

# Relatório de atividades: Uso do classificador Bayesiano

David Clifte\*

2015, v-1.0

## Resumo

Este trabalho apresenta os resultados obtidos ao aplicar o classificador bayessiano e submete-lo a 5 diferentes funções de densidade de probabilidade. Para a verificação da acurácia bem como das regiões de decisão são apresentados os resultados obtidos com três base de dados diferentes. A implementação foi feita no Matlab<sup>TM</sup>

**Palavras-chaves:** Bayes. Reconhecimento de padrões.

## Introdução

### 0.1 Estimativa de função de densidade de probabilidade

A estimativa da distribuição de dados é importante pois permite utilizar uma representação compacta dos dados e ainda sim manter as informações mais relevantes da base. Existem basicamente três abordagens para estimar a função de densidade de probabilidade de um sinal: paramétrica, não-paramétrica e semiparamétrica. O sucesso destas representações dependem do modelo que tem sido definido ([PATTERN..., 2001](#)).

#### 0.1.1 Método não paramétrico

Os métodos não-paramétricos não fazem nenhuma consideração da distribuição de probabilidade dos dados. Em geral, estes métodos se caracterizam por conseguir uma estimativa adequada para qualquer conjunto de dados que recebem como entrada. O uso de métodos não paramétricos para a estimativa das funções de densidade de probabilidade vem da falta de informações a priori sobre a função de densidade de probabilidade dos dados. Como exemplo de método não paramétrico temos a janela de Parzen.

---

\*cliftedavid@gmail.com

### 0.1.2 Método paramétrico

A abordagem paramétrica é geralmente usada quando a distribuição dos dados é conhecida antecipadamente ou quando os dados são simples de forma que permitam ser modelados usando uma distribuição conhecida, por exemplo gaussiana, Gamma, Laplace, etc

### 0.1.3 Método semi-paramétrico

A abordagem semiparamétrica combina a flexibilidade da abordagem não-paramétrica e a eficiência na avaliação dos parâmetros da abordagem paramétrica. Estes modelos utilizam um número de funções base que são sempre menores que o conjunto de treinamento. O uso dos modelos semiparamétricos baseados em gaussianas, GMM, tem se apresentado como uma ferramenta amplamente usada na estimativa da PDF de qualquer sinal.

## 0.2 Janela de Parzen–Rosenblatt

A janela de Parzen, Parzen-Rosenblatt, é um método utilizado para estimar a função de densidade de probabilidade  $p(x)$  com base nas amostras presentes na base de treinamento. A ideia básica da janela de parzen é contar quantas amostras influenciam em determinada região R (Janela). A influência de cada amostra é definida a priori com a definição do Kernel. A utilização de kernels do tipo hipercubos ou Gaussianas multivariadas são as mais comuns e ao longo deste trabalho são utilizados apenas kernels de gaussianas multivariadas.

Definimos a influencia de uma amostra  $x$  como  $p(x)$ . Para o caso da gaussiana multivariada temos:

$$p(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x - \mu)^t \Sigma^{-1} (x - \mu)\right] \quad (1)$$

Portanto a influencia total  $P(x)$  é computada como:

$$P(x) = \frac{1}{N} \sum_{i=1}^N p(x_n) \quad (2)$$

Apesar da gaussiana multivariada aceitar qualquer matriz de covariância, contanto que esta seja regular e simétrica, neste trabalho são utilizadas apenas matriz que possuem uma configuração  $\sigma^2 I$  ou seja, uma matriz diagonal de covariância igual a  $\sigma$ .

## 0.3 Gaussian Misture Models

### 0.3.1 Introdução

Considerando um conjunto de dados  $X = x_1, x_2, \dots, x_n | x \in R$ , a PDF dos dados pode ser aproximada por uma família  $F$  de funções de distribuição

de probabilidades em R. Em algoritmos dedicados à estimativa da PDF, o problema é encontrar a função de distribuição  $f(x) \in F$  que melhor gere os dados de entrada.

$$f(x, \Theta) = \sum_{k=1}^k P_k g(x, \mu_k, \sigma_k) \quad (3)$$

$\Theta$  é o conjunto de parâmetros do conjunto de funções que devem ser estimados durante a fase de treinamento. Desta forma para gaussiana temos

$$\Theta = \begin{bmatrix} \mu_1 & \sigma_1 \\ \dots & \dots \\ \mu_k & \sigma_k \end{bmatrix} \quad (4)$$

$\Theta$  pode ser estimado utilizando o Algoritmo Maximização da Expectância (EM). O algoritmo EM é um procedimento iterativo para estimar os parâmetros de uma mistura de gaussianas. Cada iteração do algoritmo EM consiste em dois processos: Expectância e Maximização. Esta aproximação se consegue através do cálculo da probabilidade de pertinência de um ponto às funções de distribuições na fase de expectância. Na fase de maximização são estimados os parâmetros que maximizam cada função de distribuição, ponderadas com os valores calculados na fase de expectância.

Na figura 1 é apresentado o resultado da aproximação da base da íris. São exibidas três das quatro combinações possíveis das características da base, essas combinações são utilizadas para realizar o treinamento do GMM.

## 1 Classificador de Bayes com opção de rejeição

### 1.1 Introdução

Dada uma classificação entre M classes, o classificador de Bayes faz a seleção da classe de um dado  $x$  com base na probabilidade de  $w_i$  dado um  $x$ ,  $P(w_i|x)$ . Assim temos:

$$x \in w_i \iff P(w_i|x) \geq P(w_j|x) \forall i \neq j \quad (5)$$

### 1.2 Opção de Rejeição

Considerando o classificar de bayes um padrão é escolhido em detrimento a outro de acordo com sua probabilidade a posteriori, A opção de rejeição sugere que de acordo com este valor de probabilidade calculado a classificação pode ser rejeitada pois a mesma reflete também o grau de confiança na classificação. Desta forma podemos definir um limiar para este grau de confiança e assim caso a probabilidade a posteriori seja menor que este limiar

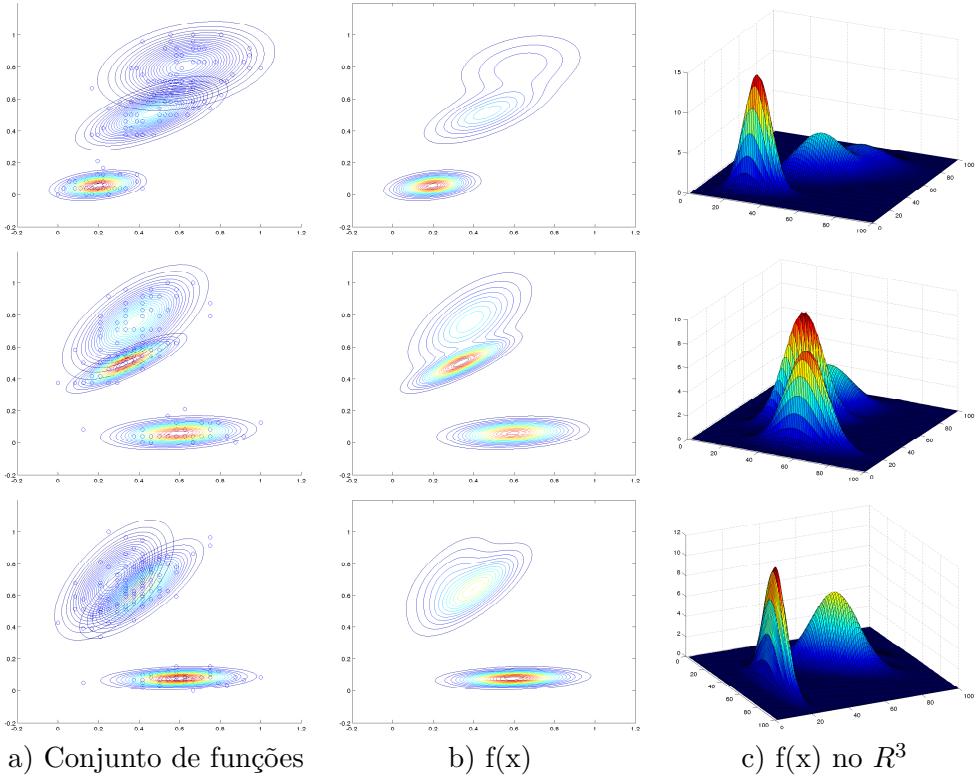


Figura 1 – Resultados obtidos ao utilizar a mistura de gaussianas para modelar a base de dados da íris. Na primeira linha temos o resultado do treinamento utilizando as informações comprimento da sepala e largura da pétala. Na linha seguinte temos largura da sépala e da pétala e na terceira linha o comprimento da sépala e largura da pétala

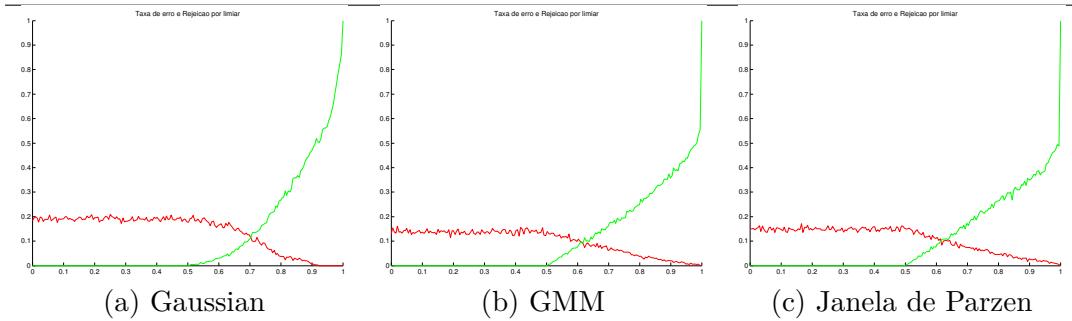
a amostra pode ser classificada para a classe de rejeição. Temos a seguinte regra de decisão para um problema com duas classes:

$$x \in \begin{cases} w_1, & \text{se } P(w_1|x) > \beta \\ w_2, & \text{se } P(w_2|x) > \beta \\ w_r, & \text{caso contrário} \end{cases} \quad (6)$$

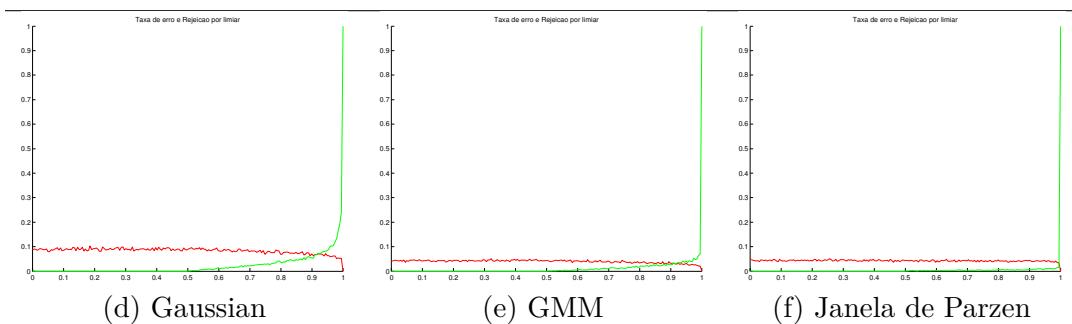
### 1.3 Avaliação de um classificador com Opção de Rejeição

Algumas das métricas importantes ao utilizarmos um classificador com opção de rejeição são taxa de erro e taxa de acurácia em função do limiar de rejeição. A análise destas duas métricas permite identificar qual o grau de confiança ao realizar uma classificação, bem como evitar erros ao rejeitar amostras duvidosas. Na figura 2, são apresentadas as curvas obtidas ao variar o valor desse limiar para as quatro bases em análise neste trabalho.

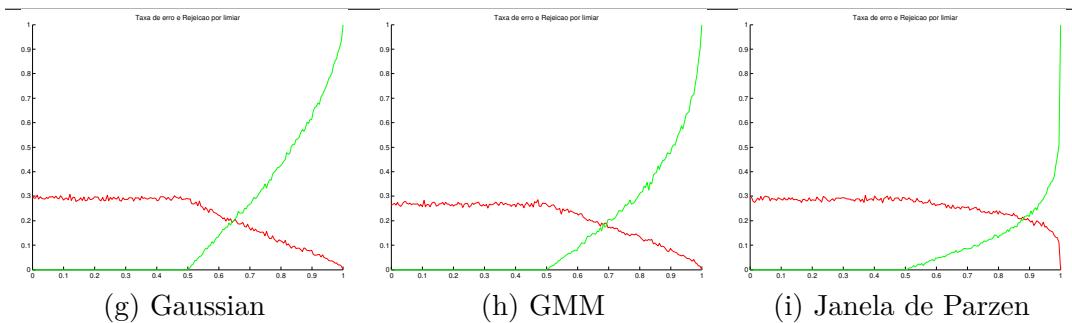
### Coluna Vertebral



### Câncer de mama



### Diabetes



### Haberman

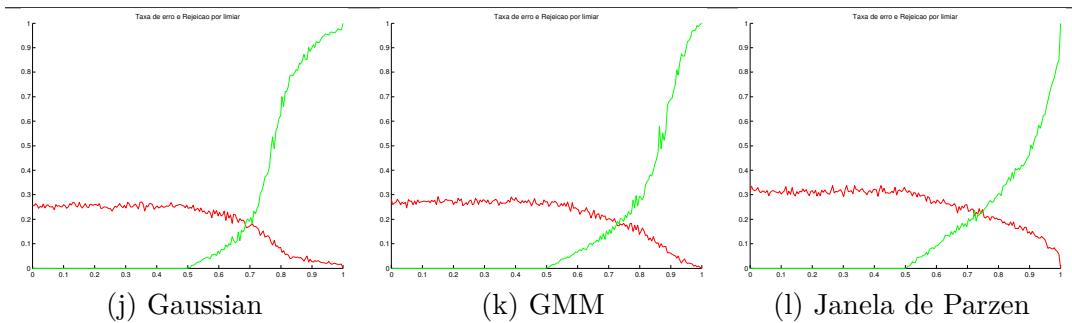


Figura 2 – Taxa de Erro e Taxa de Rejeição em função do limiar de rejeição utilizando a função gaussiana como PDF. Em verde a taxa de rejeição em vermelho a taxa de erro.

Na figura 2, os limiares de rejeição  $\beta$  variam de 0 a 1. Podemos perceber que a taxa de rejeição aumenta somente a partir de 0,5. Isso ocorre devido a tomada de decisão ser feita apenas para duas classes, dessa forma, para um limiar inferior a 50% sempre haverá uma classe com uma probabilidade a portiereira maior.

O desempenho de um classificador com capacidade de rejeição pode ser descrito através de uma curva que leva em consideração a taxa de classificação (acurácia) em relação à sua taxa de rejeição. Esta representação é denominada curva A-R (Accuracy-Reject Curve), em que cada valor correspondente a uma taxa de erro e a uma taxa de rejeição que depende do custo de rejeição  $W_r$  (SINPATCO..., 2011). Esses valores são obtidos à partir da minimização do erro empírico que é calculado através da soma ponderada do erro e da taxa de rejeição obtidos para um valor de limiar de rejeição. Temos o erro empírico definido a seguir.

$$E_{empírico}(\beta) = E(\beta) + W_r R(\beta) \quad (7)$$

Dessa forma podemos obter o erro empírico para cada uma das bases, figura 3 em função do valor de  $\beta$  e de  $W_r$ . Na figura são exibidos apenas os  $W_r$  que geram um par  $(E(\beta), R(\beta))$  distintos. Esse par ordenado é utilizado para gerar os gráficos da curva A-R exibido na figura 4.

Os mínimos de cada  $W_r$  são exibidos na tabela 1

### 1.3.1 Impacto na região de decisão

Ao optarmos por um classificador com opção de rejeição criamos uma nova classe que acolherá os dados que não puderam ser discriminados para alguma das outras duas classes, dessa forma, a classe de rejeição também define uma região de decisão. Essa região tem sua área controlada pelo valor do limiar de rejeição e sempre inicializa na região de interseção das outras duas classes. Nas figuras 5,6 e 7 são exibidas as regiões de decisões controladas com o limiar de rejeição utilizando 3 diferentes funções de densidade de probabilidade.

### 1.3.2 Busca por tamanho de janela ótima de Parzen

Abaixo, figura ?? são exibidos os resultados obtidos ao realizar a busca pelo tamanho da janela ótima de parzen para cada base de dados. Foram considerados todos os atributos disponíveis em cada base, estes foram normalizados e codificados apropriadamente assim como dito na sessão ??.

A linha ao centro de cada gráfico é a acurácia as outras duas são os limites. Os limites são determinados a uma distância de um  $\sigma^2$  (desvio padrão) acima e abaixo da acurácia média. Pode-se perceber que indiferentemente da base de dados o aumento considerável da janela de parzen causa uma redução da acurácia do classificador. Isso é causado por uma maior interferência dos vizinhos no cálculo da probabilidade condicional o que faz com que seja levado em conta apenas as probabilidades a priori para a determinação da classe para

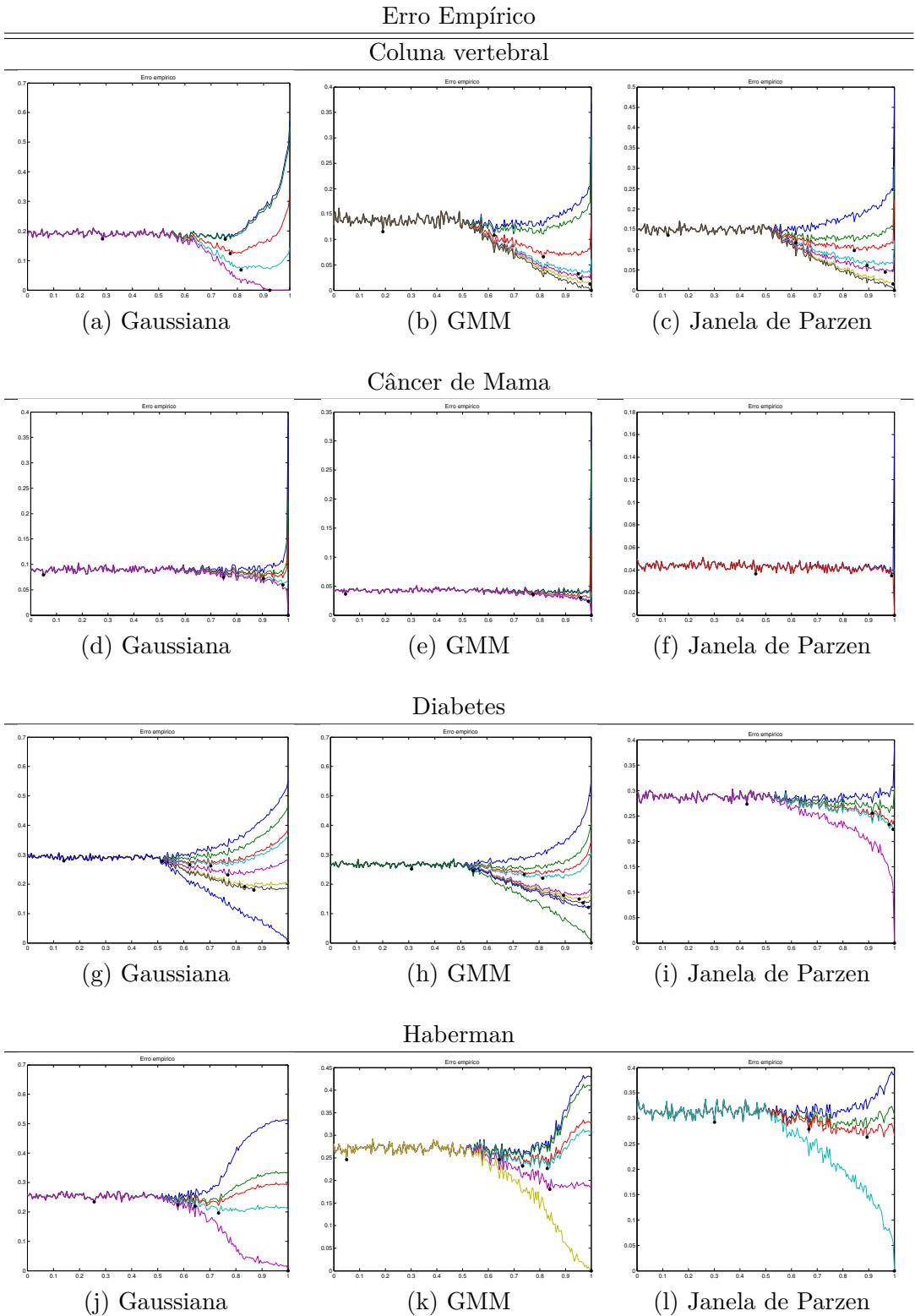
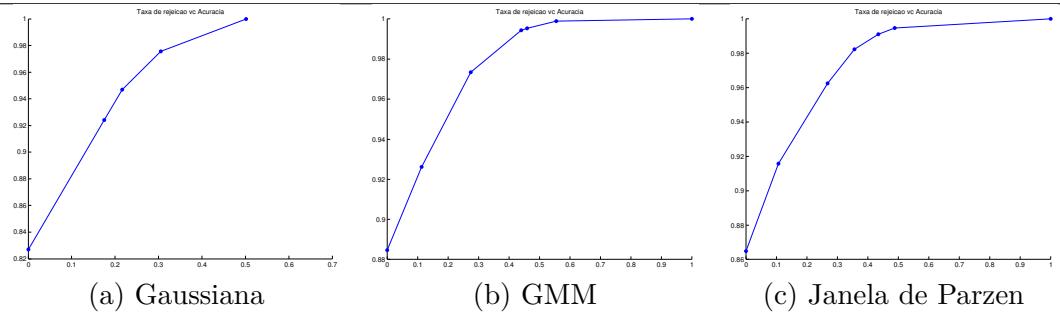


Figura 3 – Cada curva em cada um dos gráficos reflete o erro empírico em função de algum  $W_r$ . O valor de cada  $W_r$  utilizado para gerar as curvas podem ser visualizado com mais detalhes na tabela 1.

### Erro Empírico

#### Coluna vertebral

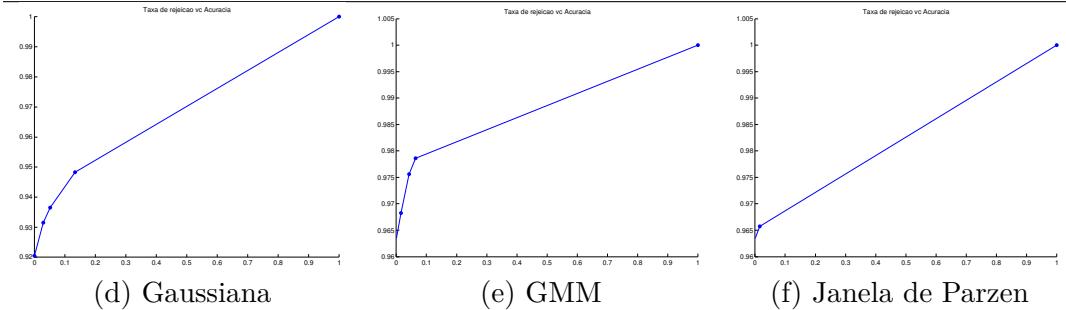


(a) Gaussiana

(b) GMM

(c) Janela de Parzen

#### Câncer de Mama

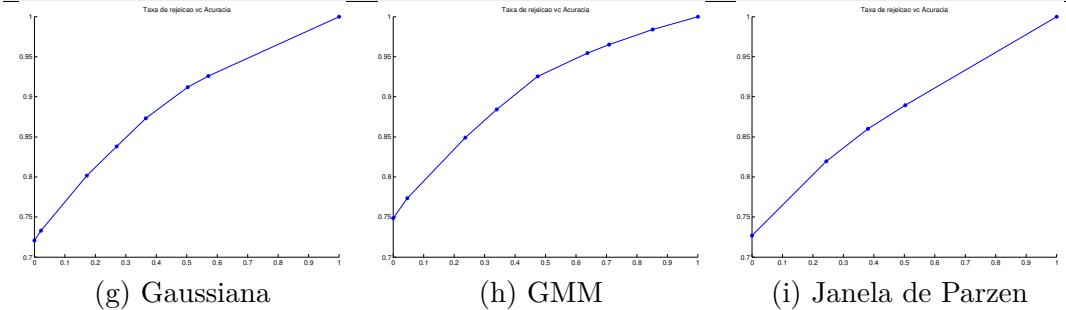


(d) Gaussiana

(e) GMM

(f) Janela de Parzen

#### Diabetes

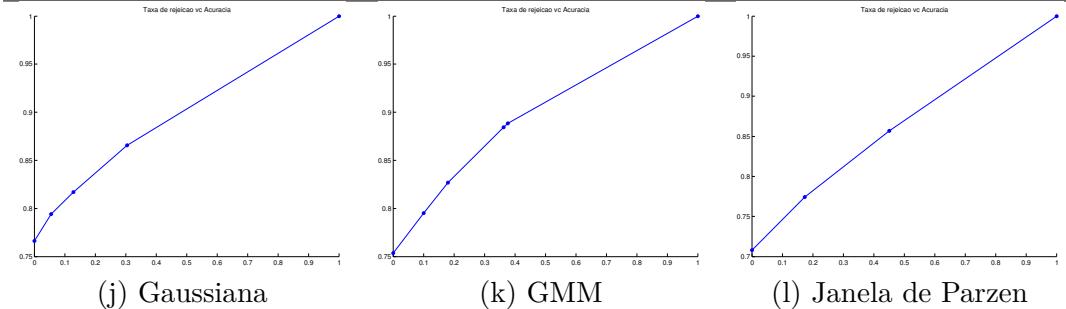


(g) Gaussiana

(h) GMM

(i) Janela de Parzen

#### Haberman



(j) Gaussiana

(k) GMM

(l) Janela de Parzen

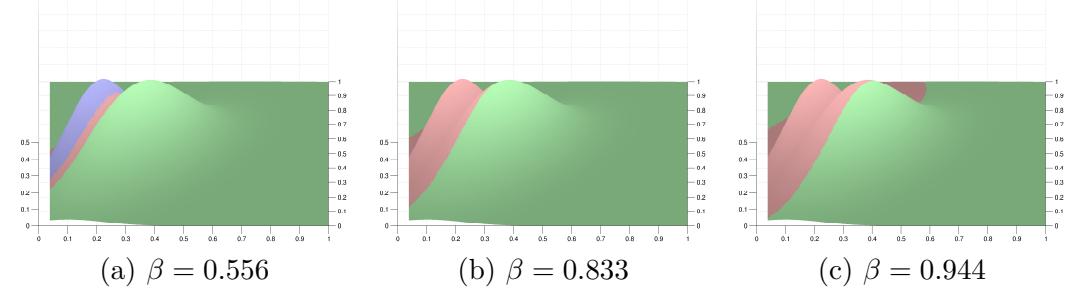
Figura 4 – Cada curva em cada um dos gráficos reflete o erro empírico em função de algum  $W_r$ . O valor de cada  $W_r$  utilizado para gerar as curvas podem ser visualizado com mais detalhes na tabela 1.

### PDF Gaussiana com RejOption

---

#### Coluna vertebral

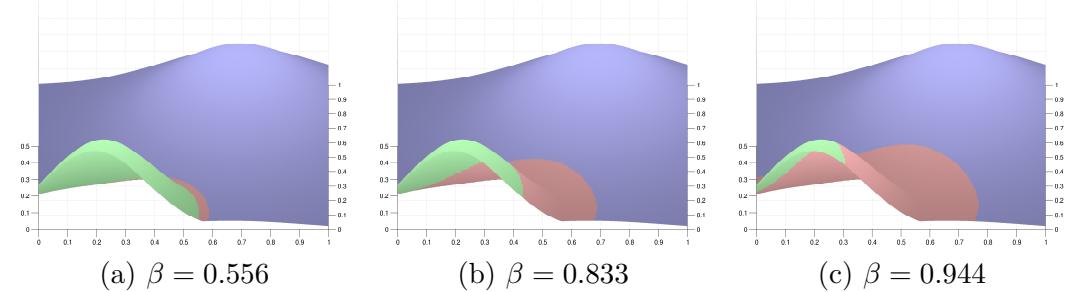
---




---

#### Câncer de Mama

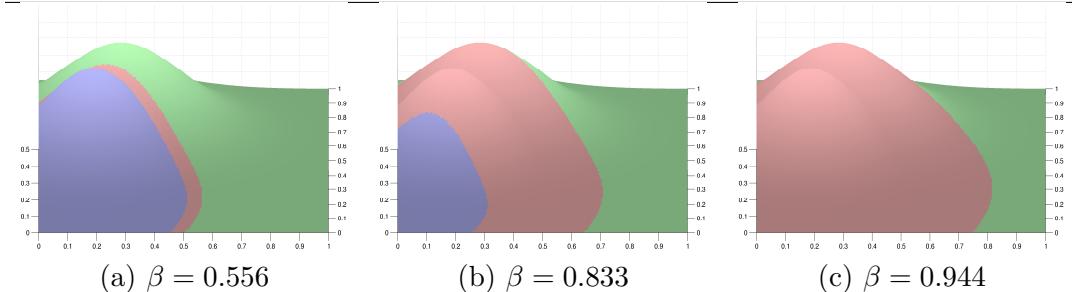
---




---

#### Diabetes

---




---

#### Haberman

---

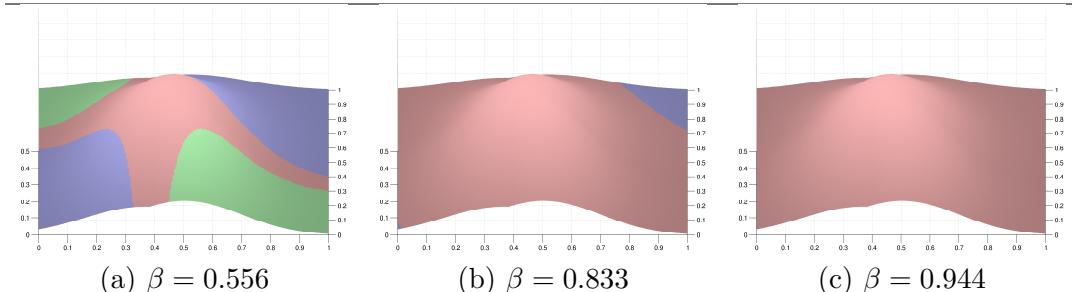


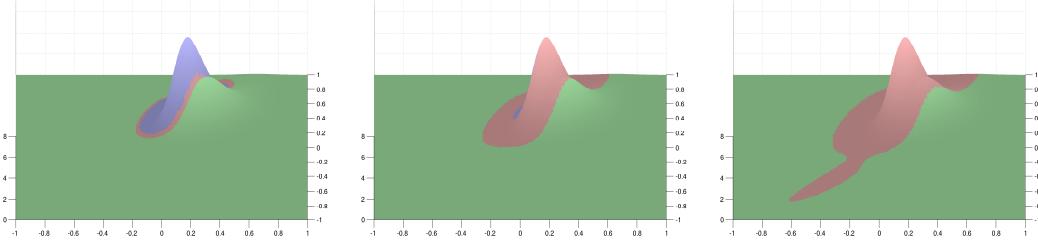
Figura 5 – Região de decisão formada utilizando três diferentes funções de densidade de probabilidade. Foram testados três valores de limiar de rejeição  $\beta$  para as três bases.

### PDF GMM com RejOption

---

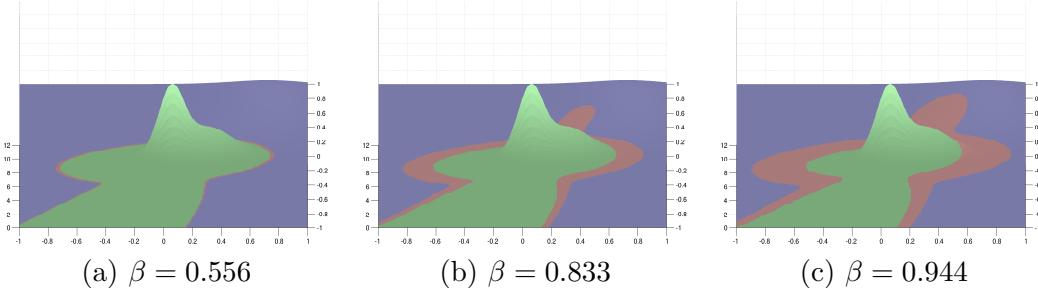
#### Coluna vertebral

---



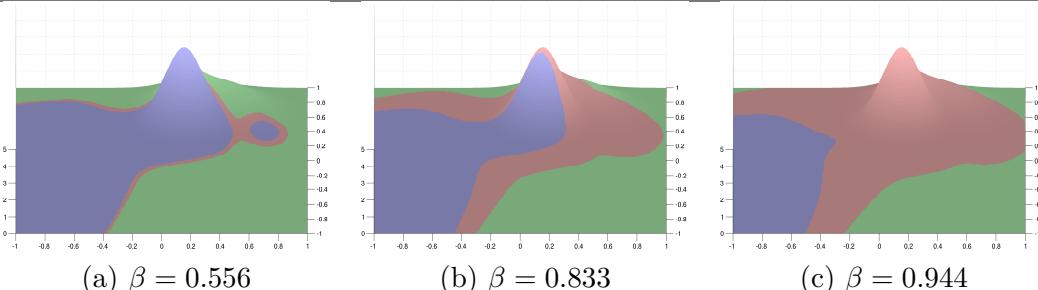
#### Câncer de Mama

---



#### Diabetes

---



#### Haberman

---

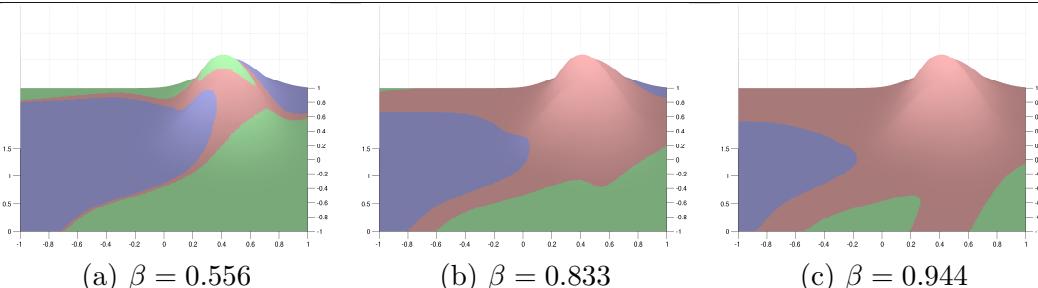
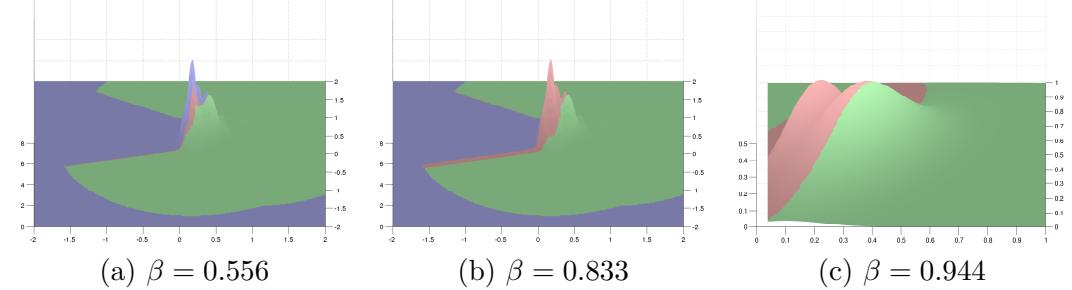


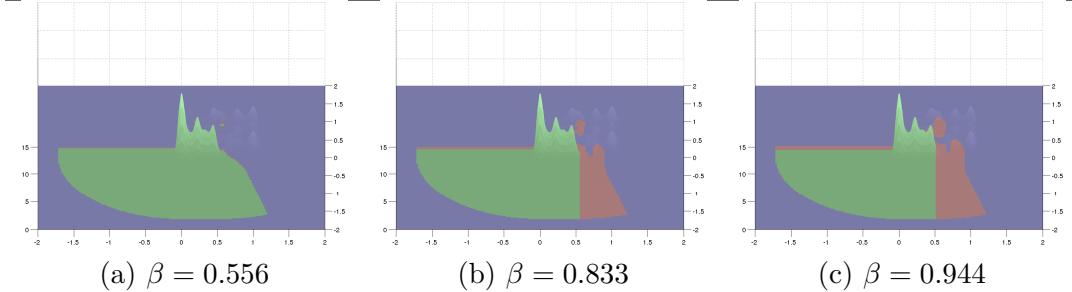
Figura 6 – Região de decisão formada utilizando mistura de gaussianas como função de densidade de probabilidade. Para todos os exemplos foram utilizadas 3 gaussianas para modelar cada classe. Foram testados 3 valores de limiar de rejeição  $\beta$  para as três bases.

PDF Janela de Parzen com RejOption

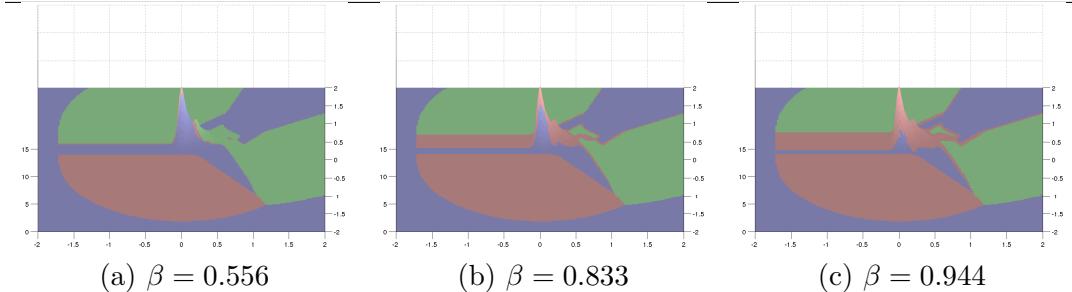
Coluna vertebral



Câncer de Mama



Diabetes



Haberman

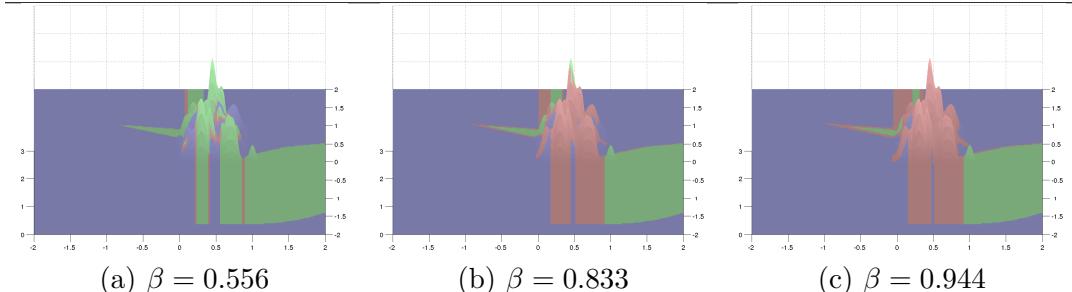


Figura 7 – Região de decisão formada utilizando janela de Parzen como função de densidade de probabilidade. Foram testados três valores de limiar de rejeição  $\beta$  para as três bases.

Erro Empírico			
Coluna vertebral			
$W_r$	$\beta_o$	$A(\beta_o)$	$R(\beta_o)$
0.571	0.286	0.827	0
0.551	0.754	0.924	0.175
0.327	0.774	0.947	0.217
0.143	0.814	0.976	0.305
0	0.925	1	0.502
0.367	0.191	0.885	0
0.306	0.623	0.926	0.113
0.143	0.814	0.974	0.274
0.0612	0.95	0.994	0.44
0.0408	0.96	0.995	0.459
0.0204	0.995	0.999	0.556
0	1	1	1
(a) Gaussiana			
(b) GMM			
(c) Janela de Parzen			
Câncer de Mama			
$W_r$	$\beta_o$	$A(\beta_o)$	$R(\beta_o)$
0.388	0.0503	0.92	0
0.224	0.749	0.931	0.0297
0.163	0.905	0.937	0.0524
0.0612	0.98	0.948	0.134
0	1	1	1
0.327	0.0452	0.963	0
0.286	0.774	0.968	0.0156
0.143	0.96	0.976	0.0423
0.0408	0.99	0.979	0.0639
0	1	1	1
(d) Gaussiana			
(e) GMM			
(f) Janela de Parzen			
Diabetes			
$W_r$	$\beta_o$	$A(\beta_o)$	$R(\beta_o)$
0.551	0.141	0.721	0
0.469	0.518	0.733	0.0219
0.388	0.623	0.802	0.172
0.367	0.704	0.838	0.27
0.286	0.769	0.873	0.366
0.204	0.834	0.912	0.504
0.184	0.869	0.926	0.572
0	1	1	1
0.551	0.312	0.749	0
0.408	0.548	0.774	0.0462
0.347	0.744	0.849	0.237
0.306	0.814	0.884	0.34
0.184	0.894	0.925	0.474
0.163	0.955	0.955	0.638
0.143	0.97	0.965	0.709
0.122	0.99	0.984	0.852
0	1	1	1
(g) Gaussiana			
(h) GMM			
(i) Janela de Parzen			
Haberman			
$W_r$	$\beta_o$	$A(\beta_o)$	$R(\beta_o)$
0.51	0.256	0.766	0
0.327	0.578	0.794	0.0553
0.286	0.643	0.817	0.128
0.204	0.734	0.866	0.305
0	1	1	1
0.429	0.0503	0.754	0
0.408	0.643	0.795	0.101
0.327	0.734	0.827	0.181
0.306	0.829	0.884	0.363
0.184	0.839	0.888	0.377
0	1	1	1
(j) Gaussiana			
(k) GMM			
(l) Janela de Parzen			

Tabela 1 – Cada tabela exibe a acurácia e a taxa de rejeição ótimos encontrado para algum limiar  $\beta_o$  em função de  $W_r$  para cada uma das bases utilizando três diferentes PDFs.

um dado  $x$ . Isso pode ser observado quando a variância, largura da janela de parzen, é maior que 2 na base da íris ou maior que 0.7 na base da coluna vertebral.

Os pontos máximos destas curvas nos indicam o valor ótimo da janela de parzen. Desta forma temos na tabela ?? o valor de acurácia e tamanho da janela para as três bases.

## 2 Segmentação de imagem utilizando o classificador de bayes

Abaixo é apresentado o resultado da segmentação de uma imagem utilizando o classificador Bayessiano, desta vez utilizando janela de parzen como função de densidade probabilidade. Foram realizados três testes com três variancias da janela, valor de  $h$ , distintas. Os valores testados foram,

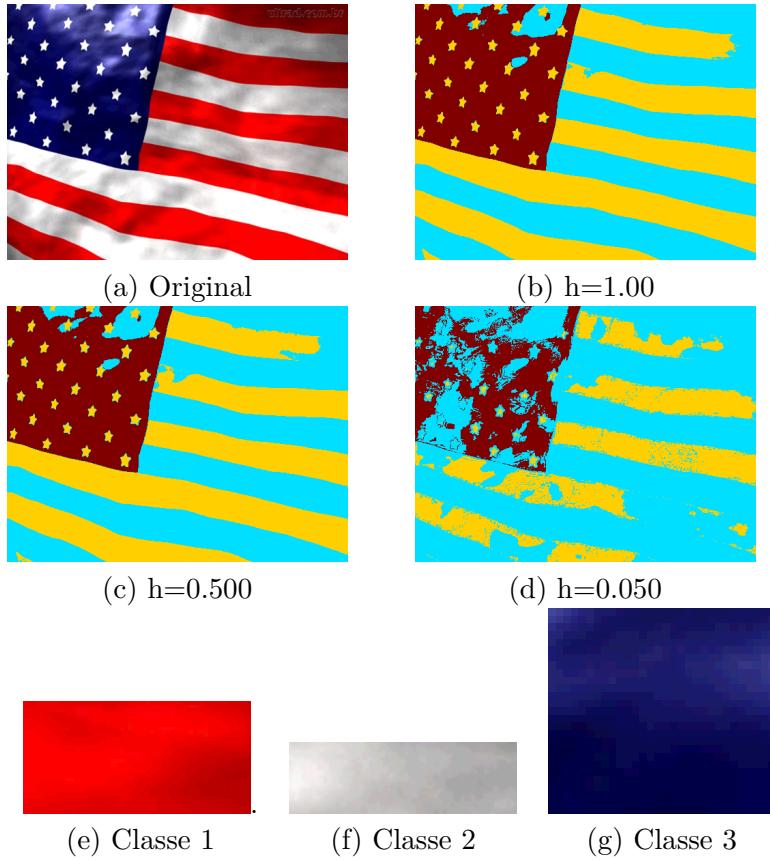


Figura 8 – Resultado da segmentação utilizando o classificador Bayessiano com janela de parzen. Com o aumento do valor de  $h$  pode-se notar que a segmentação se torna mais suave.

0.05, 0.5 e 1. Com o aumento do valor de  $h$  houve uma maior suavização.

A segmentação de uma imagem é realizada levando em consideração a intensidade dos pixels nas três componentes RGB. A definição dos dados de uma classe é feita a partir da seleção de  $k$  regiões da imagem, não necessariamente de mesma área  $A_i$ . A área é utilizada para calcular a probabilidade a priori por isso classes com uma diferença de área muito grande podem impactar na classificação. Assim temos:

$$P(w_i) = \frac{A_i}{\sum_{i=1}^k A_j} \quad (8)$$

### Considerações finais

Neste trabalho foram apresentados os resultados obtidos ao testar várias funções de densidade de probabilidade junto ao classificador bayessiano. Os resultados podem ser visualizados na tabela ???. Podemos observar uma

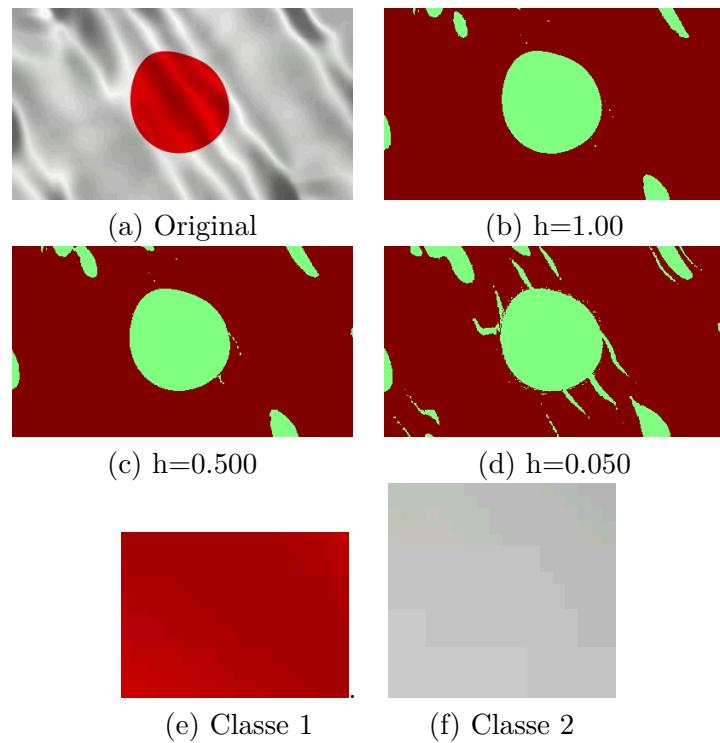


Figura 9 – Resultado da segmentação utilizando o classificador Bayessiano com janela de parzen. Com o aumento do valor de  $h$  pode-se notar que a segmentação se torna mais suave.

superioridade dos métodos não lineares frente aos lineares, bem como uma maior acurácia para métodos que consideram melhor a modelagem dos dados, quando se utiliza matrizes de covariância diferentes para cada classe de problema.

## Referências

PATTERN Classification. [S.l.]: Wiley, 2001. Citado na página [1](#).

SINPATCO II: NOVAS ESTRATÉGIAS DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE PATOLOGIAS DA COLUNA VERTEBRAL. [S.l.]: UNIVERSIDADE FEDERAL DO CEARÁ, 2011. Citado na página [6](#).