



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

Relatório de experimento

K-NN e DMC

Alan Rabelo Martins

2019

1 Introdução

O K-NN (K-vizinhos mais próximos) e o DMC são algoritmos de classificação conhecidos por sua implementação simples e resultados satisfatórios. Este relatório descreve sua implementação e resultados aplicados aos *datasets* IrisFlower, Problemas na coluna e Diagnóstico de câncer de mama.

2 Metodologia

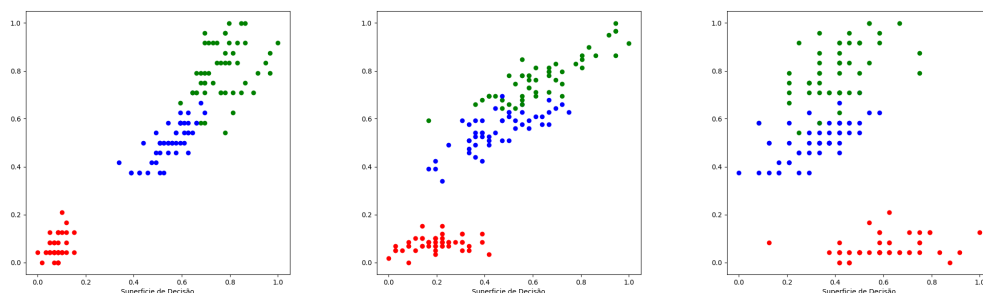
Os algoritmos foram desenvolvidos em Python 3.6 utilizando a IDE Pycharm e apenas a biblioteca numpy e a matplotlib. Para avaliação do K-nn foram utilizados 5 folds com 20 rodadas de teste para cada um dos valores de K sendo K igual aos valores ímpares no intervalo compreendido entre 1 e 15.

Já para o DMC foram utilizados os mesmos 5-folds com 20 rodadas de teste.

3 Resultados

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, 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. A matriz de confusão é baseada na primeira realização de ambos os algoritmos.

3.1 Dataset da Íris



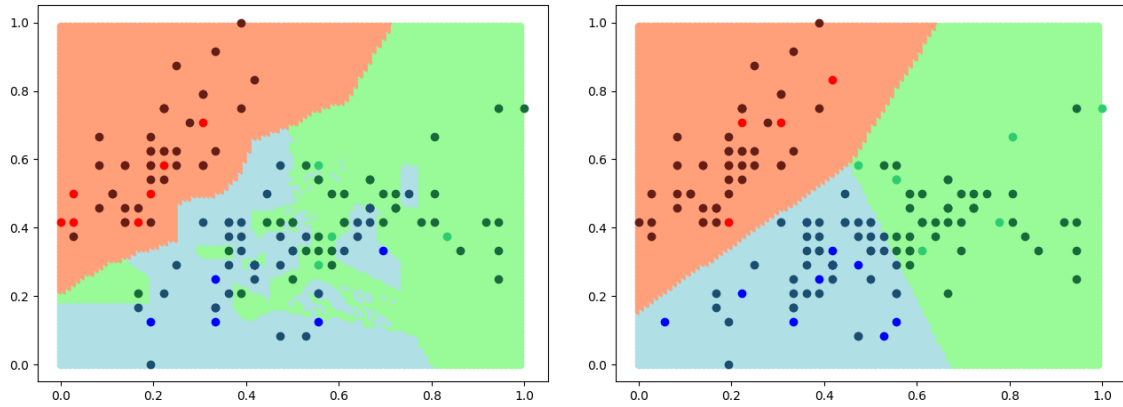
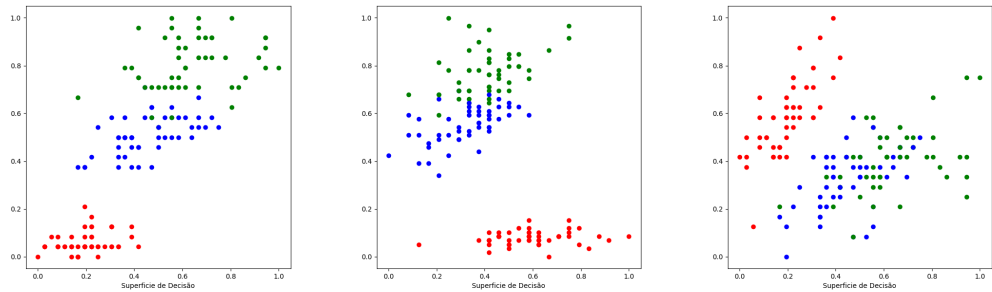


Tabela 1 – mapa de cores do dataset da iris normalizado utilizando o (a) K-NN e (b) DMC



.	Setosa	Versicolor	Virginica
Setosa	9	0	0
Versicolor	0	6	1
Virginica	0	1	13

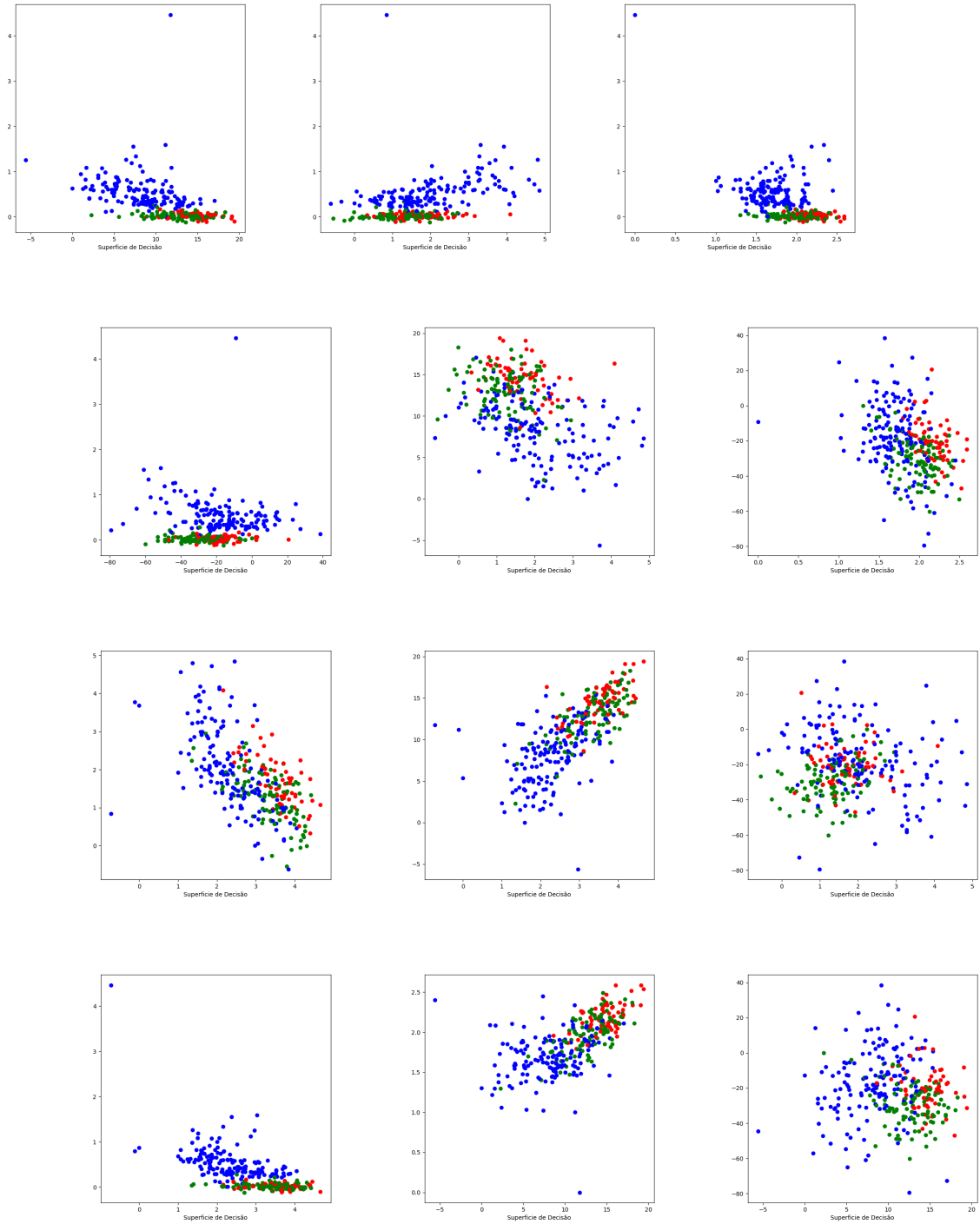
Tabela 2 – Matriz de confusão do DMC

.	Setosa	Versicolor	Virginica
Setosa	6	0	0
Versicolor	0	8	5
Virginica	0	2	6

Tabela 3 – Matriz de confusão do KNN

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.

3.2 Dataset da Coluna vertebral



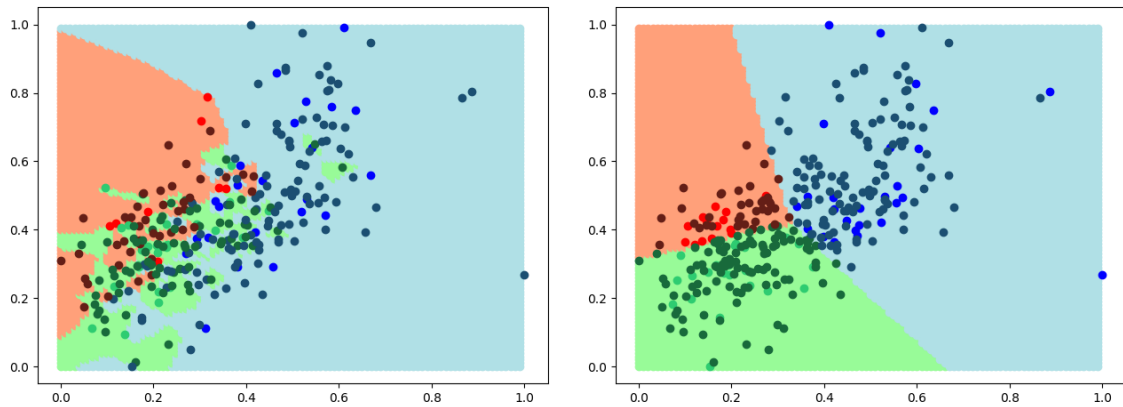
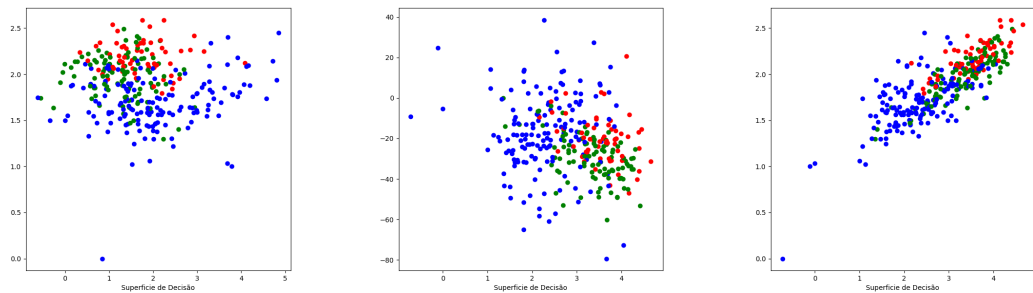


Tabela 4 – Mapa de cores do dataset normalizado da coluna vertebral utilizando o (a) K-NN e (b) DMC



.	Hernia	Spondylolisthesis	Normal
Hernia	8	1	4
Spondylolisthesis	0	23	4
Normal	3	3	16

Tabela 5 – Matriz de confusão do DMC

.	Hernia	Spondylolisthesis	Normal
Hernia	7	0	6
Spondylolisthesis	2	18	9
Normal	5	4	5

Tabela 6 – Matriz de confusão do KNN

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.

3.3 Dataset artificial

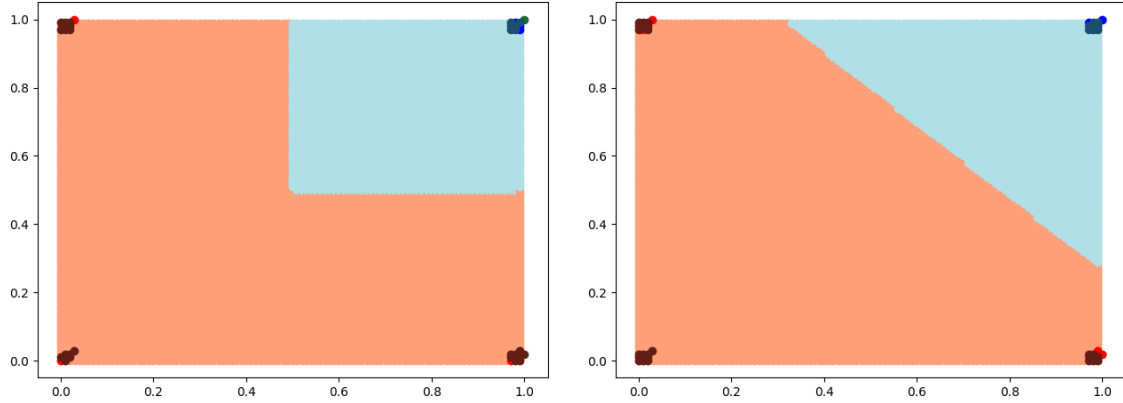


Tabela 7 – Mapa de cores do dataset artificial normalizado utilizando o (a) K-NN e (b) DMC

.	True	False
True	3	0
False	0	5

Tabela 8 – Matriz de confusão do DMC

.	True	False
True	1	0
False	0	6

Tabela 9 – Matriz de confusão do KNN

Para o dataset artificial, equivalente ao problema do AND lógico, os valores de acurácia tanto para o DMC quanto para o K-NN ficaram em 100% e desvio padrão em 0.

4 Conclusão

Concluimos que os algoritmos geram bons resultados apesar da acurácia mais baixa encontrada nos dados da coluna vertebral para o DMC, mostrando ser algoritmos muito eficientes, tendo em vista a simplicidade de implementação e velocidade de treinamento.