



Máxima Verossimilhança Gaussiana

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

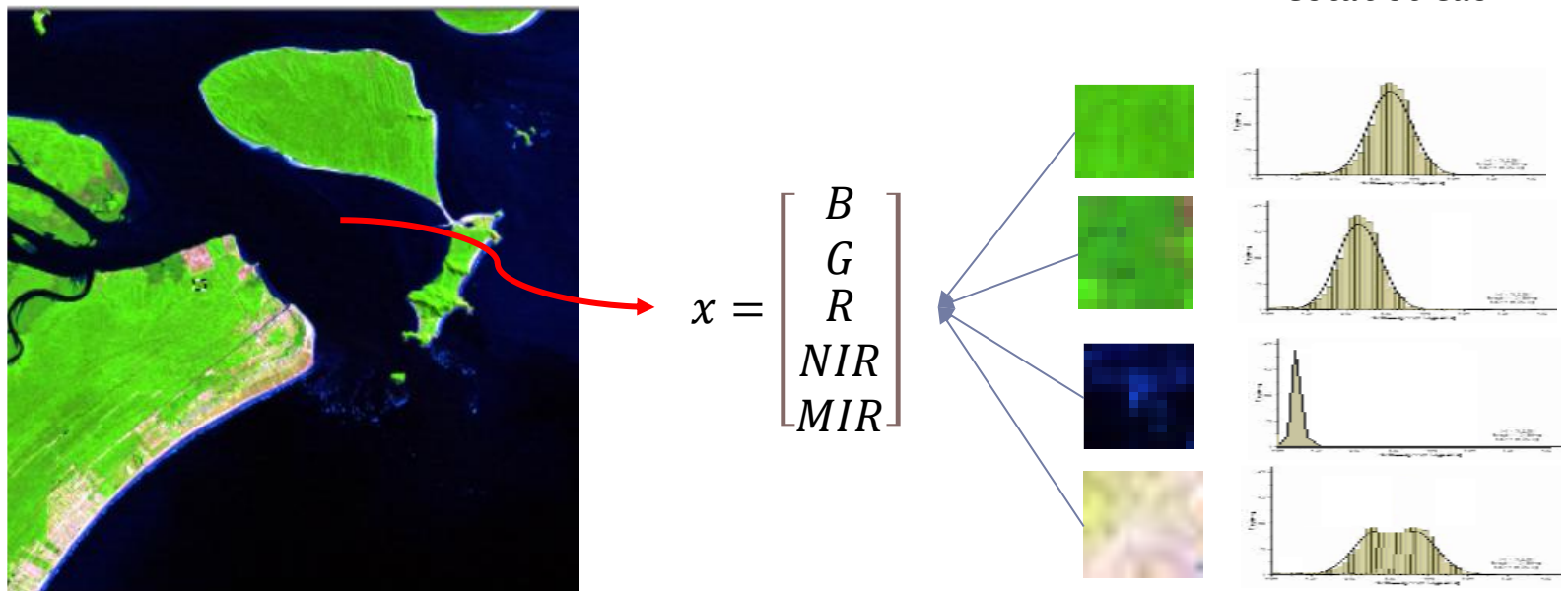
Qual a probabilidade de uma determinada classe ocorrer em cada posição da imagem?

Máxima Verossimilhança

Objetivo: Dado um pixel “x” (ou seja, para um pixel da imagem, do qual conhecemos os valores digitais) definir a classe mais provável w_i , dentre um conjunto de classes possíveis, com base nas estatísticas das amostras:

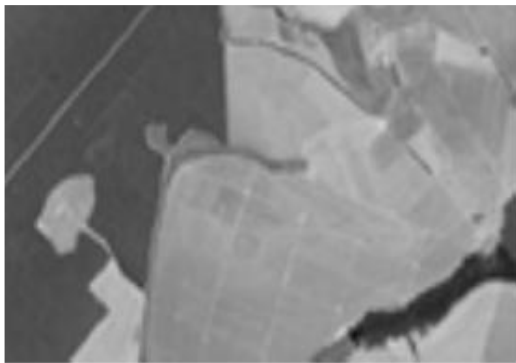
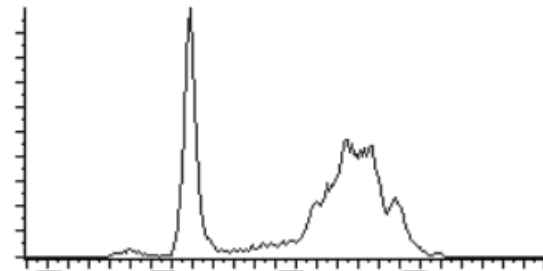
$$[W = w_1, w_2, \dots, w_n].$$

estatísticas

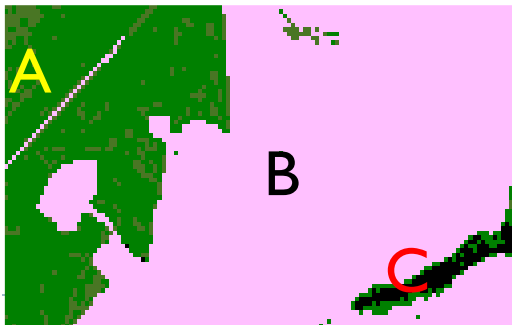
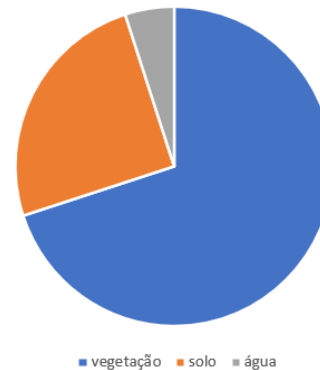


Claro que...

Na imagem digital, todos os valores digitais podem ocorrer, porém com diferentes probabilidades $p(x)$



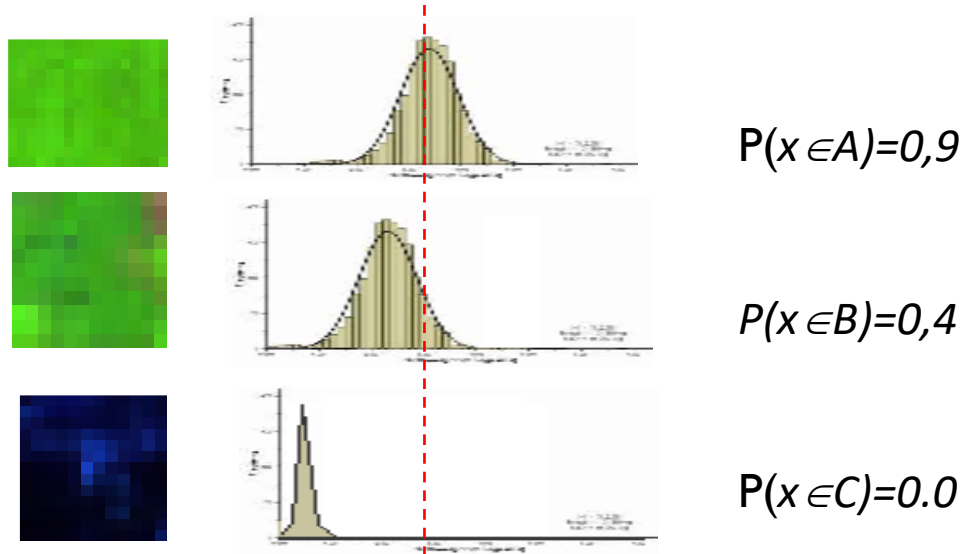
As classes não ocorrem com a mesma frequência, logo existe diferente probabilidade de cada classe ocorrer $p(w)$



Máxima Verossimilhança

Com base nas amostras, estima-se 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

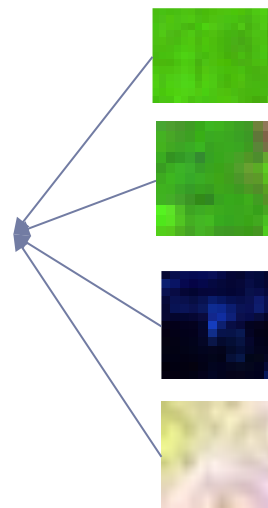
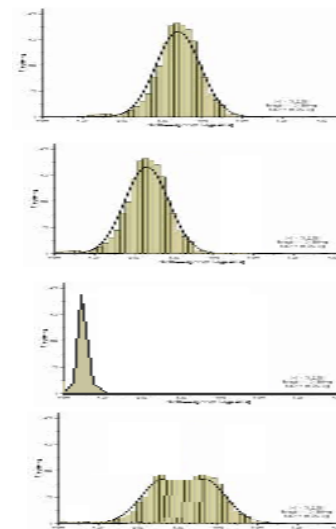
Então, queremos saber a probabilidade de uma determinada classe ocorrer, dado que o valor digital “x” ocorre $P(w/x)$.

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

estatísticas



DADO

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


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

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

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

...

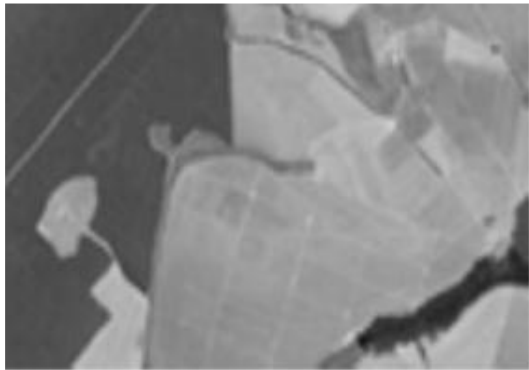
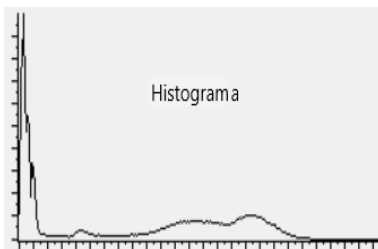
Os dois eventos ocorrem ...

Qual é probabilidade de uma classe " w_k " ocorrer em um pixel com valor x

Para isto, o valor x deve ocorrer e a classe= w_k deve ocorrer

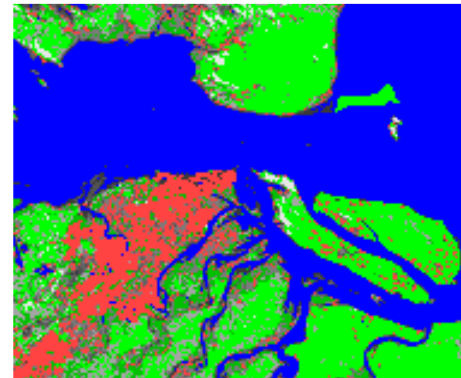
$P(\text{pixel}=x \text{ e classe}=w_k)$

O pixel tem valor " x "



$P(\text{pixel}=x)$

A classe é " w_k "

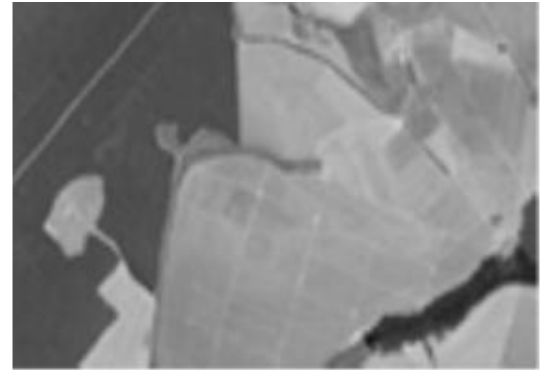


classe= w_k

Se conhecemos o valor x

A probabilidade de um pixel com valor conhecido " x " seja da classe " w_k " pode ser descrita pela probabilidade condicional da classe " w_k " **dado que** o pixel " x " ocorre

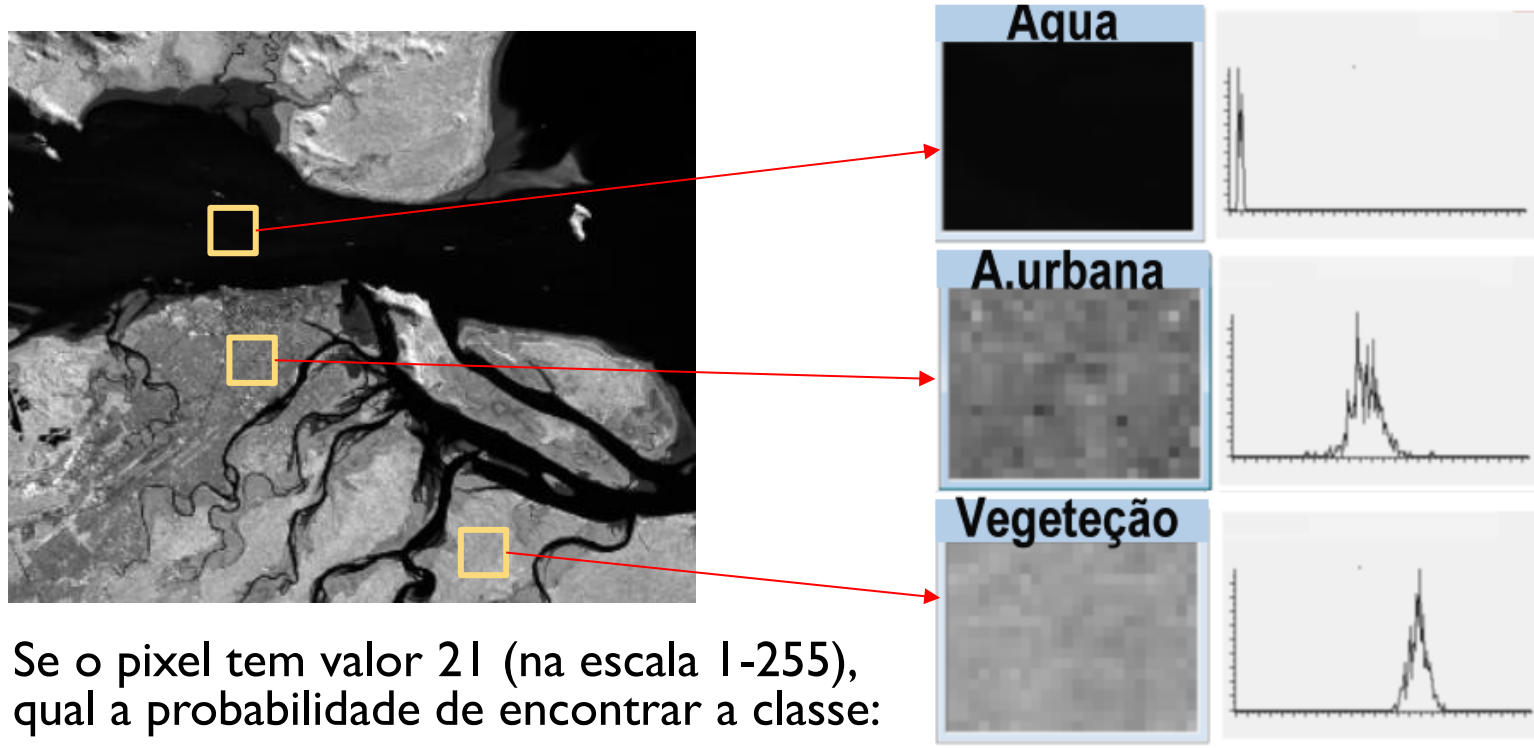
$$P(x, w_k) = p(w_k | x) * p(x)$$



- ▶ Se o pixel tem valor= x , qual a probabilidade dele ser da classe " w_k "?
-

Se conhecemos “x”

Se pixel tem valor x, qual a probabilidade da classe “k” ocorrer nesse local?



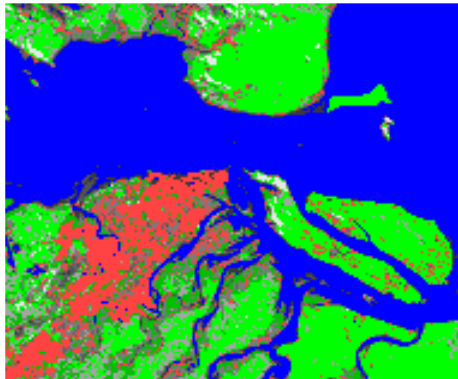
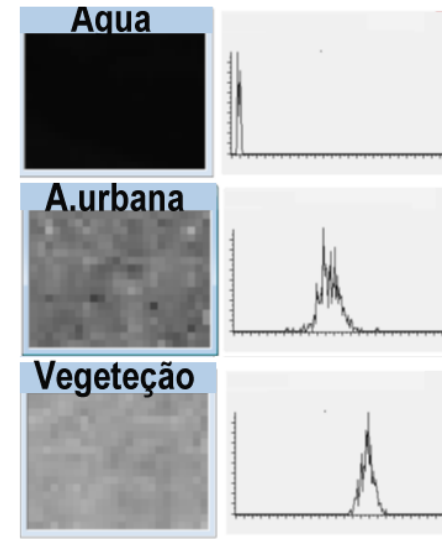
Se o pixel tem valor 21 (na escala 1-255), qual a probabilidade de encontrar a classe:

- ▶ Vegetação $p(w=\text{Veg} \mid x=21)$
- ▶ Área urbana $p(w=\text{Urb} \mid x=21)$
- ▶ Água? $P(w=\text{Agua} \mid x=21)$

Se conhecemos a classe...

Se em uma posição da imagem é encontrada a classe “w”, qual é a probabilidade do pixel ter o valor x?

$$p(x | w)$$



Se for encontrada a classe água, qual a probabilidade do pixel ter valor 200?

$$p(x=200 | w=\text{Água})$$

Se soubermos qual é a classe...

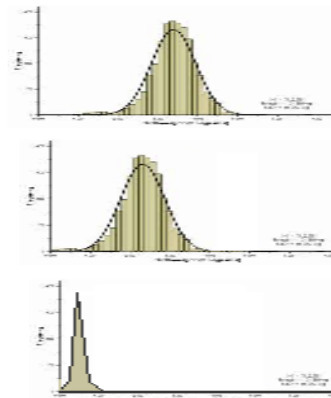
Se sabemos que o pixel é da classe “ w_k ”, e como temos as estatísticas das amostras, podemos calcular a mesma probabilidade condicional do evento:

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?



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

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

$P(x, \text{dado ÁGUA})$



Queremos prob(pixel=x ao mesmo tempo que classe= w_k)

► Igualando:

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

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

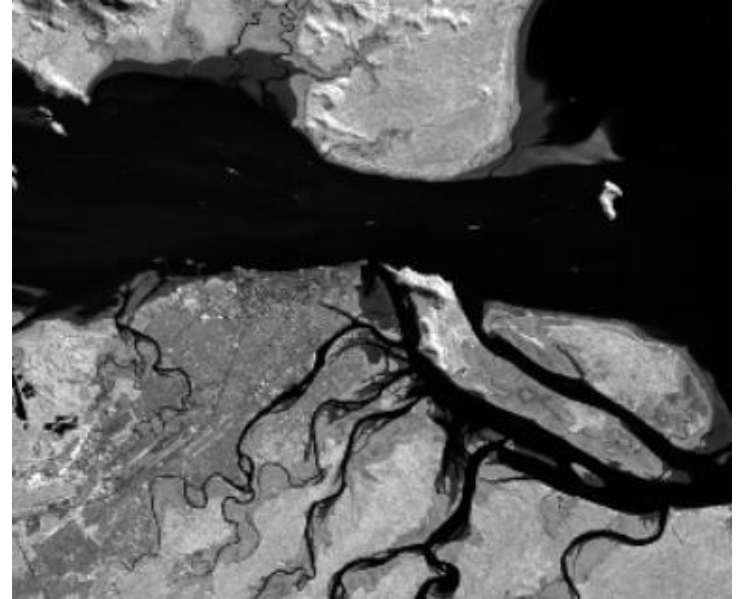
$$p(w_k | x) = \frac{p(x | w_k) * p(w_k)}{p(x)}$$

Classificação por probabilidade

Se, dado o valor do pixel, podemos calcular a probabilidade de cada classe com a regra

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

Podemos calcular a probabilidade associada a todas as classes para cada pixel



E seria razoável classificar o pixel na classe mais provável.

Pixel x pertence à classe W_A se:

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

Probabilidade da classe

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.

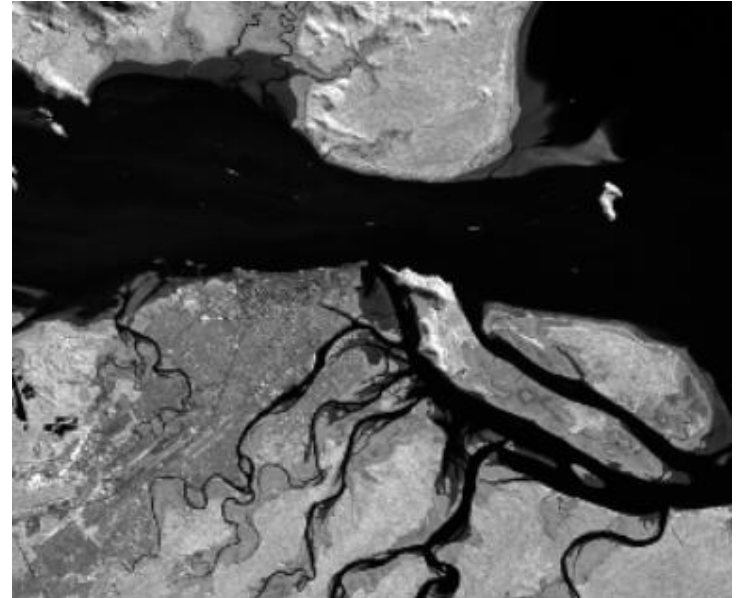
► Explicados a seguir...

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

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

$p(x)$: 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

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

Regra modificada

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:

$$G(w_A|x) > G(w_B |x)$$

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

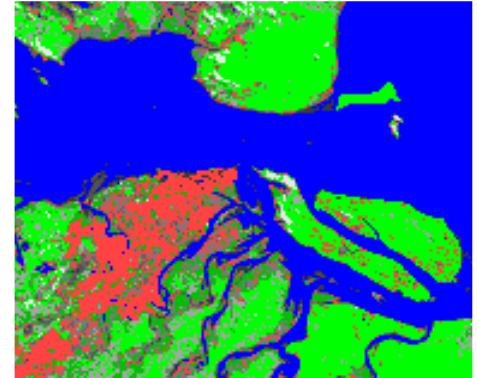
$P(w)$: Probabilidade “a priori”

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

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

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

O que é mais provável, vegetação ou água?

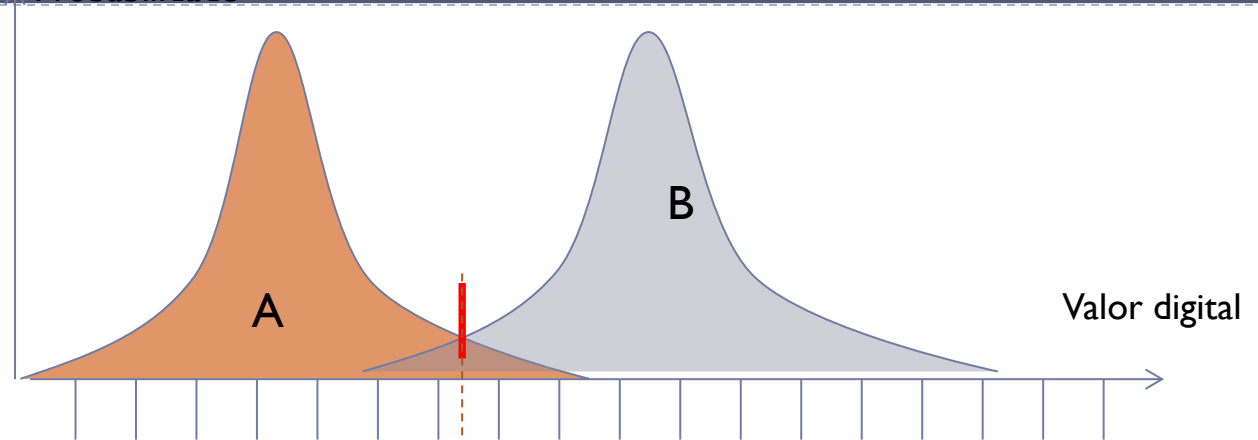


Para saber isto, deveríamos conhecer quantos pixels de cada classe tem na imagem, mas a verdadeira classificação dos pixels é desconhecida (é o que queremos).

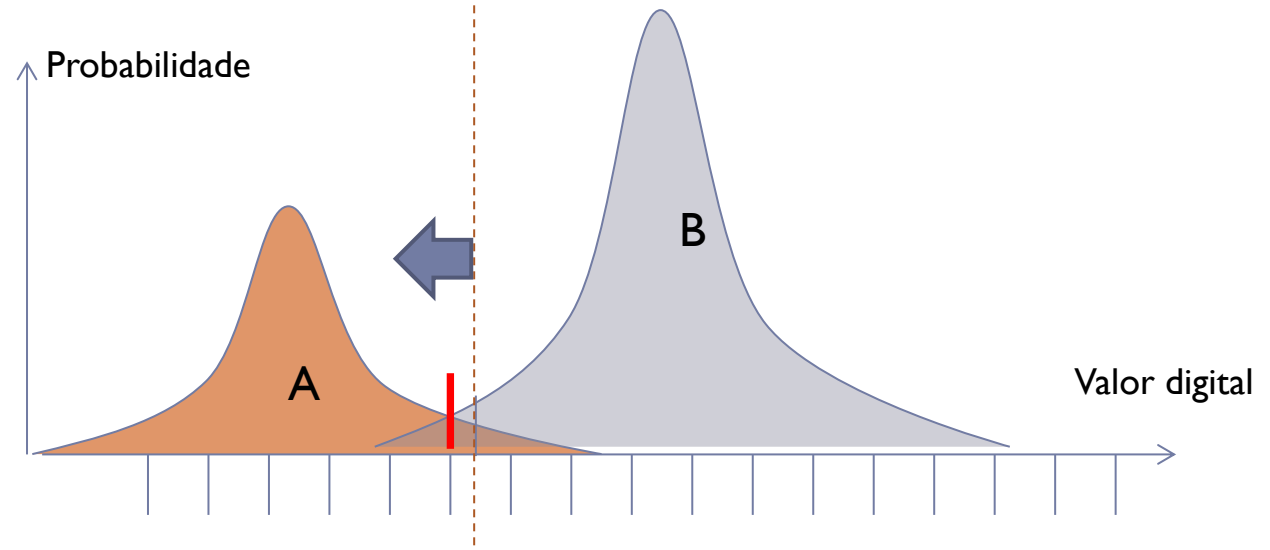
Para não favorecer nenhuma classe injustamente, o valor de $p(x)$ é igual para todas as classes ($p(w) = 1$).

Efeito da Probabilidade a priori

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



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



Resumindo e simplificando...

Se a “**probabilidade a priori**” é considerada constante, e é assumido que $p(x)$ pode ser descartado (constante) , para fins de comparação ...

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

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

A regra de classificação se simplifica para, Atribuir o pixel “ x ” à classe w_j , se:

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

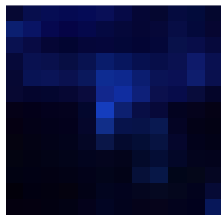
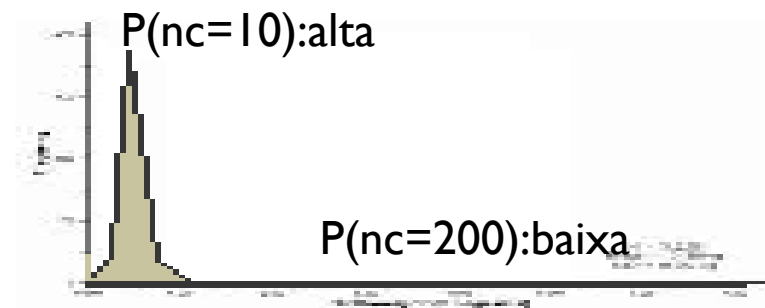
Assim, apenas a distribuição dos pixels dentro da classe é analisada (função densidade de probabilidade da classe). Neste caso, não se calcula o verdadeiro valor da probabilidade, mas sim um valor que permite discriminar as classes, por isso se fala em função discriminante.

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

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

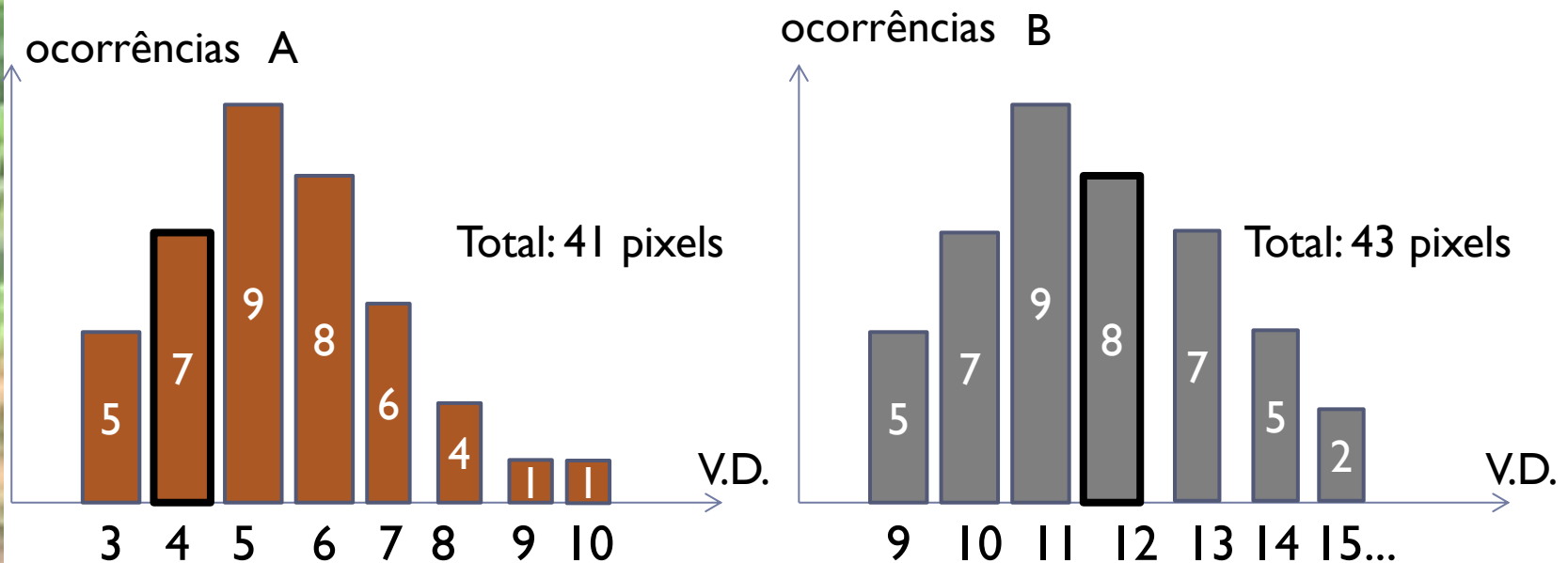
Que valores são mais frequentes em uma classe?
Quais são pouco frequentes?

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”

Aproximação Normal

Na natureza, a distribuição de muitos fenómenos podem ser adequadamente modelada com a distribuição Normal (Gaussiana).

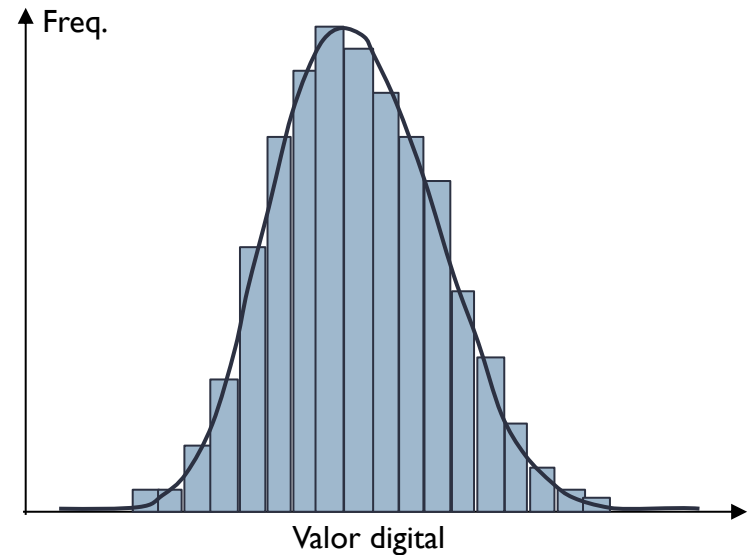
A variação de muitas classes em uma imagem de satélite seguem uma distribuição Normal, com média μ e variância σ .

Para uma banda vale:

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

Como a média e a variância populacional (μ, σ) são aproximadas pela média e variância da amostra (m, S):

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

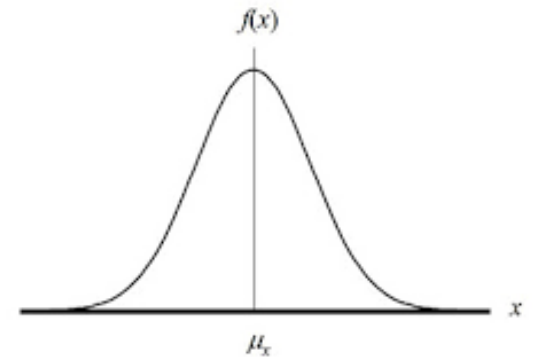


$P(x/w)$: função densidade de probabilidade

A função densidade de probabilidade descreve a variação dos valores dentro de uma classe.

Para usar a aproximação da distribuição normal, a média e a variância são calculadas a partir das amostras.

$$p(x|w) = \frac{1}{s\sqrt{2\pi}} e^{-0.5 \left(\frac{(x-m)}{s} \right)^2}$$

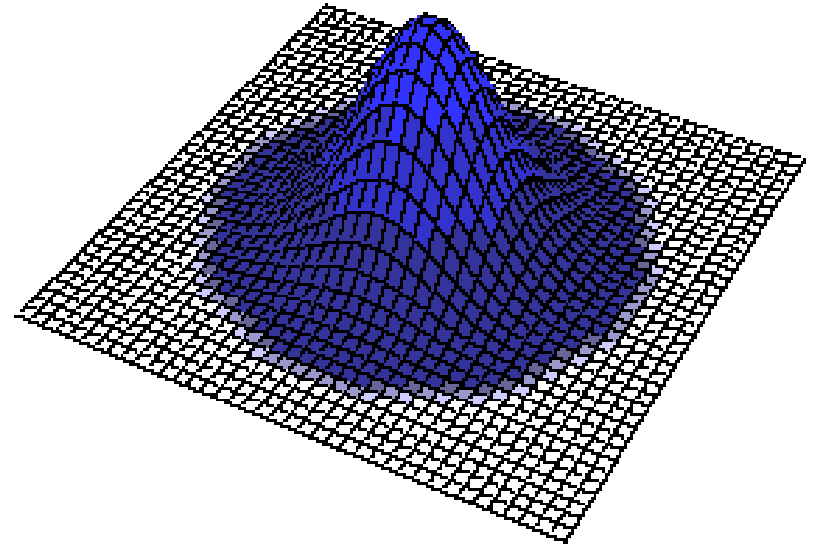


.

Caso multidimensional

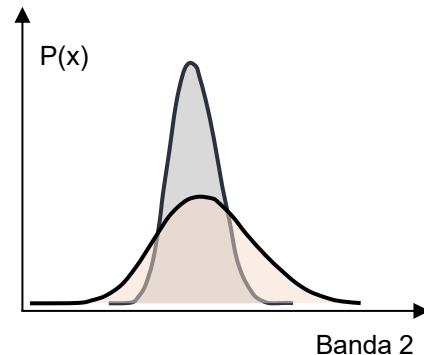
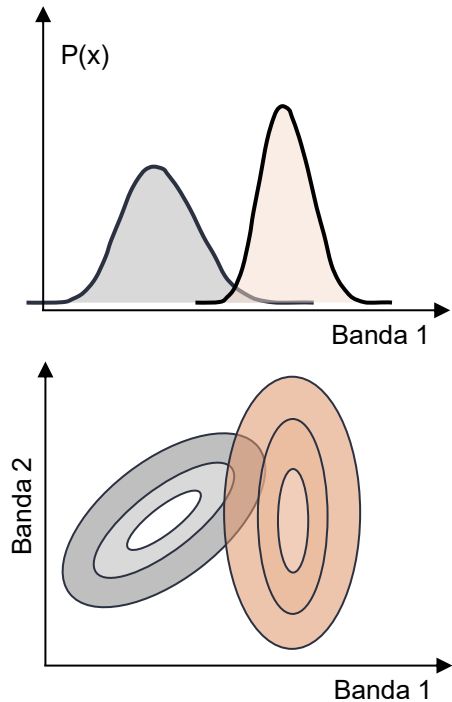
- ▶ No caso multidimensional ($\vec{x}$: valor em várias bandas espectrais em sensoriamento remoto), a expressão é escrita em função do vetor de médias ($\vec{m}$) e a variância pela matriz variância-covariância (C), para considerar a dispersão dos dados em torno da média em todas as bandas. Assim:

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



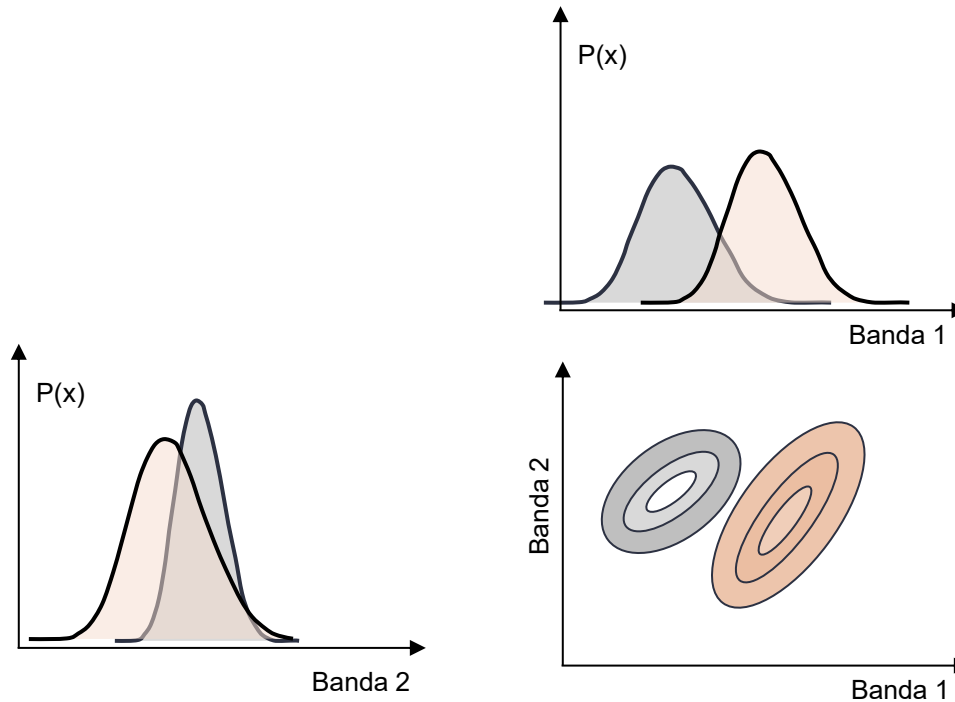
Diferença

A inclusão da matriz variância covariância é fundamental para levar em consideração a dispersão dos dados em torno das médias, visto que bandas espectral mente próximas apresentam alta correlação.



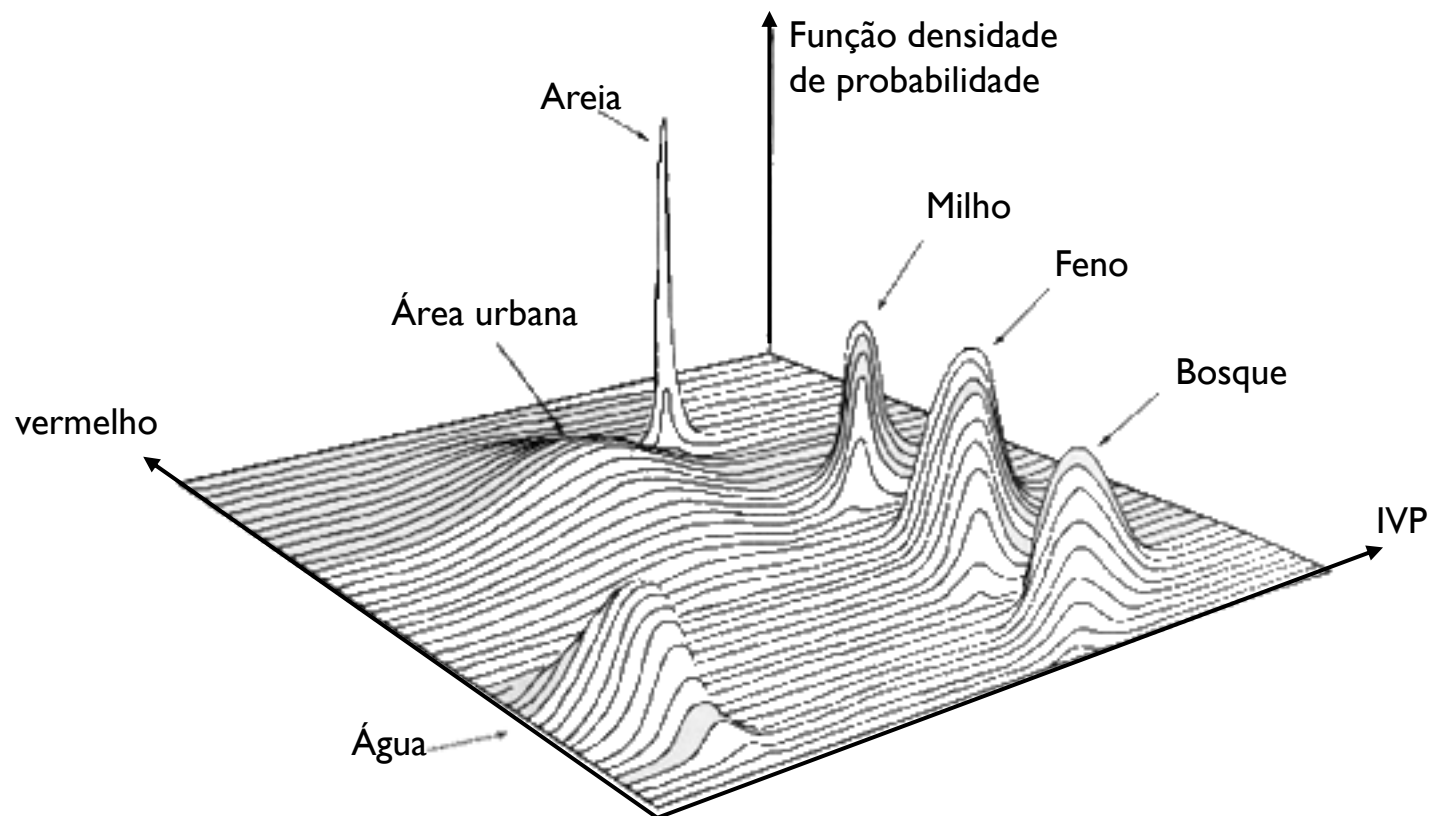
Diferença

A inclusão da matriz variância covariância facilita a separação de classes, pois leva em consideração a rotação dos dados no espaço espectral.



Visualização 2D

- ▶ Considerando duas bandas: A distribuição das classes pode ser diferente em termos de médias e dispersão.
- ▶ Quando a dispersão é muito grande, pode ocorrer superposição de classes.

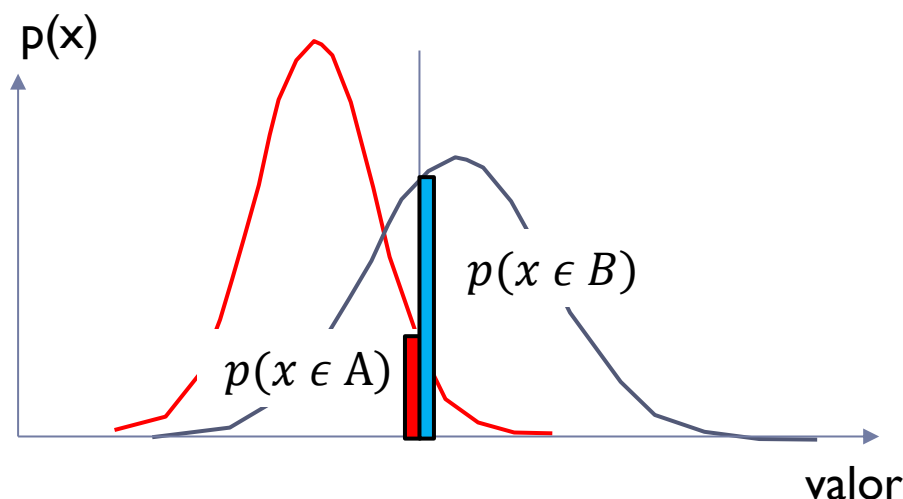


Processo

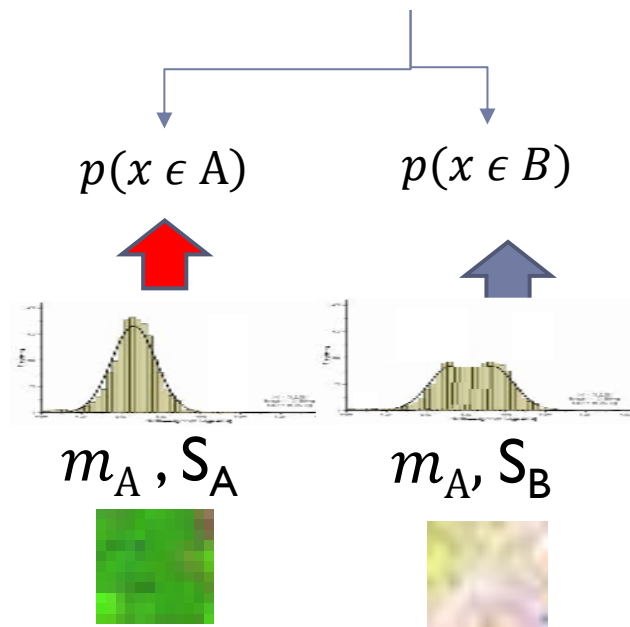
Calcular ao vetor de médias e a matriz variância-covariância das classes.

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 *

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

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

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

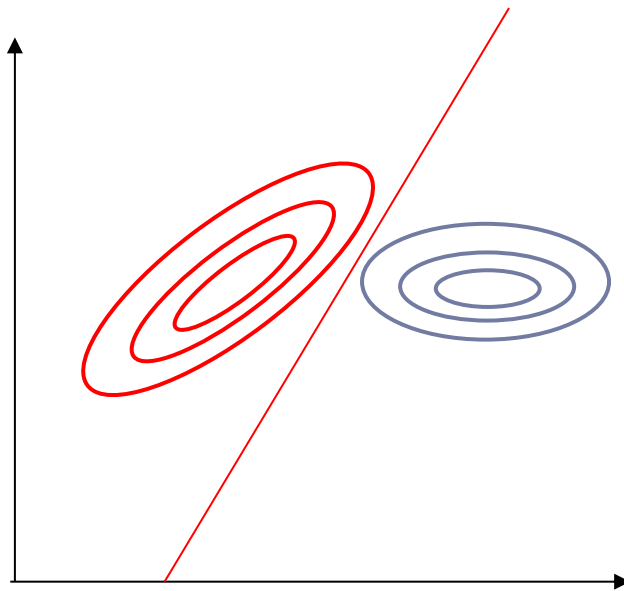
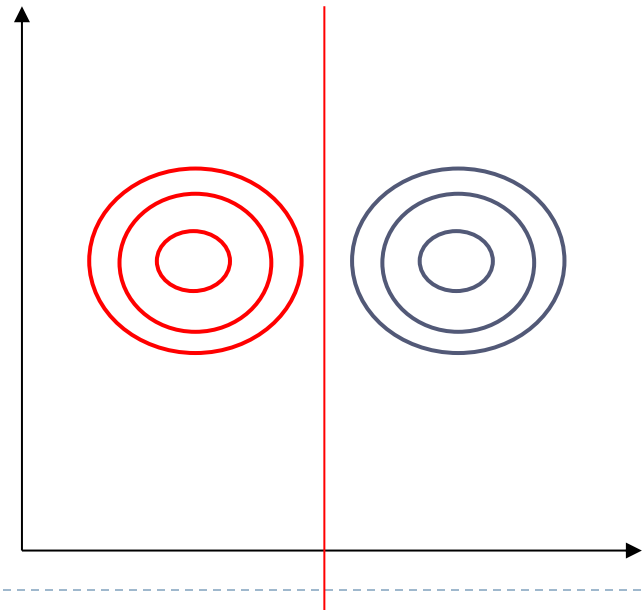
$$G = \log\left(\frac{1}{|S|^{1/2} \sqrt{2\pi}}\right) - 0.5 (x-m)^T S^{-1} (x-m)$$

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

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

MaxVer vs MinDist.Eucl.

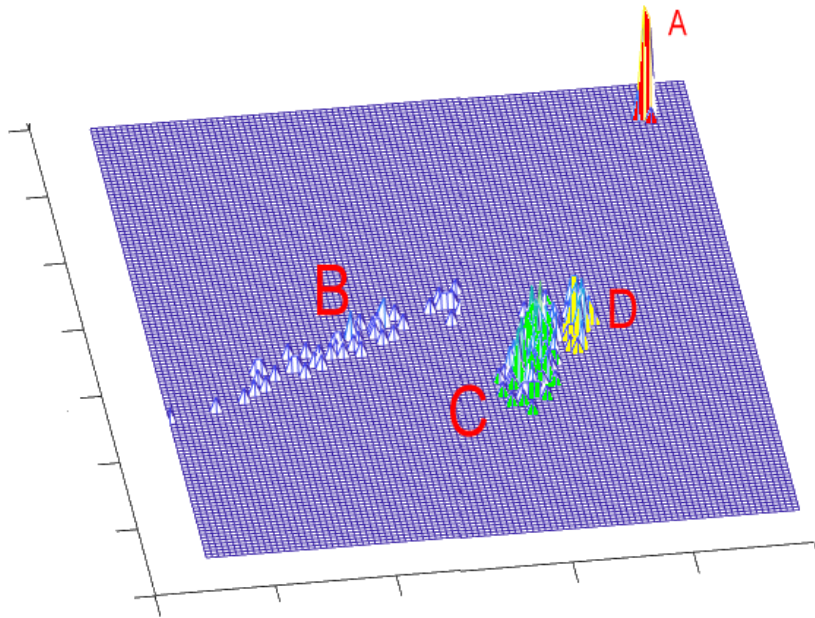
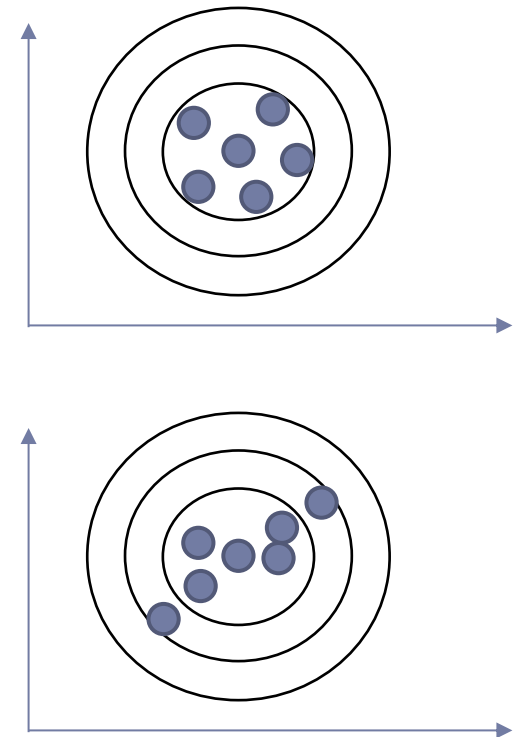
- ▶ Os dois métodos usam o vetor de médias para descrever as classes (posição no espaço dos valores digitais)
- ▶ O método da Máxima Verossimilhança considera a dispersão dos dados em torno da média (variâncias e covariâncias)
- ▶ Com isto, a fronteira de decisão é diferente para os dois métodos.



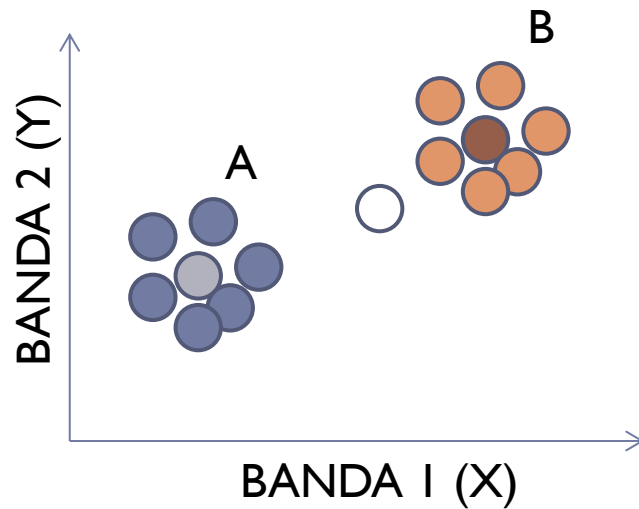
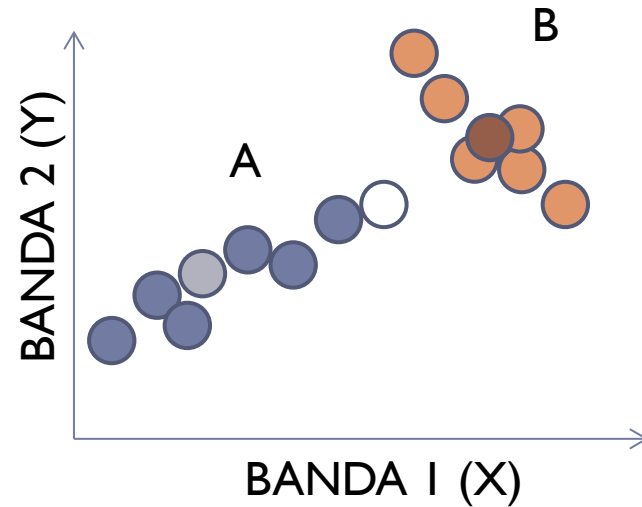
Distância Mínima

► 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) ?

Simplificando mais,,,

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

A função discriminante pode ser ainda mais simplificada.

O que ocorre se todas na matriz variância-covariância todas as variâncias são iguais e as covariâncias todas nulas, PARA TODAS AS CLASSES?

$$S = \begin{bmatrix} v^2 & 0 & 0 \\ 0 & v^2 & 0 \\ 0 & 0 & v^2 \end{bmatrix} = v^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$-\log(|S|^{1/2})$: o determinante é igual a $(v^2)^n$, igual para todas as classes, logo este termo é constante e pode ser desconsiderado.

$$G = -0.5 (x - m)^T S^{-1}(x - m)$$

$G = (x - m)^T (x - m) \dots$ Que é a distância Euclidiana.

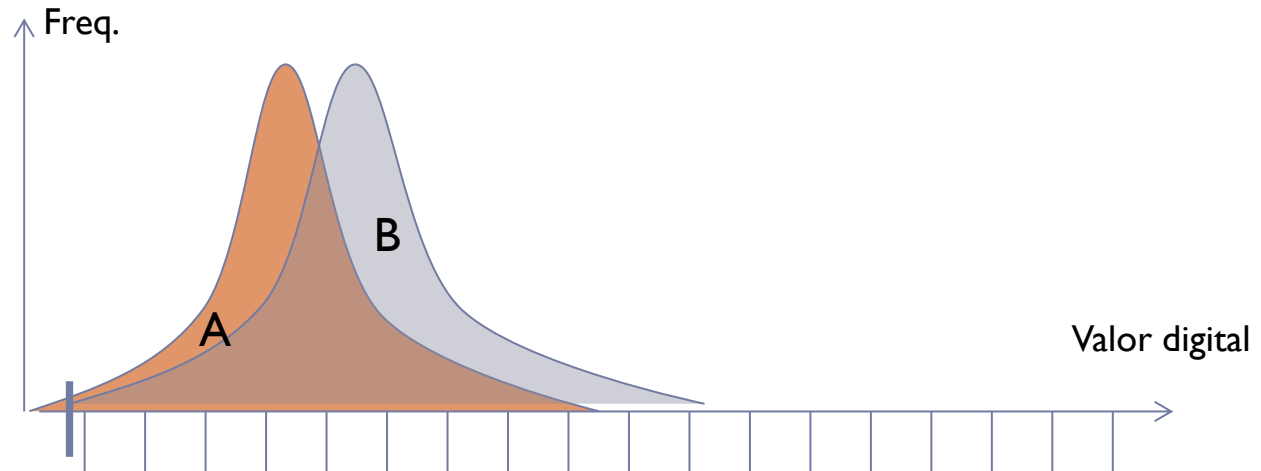
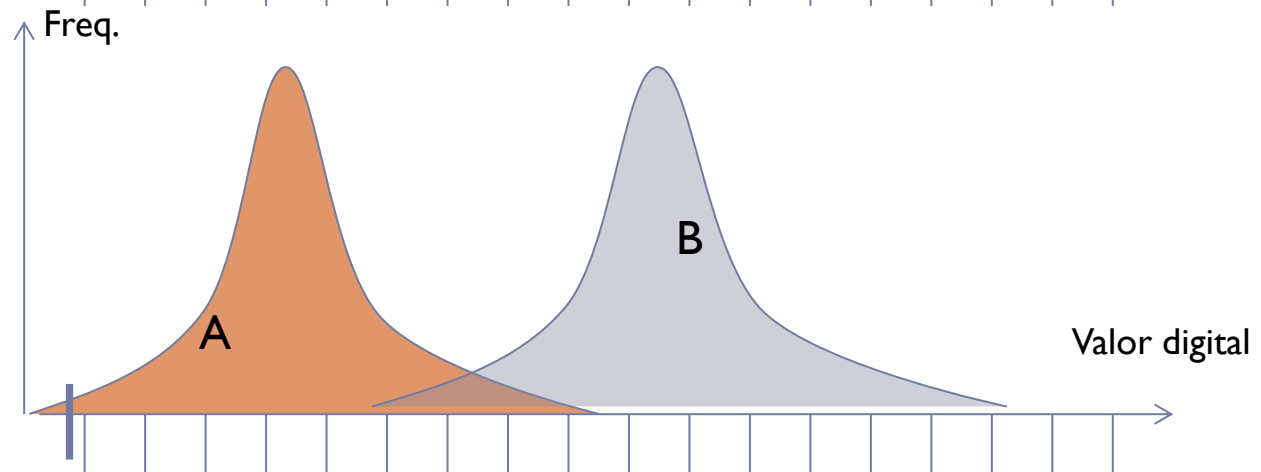


Separabilidade

- ▶ A separabilidade, ou distância, entre duas classes com distribuição NORMAL pode ser estimada a partir dos vetores de médias e a matriz Variância-Covariância.
-

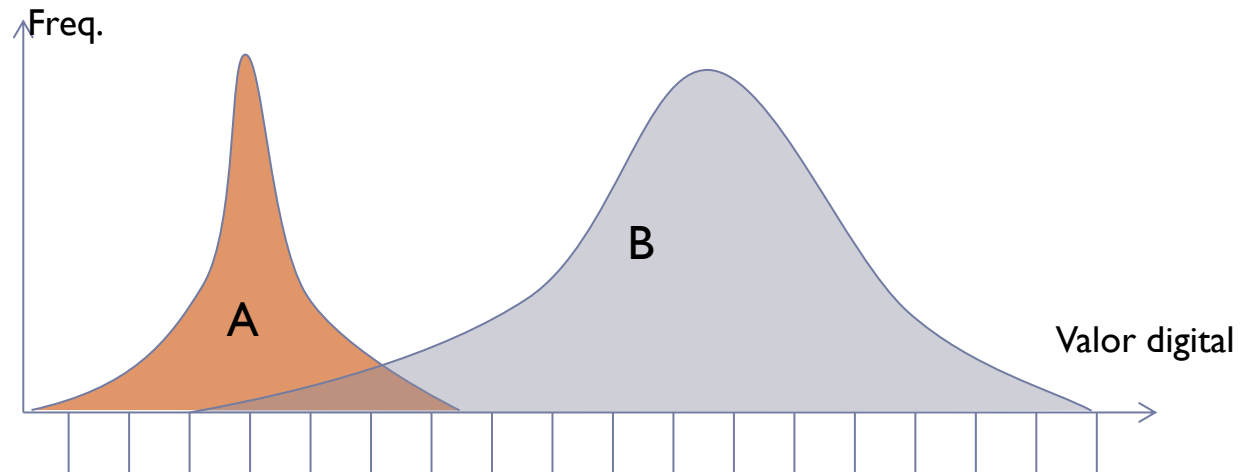
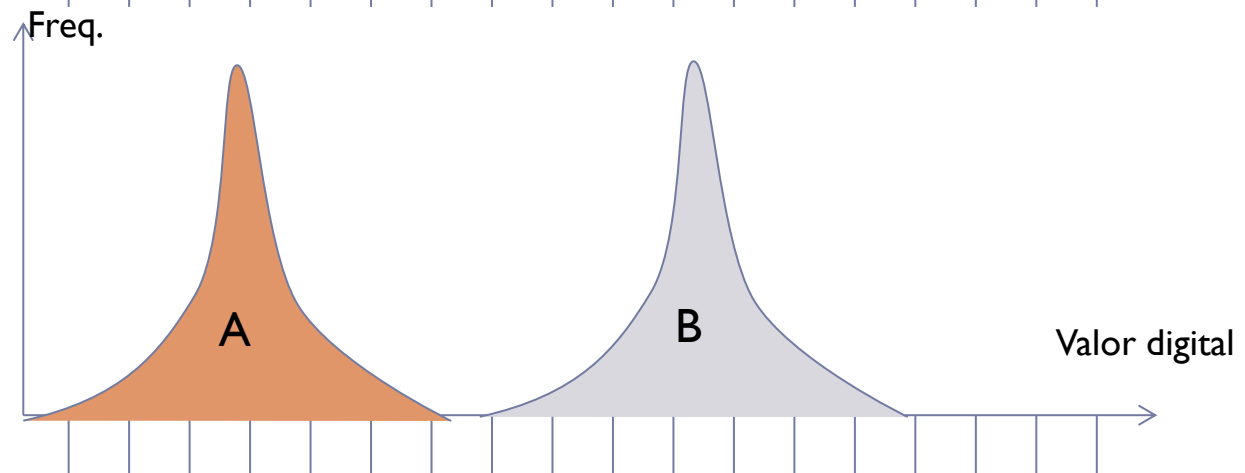
As Classes são separáveis?

- A médias são próximas?



As Classes são separáveis?

► E as variâncias?



Divergência

- ▶ Utilizando os vetores de médias (m) e as matrizes variância-covariância (V) de duas classes, a divergência entre as classes pode ser entendida como a soma de dois fatores.

- ▶ $Div(A, B) = \frac{1}{2} \{ P1 + P2 \}$

$P1$, 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:

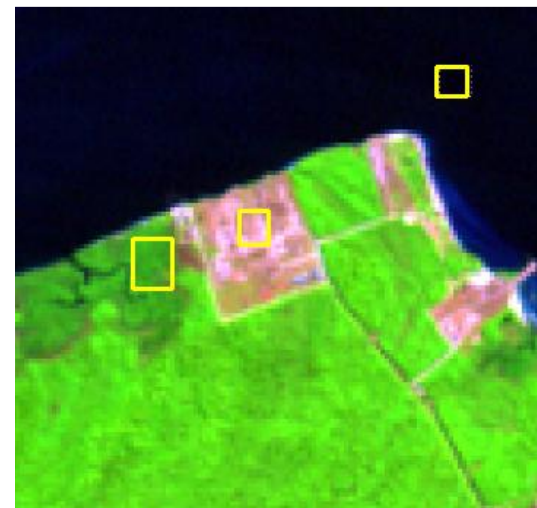
- ▶ Para fins de comparação, a divergência pode ser normalizada a uma faixa de variação determinada.

$$DT(A, B) = 2000 \left\{ 1 - e^{\frac{Div(A, B)}{8}} \right\}$$

- ▶ 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