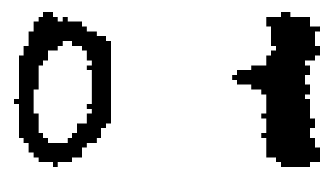
Teoria de Decisão de Bayes

Modelo Probabilístico das Observações

- Probabilidades a priori de cada classe: $p(\varpi_i)$ com $i = 1, \ldots, c$
- Traduz o conhecimento existente sobre a classe ϖ quando não se possui nenhuma informação adicional.
  - Exemplo de classificação de dois dígitos manuscritos:

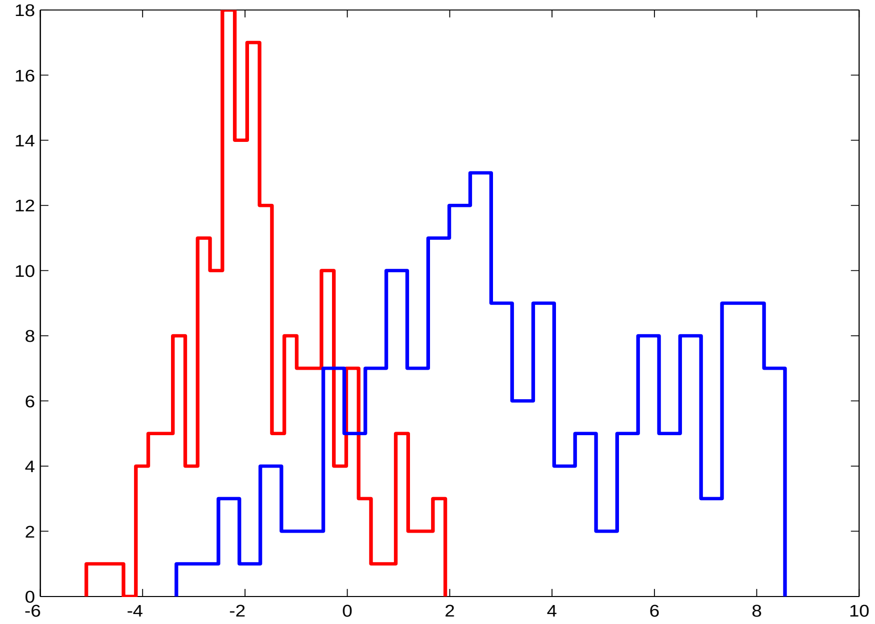


Regra de decisão:

$$\hat{\varpi} = \varpi_0 \quad \text{se} \;\; p(\varpi_0) > p(\varpi_1)$$

Modelo Probabilístico das Observações

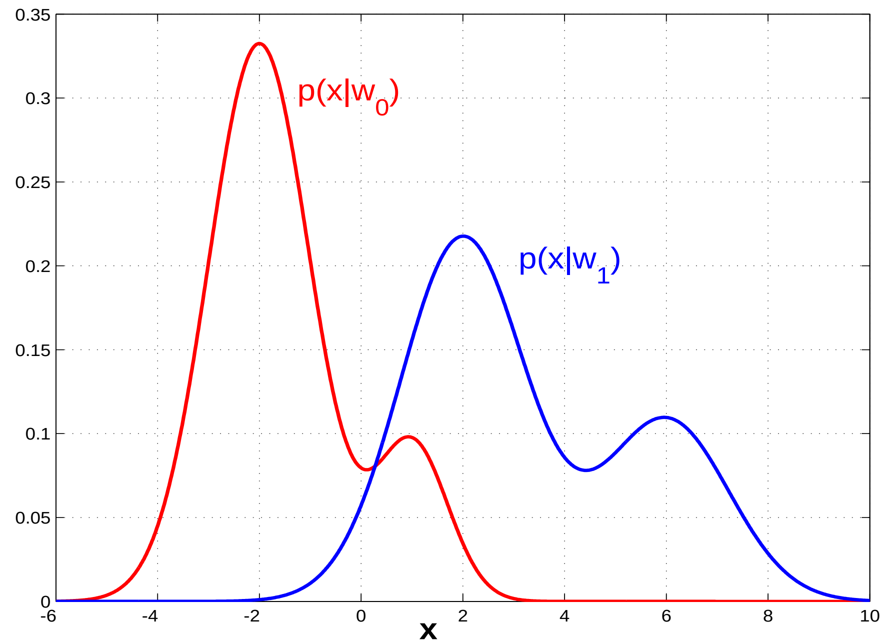
- Probabilidades a priori de cada classe: $p(\varpi_i)$ com $i = 1, \ldots, c$
- Probabilidades condicionadas $p(\mathbf{x}|\varpi_i)$ com $i = 1, \ldots, c$
  - Exemplo de classificação de dois dígitos manuscritos:



Histograma (aproximação das probabilidades condicionadas)

Modelo Probabilístico das Observações

- Probabilidades a priori de cada classe: $p(\varpi_i)$ com $i = 1, \ldots, c$
- Probabilidades condicionadas $p(\mathbf{x}|\varpi_i)$ com $i = 1, \ldots, c$
  - Exemplo de classificação de dois dígitos manuscritos:



Funções de Densidade de Probabilidade Condicionadas $p(\mathbf{x}|\varpi_i)$

Modelo Probabilístico das Observações

- Probabilidades a priori de cada classe: $p(\varpi_i)$ com $i = 1, \ldots, c$
- Probabilidades condicionadas $p(\mathbf{x}|\varpi_i)$ com $i = 1, \ldots, c$
  - Exemplo de classificação de dois dígitos manuscritos:

Regra de decisão:

$$\hat{\varpi} = \varpi_0 \quad \text{se} \;\; p(\mathbf{x}|\varpi_0) > p(\mathbf{x}|\varpi_1)$$

(Classificador de Máxima Verosimilhança)

Modelo Probabilístico das Observações

- Probabilidades a priori de cada classe: $p(\varpi_i)$ com $i = 1, \ldots, c$
- Probabilidades condicionadas $p(\mathbf{x}|\varpi_i)$ com $i = 1, \ldots, c$
- Como combinar a informação contida nas densidades a priori e nas densidades condicionadas de modo a minimizar a probabilidade de erro?

Solução: Classificador MAP

Classificador de Máximo a Posteriori

- Pressuposto: Conhecimento completo de todas as densidades
- Critério adoptado: Minimizar a probabilidade de erro
- Regra de decisão: calcular a probabilidade de ocorrência de cada classe *dada a observação* $\mathbf{x}$, e escolher a classe mais provável:

$$\hat{\varpi} = \varpi_i \;\; \text{se} \;\; p(\varpi_i|\mathbf{x}) > p(\varpi_k|\mathbf{x}),\ \forall\, k \neq i$$

- Lei de Bayes:

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

$$p(\mathbf{x}) = \sum_{i=1}^{c} p(\mathbf{x}|\varpi_i)p(\varpi_i)$$

onde $p(\mathbf{x})$ é a probabilidade das observações (factor de escala)

Classificador de Máximo a Posteriori

- Pressuposto: Conhecimento completo de todas as densidades
- Critério adoptado: Minimizar a probabilidade de erro
- Regra de decisão: calcular a probabilidade de ocorrência de cada classe *dada a observação* $\mathbf{x}$, e escolher a classe mais provável:

$$\hat{\varpi} = \varpi_i \;\text{ se }\; p(\varpi_i|\mathbf{x}) > p(\varpi_k|\mathbf{x}),\; \forall\, k \neq i$$

- Lei de Bayes:

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

- Funções discriminantes:

$$g_i(\mathbf{x}) = p(\varpi_i|\mathbf{x})$$

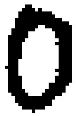
$$\text{ou} \quad g_i(\mathbf{x}) = p(\mathbf{x}|\varpi_i)p(\varpi_i)$$

  - $\hat{\varpi} = \varpi_i \;\text{ se }\; g_i(\mathbf{x}) > g_k(\mathbf{x}),\; \forall\, k \neq i$

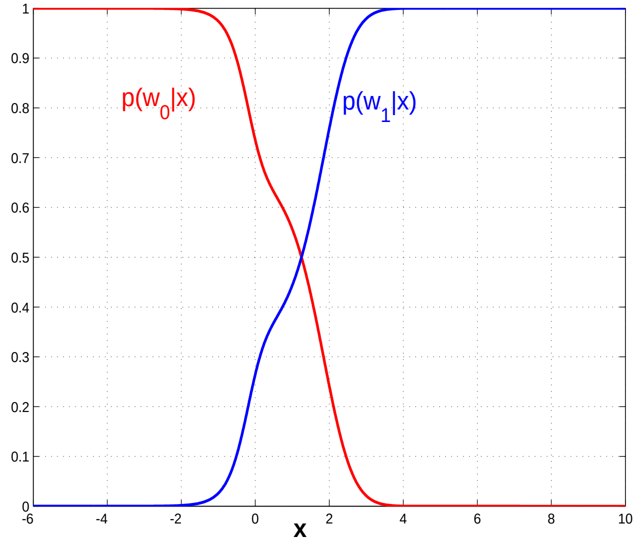
Classificador de Máximo a Posteriori

- Exemplo de classificação de dois dígitos manuscritos

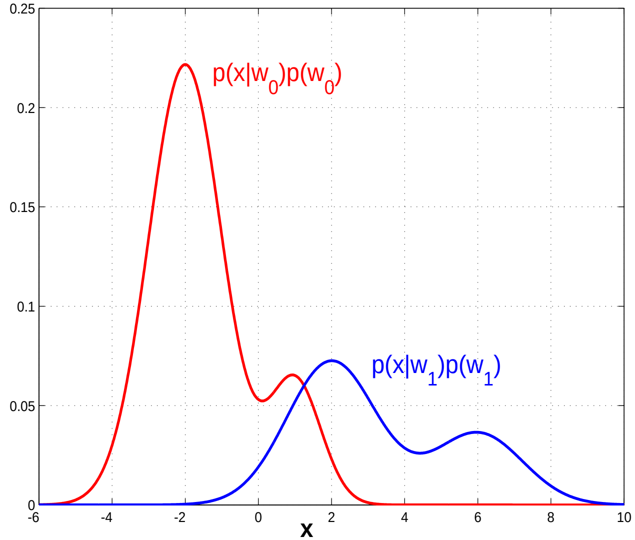
$$p(\varpi_0) = \frac{2}{3}$$



$$p(\varpi_1) = \frac{1}{3}$$



$$g_i(x) = p(\varpi_i|x)$$



$$g_i(x) = p(x|\varpi_i)p(\varpi_i)$$

Classificador de Máximo a Posteriori

Probabilidade de Erro

- $$p_{ik} = p(\hat{\varpi} = \varpi_k | \mathbf{x} \in \varpi_i) = \int_{\mathcal{S}_k} p(\mathbf{x}|\varpi_i) d\mathbf{x}$$

- Matriz de confusão:

$$\mathbf{P} = \begin{array}{c} \\ \varpi_1 \\ \varpi_2 \\ \vdots \\ \varpi_c \end{array} \begin{array}{c} \begin{array}{cccc} \hat{\varpi}_1 & \hat{\varpi}_2 & \cdots & \hat{\varpi}_c \end{array} \\ \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1c} \\ p_{21} & p_{22} & \cdots & p_{2c} \\ \vdots & & \ddots & \vdots \\ p_{c1} & p_{c2} & \cdots & p_{cc} \end{bmatrix} \end{array}$$

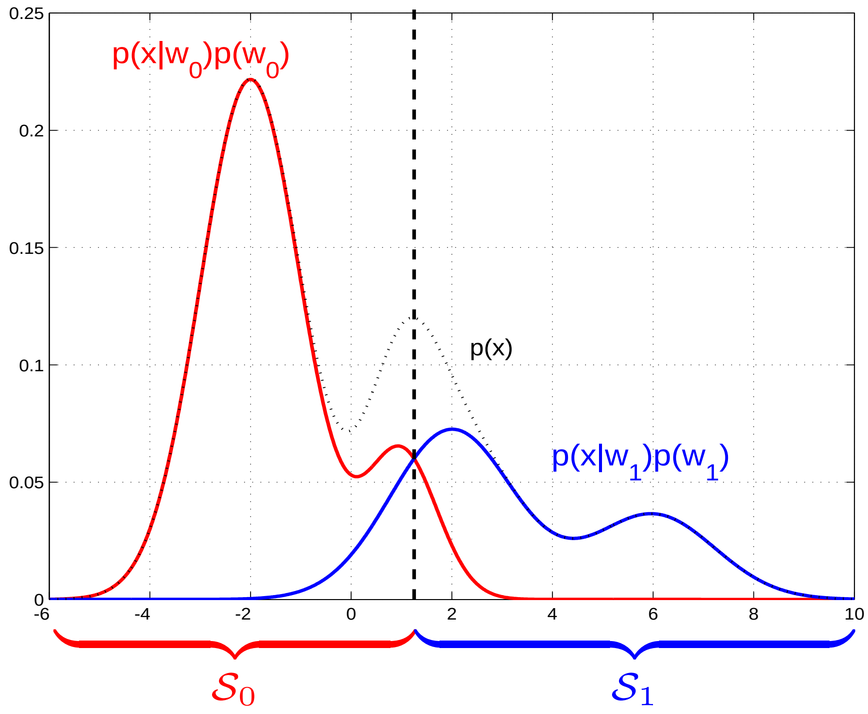
- Probabilidade total do erro:

$$\mathcal{P}_{\text{erro}} = \sum_{i=1}^{c} \left(\underbrace{1 - \overbrace{\int_{\mathcal{S}_i} p(\mathbf{x}|\varpi_i) d\mathbf{x}}^{p_{ii}}}_{\mathcal{P}_{\text{erro}}(\varpi_i)} \right) p(\varpi_i) = 1 - \sum_{i=1}^{c} p_{ii} p(\varpi_i)$$

Classificador de Máximo a Posteriori

Probabilidade de Erro

- Exemplo de duas classes (ϖ_0 e ϖ_1)

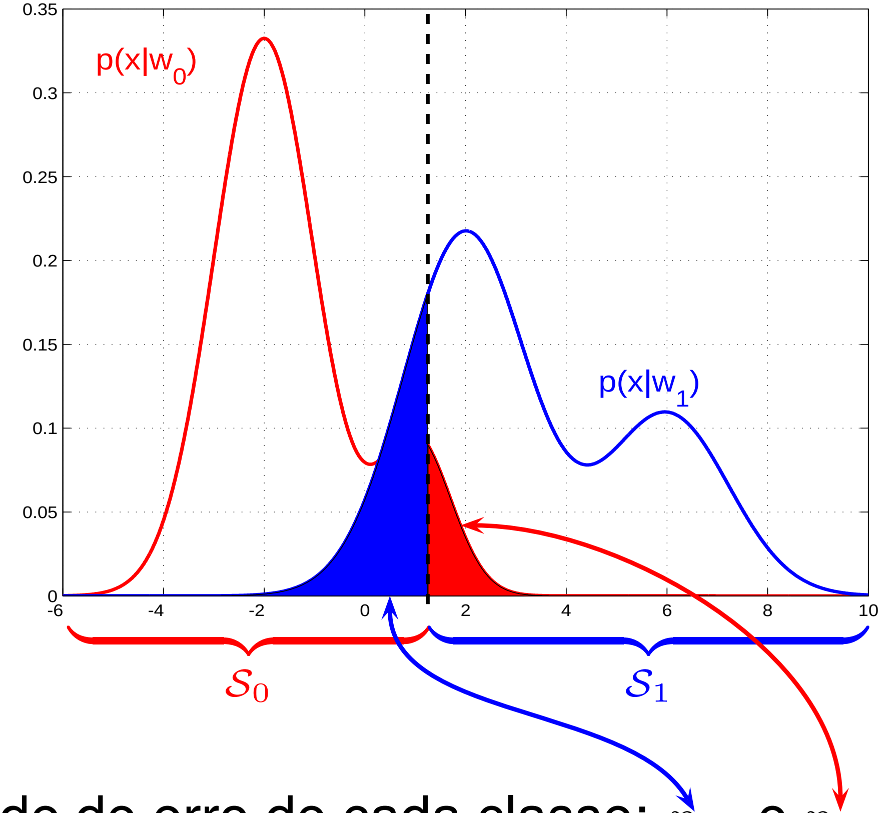


Ponto de fronteira: $p(x|\varpi_0)p(\varpi_0) = p(x|\varpi_1)p(\varpi_1)$

Classificador de Máximo a Posteriori

Probabilidade de Erro

- Exemplo de duas classes (ϖ_0 e ϖ_1)

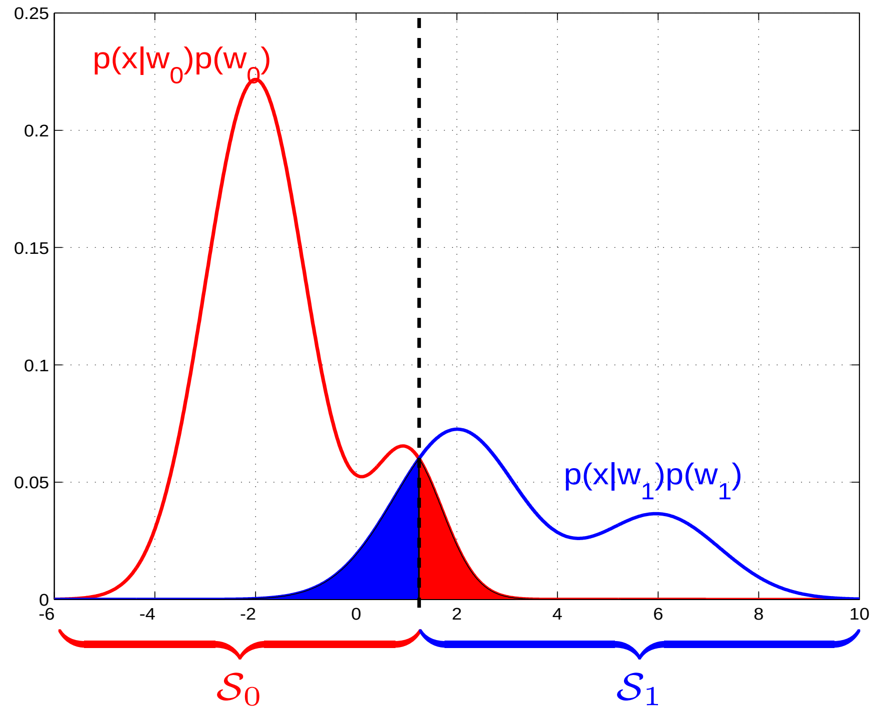


Probabilidade de erro de cada classe: p_{10} e p_{01}

Classificador de Máximo a Posteriori

Probabilidade de Erro

- Exemplo de duas classes (ϖ_0 e ϖ_1)

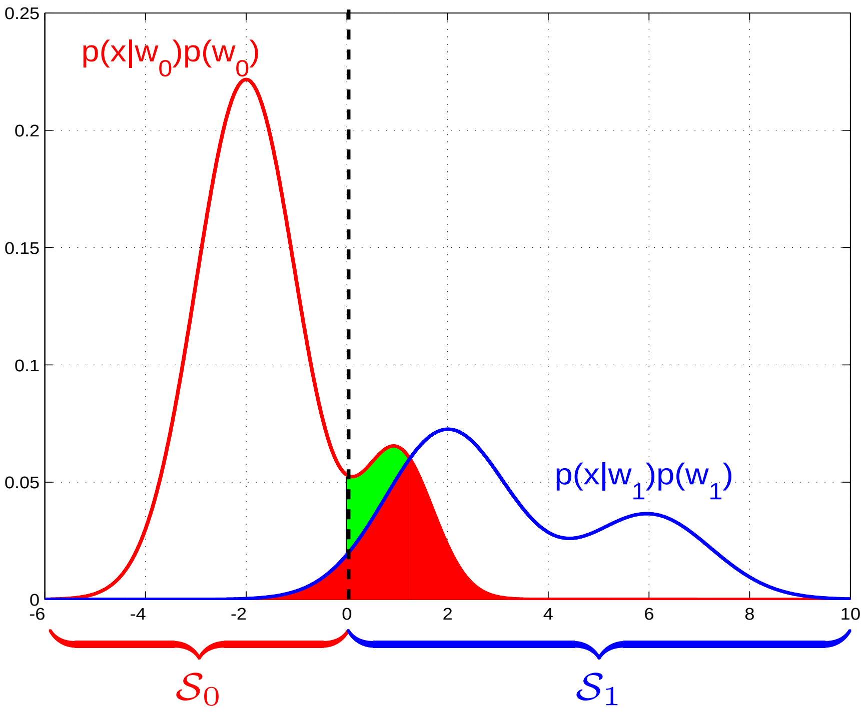


Probabilidade total de erro $p_{01}p(\varpi_0) + p_{10}p(\varpi_1)$ do classificador de MAP

Classificador de Máximo a Posteriori

Probabilidade de Erro

- Exemplo de duas classes (ϖ_0 e ϖ_1)

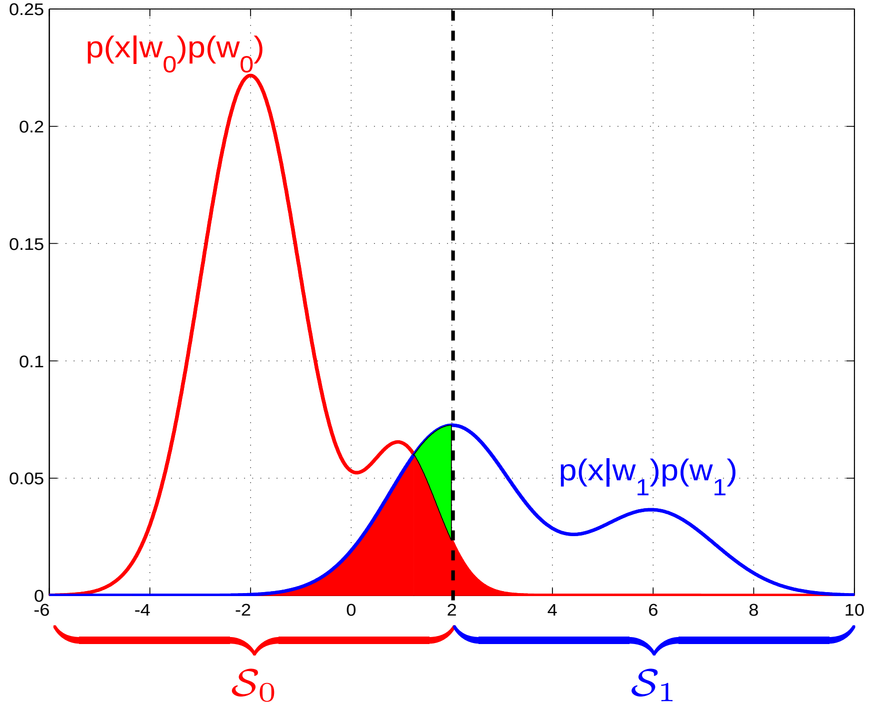


Probabilidades total de erro aumenta para outros pontos de fronteira

Classificador de Máximo a Posteriori

Probabilidade de Erro

- Exemplo de duas classes (ϖ_0 e ϖ_1)

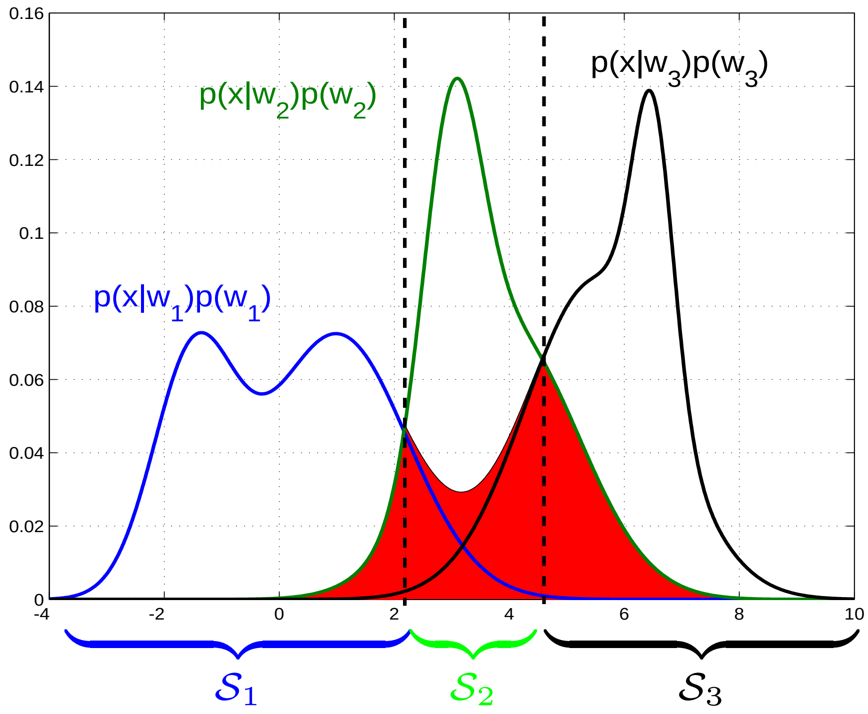


Probabilidades total de erro aumenta para outros pontos de fronteira

Classificador de Máximo a Posteriori

Probabilidade de Erro

- Exemplo de três classes ϖ_1, ϖ_2, e ϖ_3



Probabilidades total de erro do classificador de MAP

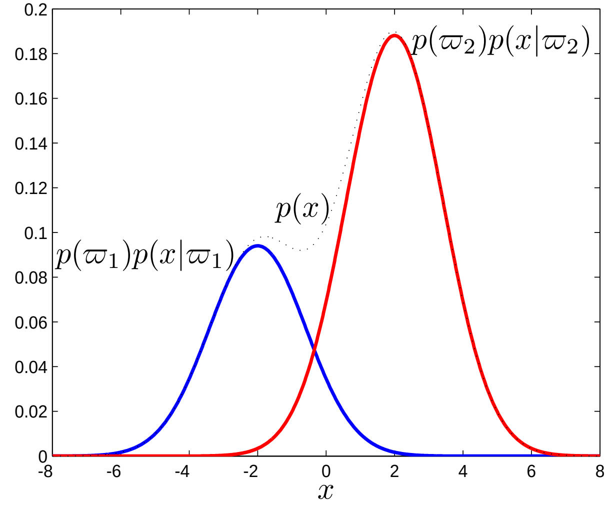
Classificador de Máximo a Posteriori

Densidades gaussianas 1D:regiões e fronteiras de decisão

- Para dados 1D em que as densidades das classes são gaussianas com variâncias idênticas, as fronteiras de decisão resultam num único ponto.

Exemplo:

$$p(\varpi_1) = \tfrac{1}{3},\ p(\varpi_2) = \tfrac{2}{3}$$
$$p(x|\varpi_1) = \mathcal{N}(-2,2) = \frac{1}{\sqrt{4\pi}} \exp(-\tfrac{1}{4}(x+2)^2)$$
$$p(x|\varpi_2) = \mathcal{N}(+2,2) = \frac{1}{\sqrt{4\pi}} \exp(-\tfrac{1}{4}(x-2)^2)$$



ponto de fronteria:

$$
\begin{aligned}
& p(\varpi_1|x) = p(\varpi_2|x) \\
\Leftrightarrow\quad & p(\varpi_1)p(x|\varpi_1) = p(w_2)p(x|\varpi_2) \\
\Leftrightarrow\quad & \frac{1/3}{\sqrt{4\pi}} \exp(-\tfrac{1}{4}(x+2)^2) = \frac{2/3}{\sqrt{4\pi}} \exp(-\tfrac{1}{4}(x-2)^2) \\
\Leftrightarrow\quad & -\tfrac{1}{4}(x+2)^2 = -\tfrac{1}{4}(x-2)^2 + \ln 2 \\
\Leftrightarrow\quad & x^2 + 4x + 4 = x^2 - 4x + 4 - 4\ln 2
\end{aligned}
$$

$$x = -0.35$$
$$\mathcal{S}_1 = [-\infty, -0.35]$$
$$\mathcal{S}_2 = [-0.35, +\infty]$$

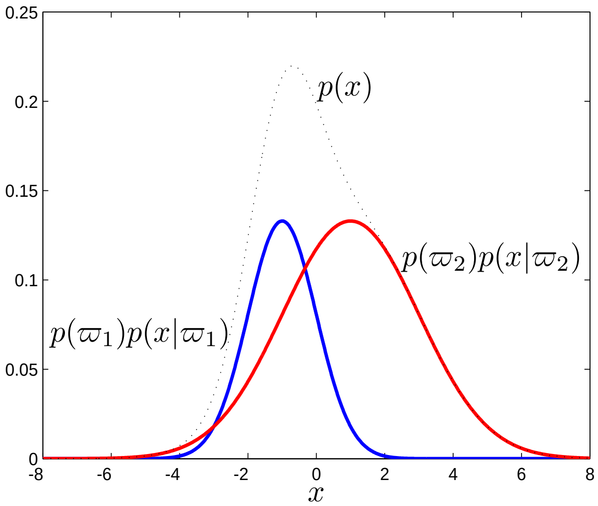
Classificador de Máximo a Posteriori

Densidades gaussianas 1D:regiões e fronteiras de decisão

- Para dados 1D em que as densidades das classes são gaussianas com variâncias distintas, as fronteiras de decisão resultam em dois pontos.

Exemplo:

$$p(\varpi_1) = \tfrac{1}{3},\ p(\varpi_2) = \tfrac{2}{3}$$
$$p(x|\varpi_1) = \mathcal{N}(-1,1) = \tfrac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}(x+1)^2)$$
$$p(x|\varpi_2) = \mathcal{N}(+1,4) = \tfrac{1}{\sqrt{8\pi}} \exp(-\tfrac{1}{8}(x-1)^2)$$



pontos de fronteria:

$$
\begin{aligned}
& p(\varpi_1|x) = p(\varpi_2|x) \\
\Leftrightarrow\quad & p(\varpi_1)p(x|\varpi_1) = p(\varpi_2)p(x|\varpi_2) \\
\Leftrightarrow\quad & \tfrac{1/3}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}(x+1)^2) = \tfrac{2/3}{\sqrt{8\pi}} \exp(-\tfrac{1}{8}(x-1)^2) \\
\Leftrightarrow\quad & -\tfrac{1}{2}(x+1)^2 = -\tfrac{1}{8}(x-1)^2 \\
\Leftrightarrow\quad & 4x^2 + 8x + 4 = x^2 - 2x + 1
\end{aligned}
$$

$$x = \{-3, -\tfrac{1}{3}\}$$
$$\mathcal{S}_1 = [-3, -\tfrac{1}{3}]$$
$$\mathcal{S}_2 = [-\infty, -3] \cup [-\tfrac{1}{3}, +\infty]$$

Classificador de Máximo a Posteriori

Densidades gaussianas 1D:

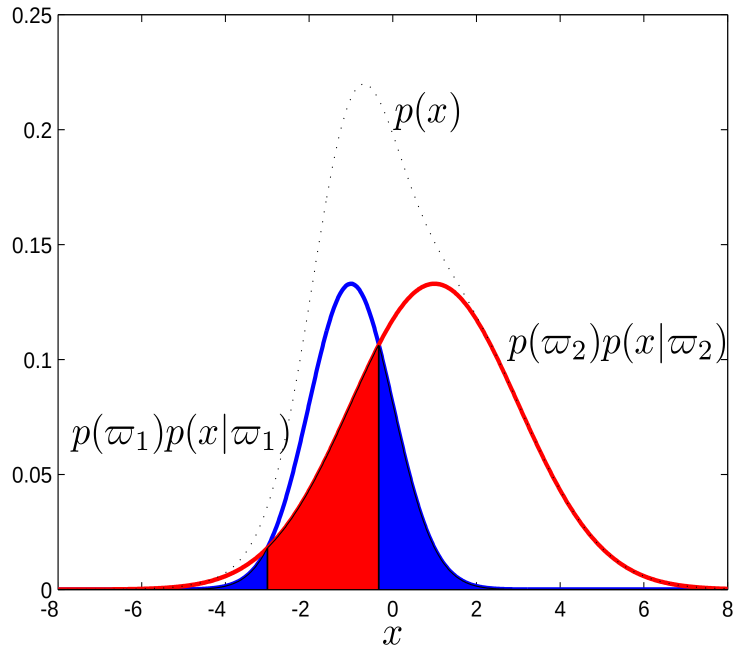
- Probabilidade de erro.

Exemplo:

$p(\varpi_1) = \frac{1}{3},\ p(\varpi_2) = \frac{2}{3}$

$p(x|\varpi_1) = \mathcal{N}(-1,1) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(x+1)^2)$

$p(x|\varpi_2) = \mathcal{N}(+1,4) = \frac{1}{\sqrt{8\pi}} \exp(-\frac{1}{8}(x-1)^2)$



pontos de fronteria:

$x = \{-3, -\frac{1}{3}\}$

$\mathcal{S}_1 = [-3, -\frac{1}{3}]$

$\mathcal{S}_2 = [-\infty, -3] \cup [-\frac{1}{3}, +\infty]$

probabilidade total de erro:

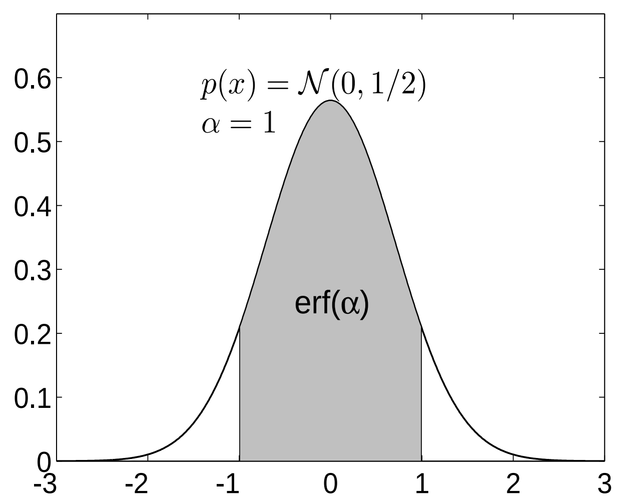
$$\mathcal{P}_{\text{erro}} = p(\varpi_1) \int_{\mathcal{S}_2} p(x|\varpi_1)dx + p(\varpi_2) \int_{\mathcal{S}_1} p(x|\varpi_2)dx = p(\varpi_1)p_{12} + p(\varpi_2)p_{21}$$

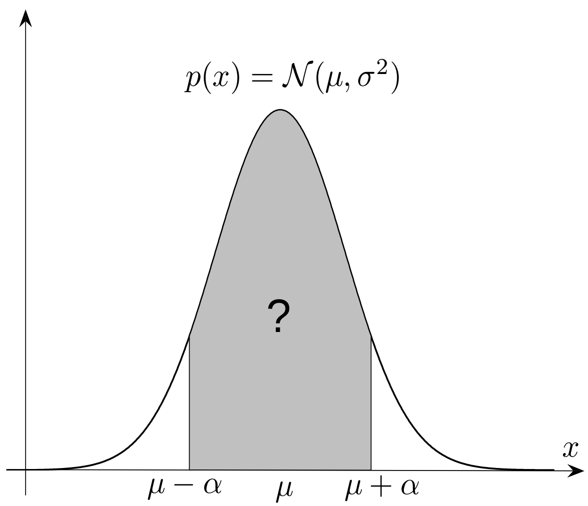
Classificador de Máximo a Posteriori

Densidades gaussianas 1D:

- Cálculo de áreas de gaussianas
- Função de erro: $\mathrm{erf}(\alpha) = \int_{-\alpha}^{+\alpha} \frac{1}{\sqrt{\pi}} \exp\left(-x^2\right) dx$

  área de $[-\alpha, +\alpha]$, duma gaussiana $\mathcal{N}(0, \frac{1}{2})$





Classificador de Máximo a Posteriori

Densidades gaussianas 1D:

- Cálculo de áreas de gaussianas:

$$\int_{\mu-\alpha}^{\mu+\alpha} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx$$

Mudança de variável:

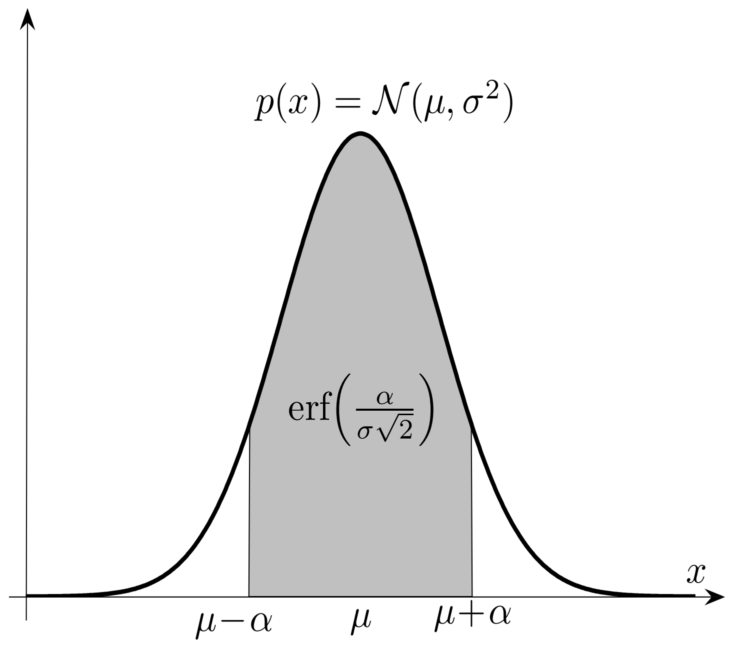
$$
\begin{array}{lll}
y = \frac{x-\mu}{\sqrt{2\sigma^2}} & & \frac{dy}{dx} = \frac{1}{\sqrt{2}\sigma} \\
 & \Longleftrightarrow & dx = \sqrt{2}\sigma dy \\
\text{intervalos} & & x_{\text{limites}} = [\mu-\alpha, \mu+\alpha] \\
 & \Longleftrightarrow & y_{\text{limites}} = [\frac{-\alpha}{\sqrt{2}\sigma}, \frac{+\alpha}{\sqrt{2}\sigma}] \\
\text{integral} & & \int_{-\alpha+\mu}^{\mu+\alpha} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\
 & \Longleftrightarrow & \int_{\frac{-\alpha}{\sqrt{2}\sigma}}^{\frac{-\alpha}{\sqrt{2}\sigma}} \frac{1}{\sqrt{\pi}} \exp(-y^2) dy = \text{erf}\left(\frac{\alpha}{\sqrt{2}\sigma}\right)
\end{array}
$$

Classificador de Máximo a Posteriori

Densidades gaussianas 1D:

- Cálculo de áreas de gaussianas:

$$\operatorname{erf}\left(\frac{\alpha}{\sqrt{2}\sigma}\right) = \int_{\mu-\alpha}^{\mu+\alpha} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx$$



Classificador de Máximo a Posteriori

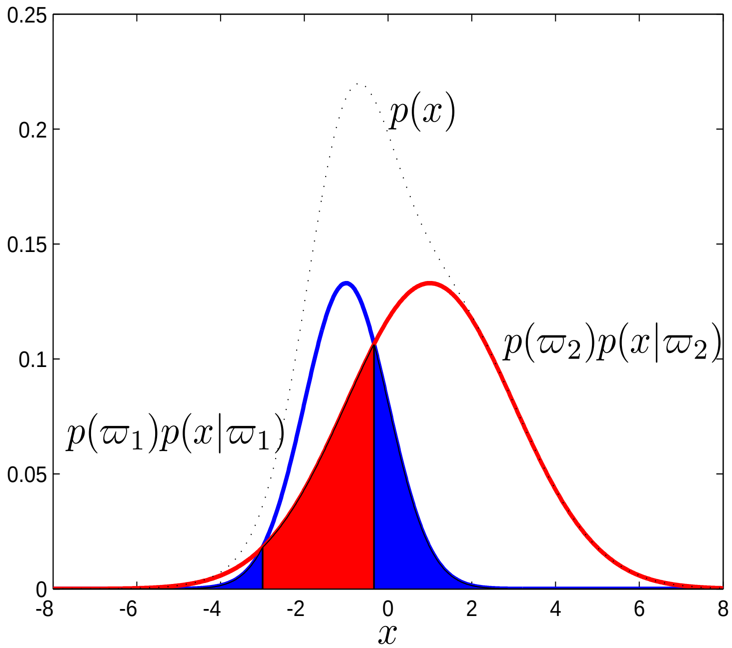
Densidades gaussianas 1D:

Exemplo:

$$p(\varpi_1) = \frac{1}{3},\ p(\varpi_2) = \frac{2}{3}$$

$$p(x|\varpi_1) = \mathcal{N}(-1,1) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(x+1)^2)$$

$$p(x|\varpi_2) = \mathcal{N}(+1,4) = \frac{1}{\sqrt{8\pi}} \exp(-\frac{1}{8}(x-1)^2)$$



pontos de fronteria:

$$x = \{-3, -\frac{1}{3}\}$$

$$\mathcal{S}_1 = [-3, -\frac{1}{3}]$$

$$\mathcal{S}_2 = [-\infty, -3] \cup [-\frac{1}{3}, +\infty]$$

probabilidade total de erro:

$$\mathcal{P}_{\text{erro}} = p(\varpi_1)\int_{\mathcal{S}_2} p(x|\varpi_1)dx + p(\varpi_2)\int_{\mathcal{S}_1} p(x|\varpi_2)dx = p(\varpi_1)p_{12} + p(\varpi_2)p_{21}$$

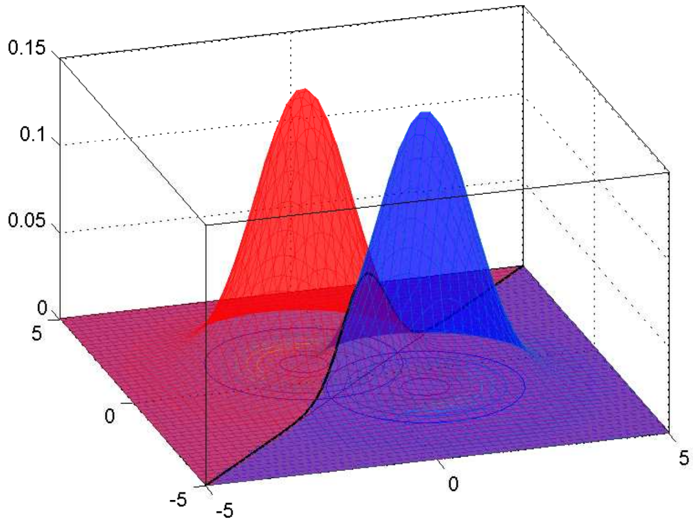
$$p_{12} = \frac{1}{2}\left(1 - \text{erf}\left(\frac{1-1/3}{\sqrt{2}}\right)\right) + \frac{1}{2}\left(1 - \text{erf}\left(\frac{2}{\sqrt{2}}\right)\right)$$

$$p_{21} = \frac{1}{2}\left(1 - \text{erf}\left(\frac{1+1/3}{2\sqrt{2}}\right)\right) - \frac{1}{2}\left(1 - \text{erf}\left(\frac{4}{2\sqrt{2}}\right)\right)$$

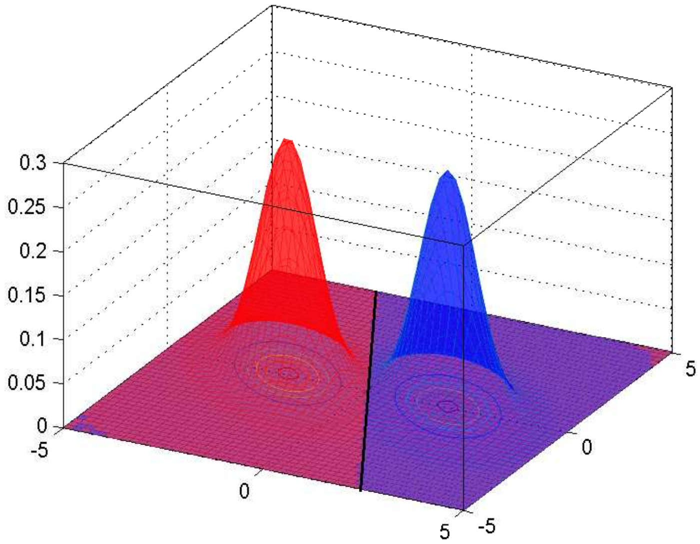
Classificador de Máximo a Posteriori

Densidades gaussianas 2D:regiões e fronteiras de decisão

- Quando as matrizes de covariância são idênticas, as fronteiras de decisão são lineares



$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

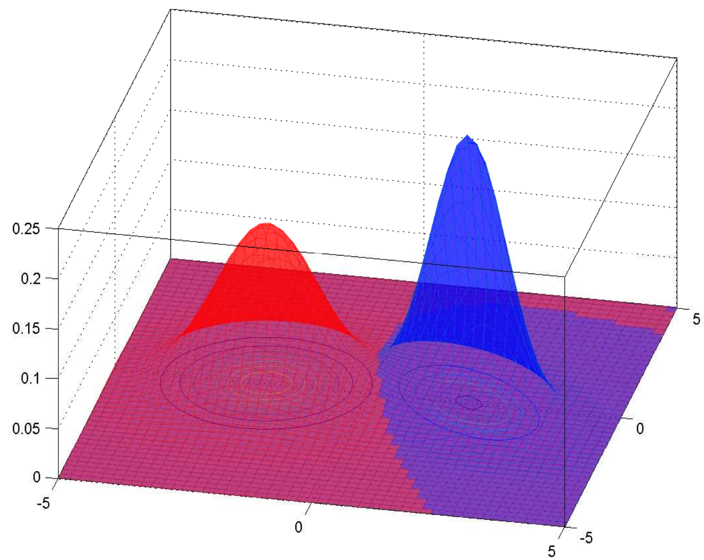


$$\Sigma = \begin{bmatrix} 0.68 & -0.32 \\ -0.32 & 0.68 \end{bmatrix}$$

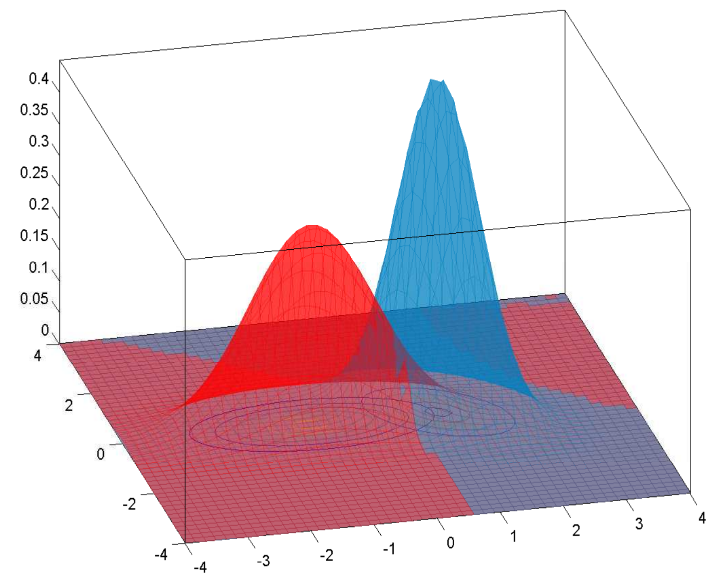
Classificador de Máximo a Posteriori

Densidades gaussianas 2D:regiões e fronteiras de decisão

- Quando as matrizes de covariância são diferentes, as fronteiras de decisão são curvas



$$\color{red}{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \color{blue}{\Sigma} = \begin{bmatrix} 0.68 & -0.32 \\ -0.32 & 0.68 \end{bmatrix}$$



$$\color{red}{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 0.25 \end{bmatrix} \quad \color{blue}{\Sigma} = \begin{bmatrix} 0.34 & -0.16 \\ -0.16 & 0.34 \end{bmatrix}$$

Classificador de Máximo a Posteriori

Exercícios:

Considere que x é uma observação gerada por uma de duas classes $\Omega = \{\varpi_1, \varpi_2\}$, de acordo com:

$$p(x|\varpi_1) = e^{-x}u(x)$$
$$p(x|\varpi_2) = \frac{1}{3}\left(u(x) - u(x-3)\right)$$

onde $u(x)$ é a função escalão. As probabilidades *a priori* de cada classe são $p(\varpi_1) = \frac{e}{e+3}$ e $p(\varpi_2) = \frac{3}{e+3}$. Para o classificador de MAP, determine:

1. As regiões de decisão.
2. A matriz de confusão.
3. A probabilidade total do erro.

Classificador de Máximo a Posteriori

Exercícios:

Seja x uma variável aleatória gerada por uma de duas classes ϖ_1 e ϖ_2 com as seguintes distribuições:

$$\begin{array}{ll} p(x|\varpi_1) = \mathcal{N}(-1, 2) & p(\varpi_1) = 0.7 \\ p(x|\varpi_2) = \mathcal{N}(1, \frac{1}{2}) & p(\varpi_2) = 0.3 \end{array}$$

Para o classificador de MAP, determine:

1. As regiões de decisão.
2. A matriz de confusão.
3. A probabilidade total do erro.

Classificador de Bayes

- Classificador de MAP minimiza a probabilidade de erro de classificação.
- Critério nem sempre recomendável visto que diferentes tipos de erros podem ter consequências variadas.
  - Sistema de detecção de incêndios florestais.
  - Sistema de detecção de aviões baseados em radar.
  - Sistema de despistagem de tumores em imagens médicas.
- Preferível haver menos *falhas de detecção* em detrimento do aumento da probabilidade de *falsos alarmes*.
- **Solução:** Atribuir custos a cada decisão.

Classificador de Bayes

- Pressuposto: Conhecimento completo de todas as densidades
- Critério adoptado: Minimizar o custo médio ou risco
- Atribuição de custos: $\lambda(\hat{\varpi}_j|\varpi_i) = \lambda_{ij}$ é o custo de classificar $\mathbf{x}$ na classe ϖ_j, dado que foi a classe ϖ_i que gerou $\mathbf{x}$
  - *A matriz de custo,* Λ, representa todas as combinações possíveis de $\hat{\varpi}$ e ϖ:

$$\mathbf{\Lambda} = \begin{array}{c} \\ \varpi_1 \\ \varpi_2 \\ \vdots \\ \varpi_c \end{array} \begin{array}{c} \begin{array}{cccc} \hat{\varpi}_1 & \hat{\varpi}_2 & \cdots & \hat{\varpi}_c \end{array} \\ \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1c} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2c} \\ \vdots & & \ddots & \vdots \\ \lambda_{c1} & \lambda_{c2} & \cdots & \lambda_{cc} \end{bmatrix} \end{array}$$

Classificador de Bayes

- Pressuposto: Conhecimento completo de todas as densidades
- Critério adoptado: Minimizar o custo médio ou risco
- λ_{ij}: custo de decidir que $\hat{\varpi} = \varpi_j$, quando $\varpi = \varpi_i$
- Custo associado a $\hat{\varpi} = \varpi_i$:

$$\lambda_i(\mathbf{x}) = \sum_{k=1}^{c} \lambda_{ki} p(\varpi_k|\mathbf{x}) = \sum_{k=1}^{c} \lambda_{ki} \frac{p(\mathbf{x}|\varpi_k)p(\varpi_k)}{p(\mathbf{x})}$$

- Regra de decisão: dada a observação $\mathbf{x}$, calcular o custo, $\lambda_i(\mathbf{x})$, associado a cada decisão $\hat{\varpi} = \varpi_i$, e escolher o a classe correspondente ao **menor** custo (oposto de funções discriminantes):

$$\hat{\varpi} = \varpi_i \;\; \text{se} \;\; \lambda_i(\mathbf{x}) < \lambda_k(\mathbf{x}), \; \forall \; k \neq i$$

- Funções discriminantes: $g_i(\mathbf{x}) = -\lambda_i(\mathbf{x})$

Classificador de Bayes

- Pressuposto: Conhecimento completo de todas as densidades
- Critério adoptado: Minimizar o custo médio ou risco
- λ_{ij}: custo de decidir que $\hat{\varpi} = \varpi_j$, quando $\varpi = \varpi_i$
- Custo associado a $\hat{\varpi} = \varpi_i$: $$\lambda_i(\mathbf{x}) = \sum_{k=1}^{c} \lambda_{ki} p(\varpi_k|\mathbf{x})$$
- Regra de decisão: $\hat{\varpi} = \varpi_i \ \text{ se } \ \lambda_i(\mathbf{x}) < \lambda_k(\mathbf{x}), \ \forall\, k \neq i$
- Risco associado a $\hat{\varpi} = \varpi_i$:

$$\mathcal{R}_i = \sum_{k=1}^{c} \int_{\mathcal{S}_i} \lambda_{ki} p(\mathbf{x}|\varpi_k) p(\varpi_k) \mathrm{d}\mathbf{x} = \int_{\mathcal{S}_i} \lambda_i(\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x}$$

- Risco Total: $$\mathcal{R} = \sum_{i=1}^{c} \mathcal{R}_i = \sum_{i=1}^{c} \int_{\mathcal{S}_i} \lambda_i(\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x}$$

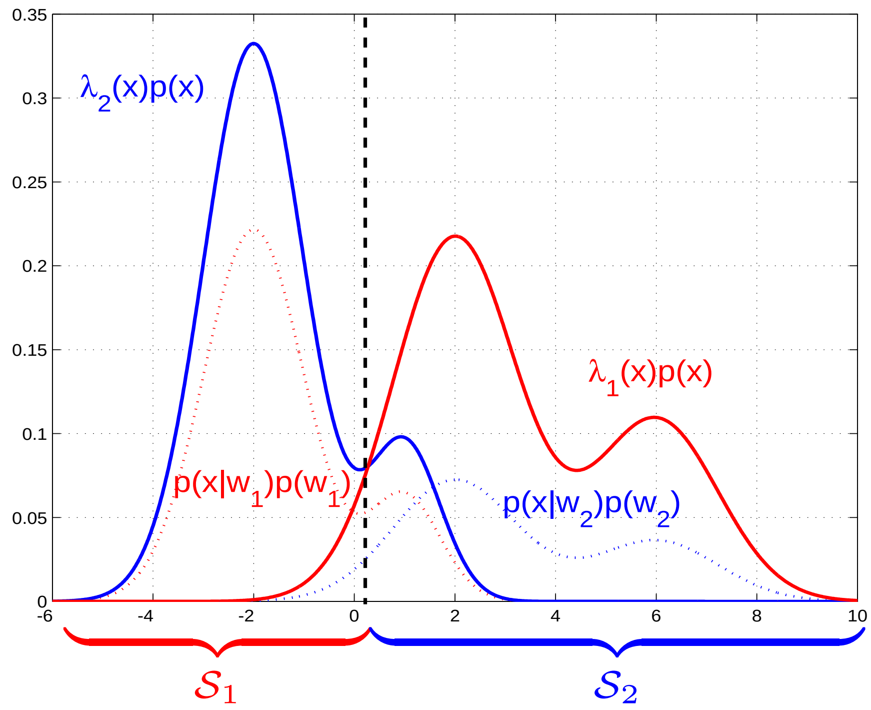
Classificador de Bayes

Exemplo de duas classes:

- Custo da decisão $\hat{\varpi} = \varpi_i$:

$$\lambda_1(\mathbf{x}) = 3.0 p(\varpi_2|\mathbf{x})$$
$$\lambda_2(\mathbf{x}) = 1.5 p(\varpi_1|\mathbf{x})$$

$$\Lambda = \begin{bmatrix} 0 & 1.5 \\ 3.0 & 0 \end{bmatrix}$$



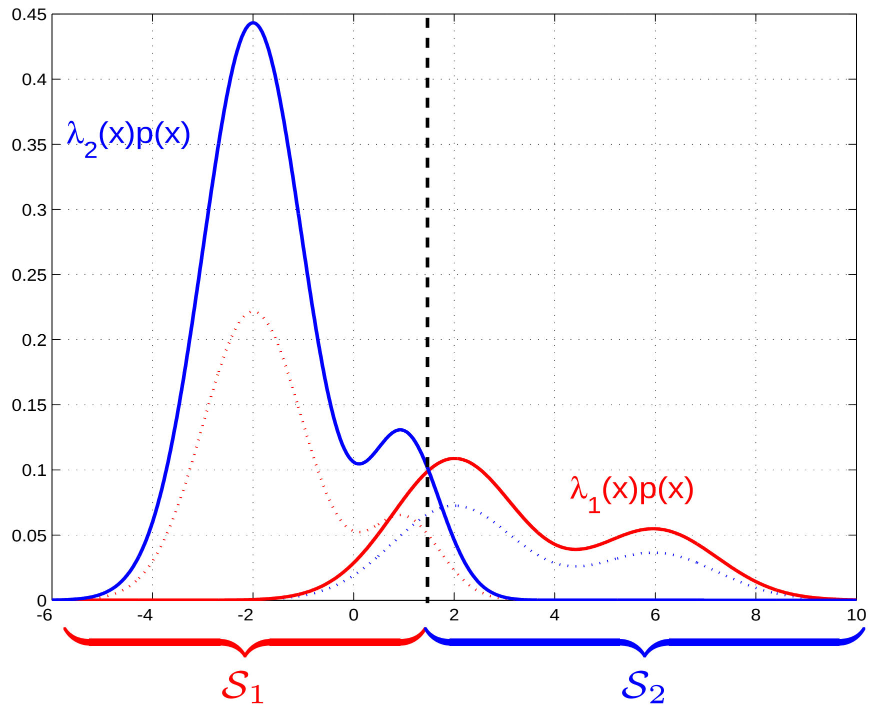
Classificador de Bayes

Exemplo de duas classes:

- Custo da decisão $\hat{\varpi} = \varpi_i$:

$$\lambda_1(\mathbf{x}) = 1.5p(\varpi_2|\mathbf{x})$$
$$\lambda_2(\mathbf{x}) = 3.0p(\varpi_1|\mathbf{x})$$

$$\Lambda = \begin{bmatrix} 0 & 3.0 \\ 1.5 & 0 \end{bmatrix}$$



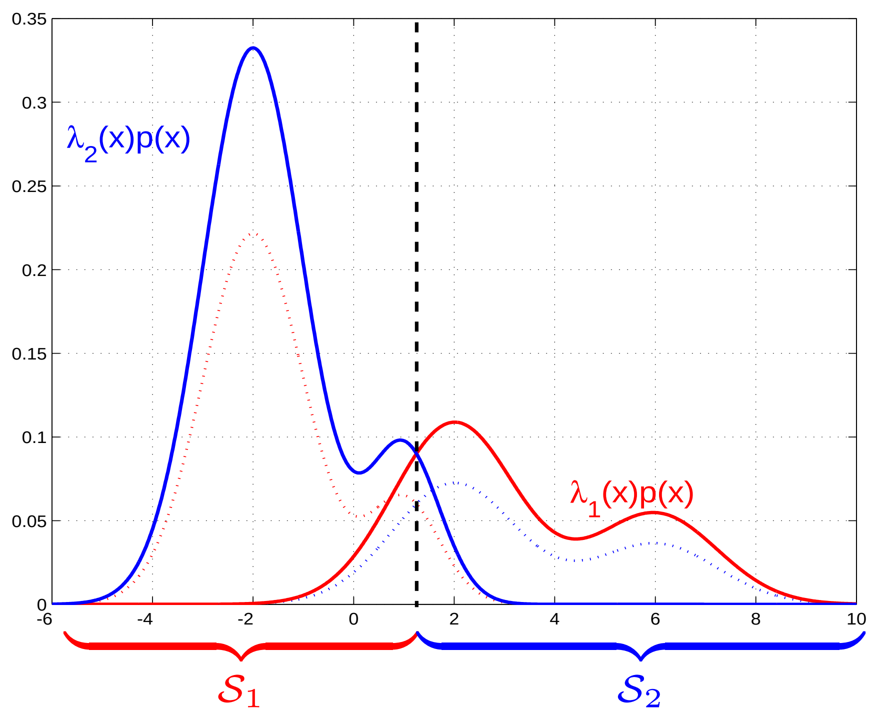
Classificador de Bayes

Exemplo de duas classes:

- Custo da decisão $\hat{\varpi} = \varpi_i$:

$$\lambda_1(\mathbf{x}) = 1.5p(\varpi_2|\mathbf{x})$$
$$\lambda_2(\mathbf{x}) = 1.5p(\varpi_1|\mathbf{x})$$

$$\Lambda = \begin{bmatrix} 0 & 1.5 \\ 1.5 & 0 \end{bmatrix}$$



Classificador de Bayes

Exemplo de duas classes:

- Generalizando, o custo da decisão $\hat{\varpi} = \varpi_i$ é:

$$\lambda_1(\mathbf{x}) = \lambda_{11} p(\varpi_1|\mathbf{x}) + \lambda_{21} p(\varpi_2|\mathbf{x})$$
$$\lambda_2(\mathbf{x}) = \lambda_{12} p(\varpi_1|\mathbf{x}) + \lambda_{22} p(\varpi_2|\mathbf{x})$$

- Regra de decisão: $\hat{\varpi} = \varpi_1 \;\; \text{se} \;\; \lambda_1(\mathbf{x}) < \lambda_2(\mathbf{x})$

$$(\lambda_{21} - \lambda_{22})\, \varpi_2|\mathbf{x}) < (\lambda_{12} - \lambda_{11})\, p(\varpi_1|\mathbf{x})$$

- Geralmente $\lambda_{12} - \lambda_{11}$ e $\lambda_{21} - \lambda_{22}$ são valores positivos. Isto é, o custo é mais elevado quando se comete um erro.
- Para $(\lambda_{12} - \lambda_{11}) > 0$ então $\hat{\varpi} = \varpi_1$ se:

$$\frac{p(\mathbf{x}|\varpi_1)}{p(\mathbf{x}|\varpi_2)} > \frac{(\lambda_{21} - \lambda_{22})\, p(\varpi_2)}{(\lambda_{12} - \lambda_{11})\, p(\varpi_1)}$$

Classificador de Bayes

Exemplo: Filtro de SPAM

- Temos duas classes:
  - ϖ_1 mail normal com $p(\varpi_1) = 0.5$
  - ϖ_2 mail SPAM com $p(\varpi_2) = 0.5$
- Temos duas acções associadas ao resultado do classificador: se for um mail normal guardamos o mail, caso contrário apagamos o mail.
- Matriz de custos: $$\Lambda = \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}$$

  É pior apargar um mail normal do que guardar um mail SPAM.
- Recebemos um mail que tem um vector de características $\mathbf{x}$ com $p(\mathbf{x}|\varpi_1) = 0.3$ e com $p(\mathbf{x}|\varpi_2) = 0.8$. Qual das acções é tomada pelo classificador de Bayes?

Classificador de Bayes

- Classificador de Bayes **é o classificador óptimo** em termos da minimização do risco
- O classificador de MAP é um caso particular do classificador de Bayes

Necessário ter o conhecimento completo de todas as densidades