



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 Gaussiano com
Janela de Parzen

Alan Rabelo Martins

2019

1 Introdução

Este relatório descreve a implementação e testes do algoritmo de Janelas de Parzen para estimação não-paramétrica. A seção 2 descreve a metodologia de treino, teste e execução. A seção 3 exhibe 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 janela de parzen foram utilizados 5 folds com 30 rodadas de teste e computadas a acurácia e o desvio padrão. 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 Janela de Parzen

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.

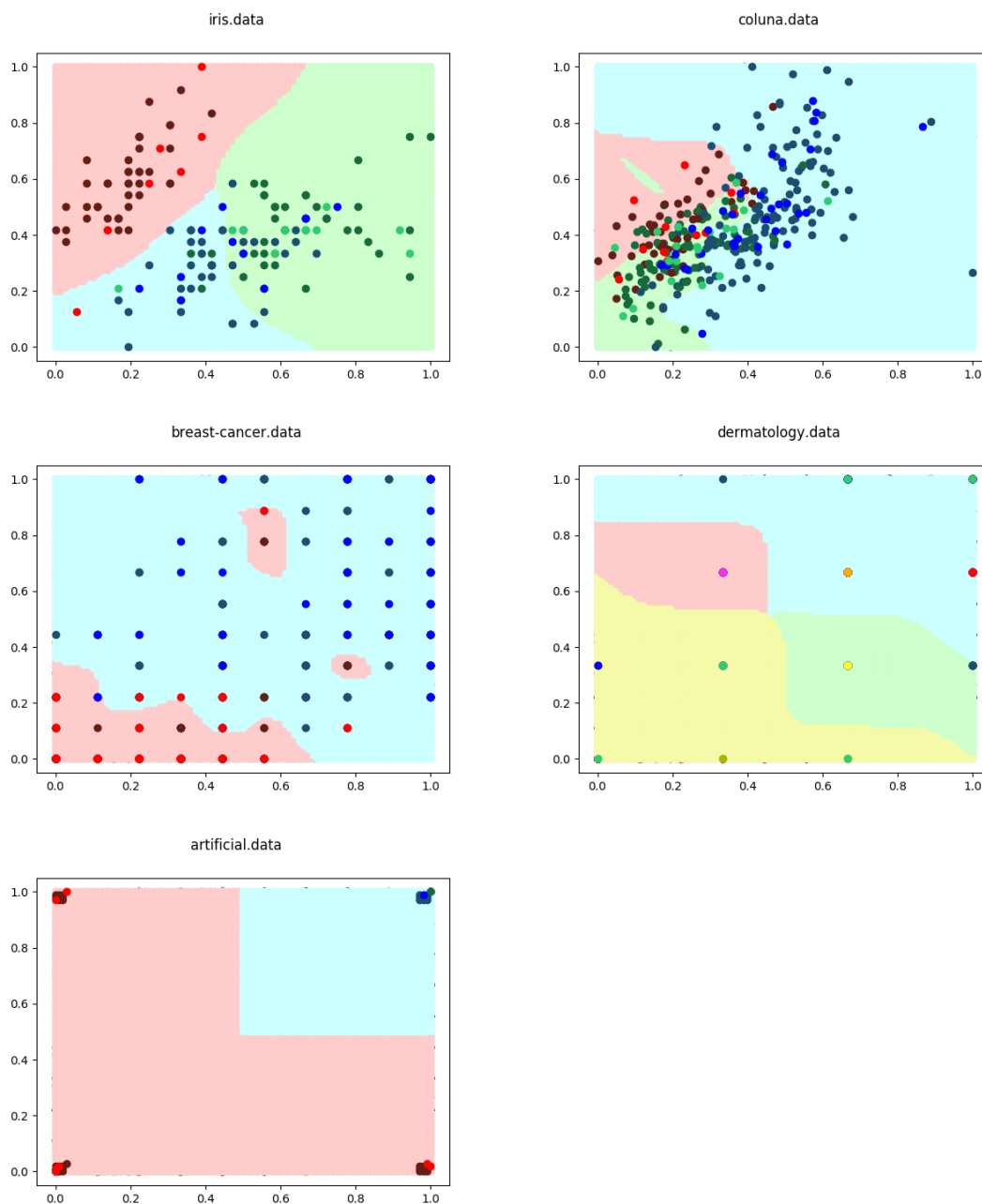


Tabela 1 – Superfícies de decisão. Da esquerda para a direita, de cima para baixo: Íris, Coluna vertebral, Câncer de Mama, Dermatologia e Artificial representando o problema do AND lógico

Dataset	Acurácia	Desvio Padrão
Iris	98.00%	0.02
Câncer	98.00%	0.01
Coluna	79.00%	0.03
Dermatology	94.00%	0.04
Artificial	95.00%	0.07

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	58	2
False	2	75

Coluna Vertebral	0	1	2
0	7	0	5
1	0	25	5
2	2	2	16

Dermatology	0	1	2	3	4
0	12	0	0	0	3
1	0	27	0	0	0
2	0	0	17	0	0
3	0	0	0	4	0
4	0	0	0	0	6

Artificial	0	1
0	5	0
1	0	3

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, 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

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

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