



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

Relatório de experimento

Trabalho 6
Classificador Bayesiano com Mistura de Gaussianas

Alan Rabelo Martins

2019

1 Introdução

Este relatório descreve a implementação e testes do algoritmo de Mistura de Gaussianas para estimação semi-paramétrica. A seção 2 descreve a metodologia de treino, teste e execução. A seção 3 exibe os resultados e a seção 4 a comparação com algoritmos anteriores e conclusão.

2 Metodologia

O algoritmo foi desenvolvido em Python 3.6 utilizando a IDE Pycharm e apenas a biblioteca numpy e a matplotlib. Para avaliação do classificador com Mistura de Gaussianas foram utilizados 5 folds com 20 rodadas de teste e computadas a acurácia e o desvio padrão. Não foi utilizado Cross-validation porque como temos os dados de classe de cada entrada o valor de k (número de classes) foi configurado baseado nos dados. Para execução dos algoritmos foi utilizado um iMac com processador core i3 8100 com 3.6Ghz e 16Gb de memória RAM.

3 Resultados do Classificador

Após aplicar a metodologia nos *datasets* podemos verificar o resultado do experimento em termos de acurácia e desvio padrão exibidos nas seções abaixo. 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.

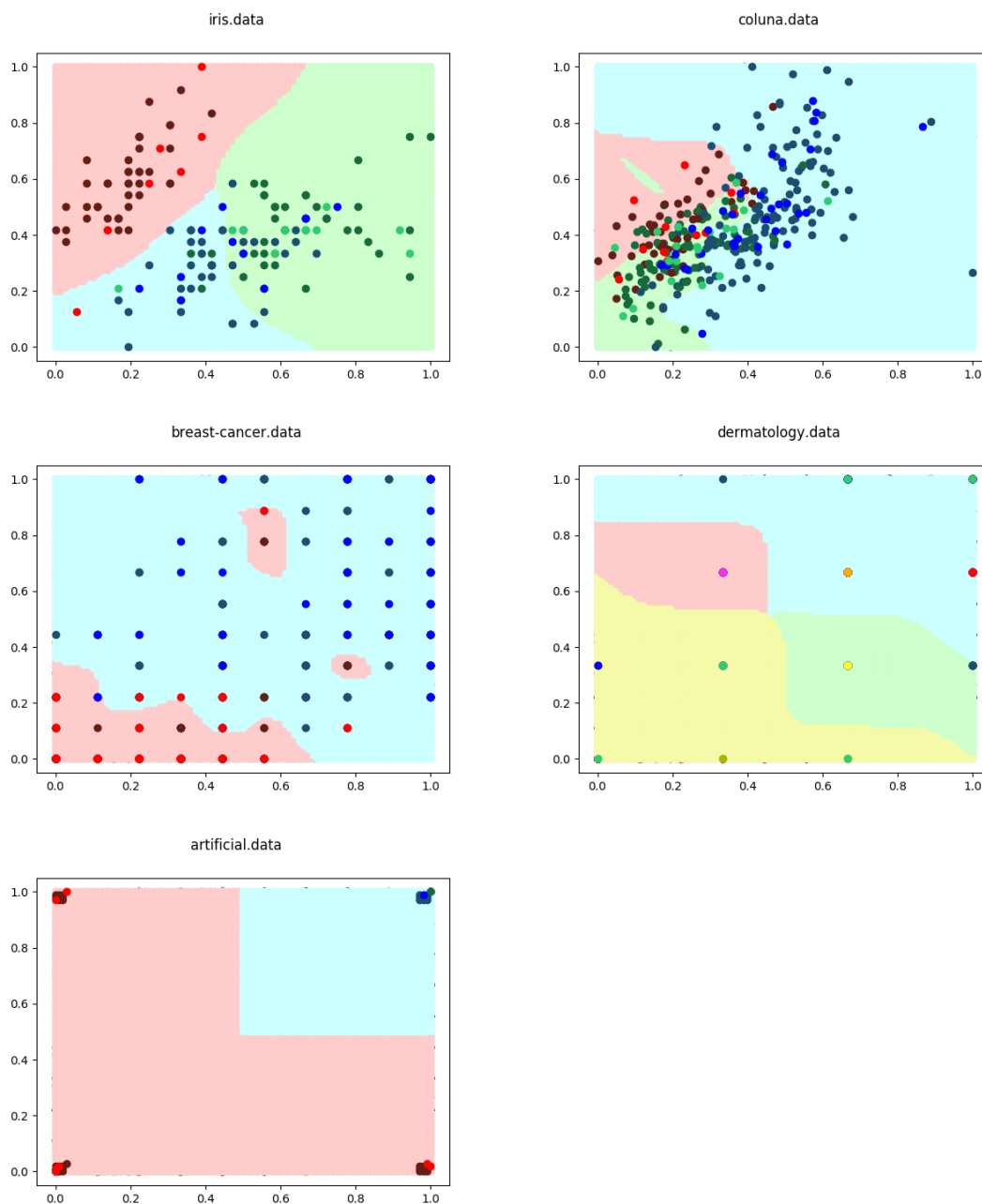


Tabela 1 – Superfícies de decisão. Da esquerda para a direita, de cima para baixo: Íris, Coluna vertebral, Câncer de Mama, Dermatologia e Artificial representando o problema do AND lógico

Dataset	Acurácia	Desvio Padrão
Iris	93.67%	0.05
Câncer	96.25%	0.02
Coluna	65.32%	0.07
Dermatology	86.74%	0.04
Artificial	85.50%	0.03

Tabela 2 – Acurácia e Desvio padrão dos testes

Iris	0	1	2
0	7	0	0
1	0	9	0
2	0	0	14

Câncer de Mama	True	False
True	56	4
False	0	77

Coluna Vertebral	0	1	2
0	6	0	6
1	10	15	5
2	2	2	16

Dermatology	0	1	2	3	4
0	12	0	0	0	3
1	0	25	0	2	0
2	3	0	14	0	0
3	0	0	0	4	0
4	0	0	0	1	5

Artificial	0	1
0	7	3
1	2	8

4 Conclusão

A partir dos resultados podemos constatar que o algoritmo gera bons resultados apesar da acurácia mais baixa encontrada nos dados da coluna vertebral se comparado aos outros algoritmos bayesianos, mostrando ser um algoritmo muito eficiente quando não temos dados de classes diferentes sobrepostos.

Iris Dataset	Acurácia	Desvio Padrão
KNN	93%	0.11
DMC	94.8%	0.34
Bayesiano	97.33%	0.03
Parzen	98.00%	0.02
Mistura de Gaussianas	93.67%	0.05

Tabela 3 – 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
Parzen	79.00%	0.03
Mistura de Gaussianas	65.32%	0.07

Tabela 4 – Comparação entre Acurácia e Desvio Padrão entre diferentes algoritmos desenvolvidos aplicando o dataset da coluna vertebral