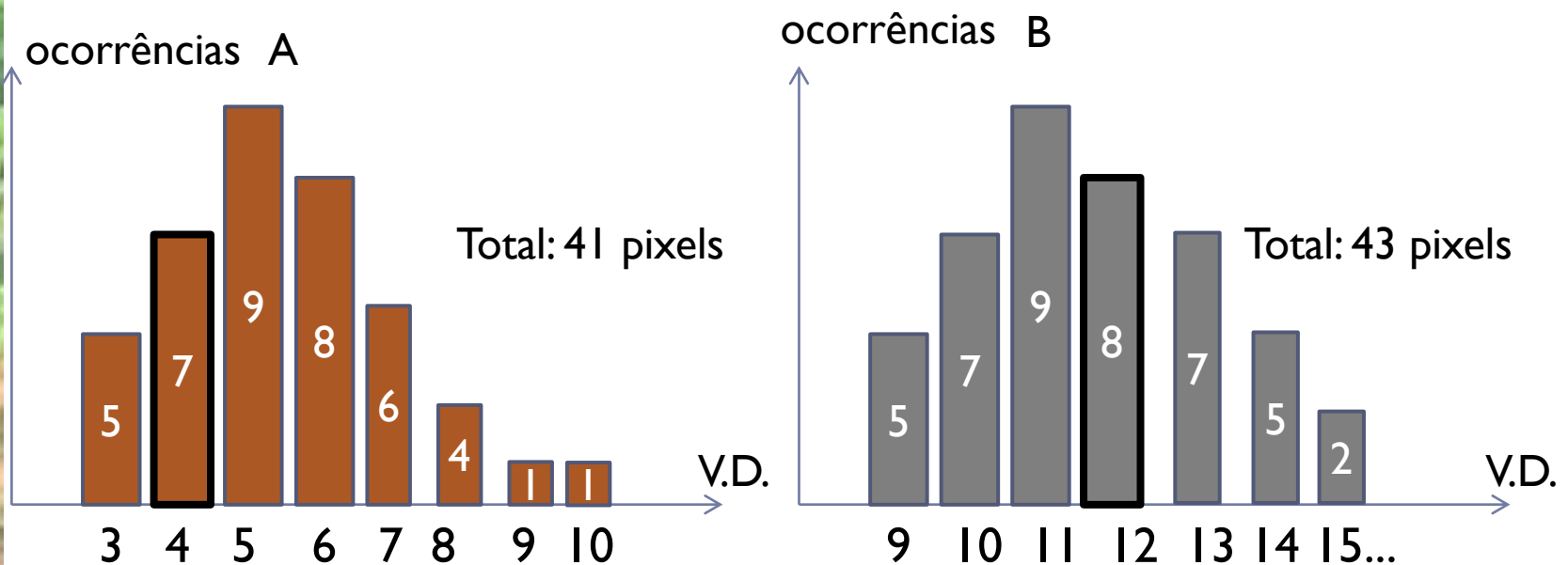
Que valores são mais frequentes em uma classe? Quais são pouco frequentes?

Cada classe tem uma distribuição própria.



$p(x|w)$ dada uma classe, qual a $prob(x)$?

- ▶ Dado que temos definida a classe, qual a probabilidade de encontrar um pixel com determinado valor?

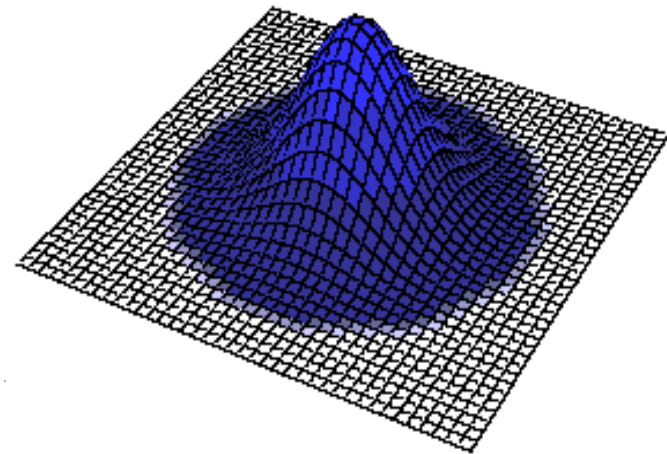


Qual a probabilidade de encontrar um pixel com valor=9 na classe A?
E na B?

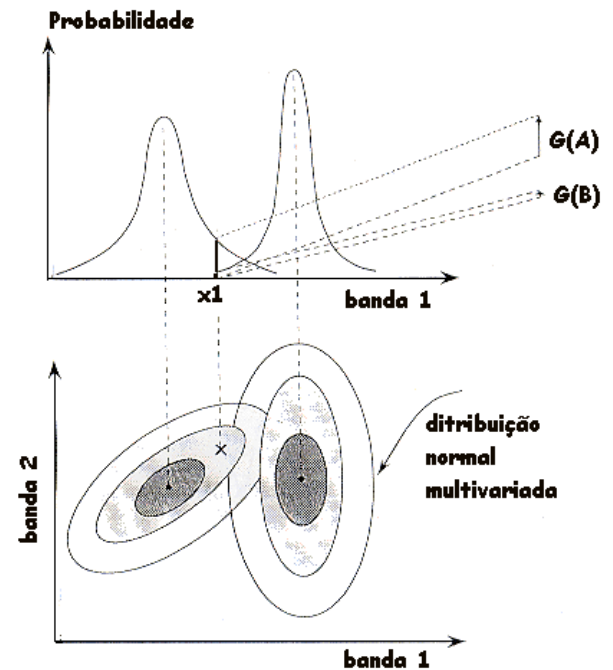
$p(x|w)$ = distribuição dos pixels considerando apenas os membros de uma classe "w"

- ▶ No caso multidimensional, o que se encontra geralmente na classificação de várias bandas espectrais em sensoriamento remoto, a expressão muda, substituindo-se a média pelo vetor de médias (m) e a variância pela matriz variância-covariância (C), para considerar a dispersão dos dados em torno da média em todas as bandas. Assim:

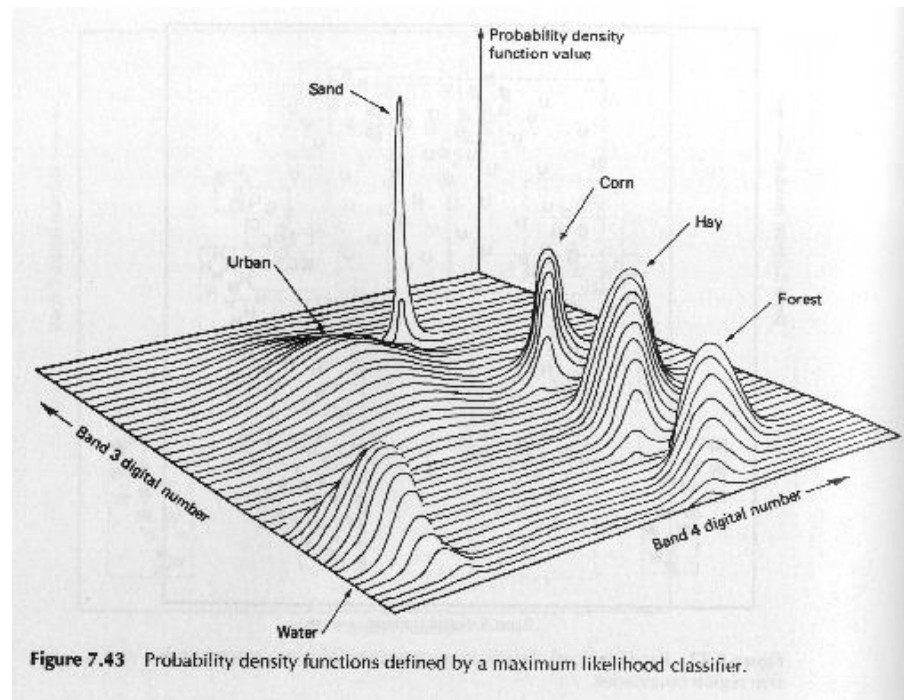
$$p(x|w) = \frac{1}{|C|^{1/2} \sqrt{2\pi}} e^{-0.5 (x-m)^T C^{-1} (x-m)}$$



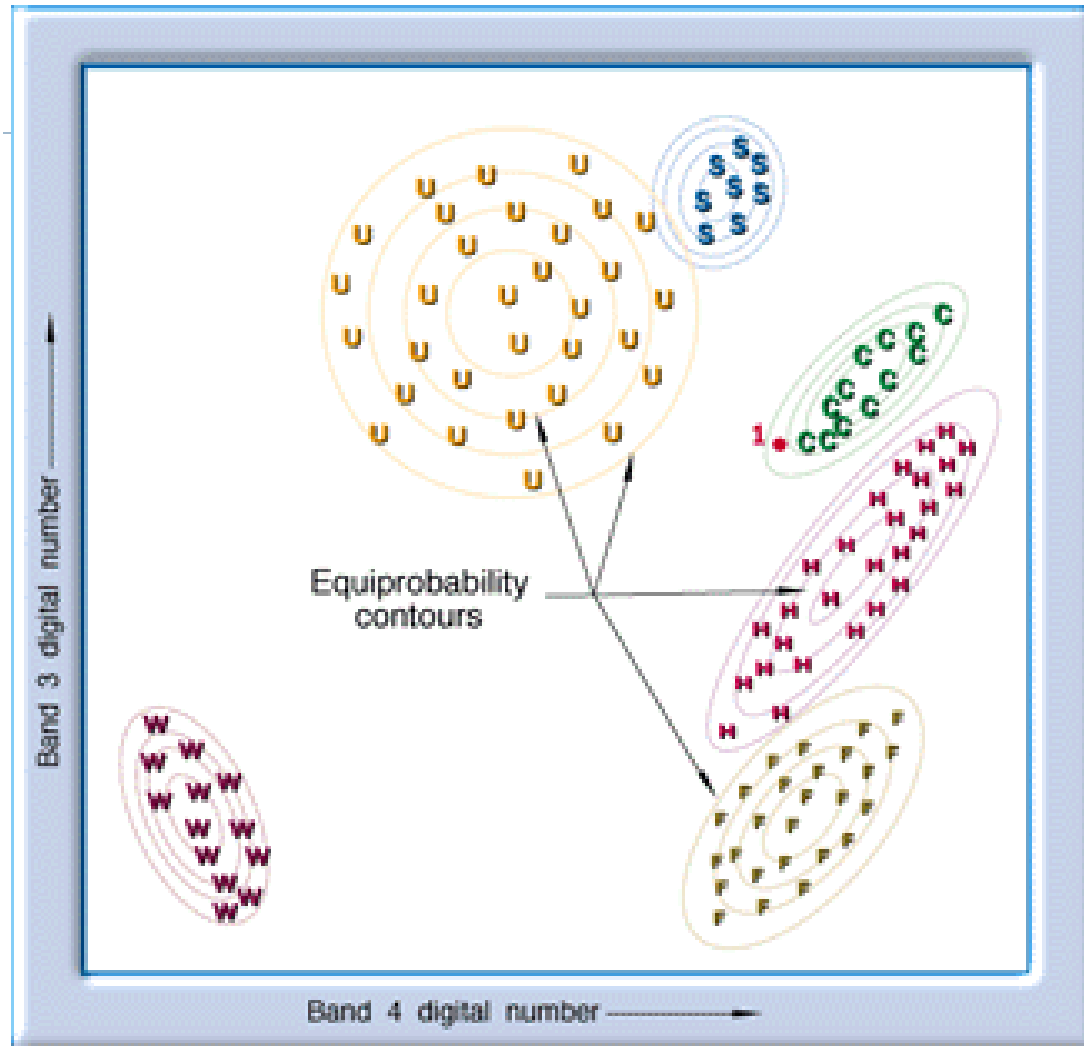
A inclusão da matriz variância covariância é fundamental para levar em consideração a dispersão dos dados em torno das médias, que pode ser paralela aos eixos, com intensidades diferentes, e apresentar correlação entre bandas.



► Swain & Davies



- Considerando duas bandas: a probabilidade varia em função das leituras de cada banda para cada classe e cada classe pode ter uma distribuição diferente.



- ▶ Contornos de mesma probabilidade variam em função da matriz var-covariância



Resumindo e simplificando...

Se a “**probabilidade a priori**” é considerada constante

E é assumido que $p(x)$ pode ser descartado, para fins de comparação ...

Atribuir o pixel “ x ” à classe w_j , se:

$$P(x \in w_j) > P(x \in w_k)$$
$$\frac{p(x|w_j) * p(w_j)}{p(x)} > \frac{p(x|w_k) * p(w_k)}{p(x)}$$
$$p(x|w_j) * p(w_j) > p(x|w_k) * p(w_k)$$
$$p(x|w_j) > p(x|w_k)$$

Se simplifica para, Atribuir o pixel “ x ” à classe w_j , se:

$$p(x|w_i) > p(x|w_k)$$

Assim, apenas a distribuição dos pixels dentro da classe é analisada. Neste caso, não se calcula o verdadeiro valor da probabilidade, mas sim um valor que permite discriminar as classes, por isso se fala em função discriminante.

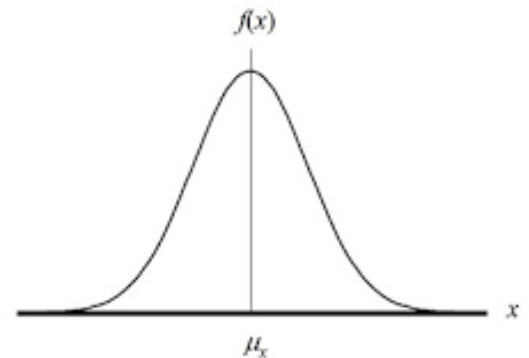
$P(x/w)$: função densidade de probabilidade

O problema é estimar a função densidade de probabilidade de cada classe. Estimar a variação das probabilidades dos valores digitais dentro de cada classe.

Uma função pode ser usada para modelar a distribuição dos pixels em cada classe, com base nas amostras de treinamento.

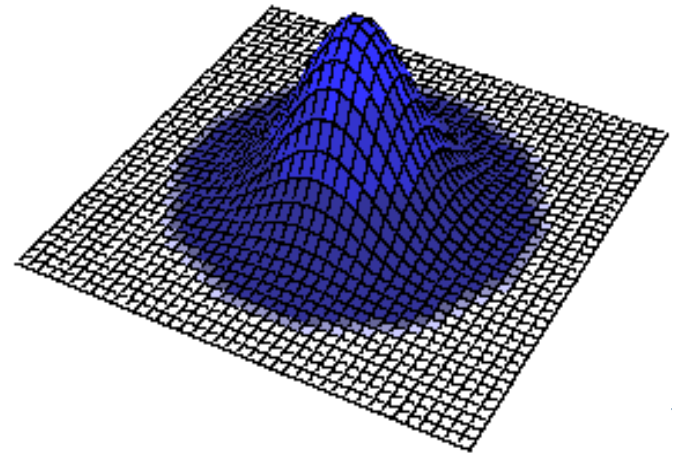
Na maioria dos casos, as classes seguem uma distribuição Gaussiana, ou Normal, que pode ser descrita, no caso unidimensional, com a média μ e a variância σ da classe.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5 \left(\frac{(x-\mu)}{\sigma}\right)^2}$$

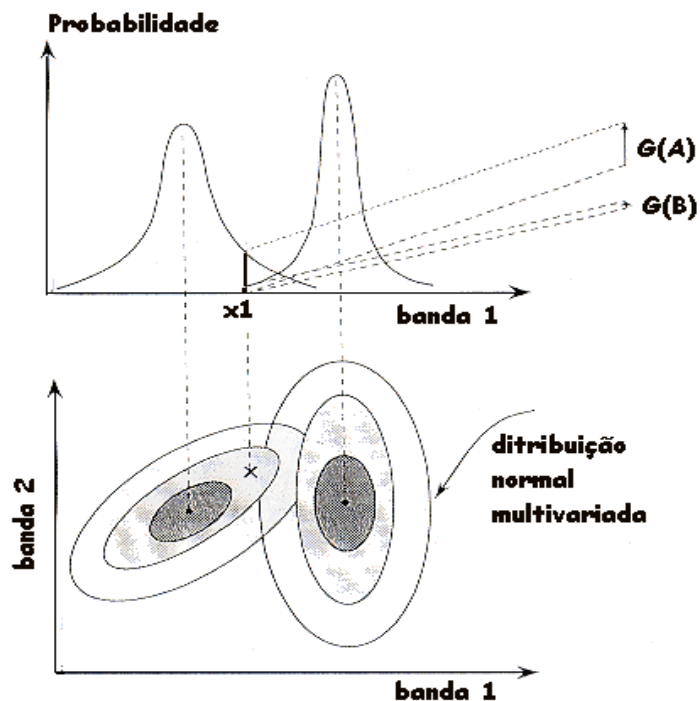


- ▶ No caso multidimensional, a expressão muda, substituindo-se a média pelo vetor de médias (m) e a variância pela matriz variância-covariância (C), para considerar a dispersão dos dados em torno da média em todas as bandas:

$$p(x|w) = \frac{1}{|C|^{1/2} \sqrt{2\pi}} e^{-0.5 (x-m)^T C^{-1} (x-m)}$$



- ▶ A inclusão da matriz variância covariância é fundamental para levar em consideração a dispersão dos dados em torno das médias, que pode ser paralela aos eixos, ou apresentar correlação entre bandas.

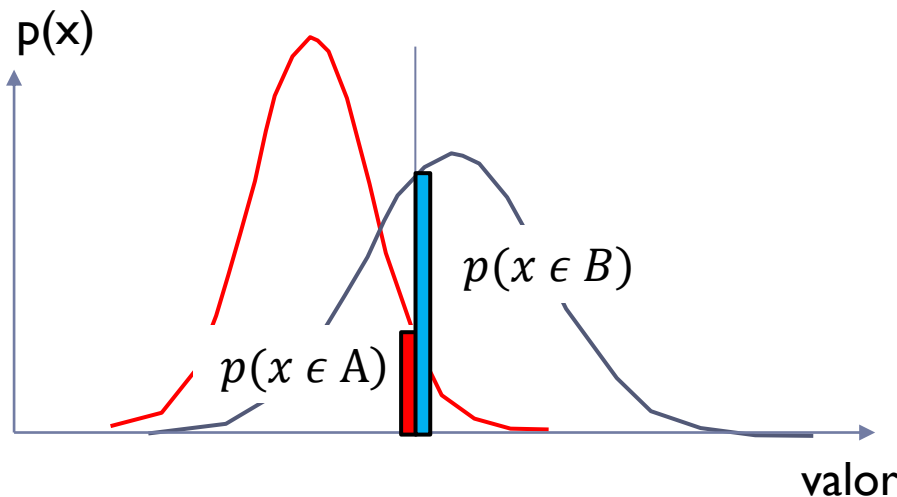


processo

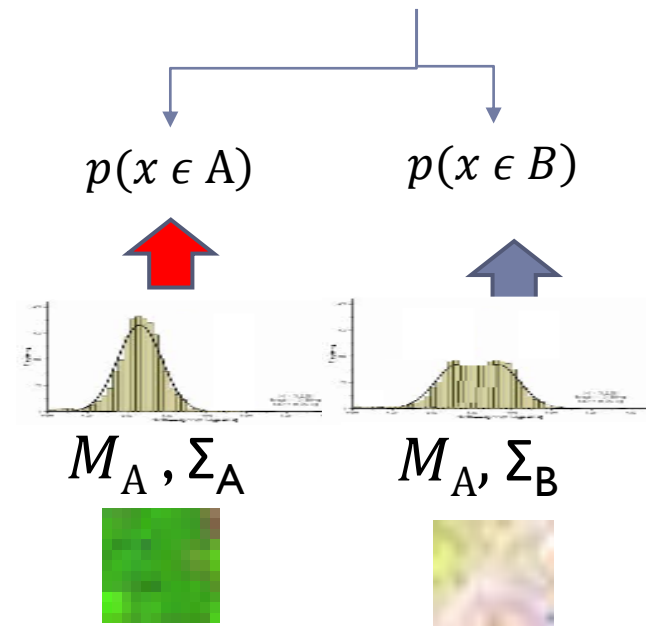
Calcular ao vetor de médias e a matriz variância-covariância das classes M e Σ .

Para cada pixel:

- Calcular a probabilidade que associa este pixel a cada uma das classes
- Comparar as probabilidades e selecionar a classe mais provável.



$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$



Implementação *

$$p(x|w) = \frac{1}{|C|^{1/2} \sqrt{2\pi}} e^{-0.5 (x-m)^T C^{-1} (x-m)}$$

Na realidade, não se calcula toda a expressão, pois existem alguns fatores constantes que podem ser desconsiderados.

O cálculo é mais eficiente usando a representação logarítmica

$$G = \log(p(x|w))$$

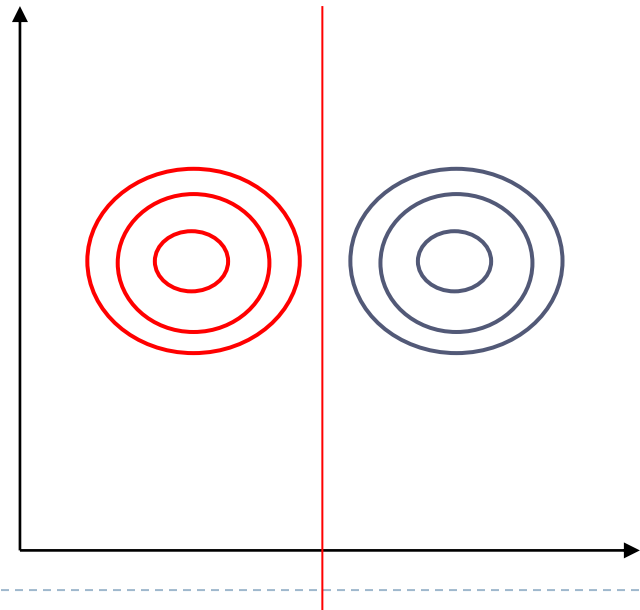
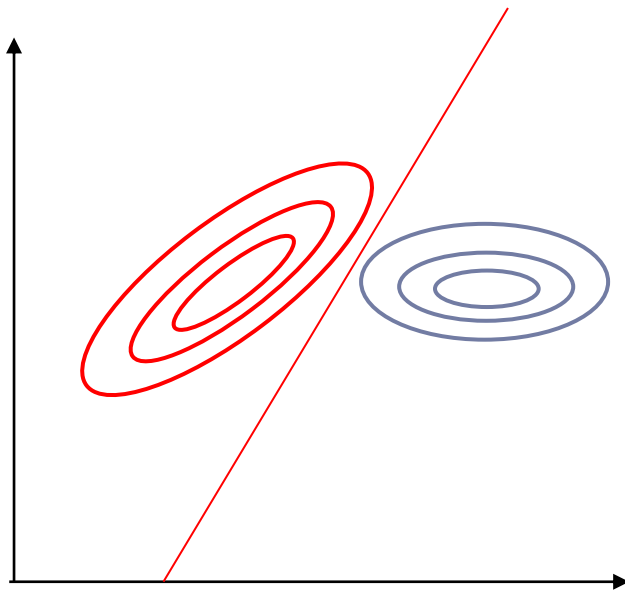
$$G = \log\left(\frac{1}{|C|^{1/2} \sqrt{2\pi}}\right) - 0.5 (x - m)^T C^{-1} (x - m)$$

$$G = -\log(|C|^{1/2}) - \log(\sqrt{2\pi}) - 0.5 (x - m)^T C^{-1} (x - m)$$

$$G = -\log(|C|^{1/2}) - 0.5 (x - m)^T C^{-1} (x - m)$$

MaxVer vs MinDist.Eucl.

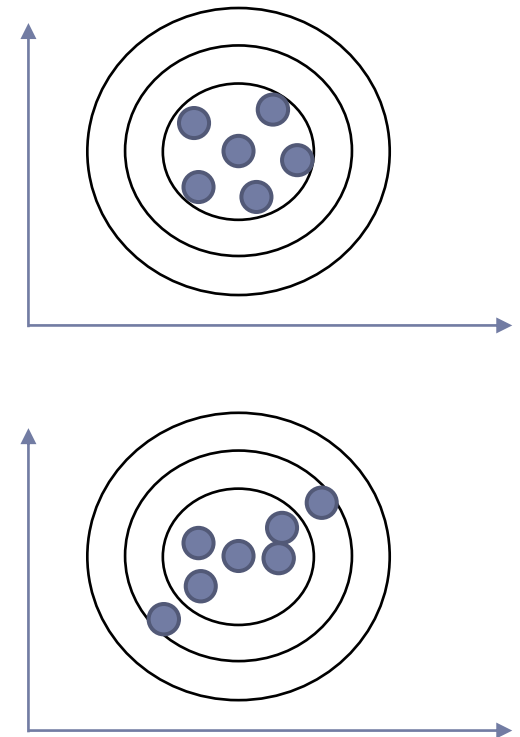
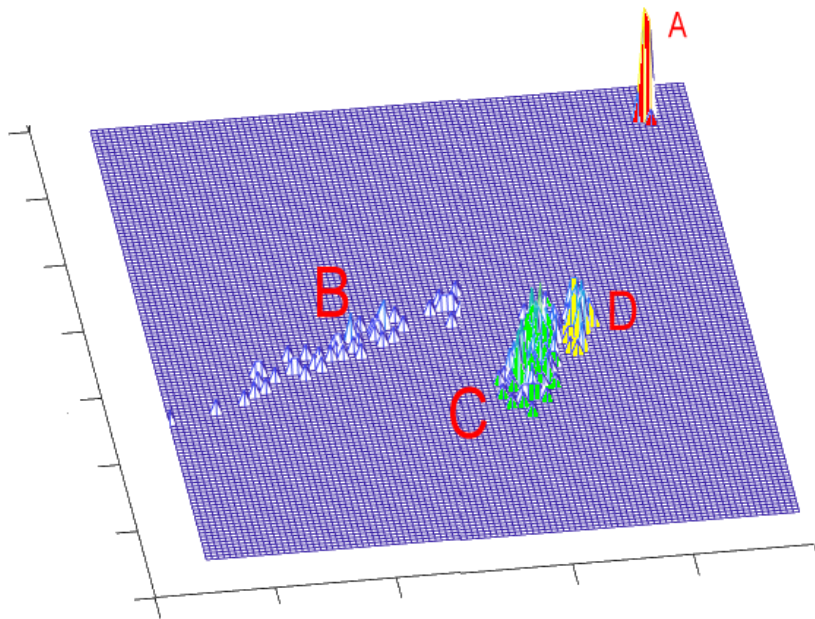
- ▶ Os dois métodos usam o vetor de médias para descrever as classes (posição no espaço dos valores digitais)
- ▶ O método da Máxima Verossimilhança considera a dispersão dos dados em torno da média (variâncias e covariâncias)
- ▶ Com isto, a fronteira de decisão é diferente para os dois métodos.



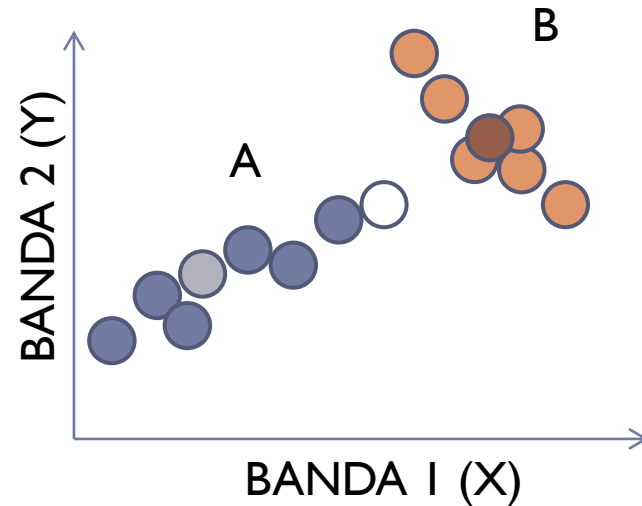
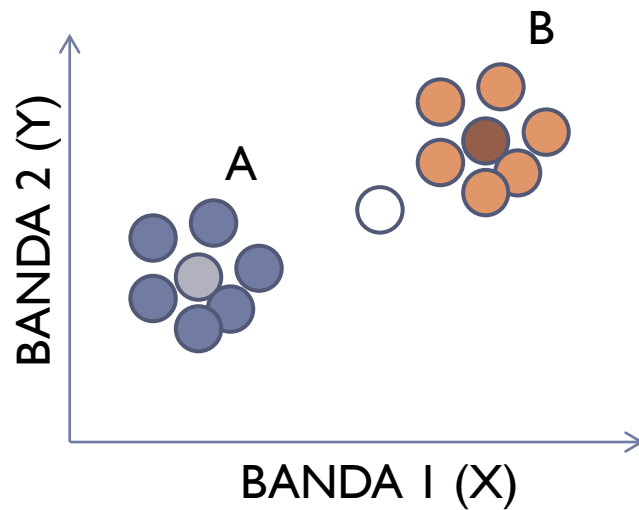
Distância Mínima

▶ Contra

- ▶ Assume que a variação dos valores digitais é igual em todas as direções, o que nem sempre é verdade



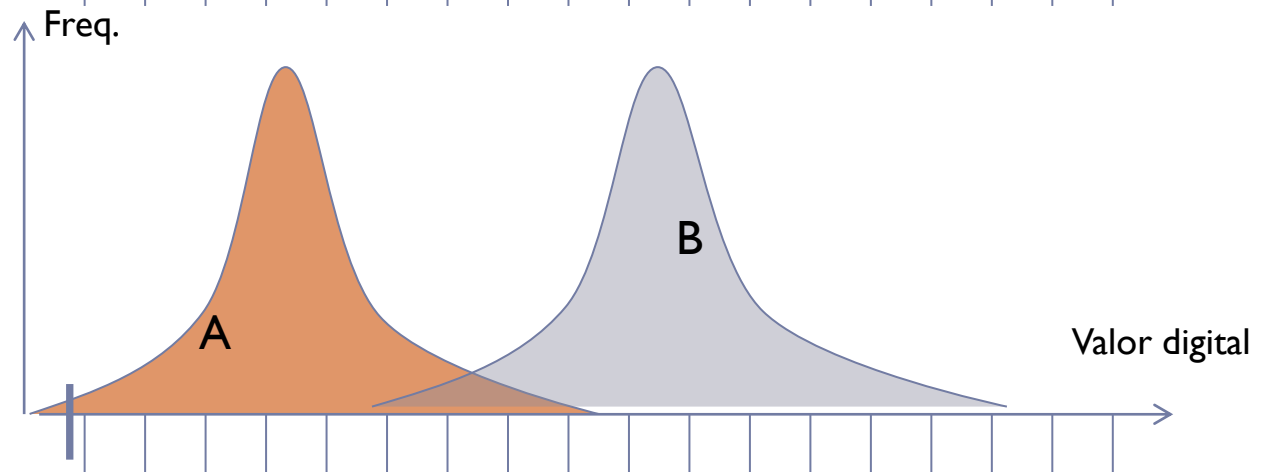
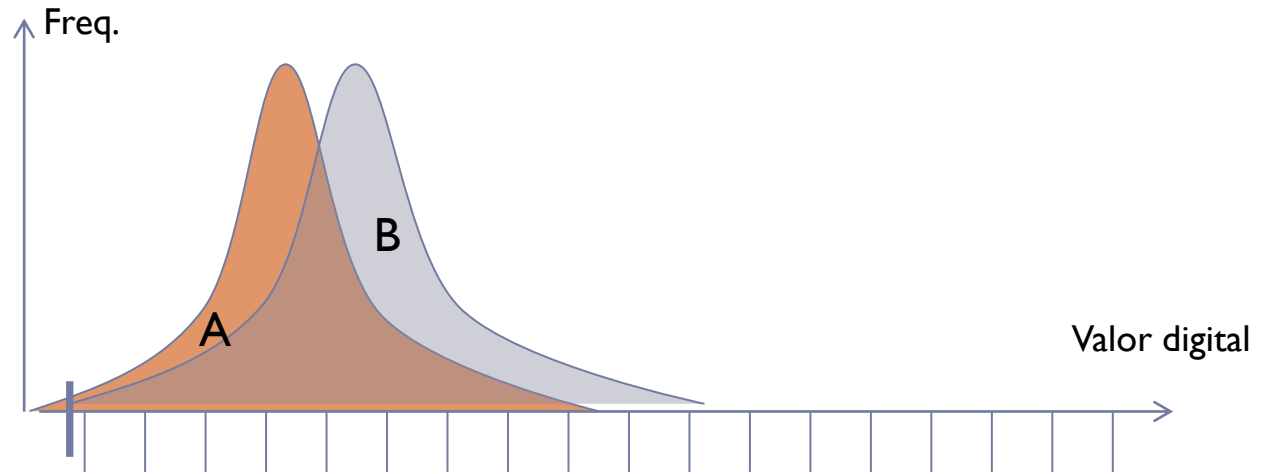
Considere os 2 casos



Qual é a decisão mais correta para classificar o pixel (Ponto Branco) ?

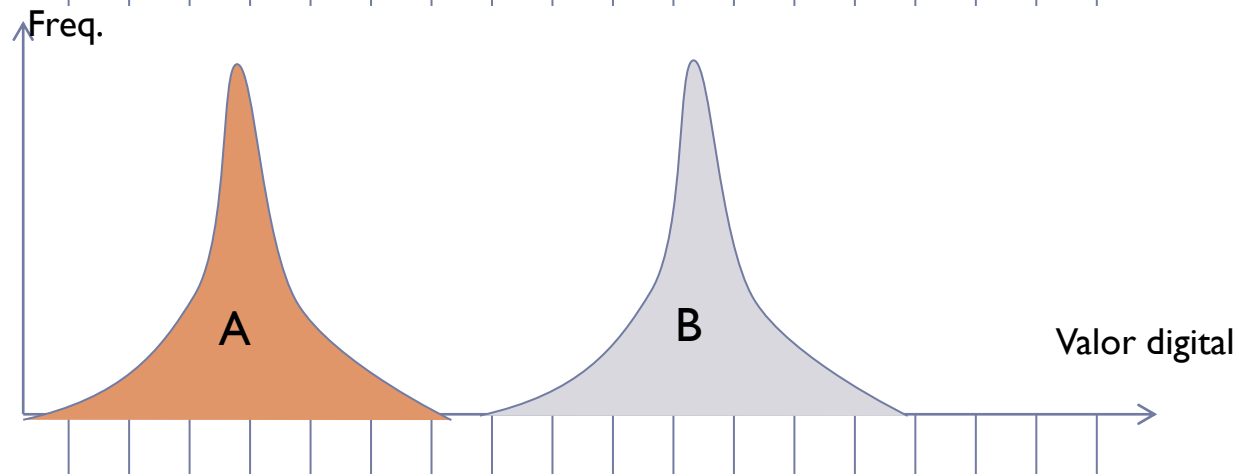
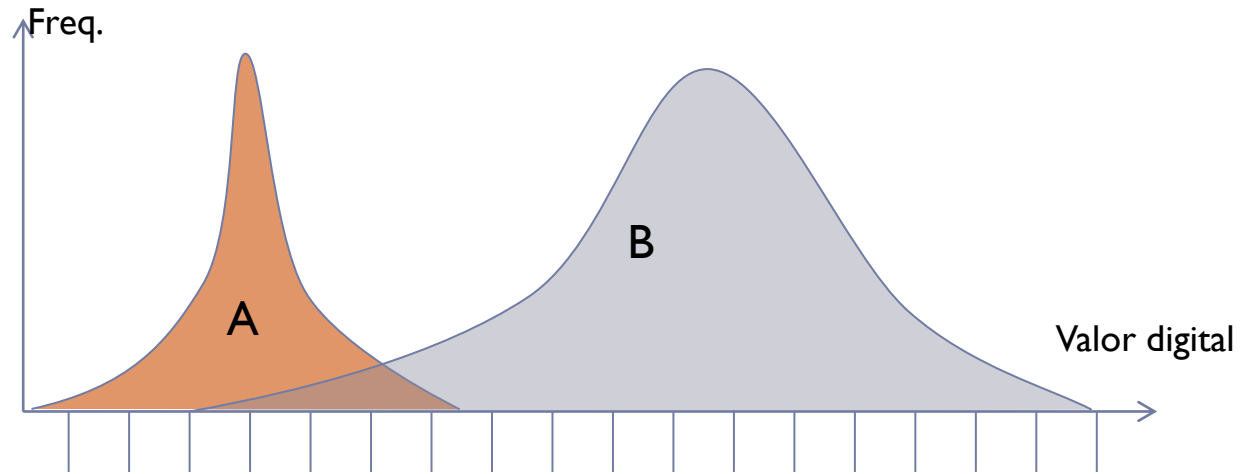
As Classes são separáveis?

- ▶ A médias são próximas?



As Classes são separáveis?

► E as variâncias?



Divergência

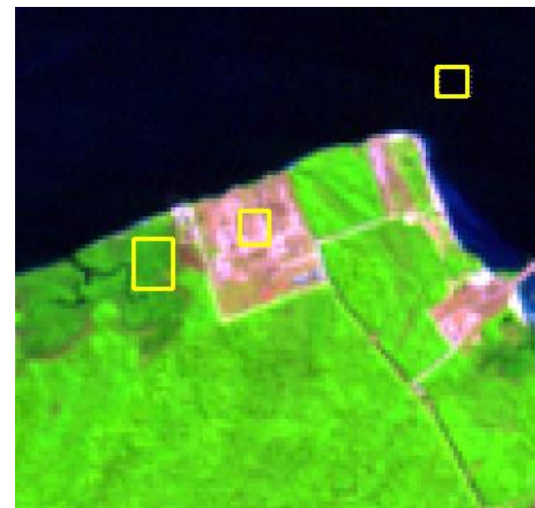
- ▶ Utiliza os vetores de médias (m), bem como as matrizes variância-covariância (V) de cada classe, estimadas a partir das amostras. A divergência pode ser entendida como a soma de duas medidas.
 - ▶ $Div(A, B) = \frac{1}{2} \{ P1 + P2 \}$
 - ▶ A primeira, $P1$, é a diferença normalizada entre as matrizes variância-covariância (V).
 - ▶ $P1 = Tr\{ (V_A - V_B) (V_B^{-1} - V_A^{-1}) \}$
 - ▶ A segunda parte, $P2$, leva em consideração a diferença entre os vetores de médias.
 - ▶ $P2 = Tr\{ (V_B^{-1} - V_A^{-1}) (m_A - m_B)(m_A - m_B)^t \}$
 - ▶ Quanto maior o valor da divergência, maior a separabilidade entre classes.
-

Divergência Transformada:

- ▶ Para fins de comparação, a divergência pode ser normalizada a uma faixa de variação determinada.
 - ▶ $DT(A, B) = 2000 [1 - \exp(Div(A, B)/8)]$
 - ▶ Alguns autores consideram que uma separabilidade aceitável é atingida quando os valores da divergência se aproximam de 2000 (Chuvieco [1990]).
-

exercício

- ▶ Selecione amostras de diferentes classes em uma imagem e calcule a divergência transformada entre os pares de classes, considerando diferentes combinações de bandas.
- ▶ Qual o menor conjunto de bandas que é suficiente para separar suas classes?



Exercício

- ▶ Com as mesmas amostras do exercício anterior, classifique a imagem usando o método da distância mínima Euclidiana e compare os resultados



Ex:
Água
Mata Atlântica
Áreas urbanas
Areia
Restinga
Solo úmido
Solo seco.... Etc etc