



Instituto Federal de Educação, Ciência e Tecnologia do Ceará  
Programa de Pós-Graduação em Ciência da Computação  
Departamento de Telemática

Relatório de experimento

Trabalhos 2 e 3

Classificador Bayesiano Gaussiano Puro e  
com discriminantes lineares e quadráticos

*Alan Rabelo Martins*

2019

## 1 Introdução

A função densidade de probabilidade Gaussiana é completamente especificada por dois parâmetros: média  $\mu$  e variância  $\sigma$ , enquanto  $\sigma^2$  é denominado desvio padrão.

## 2 Metodologia

Os algoritmos foram desenvolvidos em Python 3.6 utilizando a IDE Pycharm e apenas a biblioteca numpy e a matplotlib. Para avaliação do classificador Bayesiano foram utilizados 5 folds com 30 rodadas de teste e computadas a acurácia e o desvio padrão.

Os resultados da seção 3 se referem o Classificador Bayesiano Puro enquanto que a seção 4 trata dos Classificadores Bayesianos com discriminante linear e quadráticos utilizando uma matriz de covariância para cada classe e comparando os resultados com a matriz de covariância igual a média das matrizes das classes.

## 3 Resultados do Classificador Puro

Após aplicar a metodologia nos datasets podemos verificar o resultado do experimento em termos de acurácia e desvio padrão exibidos nas seções abaixo. Para cálculo de acurácia dos algoritmos utilizados na comparação, tanto para o DMC quanto para o KNN foram utilizadas 25 realizações. A avaliação se dá pelo valor de acurácia dos modelos e seu respectivo desvio padrão. Tanto a matriz de confusão quanto a superfície de decisão foram baseadas na primeira realização pois, tendo em vista que o desvio padrão se manteve muito estável, não há diferença visível entre as realizações.

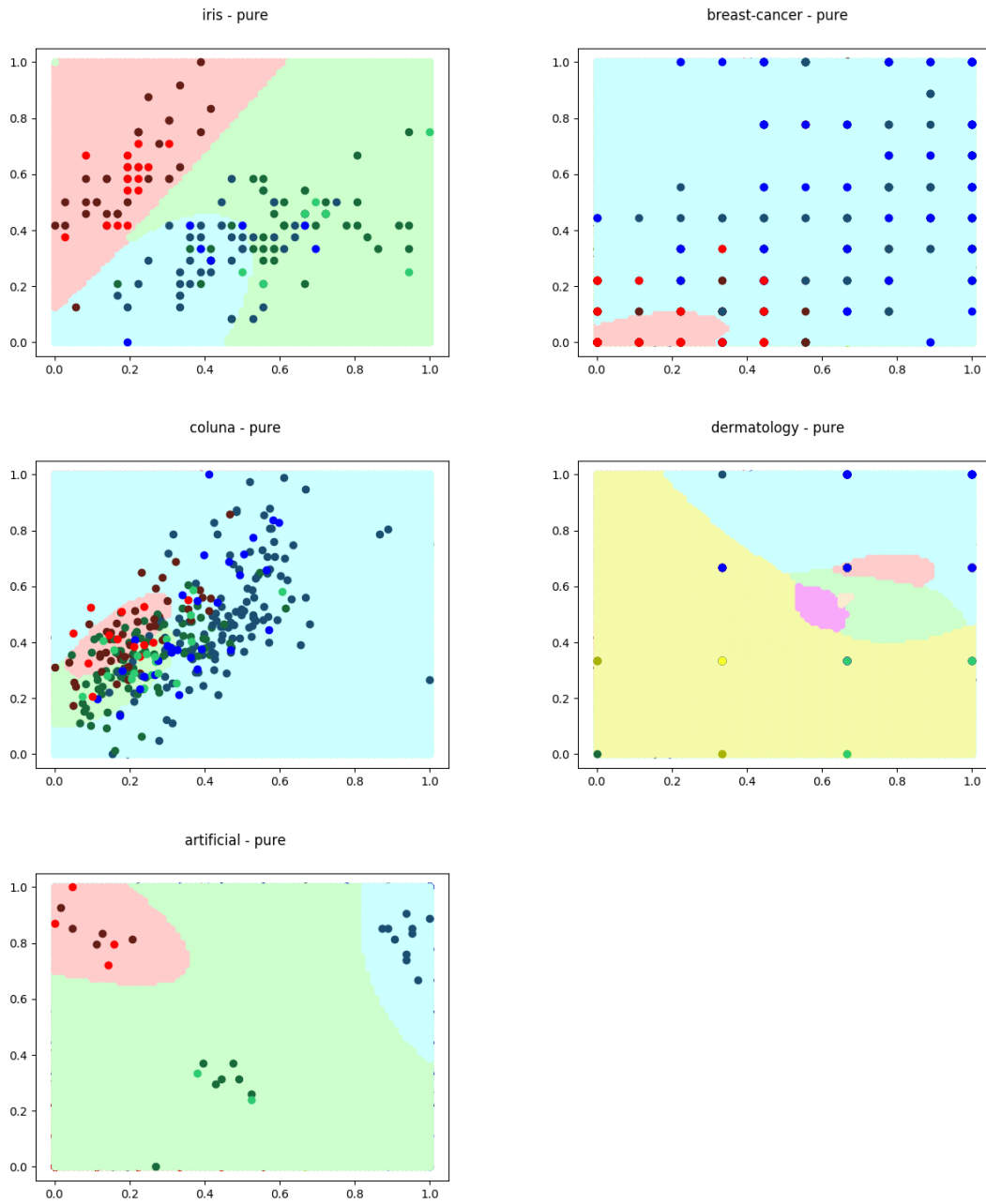


Tabela 1 – Superfícies de decisão. Da esquerda para a direita, de cima para baixo: Íris, Câncer de Mama, Coluna vertebral, Dermatologia

Dataset	Acurácia	Desvio Padrão
Iris	89.22%	0.05
Câncer	75.78%	0.03
Coluna	61.84%	0.14
Dermatology	46.27%	0.05
Artificial	92.22%	0.1

Tabela 2 – Acurácia e Desvio padrão dos testes

Iris	0	1	2
0	14	1	0
1	0	4	0
2	0	0	11

Câncer de Mama	True	False
True	3	0
False	0	5

Coluna Vertebral	0	1	2
0	11	1	2
1	6	10	14
2	4	0	13

Dermatology	0	1	2	3	4
0	1	9	1	0	0
1	0	23	0	0	0
2	0	11	2	0	0
3	0	7	0	7	0
4	0	3	0	0	4

Artificial	0	1	2
0	3	0	0
1	0	5	0
2	1	0	4

Para efeito de comparação no dataset da Íris os valores de acurácia para o K-NN ficaram em 93%, enquanto que para o DMC ficaram em 94.8% com um desvio padrão de 0,11 para o K-NN e 0.34 para o DMC. Já no dataset da Coluna vertebral, os valores de acurácia para o K-NN ficaram em 90%, enquanto que para o DMC ficaram em 76.8% com um desvio padrão de 0,11 para o K-NN e 0.7 para o DMC.

## 4 Resultados e comparação dos classificadores lineares e quadráticos

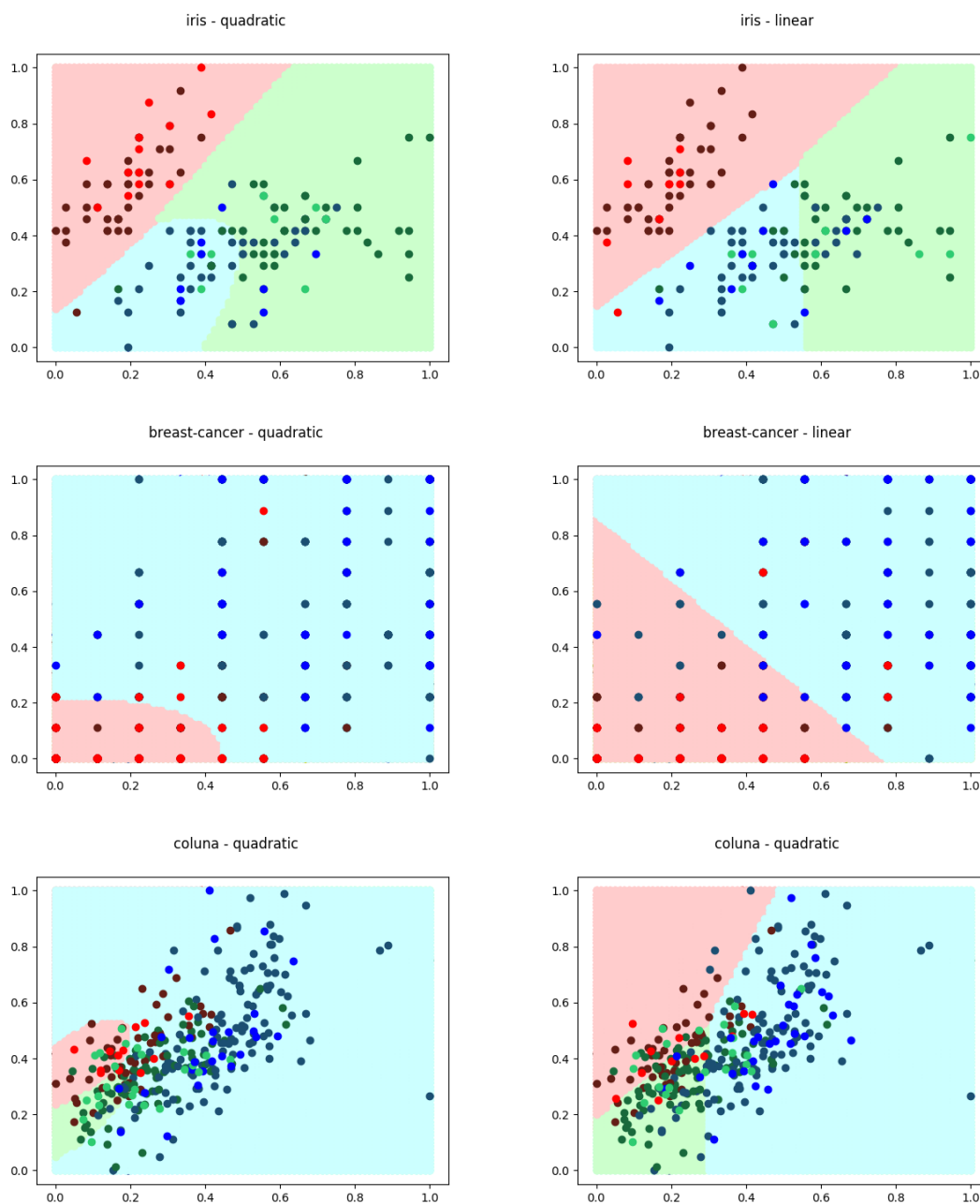


Tabela 3 – Superfícies de decisão. De cima para baixo: Íris, Câncer de Mama, Coluna vertebral

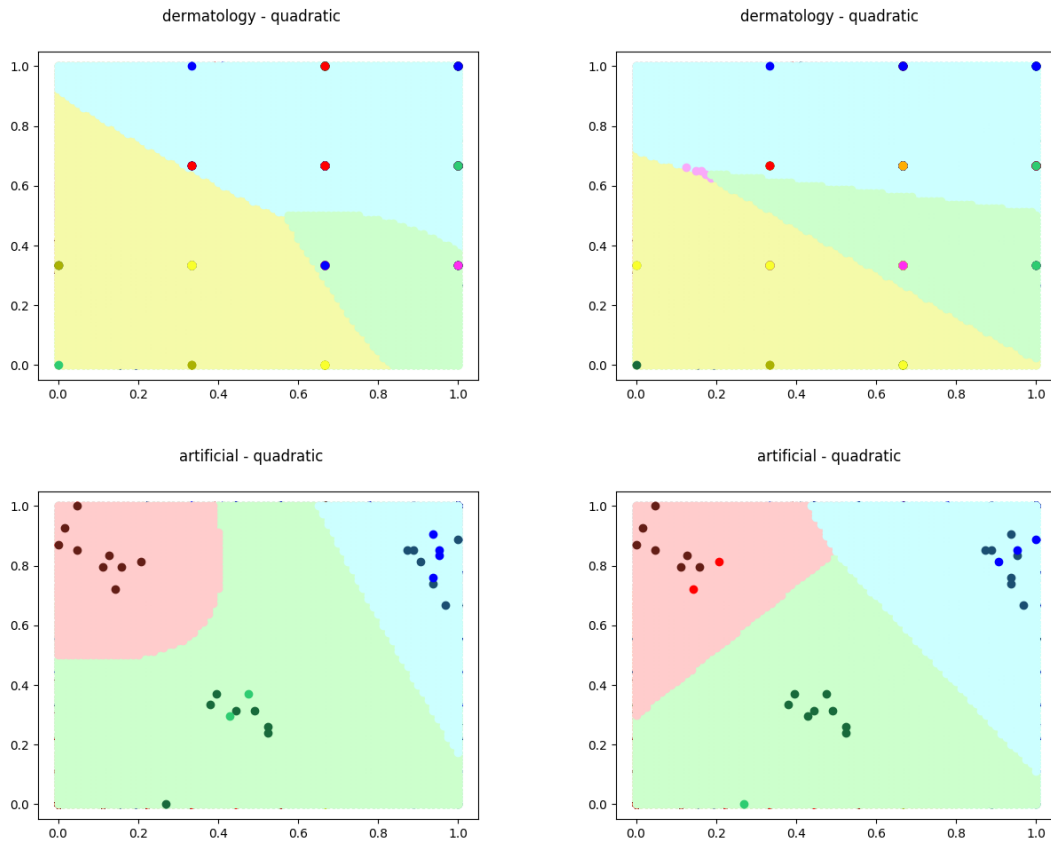


Tabela 4 – Superfícies de decisão. De cima para baixo: Dermatologia e Base de dados artificial

Dataset	Acurácia Linear	Desvio Padrão Linear
Acurácia Quadrática	Desvio Padrão Quadrático	
Iris	72.67%	0.13
96.22%	0.03	
Câncer	82.46%	0.05
47.91%	0.05	
Coluna	33.71%	0.13
59.52%	0.11	
Dermatology	46.27%	0.05
33.24%	0.07	
Artificial	60.56%	0.24
96.11%	0.07	

Tabela 5 – Acurácia e Desvio padrão dos testes

Iris	0	1	2
0	14	1	0
1	0	4	0
2	0	0	11

Câncer de Mama	True	False
True	3	0
False	0	5

Coluna Vertebral	0	1	2
0	11	1	2
1	6	10	14
2	4	0	13

Dermatology	0	1	2	3	4
0	1	9	1	0	0
1	0	23	0	0	0
2	0	11	2	0	0
3	0	7	0	7	0
4	0	3	0	0	4

Artificial	0	1	2
0	3	0	0
1	0	5	0
2	1	0	4

Para efeito de comparação no dataset da Íris os valores de acurácia para o K-NN ficaram em 93%, enquanto que para o DMC ficaram em 94.8% com um desvio padrão de 0,11 para o K-NN e 0.34 para o DMC. Já no dataset da Coluna vertebral, os valores de acurácia para o K-NN ficaram em 90%, enquanto que para o DMC ficaram em 76.8% com um desvio padrão de 0,11 para o K-NN e 0.7 para o DMC.

## 5 Conclusão

A partir dos resultados podemos constatar que o algoritmo gera bons resultados apesar da acurácia mais baixa encontrada nos dados da coluna vertebral e do *dermatology*, mostrando ser um algoritmo muito eficiente, tendo em vista a velocidade de treinamento.

Iris Dataset	Acurácia	Desvio Padrão
KNN	93%	0.11
DMC	94.8%	0.34
Bayesiano	97.33%	0.03

Tabela 6 – Comparação entre Acurácia e Desvio Padrão entre diferentes algoritmos desenvolvidos aplicando o dataset da íris

---

Coluna Vertebral	Acurácia	Desvio Padrão
KNN	90%	0.11
DMC	76.8%	0.07
Bayesiano	79.84%	0.21

Tabela 7 – Comparação entre Acurácia e Desvio Padrão entre diferentes algoritmos desenvolvidos aplicando o dataset da coluna vertebral