

# Relatório: Algoritmo KNN

Tiago Alexandre Francisco de Queiroz

O trabalho foi feito analisando os dados inseridos no arquivo iris.txt. Inicialmente utilizei o  $k = 10$ , treinamento = 120, e teste = 30; repetindo o algoritmo 10 vezes, foram encontradas algumas informações:

Teste 1 : acurácia = 28/30  
Teste 2 : acurácia = 29/30  
Teste 3 : acurácia = 29/30  
Teste 4 : acurácia = 30/30  
Teste 5 : acurácia = 30/30  
Teste 6 : acurácia = 29/30  
Teste 7 : acurácia = 30/30  
Teste 8 : acurácia = 30/30  
Teste 9 : acurácia = 29/30  
Teste 10 : acurácia = 30/30

Porcentagem de acertos: 98%  
Desvio Padrão: 0,66

Mantendo os valores de treinamento e teste, e utilizando o  $k = 50$  temos os seguintes resultados:

Teste 1 : acurácia = 26/30  
Teste 2 : acurácia = 28/30  
Teste 3 : acurácia = 29/30  
Teste 4 : acurácia = 27/30  
Teste 5 : acurácia = 28/30  
Teste 6 : acurácia = 30/30  
Teste 7 : acurácia = 28/30  
Teste 8 : acurácia = 27/30  
Teste 9 : acurácia = 28/30  
Teste 10 : acurácia = 27/30

Porcentagem de acertos: 92,67%  
Desvio Padrão: 1,08

Fazendo um último teste utilizando o  $k = 100$ , temos os seguintes resultados:

Teste 1 : acurácia = 24/30  
Teste 2 : acurácia = 14/30  
Teste 3 : acurácia = 21/30  
Teste 4 : acurácia = 11/30  
Teste 5 : acurácia = 17/30  
Teste 6 : acurácia = 9/30  
Teste 7 : acurácia = 25/30  
Teste 8 : acurácia = 9/30  
Teste 9 : acurácia = 6/30  
Teste 10 : acurácia = 18/30

Porcentagem de acertos: 51,33%

Desvio Padrão: 6.31

Portanto, observa-se que o melhor valor de  $K$  entre esses 3 é  $K = 10$ . Porque ao aumentarmos a quantidade de vizinhos vemos que a classificação se torna mais imprecisa.