

Aluno: Paulo Henrique Fernandes Leite

Professor Orientador: Willian Yukio Honda

Projeto Integrador V: Computação Científica

19 de março de 2017

1. Conjunto de dados utilizado

Neste projeto foram utilizados os seguintes conjuntos de dados do UCI Machine Learning Repository:

- Iris - composto por 3 classes e 4 atributos
- Adult - composto por 2 classes e 14 atributos
- Wine - composto por 3 classes e 13 atributos
- Breast Cancer Wisconsin - composto por 2 classes e 9 atributos
- Wine Quality - composto por 11 classes e 11 atributos
- Abalone - composto por 29 classes e 8 atributos

Os conjuntos podem ser encontrados em: <http://archive.ics.uci.edu/ml/>

2. Versão do algoritmo k-NN

Para o projeto, foram implementados quatro versões do algoritmo k-NN, sendo elas, 1-NN, $(m+2)$ -NN, $(m*10 + 1)$ -NN, onde m é igual ao total de classes do conjunto de dados em execução, se este total for ímpar ou m é este total mais 1, se for par, e por fim, $(q/2 + 1)$ -NN se $q/2$ for par ou $(q/2)$ -NN se $q/2$ for ímpar, onde q é a quantidade de instâncias (linhas) do conjunto de dados em execução.

O algoritmo k-NN basicamente é utilizado para encontrar o vizinho mais próximo de um determinado conjunto de dados, onde k remetesse a quantos vizinhos o algoritmo deve encontrar.

Para os conjuntos usados no projeto, os algoritmos ficaram da seguinte forma:

Iris.csv com 150 elementos: 1-knn, 5-knn, 31-knn e 76-knn

Tempo total para obter os vizinhos: 0.021 segundos

Wine.csv com 178 elementos: 1-knn, 5-knn, 31-knn e 90-knn

Tempo total para obter os vizinhos: 0.041 segundos

Cancer.csv com 683 elementos: 1-knn, 5-knn, 31-knn e 341-knn

Tempo total para obter os vizinhos: 0.25 segundos

Quality.csv com 1599 elementos: 1-knn, 13-knn, 111-knn e 799-knn

Tempo total para obter os vizinhos: 1.035 segundos

Abalone.csv com 4177 elementos: 1-knn, 31-knn, 291-knn e 76-knn

Tempo total para obter os vizinhos: 6.26 segundos

Adult.csv com 30162 elementos: 1-knn, 5-knn, 31-knn e 15082-knn

Tempo total para obter os vizinhos: 315.39 segundos

3. Cross-Validation e Resultados

1-knn in iris.csv data set with 150 elements:

matrix confusion: {'3 | 2': 5, '1 | 1': 50, '2 | 1': 10, '3 | 3': 45, '2 | 2': 40}

sample error: {'iris.csv p10': 0.0, 'iris.csv p8': 0.0, 'iris.csv p9': 0.0, 'iris.csv p6': 0.0,
'iris.csv p7': 5.0, 'iris.csv p4': 10.0, 'iris.csv p5': 0.0, 'iris.csv p2': 0.0, 'iris.csv p3': 0.0,
'iris.csv p1': 0.0}

cross validation error: 1.5

accuracy: 0.9

5-knn in iris.csv data set with 150 elements:

matrix confusion: {'3 | 2': 55, '3 | 3': 195, '2 | 3': 15, '1 | 1': 250, '2 | 1': 50, '2 | 2': 185

sample error: {'iris.csv p10': 15.0, 'iris.csv p8': 0.0, 'iris.csv p9': 15.0, 'iris.csv p6': 15.0,
'iris.csv p7': 25.0, 'iris.csv p4': 50.0, 'iris.csv p5': 0.0, 'iris.csv p2': 0.0, 'iris.csv p3': 0.0,
'iris.csv p1': 0.0}

cross validation error: 12.0

accuracy: 0.84

31-knn in iris.csv data set with 150 elements:

matrix confusion: {'3 | 3': 955, '3 | 2': 595, '2 | 2': 980, '2 | 3': 280, '1 | 1': 1510, '1 | 2': 30,
'1 | 3': 10, '2 | 1': 290}

sample error: {'iris.csv p10': 150.0, 'iris.csv p8': 165.0, 'iris.csv p9': 150.0, 'iris.csv p6':
105.0, 'iris.csv p7': 180.0, 'iris.csv p4': 320.0, 'iris.csv p5': 105.0, 'iris.csv p2': 30.0,
'iris.csv p3': 0.0, 'iris.csv p1': 0.0}

cross validation error: 120.5

accuracy: 0.740860215054

76-knn in iris.csv data set with 150 elements:

matrix confusion: {'3 | 1': 35, '3 | 3': 1700, '3 | 2': 2065, '2 | 2': 1480, '2 | 3': 1790, '1 | 1': 1790, '1 | 2': 1780, '1 | 3': 230, '2 | 1': 530}

sample error: {'iris.csv p10': 630.0, 'iris.csv p8': 615.0, 'iris.csv p9': 630.0, 'iris.csv p6': 615.0, 'iris.csv p7': 605.0, 'iris.csv p4': 875.0, 'iris.csv p5': 615.0, 'iris.csv p2': 615.0, 'iris.csv p3': 615.0, 'iris.csv p1': 615.0}

cross validation error: 643.0

accuracy: 0.435964912281

1-knn in wine.csv data set with 178 elements:

matrix confusion: {'3 | 2': 6, '1 | 1': 59, '2 | 1': 9, '3 | 3': 34, '2 | 2': 62}

sample error: {'wine.csv p10': 0.0, 'wine.csv p4': 9.0, 'wine.csv p5': 0.0, 'wine.csv p6': 0.0, 'wine.csv p7': 0.0, 'wine.csv p1': 0.0, 'wine.csv p2': 0.0, 'wine.csv p3': 0.0, 'wine.csv p8': 6.0, 'wine.csv p9': 0.0}

cross validation error: 1.5

accuracy: 0.911764705882

5-knn in wine.csv data set with 178 elements:

matrix confusion: {'3 | 2': 81, '1 | 1': 295, '2 | 1': 45, '3 | 3': 119, '2 | 2': 310}

sample error: {'wine.csv p10': 0.0, 'wine.csv p4': 45.0, 'wine.csv p5': 0.0, 'wine.csv p6': 0.0, 'wine.csv p7': 0.0, 'wine.csv p1': 0.0, 'wine.csv p2': 0.0, 'wine.csv p3': 0.0, 'wine.csv p8': 30.0, 'wine.csv p9': 51.0}

cross validation error: 12.6

accuracy: 0.851764705882

31-knn in wine.csv data set with 178 elements:

matrix confusion: {'3 | 3': 624, '3 | 2': 544, '2 | 2': 1386, '2 | 3': 141, '1 | 1': 1651, '1 | 2': 178, '3 | 1': 72, '2 | 1': 674}

sample error: {'wine.csv p10': 153.0, 'wine.csv p4': 278.0, 'wine.csv p5': 136.0, 'wine.csv p6': 119.0, 'wine.csv p7': 136.0, 'wine.csv p1': 0.0, 'wine.csv p2': 85.0, 'wine.csv p3': 85.0, 'wine.csv p8': 328.0, 'wine.csv p9': 289.0}

cross validation error: 160.9

accuracy: 0.694686907021

90-knn in wine.csv data set with 178 elements:

matrix confusion: {'3 | 1': 808, '3 | 3': 909, '3 | 2': 1883, '2 | 2': 2930, '2 | 3': 987, '1 | 1': 2516, '1 | 2': 2497, '1 | 3': 297, '2 | 1': 2473}

sample error: {'wine.csv p10': 1139.0, 'wine.csv p4': 780.0, 'wine.csv p5': 952.0, 'wine.csv p6': 765.0, 'wine.csv p7': 714.0, 'wine.csv p1': 816.0, 'wine.csv p2': 850.0, 'wine.csv p3': 816.0, 'wine.csv p8': 957.0, 'wine.csv p9': 1156.0}

cross validation error: 894.5

accuracy: 0.415359477124

1-knn in cancer.csv data set with 683 elements:

matrix confusion: {'TN': 38, 'FP': 30, 'FN': 198, 'TP': 414}

sample error: {'cancer.csv p10': 10.0, 'cancer.csv p7': 19.0, 'cancer.csv p6': 16.0, 'cancer.csv p5': 33.0, 'cancer.csv p4': 38.0, 'cancer.csv p3': 26.0, 'cancer.csv p2': 27.0, 'cancer.csv p1': 34.0, 'cancer.csv p9': 22.0, 'cancer.csv p8': 11.0}

cross validation error: 23.6

sensitivity: 0.676470588235

specificity: 0.558823529412

precision: 0.932432432432

recall: 0.676470588235

5-knn in cancer.csv data set with 683 elements:

matrix confusion: {'TN': 190, 'FP': 150, 'FN': 990, 'TP': 2070}

sample error: {'cancer.csv p10': 50.0, 'cancer.csv p7': 95.0, 'cancer.csv p6': 80.0, 'cancer.csv p5': 165.0, 'cancer.csv p4': 190.0, 'cancer.csv p3': 130.0, 'cancer.csv p2': 135.0, 'cancer.csv p1': 170.0, 'cancer.csv p9': 110.0, 'cancer.csv p8': 55.0}

cross validation error: 118.0

sensitivity: 0.676470588235

specificity: 0.558823529412

precision: 0.932432432432

recall: 0.676470588235

31-knn in cancer.csv data set with 683 elements:

matrix confusion: {'TN': 1178, 'FP': 930, 'FN': 6138, 'TP': 12834}

sample error: {'cancer.csv p10': 310.0, 'cancer.csv p7': 589.0, 'cancer.csv p6': 496.0, 'cancer.csv p5': 1023.0, 'cancer.csv p4': 1178.0, 'cancer.csv p3': 806.0, 'cancer.csv p2': 837.0, 'cancer.csv p1': 1054.0, 'cancer.csv p9': 682.0, 'cancer.csv p8': 341.0}

cross validation error: 731.6

sensitivity: 0.676470588235

specificity: 0.558823529412

precision: 0.932432432432

recall: 0.676470588235

341-knn in cancer.csv data set with 683 elements:

matrix confusion: {'TN': 11774, 'FP': 11414, 'FN': 78776, 'TP': 129916}

sample error: {'cancer.csv p10': 5858.0, 'cancer.csv p7': 7289.0, 'cancer.csv p6': 9668.0, 'cancer.csv p5': 11539.0, 'cancer.csv p4': 11774.0, 'cancer.csv p3': 9586.0, 'cancer.csv p2': 9795.0, 'cancer.csv p1': 11594.0, 'cancer.csv p9': 7718.0, 'cancer.csv p8': 5729.0}

cross validation error: 9055.0

sensitivity: 0.622525060855

specificity: 0.507762635846

precision: 0.919238661289

recall: 0.622525060855

1-knn in quality.csv data set with 1599 elements:

matrix confusion: {'4 | 6': 21, '6 | 7': 49, '8 | 6': 13, '4 | 7': 5, '7 | 7': 4, '3 | 7': 1, '7 | 6': 152, '8 | 5': 5, '5 | 7': 100, '6 | 6': 349, '4 | 5': 27, '5 | 5': 323, '6 | 5': 233, '7 | 5': 43, '3 | 6': 4, '5 | 6': 256, '3 | 5': 5}

sample error: {'quality.csv p10': 93.0, 'quality.csv p1': 54.0, 'quality.csv p3': 85.0, 'quality.csv p2': 84.0, 'quality.csv p5': 155.0, 'quality.csv p4': 90.0, 'quality.csv p7': 90.0, 'quality.csv p6': 106.0, 'quality.csv p9': 82.0, 'quality.csv p8': 75.0}

cross validation error: 91.4

accuracy: 0.425157232704

13-knn in quality.csv data set with 1599 elements:

matrix confusion: {'6 | 6': 3375, '7 | 6': 1176, '8 | 3': 1, '8 | 5': 109, '8 | 4': 2, '8 | 7': 19, '8 | 6': 103, '5 | 4': 166, '5 | 5': 4507, '5 | 6': 3315, '5 | 7': 785, '4 | 7': 59, '5 | 3': 54, '6 | 4': 115, '6 | 5': 3889, '6 | 3': 53, '7 | 7': 256, '7 | 3': 42, '4 | 3': 7, '4 | 5': 359, '6 | 7': 771, '4 | 4': 11, '7 | 4': 20, '4 | 6': 253, '3 | 6': 46, '3 | 7': 18, '3 | 4': 4, '3 | 5': 60, '7 | 5': 1093, '3 | 3': 2}

sample error: {'quality.csv p10': 1369.0, 'quality.csv p1': 1027.0, 'quality.csv p3': 1329.0, 'quality.csv p2': 1122.0, 'quality.csv p5': 1258.0, 'quality.csv p4': 1206.0, 'quality.csv p7': 1270.0, 'quality.csv p6': 1446.0, 'quality.csv p9': 1248.0, 'quality.csv p8': 1244.0}

cross validation error: 1251.9

accuracy: 0.394339622642

111-knn in quality.csv data set with 1599 elements:

matrix confusion: {'7 | 8': 96, '6 | 6': 28282, '7 | 6': 9013, '8 | 3': 4, '8 | 5': 909, '8 | 4': 44, '8 | 7': 246, '8 | 6': 791, '5 | 4': 2021, '5 | 5': 35547, '5 | 6': 29881, '5 | 7': 7452, '4 | 7': 543, '5 | 3': 206, '4 | 6': 2352, '6 | 5': 32310, '3 | 6': 436, '3 | 7': 121, '3 | 4': 36, '4 | 5': 2781, '3 | 8': 8, '4 | 8': 26, '4 | 4': 161, '6 | 8': 240, '6 | 7': 7268, '7 | 7': 2178, '7 | 4': 588, '6 | 4': 1761, '6 | 3': 180, '7 | 3': 63, '4 | 3': 20, '3 | 5': 505, '8 | 8': 4, '7 | 5': 10151, '3 | 3': 4, '5 | 8': 262}

sample error: {'quality.csv p10': 10983.0, 'quality.csv p1': 9313.0, 'quality.csv p3': 11177.0, 'quality.csv p2': 10330.0, 'quality.csv p5': 10671.0, 'quality.csv p4': 12035.0, 'quality.csv p7': 11622.0, 'quality.csv p6': 12165.0, 'quality.csv p9': 10935.0, 'quality.csv p8': 11083.0}

cross validation error: 11031.4

accuracy: 0.374956088164

799-knn in quality.csv data set with 1599 elements:

matrix confusion: {'6 | 6': 206747, '8 | 6': 5829, '7 | 6': 65737, '8 | 3': 63, '5 | 8': 4435, '8 | 4': 392, '8 | 7': 1794, '4 | 4': 1188, '5 | 4': 15822, '5 | 5': 228653, '5 | 6': 224452, '5 | 7': 66835, '4 | 7': 5217, '5 | 3': 2324, '4 | 6': 17574, '6 | 5': 214655, '3 | 6': 3248, '3 | 7': 977, '3 | 4': 244, '4 | 5': 17847, '3 | 8': 64, '4 | 8': 351, '7 | 8': 1504, '6 | 8': 4481, '6 | 7': 61762, '7 | 7': 18440, '7 | 4': 4455, '6 | 4': 14310, '6 | 3': 2214, '7 | 3': 759, '4 | 3': 170, '3 | 5': 3422, '8 | 8': 136, '7 | 5': 68106, '3 | 3': 35, '8 | 5': 6168}

sample error: {'quality.csv p10': 81866.0, 'quality.csv p1': 78906.0, 'quality.csv p3': 82957.0, 'quality.csv p2': 77970.0, 'quality.csv p5': 79104.0, 'quality.csv p4': 82206.0, 'quality.csv p7': 84845.0, 'quality.csv p6': 86727.0, 'quality.csv p9': 77421.0, 'quality.csv p8': 83209.0}

cross validation error: 81521.1

accuracy: 0.358308734975

1-knn in abalone.csv data set with 4177 elements:

matrix confusion: {'10 | 9': 71, '11 | 9': 54, '7 | 8': 30, '10 | 8': 70, '11 | 4': 47, '26 | 7': 1, '10 | 7': 144, '11 | 7': 71, '19 | 10': 2, '19 | 11': 7, '18 | 11': 12, '18 | 10': 7, '3 | 10': 2, '3 | 11': 2, '15 | 8': 8, '10 | 5': 63, '19 | 9': 4, '10 | 4': 47, '6 | 9': 29, '6 | 8': 33, '6 | 7': 32, '6 | 5': 32, '6 | 4': 23, '25 | 11': 1, '9 | 10': 227, '13 | 10': 42, '13 | 11': 39, '8 | 5': 91, '8 | 4': 13, '8 | 7': 125, '8 | 9': 68, '8 | 8': 62, '24 | 10': 2, '2 | 4': 1, '23 | 11': 4, '15 | 4': 28, '15 | 5': 1, '7 | 10': 124, '7 | 11': 26, '29 | 4': 1, '19 | 4': 9, '15 | 9': 3, '19 | 7': 8, '8 | 11': 30, '8 | 10': 178, '15 | 7': 24, '12 | 11': 38, '12 | 10': 84, '4 | 9': 5, '4 | 8': 7, '4 | 5': 9, '4 | 4': 7, '4 | 7': 9, '11 | 5': 37, '3 | 7': 4, '3 | 4': 3, '3 | 5': 1, '18 | 9': 3, '3 | 8': 1, '3 | 9': 2, '4 | 11': 3, '4 | 10': 17, '18 | 8': 2, '20 | 8': 2, '20 | 9': 1, '20 | 4': 7, '17 | 10': 8, '20 | 7': 9, '17 | 11': 19, '5 | 8': 6, '5 | 9': 12, '15 | 10': 19,

'15 | 11': 20, '5 | 4': 12, '5 | 5': 12, '5 | 7': 22, '14 | 8': 11, '14 | 11': 34, '14 | 10': 25, '14 | 5':
1, '14 | 4': 20, '22 | 4': 2, '12 | 9': 18, '23 | 4': 3, '16 | 9': 3, '17 | 9': 3, '16 | 7': 16, '17 | 7': 6,
'17 | 4': 20, '16 | 4': 13, '12 | 7': 55, '12 | 5': 11, '12 | 4': 36, '11 | 8': 60, '7 | 9': 43, '7 | 7': 78,
'7 | 4': 30, '7 | 5': 60, '19 | 8': 2, '5 | 10': 39, '5 | 11': 12, '14 | 9': 5, '18 | 4': 11, '18 | 7': 7, '23
| 7': 2, '22 | 7': 4, '14 | 7': 30, '9 | 11': 42, '27 | 11': 2, '12 | 8': 24, '9 | 4': 35, '9 | 5': 91, '9 |
7': 149, '9 | 8': 69, '9 | 9': 75, '10 | 11': 46, '10 | 10': 190, '1 | 7': 1, '17 | 8': 2, '16 | 8': 8, '13
| 8': 19, '13 | 9': 18, '21 | 8': 1, '21 | 4': 6, '21 | 7': 3, '13 | 7': 34, '13 | 4': 43, '13 | 5': 8, '20 |
10': 2, '20 | 11': 5, '11 | 10': 181, '11 | 11': 36, '6 | 11': 21, '6 | 10': 89, '21 | 11': 3, '21 | 10':
1, '16 | 11': 15, '16 | 10': 12}

sample error: {'abalone.csv p8': 360.0, 'abalone.csv p9': 342.0, 'abalone.csv p6': 381.0,
'abalone.csv p7': 405.0, 'abalone.csv p4': 374.0, 'abalone.csv p5': 334.0, 'abalone.csv
p2': 410.0, 'abalone.csv p3': 367.0, 'abalone.csv p1': 382.0, 'abalone.csv p10': 355.0}

cross validation error: 371.0

accuracy: 0.1103117506

31-knn in abalone.csv data set with 4177 elements:

matrix confusion: {'5 | 20': 7, '24 | 9': 22, '24 | 7': 2, '27 | 17': 2, '29 | 5': 9, '13 | 10': 758,
'13 | 11': 477, '2 | 9': 6, '13 | 13': 205, '13 | 14': 99, '13 | 15': 232, '16 | 20': 1, '13 | 17': 39,
'13 | 19': 39, '2 | 7': 3, '2 | 6': 2, '2 | 5': 9, '2 | 4': 3, '7 | 19': 26, '7 | 10': 1399, '7 | 11': 652, '7
| 12': 294, '7 | 13': 379, '7 | 14': 193, '7 | 15': 330, '7 | 16': 91, '7 | 17': 26, '15 | 8': 601, '15
| 9': 645, '15 | 4': 85, '15 | 5': 282, '15 | 6': 237, '15 | 7': 354, '29 | 8': 5, '12 | 13': 284, '12 |
12': 259, '12 | 11': 591, '12 | 10': 1100, '12 | 17': 38, '12 | 16': 93, '12 | 15': 325, '12 | 14':
140, '12 | 19': 38, '22 | 10': 10, '22 | 11': 8, '22 | 13': 4, '11 | 20': 69, '5 | 8': 780, '5 | 9': 544,
'5 | 4': 48, '5 | 5': 205, '5 | 6': 345, '5 | 7': 547, '14 | 19': 34, '14 | 11': 299, '14 | 10': 452, '14
| 13': 137, '14 | 12': 157, '14 | 15': 165, '14 | 14': 64, '14 | 17': 34, '14 | 16': 58, '4 | 20': 3,

'10 | 20': 83, '5 | 19': 12, '5 | 16': 34, '5 | 17': 12, '5 | 14': 63, '5 | 15': 105, '5 | 12': 101, '5 | 13': 130, '5 | 10': 433, '5 | 11': 199, '22 | 8': 54, '22 | 9': 20, '22 | 6': 32, '22 | 7': 30, '22 | 4': 6, '22 | 5': 22, '21 | 9': 83, '21 | 8': 86, '21 | 5': 57, '21 | 4': 18, '21 | 7': 48, '21 | 6': 33, '1 | 13': 1, '1 | 10': 2, '1 | 11': 1, '16 | 19': 15, '16 | 17': 15, '16 | 16': 29, '16 | 15': 78, '16 | 14': 30, '16 | 13': 71, '16 | 12': 76, '16 | 11': 145, '16 | 10': 235, '11 | 8': 2701, '11 | 9': 3058, '26 | 8': 11, '26 | 9': 2, '26 | 6': 7, '26 | 7': 6, '11 | 6': 1078, '11 | 7': 2016, '3 | 19': 2, '3 | 14': 6, '3 | 15': 9, '3 | 16': 4, '3 | 17': 2, '3 | 10': 41, '3 | 11': 28, '3 | 12': 13, '3 | 13': 13, '9 | 6': 2293, '25 | 9': 7, '25 | 8': 4, '25 | 7': 1, '6 | 9': 1271, '6 | 8': 1844, '6 | 7': 1336, '6 | 6': 696, '6 | 5': 443, '6 | 4': 101, '24 | 12': 4, '24 | 13': 6, '24 | 10': 18, '24 | 11': 2, '24 | 16': 2, '24 | 14': 2, '24 | 15': 4, '8 | 19': 30, '29 | 4': 3, '29 | 7': 3, '29 | 6': 2, '8 | 11': 893, '8 | 10': 2049, '8 | 13': 542, '8 | 12': 463, '29 | 9': 6, '8 | 14': 276, '8 | 17': 30, '8 | 16': 138, '13 | 20': 12, '15 | 16': 39, '15 | 17': 20, '15 | 14': 42, '15 | 15': 107, '15 | 12': 102, '15 | 13': 101, '15 | 10': 338, '15 | 11': 218, '15 | 19': 20, '17 | 8': 299, '17 | 9': 375, '17 | 6': 86, '17 | 7': 155, '17 | 4': 60, '17 | 5': 186, '26 | 10': 2, '26 | 11': 1, '26 | 13': 1, '13 | 12': 237, '2 | 8': 5, '14 | 9': 798, '13 | 16': 87, '23 | 9': 50, '23 | 8': 53, '14 | 5': 215, '14 | 4': 61, '23 | 5': 29, '14 | 6': 268, '27 | 13': 2, '27 | 12': 6, '27 | 11': 10, '27 | 10': 6, '10 | 19': 46, '27 | 16': 2, '27 | 15': 6, '9 | 7': 3580, '10 | 15': 621, '10 | 14': 307, '27 | 19': 2, '10 | 16': 174, '10 | 11': 1224, '1 | 5': 1, '1 | 6': 7, '10 | 12': 554, '13 | 8': 1058, '13 | 9': 1344, '13 | 6': 406, '13 | 7': 692, '13 | 4': 137, '13 | 5': 471, '21 | 19': 3, '6 | 16': 64, '6 | 15': 239, '6 | 14': 139, '6 | 13': 263, '6 | 12': 198, '6 | 11': 413, '6 | 10': 960, '21 | 11': 31, '21 | 10': 32, '21 | 13': 9, '21 | 12': 11, '21 | 15': 12, '21 | 14': 4, '6 | 19': 21, '21 | 16': 4, '10 | 9': 3379, '10 | 8': 3870, '10 | 5': 919, '10 | 4': 204, '10 | 7': 3122, '10 | 6': 1973, '19 | 12': 37, '19 | 13': 29, '19 | 10': 96, '19 | 11': 79, '19 | 16': 13, '19 | 17': 7, '19 | 14': 13, '19 | 15': 27, '19 | 19': 7, '11 | 4': 178, '11 | 5': 714, '26 | 5': 1, '15 | 20': 2, '23 | 19': 4, '23 | 17': 4, '23 | 16': 4, '23 | 15': 12, '23 | 14': 4, '23 | 13': 6, '23 | 12': 12,

'23 | 11': 28, '23 | 10': 19, '8 | 15': 482, '3 | 8': 111, '3 | 9': 67, '4 | 9': 278, '4 | 8': 410, '4 | 5': 134, '4 | 4': 30, '4 | 7': 294, '4 | 6': 172, '3 | 6': 49, '3 | 7': 73, '3 | 4': 10, '17 | 19': 19, '29 | 11': 2, '29 | 10': 1, '18 | 19': 12, '14 | 8': 738, '23 | 7': 25, '23 | 6': 20, '3 | 5': 37, '14 | 7': 422, '6 | 20': 20, '23 | 4': 9, '16 | 9': 440, '16 | 8': 401, '16 | 7': 231, '16 | 6': 138, '16 | 5': 133, '16 | 4': 39, '1 | 8': 11, '1 | 9': 2, '8 | 20': 56, '9 | 4': 196, '9 | 5': 1001, '27 | 9': 14, '27 | 8': 8, '18 | 5': 106, '18 | 4': 33, '18 | 7': 116, '18 | 6': 71, '18 | 9': 283, '18 | 8': 211, '27 | 14': 2, '27 | 7': 2, '9 | 8': 4285, '9 | 9': 3475, '10 | 17': 46, '10 | 10': 2442, '10 | 13': 597, '1 | 7': 6, '20 | 16': 8, '20 | 17': 5, '20 | 14': 8, '20 | 15': 21, '20 | 12': 22, '20 | 13': 22, '20 | 10': 69, '20 | 11': 54, '20 | 19': 5, '18 | 15': 52, '18 | 14': 22, '18 | 17': 12, '18 | 16': 22, '18 | 11': 108, '18 | 10': 149, '18 | 13': 46, '18 | 12': 59, '17 | 14': 30, '17 | 15': 73, '17 | 16': 28, '17 | 17': 19, '17 | 10': 178, '17 | 11': 159, '17 | 12': 78, '17 | 13': 53, '25 | 19': 1, '25 | 15': 3, '25 | 14': 1, '25 | 17': 1, '25 | 16': 1, '25 | 11': 5, '25 | 10': 3, '25 | 13': 1, '25 | 12': 3, '7 | 20': 38, '8 | 5': 790, '8 | 4': 130, '8 | 7': 3247, '8 | 6': 2060, '8 | 9': 2614, '8 | 8': 3777, '9 | 20': 89, '19 | 4': 27, '19 | 5': 89, '19 | 6': 74, '19 | 7': 108, '19 | 8': 195, '19 | 9': 191, '14 | 20': 4, '4 | 19': 3, '4 | 15': 43, '4 | 14': 25, '4 | 17': 3, '4 | 16': 12, '4 | 11': 77, '4 | 10': 194, '4 | 13': 56, '4 | 12': 33, '20 | 8': 179, '20 | 9': 143, '20 | 4': 21, '20 | 5': 72, '20 | 6': 77, '20 | 7': 100, '6 | 17': 21, '12 | 9': 1625, '12 | 8': 1529, '12 | 7': 1009, '12 | 6': 624, '12 | 5': 439, '12 | 4': 119, '7 | 8': 2628, '7 | 9': 1776, '7 | 6': 1343, '7 | 7': 2071, '7 | 4': 150, '7 | 5': 725, '9 | 19': 42, '9 | 12': 575, '9 | 13': 668, '9 | 10': 2661, '9 | 11': 1236, '9 | 16': 179, '9 | 17': 42, '9 | 14': 344, '9 | 15': 662, '21 | 17': 3, '12 | 20': 33, '2 | 11': 2, '2 | 10': 1, '11 | 19': 36, '11 | 12': 459, '11 | 13': 544, '11 | 10': 2198, '11 | 11': 981, '11 | 16': 165, '11 | 17': 36, '11 | 14': 271, '11 | 15': 562}

sample error: {'abalone.csv p8': 11409.0, 'abalone.csv p9': 11226.0, 'abalone.csv p6': 11881.0, 'abalone.csv p7': 11492.0, 'abalone.csv p4': 11486.0, 'abalone.csv p5': 10933.0, 'abalone.csv p2': 12246.0, 'abalone.csv p3': 11038.0, 'abalone.csv p1': 11956.0, 'abalone.csv p10': 11236.0}

cross validation error: 11490.3

accuracy: 0.111139475516

291-knn in abalone.csv data set with 4177 elements:

matrix confusion: {'5 | 23': 56, '5 | 22': 34, '5 | 21': 45, '5 | 20': 136, '24 | 8': 68, '24 | 9': 118, '24 | 6': 2, '24 | 7': 8, '17 | 21': 38, '17 | 20': 146, '17 | 23': 54, '17 | 22': 45, '25 | 22': 1, '27 | 17': 12, '8 | 19': 788, '29 | 4': 8, '13 | 10': 9048, '13 | 11': 6017, '13 | 12': 4024, '13 | 13': 2628, '13 | 14': 1491, '13 | 15': 1588, '13 | 16': 1245, '13 | 17': 678, '13 | 18': 667, '13 | 19': 437, '2 | 7': 16, '2 | 6': 19, '2 | 5': 17, '2 | 4': 8, '1 | 21': 1, '7 | 18': 916, '7 | 19': 566, '27 | 15': 24, '7 | 10': 13797, '7 | 11': 9130, '7 | 12': 6103, '7 | 13': 3741, '7 | 14': 2367, '7 | 15': 2297, '7 | 16': 1650, '7 | 17': 751, '15 | 8': 4676, '15 | 9': 5343, '15 | 4': 229, '15 | 5': 719, '15 | 6': 1600, '15 | 7': 2842, '29 | 8': 44, '12 | 13': 3173, '12 | 12': 4984, '12 | 11': 7711, '12 | 10': 11796, '12 | 17': 841, '12 | 16': 1431, '12 | 15': 2059, '12 | 14': 2006, '12 | 19': 544, '12 | 18': 843, '2 | 22': 1, '2 | 23': 1, '2 | 20': 2, '2 | 21': 1, '22 | 18': 12, '22 | 19': 2, '22 | 14': 32, '22 | 15': 46, '22 | 16': 38, '22 | 17': 6, '22 | 10': 224, '22 | 11': 112, '22 | 12': 84, '22 | 13': 52, '27 | 9': 88, '11 | 23': 287, '11 | 22': 158, '11 | 21': 229, '11 | 20': 751, '5 | 8': 6216, '5 | 9': 5441, '5 | 4': 156, '5 | 5': 732, '5 | 6': 2718, '5 | 7': 4594, '5 | 3': 24, '14 | 19': 313, '14 | 18': 460, '6 | 18': 668, '14 | 11': 3763, '14 | 10': 5674, '14 | 13': 1745, '14 | 12': 2693, '14 | 15': 1061, '14 | 14': 967, '14 | 17': 462, '14 | 16': 797, '4 | 20': 52, '4 | 21': 23, '4 | 22': 14, '4 | 23': 23, '2 | 3': 2, '10 | 20': 729, '10 | 21': 216, '10 | 22': 150, '10 | 23': 278, '5 | 18': 308, '5 | 19': 183, '5 | 16': 575, '5 | 17': 267, '5 | 14': 727, '5 | 15': 723, '5 | 12': 2014,

'5 | 13': 1231, '5 | 10': 4383, '5 | 11': 2902, '22 | 8': 312, '22 | 9': 272, '22 | 6': 170, '22 | 7': 280, '22 | 4': 16, '22 | 5': 70, '22 | 3': 4, '21 | 20': 29, '21 | 21': 11, '21 | 22': 10, '21 | 23': 11, '21 | 9': 730, '21 | 8': 627, '21 | 5': 129, '21 | 4': 48, '21 | 7': 370, '21 | 6': 228, '21 | 3': 12, '1 | 12': 13, '1 | 13': 7, '1 | 10': 37, '1 | 11': 14, '1 | 16': 4, '1 | 17': 1, '1 | 14': 5, '16 | 18': 237, '16 | 17': 236, '16 | 16': 422, '1 | 18': 2, '16 | 14': 489, '16 | 13': 894, '16 | 12': 1365, '16 | 11': 1967, '16 | 10': 3024, '11 | 8': 24731, '11 | 9': 25434, '26 | 8': 56, '26 | 9': 39, '11 | 3': 94, '26 | 6': 33, '26 | 7': 62, '11 | 6': 9298, '11 | 7': 15987, '3 | 18': 37, '3 | 19': 23, '3 | 14': 89, '3 | 15': 99, '3 | 16': 85, '3 | 17': 25, '3 | 10': 533, '3 | 11': 350, '3 | 12': 263, '3 | 13': 164, '25 | 9': 44, '25 | 8': 27, '25 | 7': 13, '25 | 6': 2, '6 | 9': 12750, '6 | 8': 14863, '6 | 7': 10656, '6 | 6': 6092, '6 | 5': 1567, '6 | 4': 344, '6 | 3': 46, '24 | 18': 10, '24 | 19': 2, '24 | 12': 42, '24 | 13': 28, '24 | 10': 144, '24 | 11': 86, '24 | 16': 10, '24 | 17': 18, '24 | 14': 14, '24 | 15': 12, '27 | 16': 18, '29 | 20': 2, '29 | 21': 1, '29 | 22': 1, '29 | 23': 1, '29 | 5': 17, '8 | 18': 1270, '29 | 7': 16, '29 | 6': 19, '29 | 3': 2, '8 | 11': 12755, '8 | 10': 19278, '8 | 13': 5121, '8 | 12': 8251, '8 | 15': 3104, '8 | 14': 3273, '8 | 17': 1093, '8 | 16': 1954, '13 | 21': 131, '13 | 20': 416, '13 | 23': 160, '13 | 22': 112, '15 | 16': 697, '15 | 17': 338, '15 | 14': 734, '15 | 15': 824, '15 | 12': 2035, '15 | 13': 1342, '15 | 10': 4559, '15 | 11': 2981, '15 | 18': 343, '15 | 19': 205, '22 | 21': 6, '22 | 20': 4, '22 | 23': 2, '22 | 22': 2, '17 | 8': 2425, '17 | 9': 3022, '17 | 6': 683, '17 | 7': 1193, '17 | 4': 160, '17 | 5': 394, '17 | 3': 40, '26 | 10': 37, '26 | 11': 14, '26 | 12': 13, '26 | 13': 7, '26 | 14': 5, '26 | 15': 8, '26 | 16': 4, '23 | 21': 5, '26 | 18': 2, '16 | 19': 144, '1 | 15': 8, '2 | 9': 58, '2 | 8': 44, '16 | 22': 39, '16 | 15': 553, '16 | 23': 53, '14 | 9': 6408, '14 | 8': 5643, '23 | 9': 428, '23 | 8': 352, '14 | 5': 635, '23 | 6': 131, '23 | 5': 69, '14 | 6': 1767, '23 | 3': 6, '14 | 3': 40, '1 | 8': 56, '1 | 9': 39, '27 | 11': 80, '9 | 3': 70, '9 | 4': 735, '10 | 18': 1599, '9 | 6': 19048, '9 | 7': 30113, '10 | 15': 4080, '10 | 14': 4153, '10 | 17': 1485, '10 | 16': 2568, '10 | 11': 15908, '1 | 5': 9, '1 | 6': 33, '1 | 7': 62, '16 | 21': 51, '13 | 8': 9295, '13 | 9': 10680,

'13 | 3': 86, '13 | 6': 3229, '13 | 7': 5500, '13 | 4': 384, '13 | 5': 1257, '24 | 23': 4, '24 | 22':
2, '24 | 21': 4, '24 | 20': 10, '6 | 17': 513, '21 | 18': 44, '6 | 15': 1573, '6 | 14': 1609, '6 | 13':
2638, '6 | 12': 4287, '6 | 11': 6389, '6 | 10': 9340, '21 | 11': 394, '21 | 10': 577, '21 | 13':
187, '21 | 12': 276, '21 | 15': 115, '21 | 14': 99, '21 | 17': 37, '21 | 16': 111, '18 | 20': 105,
'18 | 21': 32, '18 | 22': 30, '18 | 23': 40, '23 | 23': 7, '10 | 9': 30691, '10 | 8': 33519, '23 |
20': 22, '10 | 3': 94, '26 | 17': 1, '10 | 5': 3744, '10 | 4': 691, '10 | 7': 26041, '10 | 6': 16050,
'25 | 20': 4, '16 | 20': 143, '25 | 23': 1, '19 | 12': 659, '19 | 13': 424, '19 | 10': 1376, '19 | 11':
896, '19 | 16': 222, '19 | 17': 87, '19 | 14': 226, '19 | 15': 261, '19 | 18': 100, '19 | 19': 66,
'10 | 19': 1038, '11 | 4': 561, '17 | 19': 166, '11 | 5': 2433, '26 | 5': 9, '15 | 23': 83, '15 | 22':
64, '15 | 21': 83, '15 | 20': 220, '23 | 19': 31, '23 | 18': 34, '23 | 17': 29, '23 | 16': 77, '23 |
15': 85, '23 | 14': 76, '23 | 13': 146, '23 | 12': 222, '23 | 11': 272, '23 | 10': 372, '3 | 8': 824,
'3 | 9': 721, '4 | 9': 2783, '4 | 8': 3222, '4 | 5': 459, '17 | 18': 214, '4 | 7': 2375, '4 | 6': 1440,
'3 | 6': 358, '3 | 7': 613, '3 | 4': 29, '3 | 5': 114, '29 | 11': 28, '29 | 10': 38, '29 | 13': 12, '29 |
12': 16, '29 | 15': 7, '29 | 14': 6, '29 | 17': 1, '29 | 16': 11, '29 | 19': 1, '29 | 18': 2, '18 | 19':
107, '18 | 18': 155, '4 | 4': 101, '23 | 7': 224, '4 | 3': 14, '14 | 4': 165, '23 | 22': 7, '6 | 22':
58, '14 | 7': 3538, '6 | 20': 240, '6 | 21': 63, '23 | 4': 24, '16 | 9': 3536, '16 | 8': 3096, '16 |
7': 1850, '16 | 6': 903, '16 | 5': 365, '16 | 4': 104, '16 | 3': 26, '27 | 13': 48, '27 | 12': 74, '27
| 10': 92, '8 | 20': 458, '8 | 21': 129, '8 | 22': 83, '8 | 23': 191, '19 | 23': 24, '19 | 22': 18, '19
| 21': 21, '19 | 20': 56, '9 | 5': 4375, '18 | 3': 22, '27 | 8': 54, '18 | 5': 250, '18 | 4': 88, '18 |
7': 932, '18 | 6': 508, '18 | 9': 2198, '18 | 8': 1778, '27 | 14': 24, '27 | 7': 26, '27 | 6': 4, '9 |
8': 37226, '9 | 9': 32461, '26 | 21': 1, '27 | 19': 14, '27 | 18': 12, '10 | 10': 24309, '10 | 13':
6304, '10 | 12': 9974, '29 | 9': 58, '20 | 16': 174, '20 | 17': 68, '20 | 14': 177, '20 | 15': 213,
'20 | 12': 497, '20 | 13': 323, '20 | 10': 1083, '20 | 11': 681, '20 | 18': 80, '20 | 19': 47, '27 |
22': 2, '27 | 23': 2, '27 | 20': 8, '18 | 15': 351, '18 | 14': 322, '18 | 17': 161, '18 | 16': 306,

'18 | 11': 1322, '18 | 10': 1951, '18 | 13': 619, '18 | 12': 945, '17 | 14': 460, '17 | 15': 494,
'17 | 16': 471, '17 | 17': 204, '17 | 10': 2574, '17 | 11': 1847, '17 | 12': 1354, '17 | 13': 894,
'15 | 3': 56, '25 | 19': 7, '25 | 18': 6, '25 | 15': 12, '25 | 14': 12, '25 | 17': 6, '25 | 16': 9, '25 |
11': 40, '25 | 10': 46, '25 | 13': 24, '25 | 12': 37, '7 | 21': 109, '7 | 20': 350, '7 | 23': 143, '7 |
22': 78, '8 | 3': 26, '8 | 5': 3807, '8 | 4': 559, '8 | 7': 27350, '8 | 6': 17297, '8 | 9': 25948, '8 |
8': 32262, '3 | 3': 6, '9 | 23': 276, '9 | 22': 139, '9 | 21': 208, '9 | 20': 726, '19 | 3': 18, '19 |
4': 72, '19 | 5': 225, '19 | 6': 495, '19 | 7': 893, '19 | 8': 1501, '19 | 9': 1672, '14 | 20': 279,
'14 | 21': 86, '14 | 22': 73, '14 | 23': 97, '4 | 19': 67, '4 | 18': 141, '4 | 15': 345, '4 | 14': 330,
'4 | 17': 108, '4 | 16': 267, '4 | 11': 1330, '4 | 10': 2048, '4 | 13': 551, '4 | 12': 894, '20 | 8':
1235, '20 | 9': 1309, '20 | 4': 56, '20 | 5': 200, '20 | 6': 465, '20 | 7': 849, '20 | 3': 14, '20 |
23': 17, '20 | 22': 14, '20 | 21': 20, '20 | 20': 44, '21 | 19': 29, '12 | 9': 13602, '12 | 8':
12705, '6 | 16': 1171, '12 | 3': 72, '12 | 7': 8315, '12 | 6': 4689, '12 | 5': 1392, '12 | 4': 343,
'7 | 8': 21651, '7 | 9': 18040, '7 | 6': 11131, '7 | 7': 17495, '7 | 4': 540, '7 | 5': 2866, '7 | 3':
60, '6 | 23': 101, '9 | 18': 1671, '9 | 19': 1067, '9 | 12': 10541, '9 | 13': 6592, '9 | 10': 25499,
'9 | 11': 16755, '9 | 16': 2630, '9 | 17': 1545, '9 | 14': 4359, '9 | 15': 4172, '6 | 19': 401, '12 |
22': 111, '12 | 23': 166, '12 | 20': 475, '12 | 21': 148, '2 | 19': 1, '2 | 18': 2, '2 | 13': 12, '2 |
12': 16, '2 | 11': 28, '2 | 10': 38, '2 | 17': 1, '2 | 16': 11, '2 | 15': 7, '2 | 14': 6, '11 | 18': 1424,
'11 | 19': 840, '11 | 12': 8380, '11 | 13': 5295, '11 | 10': 21170, '11 | 11': 13814, '11 | 16':
2224, '11 | 17': 1451, '11 | 14': 3464, '11 | 15': 3401, '3 | 21': 6, '3 | 20': 14, '3 | 23': 7, '3 |
22': 5}

sample error: {'abalone.csv p8': 107314.0, 'abalone.csv p9': 105212.0, 'abalone.csv p6':
111369.0, 'abalone.csv p7': 106115.0, 'abalone.csv p4': 106411.0, 'abalone.csv p5':
104436.0, 'abalone.csv p2': 112833.0, 'abalone.csv p3': 105091.0, 'abalone.csv p1':
111153.0, 'abalone.csv p10': 105950.0}

cross validation error: 107588.4

accuracy: 0.113382283864

2088-knn in abalone.csv data set with 4177 elements:

matrix confusion: {'5 | 27': 51, '5 | 26': 53, '5 | 25': 51, '5 | 24': 7, '5 | 23': 396, '5 | 22': 282, '5 | 21': 595, '5 | 20': 1439, '5 | 29': 29, '24 | 8': 554, '24 | 9': 812, '10 | 2': 152, '27 | 12': 364, '24 | 5': 2, '24 | 6': 80, '24 | 7': 224, '27 | 10': 828, '17 | 29': 25, '17 | 21': 391, '17 | 20': 851, '17 | 23': 243, '17 | 22': 237, '17 | 25': 11, '25 | 22': 6, '17 | 27': 11, '17 | 26': 48, '27 | 17': 72, '29 | 5': 35, '29 | 4': 10, '13 | 10': 71991, '13 | 11': 51621, '13 | 12': 27947, '13 | 13': 20271, '13 | 14': 13263, '13 | 15': 11266, '13 | 16': 7103, '13 | 17': 5448, '13 | 18': 3803, '13 | 19': 3281, '2 | 7': 157, '2 | 6': 96, '2 | 5': 35, '2 | 4': 10, '1 | 23': 2, '1 | 22': 2, '1 | 21': 4, '1 | 20': 8, '19 | 29': 9, '7 | 18': 7007, '7 | 19': 5290, '27 | 15': 148, '7 | 10': 124791, '7 | 11': 84557, '7 | 12': 47124, '7 | 13': 36500, '7 | 14': 23296, '7 | 15': 19339, '7 | 16': 11981, '7 | 17': 9241, '15 | 8': 30753, '15 | 9': 40610, '15 | 4': 1182, '15 | 5': 3124, '15 | 6': 10056, '15 | 7': 17164, '15 | 1': 2, '15 | 2': 2, '15 | 3': 267, '12 | 13': 26826, '12 | 12': 35730, '12 | 11': 65317, '12 | 10': 92389, '12 | 17': 7289, '12 | 16': 9249, '12 | 15': 14750, '12 | 14': 17248, '12 | 19': 4293, '12 | 18': 5189, '2 | 22': 4, '2 | 23': 4, '2 | 20': 11, '2 | 21': 4, '22 | 18': 88, '22 | 19': 76, '22 | 14': 302, '22 | 15': 272, '22 | 16': 156, '22 | 17': 114, '22 | 10': 1962, '22 | 11': 1318, '22 | 12': 728, '22 | 13': 504, '27 | 9': 748, '11 | 23': 1996, '11 | 22': 1340, '11 | 21': 2887, '11 | 20': 6635, '5 | 8': 36657, '5 | 9': 43671, '11 | 25': 235, '11 | 24': 69, '5 | 4': 2221, '5 | 5': 4955, '5 | 6': 13888, '5 | 7': 23106, '5 | 1': 21, '5 | 2': 21, '5 | 3': 504, '14 | 19': 2104, '14 | 18': 2485, '6 | 18': 4677, '14 | 11': 32991, '14 | 10': 45177, '14 | 13': 12897, '14 | 12': 17761, '14 | 15': 7195, '14 | 14': 8172, '14 | 17': 3479, '14 | 16': 4421, '4 | 29': 10, '4 | 20': 643, '4 | 21': 263, '4 | 22': 125, '4 | 23': 181, '4 | 24': 3, '4 | 25': 22, '4 | 26': 22, '4 | 27': 22, '2 | 3': 3, '27 | 18': 50, '10 | 20': 8024, '10 | 21': 3311, '10 | 22': 1560,

'10 | 23': 2362, '10 | 24': 83, '10 | 25': 261, '10 | 26': 304, '10 | 27': 261, '10 | 29': 186, '27 | 4': 2, '5 | 18': 2111, '5 | 19': 1621, '5 | 16': 3615, '5 | 17': 2746, '5 | 14': 7022, '5 | 15': 5839, '5 | 12': 14503, '5 | 13': 10982, '5 | 10': 37806, '5 | 11': 25928, '22 | 8': 1940, '22 | 9': 2338, '22 | 6': 856, '22 | 7': 1322, '22 | 4': 128, '22 | 5': 282, '22 | 3': 30, '21 | 29': 4, '21 | 25': 1, '21 | 26': 10, '21 | 27': 1, '21 | 20': 177, '21 | 21': 80, '21 | 22': 53, '21 | 23': 53, '21 | 9': 5583, '21 | 8': 4256, '21 | 5': 448, '21 | 4': 156, '21 | 7': 2307, '21 | 6': 1384, '21 | 3': 37, '1 | 12': 115, '1 | 13': 86, '1 | 10': 314, '1 | 11': 205, '1 | 16': 22, '1 | 17': 21, '1 | 14': 45, '16 | 18': 1278, '16 | 17': 1784, '16 | 16': 2278, '1 | 18': 17, '16 | 14': 4268, '16 | 13': 6691, '16 | 12': 9265, '16 | 11': 17213, '16 | 10': 23871, '9 | 4': 15005, '11 | 8': 151903, '11 | 9': 186926, '26 | 8': 311, '26 | 9': 374, '26 | 3': 6, '11 | 2': 76, '11 | 3': 1676, '26 | 6': 166, '26 | 7': 252, '11 | 6': 51676, '11 | 7': 89521, '27 | 26': 2, '3 | 18': 245, '3 | 19': 195, '3 | 14': 870, '3 | 15': 724, '3 | 16': 443, '3 | 17': 315, '3 | 10': 4940, '3 | 11': 3402, '3 | 12': 1889, '3 | 13': 1364, '25 | 9': 374, '25 | 8': 191, '25 | 5': 8, '25 | 4': 1, '25 | 7': 73, '25 | 6': 40, '27 | 29': 2, '6 | 9': 98493, '6 | 8': 85138, '6 | 7': 54479, '6 | 6': 32373, '6 | 5': 11865, '6 | 4': 5078, '6 | 3': 1125, '6 | 2': 52, '6 | 1': 52, '24 | 18': 54, '24 | 19': 46, '24 | 12': 296, '24 | 13': 234, '24 | 10': 766, '24 | 11': 538, '24 | 16': 86, '24 | 17': 84, '24 | 14': 166, '24 | 15': 140, '27 | 16': 94, '29 | 26': 1, '29 | 20': 11, '29 | 21': 4, '29 | 22': 4, '29 | 23': 4, '8 | 19': 7544, '8 | 18': 10202, '29 | 7': 157, '29 | 6': 96, '29 | 3': 3, '8 | 11': 119185, '8 | 10': 177162, '8 | 13': 52451, '8 | 12': 66183, '8 | 15': 27922, '8 | 14': 33517, '8 | 17': 13635, '8 | 16': 17041, '13 | 29': 81, '13 | 25': 60, '13 | 24': 12, '13 | 27': 60, '13 | 26': 142, '13 | 21': 1306, '13 | 20': 2849, '13 | 23': 837, '13 | 22': 693, '15 | 16': 3518, '15 | 17': 2644, '15 | 14': 6536, '15 | 15': 5620, '15 | 12': 14175, '15 | 13': 10090, '15 | 10': 36533, '15 | 11': 26283, '15 | 18': 1881, '15 | 19': 1660, '16 | 24': 1, '22 | 21': 24, '22 | 20': 54, '22 | 23': 16, '22 | 22': 16, '22 | 26': 2, '11 | 26': 281, '17 | 8': 16563, '17 | 9': 22964, '17 | 6': 4772, '17 | 7': 8254, '17 | 4': 485, '17 | 5': 1447,

'17 | 3': 110, '26 | 10': 314, '26 | 11': 205, '26 | 12': 115, '26 | 13': 86, '26 | 14': 45, '26 | 15': 46, '26 | 16': 22, '23 | 21': 60, '26 | 18': 17, '26 | 19': 12, '23 | 29': 4, '16 | 26': 43, '16 | 19': 1086, '16 | 27': 15, '1 | 15': 46, '2 | 9': 421, '2 | 8': 348, '16 | 22': 234, '16 | 15': 3722, '16 | 23': 261, '1 | 19': 12, '14 | 9': 49040, '14 | 8': 36030, '23 | 9': 3507, '23 | 8': 2430, '14 | 5': 3581, '23 | 6': 780, '23 | 5': 243, '14 | 6': 11847, '23 | 3': 21, '14 | 3': 297, '14 | 2': 3, '1 | 8': 311, '1 | 9': 374, '9 | 2': 191, '9 | 3': 3520, '10 | 19': 9140, '10 | 18': 11624, '9 | 6': 87229, '9 | 7': 142539, '10 | 15': 32744, '10 | 14': 38963, '10 | 17': 16163, '10 | 16': 20306, '10 | 11': 141308, '1 | 5': 53, '1 | 6': 166, '1 | 7': 252, '29 | 8': 348, '16 | 29': 27, '16 | 21': 438, '13 | 8': 60430, '13 | 9': 79398, '13 | 2': 14, '13 | 3': 548, '13 | 1': 14, '13 | 6': 19486, '13 | 7': 33340, '13 | 4': 2425, '13 | 5': 6175, '24 | 23': 12, '24 | 22': 8, '24 | 21': 24, '24 | 20': 42, '24 | 27': 2, '24 | 26': 2, '24 | 25': 2, '24 | 29': 2, '6 | 17': 5923, '21 | 18': 229, '6 | 15': 12694, '6 | 14': 15526, '6 | 13': 24322, '6 | 12': 31560, '6 | 11': 56493, '6 | 10': 83445, '21 | 11': 3618, '21 | 10': 4975, '21 | 13': 1318, '21 | 12': 1934, '21 | 15': 732, '21 | 14': 864, '21 | 17': 320, '21 | 16': 474, '18 | 29': 19, '18 | 20': 621, '18 | 21': 295, '18 | 22': 163, '18 | 23': 176, '18 | 25': 10, '18 | 26': 33, '23 | 23': 36, '10 | 9': 237981, '10 | 8': 200518, '23 | 20': 129, '10 | 1': 152, '10 | 3': 2961, '26 | 17': 21, '10 | 5': 28063, '10 | 4': 12793, '10 | 7': 126776, '10 | 6': 77117, '25 | 20': 20, '25 | 21': 10, '16 | 20': 915, '25 | 23': 5, '25 | 26': 1, '25 | 29': 1, '19 | 12': 4454, '19 | 13': 3096, '19 | 10': 11413, '19 | 11': 8175, '19 | 16': 1061, '19 | 17': 769, '19 | 14': 2005, '19 | 15': 1706, '19 | 18': 551, '19 | 19': 484, '11 | 1': 76, '11 | 4': 7509, '17 | 19': 981, '11 | 5': 17566, '26 | 4': 27, '26 | 5': 53, '15 | 29': 38, '15 | 27': 22, '15 | 26': 69, '15 | 25': 22, '15 | 24': 2, '15 | 23': 404, '15 | 22': 366, '15 | 21': 653, '15 | 20': 1388, '4 | 2': 10, '23 | 19': 154, '23 | 18': 164, '23 | 17': 231, '23 | 16': 334, '23 | 15': 520, '23 | 14': 585, '23 | 13': 924, '23 | 12': 1360, '23 | 11': 2565, '23 | 10': 3343, '23 | 26': 7, '3 | 8': 4820, '3 | 9': 5753, '4 | 9': 21848, '4 | 8': 18876, '4 | 5': 2653, '17 | 18': 1059, '4 | 7': 12217, '4 | 6': 7343,

'3 | 6': 1911, '3 | 7': 3072, '3 | 4': 295, '3 | 5': 675, '29 | 11': 249, '29 | 10': 353, '29 | 13': 80, '29 | 12': 134, '29 | 15': 44, '29 | 14': 61, '29 | 17': 15, '29 | 16': 34, '29 | 19': 14, '29 | 18': 10, '18 | 19': 714, '18 | 18': 798, '3 | 2': 2, '4 | 4': 1121, '3 | 1': 2, '27 | 21': 20, '4 | 1': 10, '23 | 7': 1267, '4 | 3': 256, '14 | 4': 1390, '23 | 22': 40, '6 | 27': 118, '6 | 24': 20, '6 | 25': 118, '6 | 22': 567, '14 | 7': 20405, '6 | 20': 3082, '6 | 21': 1210, '23 | 4': 88, '16 | 9': 26352, '16 | 8': 19688, '16 | 7': 11190, '16 | 6': 6462, '16 | 5': 1935, '16 | 4': 715, '16 | 3': 149, '2 | 26': 1, '16 | 25': 15, '27 | 13': 256, '9 | 1': 191, '27 | 11': 704, '8 | 24': 56, '8 | 25': 246, '8 | 26': 207, '8 | 27': 246, '8 | 20': 6732, '8 | 21': 2641, '8 | 22': 1136, '8 | 23': 1882, '19 | 23': 119, '19 | 22': 108, '19 | 21': 186, '19 | 20': 417, '19 | 27': 6, '19 | 26': 22, '19 | 25': 6, '9 | 5': 32452, '18 | 3': 83, '27 | 8': 382, '18 | 5': 1036, '18 | 4': 377, '18 | 7': 6190, '18 | 6': 3572, '18 | 9': 16575, '18 | 8': 11954, '14 | 1': 3, '27 | 14': 156, '27 | 7': 146, '27 | 6': 80, '27 | 5': 16, '27 | 19': 44, '9 | 8': 220958, '9 | 9': 257670, '26 | 21': 4, '26 | 20': 8, '26 | 23': 2, '26 | 22': 2, '1 | 3': 6, '1 | 4': 27, '10 | 10': 206078, '10 | 13': 60418, '10 | 12': 77919, '29 | 9': 421, '20 | 16': 828, '20 | 17': 616, '20 | 14': 1553, '20 | 15': 1360, '20 | 12': 3505, '20 | 13': 2466, '20 | 10': 9083, '20 | 11': 6485, '20 | 18': 450, '20 | 19': 391, '27 | 22': 12, '27 | 23': 10, '27 | 20': 40, '18 | 15': 2416, '18 | 14': 2814, '18 | 17': 1144, '18 | 16': 1526, '18 | 11': 11342, '18 | 10': 15420, '18 | 13': 4295, '18 | 12': 6113, '17 | 14': 3893, '17 | 15': 3298, '17 | 16': 2148, '17 | 17': 1498, '17 | 10': 21402, '17 | 11': 15947, '17 | 12': 8568, '17 | 13': 5898, '25 | 19': 22, '25 | 18': 25, '25 | 15': 74, '25 | 14': 78, '25 | 17': 36, '25 | 16': 47, '25 | 11': 352, '25 | 10': 414, '25 | 13': 128, '25 | 12': 182, '7 | 29': 86, '7 | 25': 167, '11 | 27': 235, '7 | 27': 167, '7 | 26': 159, '7 | 21': 1859, '7 | 20': 4684, '7 | 23': 1322, '7 | 22': 864, '8 | 1': 180, '8 | 3': 3219, '8 | 2': 180, '8 | 5': 29496, '8 | 4': 13738, '8 | 7': 124814, '8 | 6': 77393, '8 | 9': 210690, '8 | 8': 186072, '3 | 3': 66, '11 | 29': 180, '9 | 23': 2504, '9 | 22': 1594, '9 | 21': 3515, '9 | 20': 8626, '9 | 27': 302, '9 | 26': 303, '9 | 25': 302, '9 | 24': 89, '9 | 29': 193, '19 | 3': 85,

'19 | 4': 393, '19 | 5': 997, '19 | 6': 3244, '19 | 7': 5364, '19 | 8': 9545, '19 | 9': 12601, '7 | 24': 38, '14 | 29': 57, '14 | 24': 4, '14 | 25': 30, '14 | 26': 82, '14 | 27': 30, '14 | 20': 1790, '14 | 21': 856, '14 | 22': 457, '14 | 23': 504, '4 | 19': 744, '4 | 18': 1003, '4 | 15': 2722, '4 | 14': 3299, '4 | 17': 1267, '4 | 16': 1672, '4 | 11': 12308, '4 | 10': 18317, '4 | 13': 5220, '4 | 12': 6839, '20 | 8': 7830, '20 | 9': 10203, '20 | 4': 356, '20 | 5': 895, '20 | 6': 2844, '20 | 7': 4673, '20 | 3': 79, '20 | 29': 7, '20 | 27': 3, '20 | 26': 15, '20 | 25': 3, '20 | 23': 91, '20 | 22': 87, '20 | 21': 147, '20 | 20': 318, '21 | 19': 218, '12 | 9': 102594, '12 | 8': 80061, '6 | 16': 7900, '12 | 3': 830, '12 | 2': 28, '12 | 1': 28, '6 | 29': 55, '12 | 7': 46358, '12 | 6': 26905, '12 | 5': 8806, '12 | 4': 3710, '6 | 26': 107, '7 | 8': 127185, '7 | 9': 146696, '7 | 6': 51032, '7 | 7': 82952, '7 | 4': 8737, '7 | 5': 19078, '7 | 2': 103, '7 | 3': 2050, '7 | 1': 103, '6 | 23': 861, '9 | 18': 12697, '9 | 19': 9765, '9 | 12': 83333, '9 | 13': 65380, '9 | 10': 221118, '9 | 11': 150594, '9 | 16': 21799, '9 | 17': 17430, '9 | 14': 42036, '9 | 15': 35209, '12 | 29': 108, '6 | 19': 3459, '8 | 29': 126, '12 | 22': 845, '12 | 23': 1100, '12 | 20': 3692, '12 | 21': 1664, '12 | 26': 162, '12 | 27': 102, '12 | 24': 33, '12 | 25': 102, '2 | 19': 14, '2 | 18': 10, '2 | 13': 80, '2 | 12': 134, '2 | 11': 249, '2 | 10': 353, '2 | 17': 15, '2 | 16': 34, '2 | 15': 44, '2 | 14': 61, '11 | 18': 9421, '11 | 19': 7518, '11 | 12': 62773, '11 | 13': 48291, '11 | 10': 164878, '11 | 11': 113693, '11 | 16': 16499, '11 | 17': 13181, '11 | 14': 31479, '11 | 15': 26248, '3 | 21': 67, '3 | 20': 170, '3 | 23': 46, '3 | 22': 37, '3 | 25': 4, '3 | 27': 4, '3 | 26': 7, '3 | 29': 2, '18 | 27': 10}

sample error: {'abalone.csv p8': 776707.0, 'abalone.csv p9': 765614.0, 'abalone.csv p6': 796672.0, 'abalone.csv p7': 757889.0, 'abalone.csv p4': 759867.0, 'abalone.csv p5': 757884.0, 'abalone.csv p2': 803911.0, 'abalone.csv p3': 763460.0, 'abalone.csv p1': 795724.0, 'abalone.csv p10': 768951.0}

cross validation error: 774667.9

accuracy: 0.110288895321

1-knn in adult.csv data set with 30162 elements:

matrix confusion: {'FP': 1512, 'TN': 4520, 'FN': 5995, 'TP': 18133}

sample error: {'adult.csv p6': 729.0, 'adult.csv p7': 2269.0, 'adult.csv p4': 758.0,
'adult.csv p5': 753.0, 'adult.csv p2': 727.0, 'adult.csv p3': 708.0, 'adult.csv p1': 781.0,
'adult.csv p8': 771.0, 'adult.csv p9': 2251.0, 'adult.csv p10': 768.0}

cross validation error: 1051.5

sensitivity: 0.751533488064

specificity: 0.749336870027

precision: 0.923033850853

recall: 0.751533488064

5-knn in adult.csv data set with 30162 elements:

matrix confusion: {'FP': 13551, 'TN': 16609, 'FN': 40720, 'TP': 79920}

sample error: {'adult.csv p6': 8319.0, 'adult.csv p7': 5257.0, 'adult.csv p4': 8290.0,
'adult.csv p5': 5275.0, 'adult.csv p2': 5197.0, 'adult.csv p3': 5140.0, 'adult.csv p1':
5359.0, 'adult.csv p8': 3855.0, 'adult.csv p9': 6797.0, 'adult.csv p10': 3840.0}

cross validation error: 5732.9

sensitivity: 0.662466843501

specificity: 0.550696286472

precision: 0.85502455307

recall: 0.662466843501

31-knn in adult.csv data set with 30162 elements:

matrix confusion: {'FP': 88946, 'TN': 98046, 'FN': 264437, 'TP': 483531}

sample error: {'adult.csv p6': 42853.0, 'adult.csv p7': 26201.0, 'adult.csv p4': 50498.0,
'adult.csv p5': 33913.0, 'adult.csv p2': 33471.0, 'adult.csv p3': 47548.0, 'adult.csv p1':
35843.0, 'adult.csv p8': 32745.0, 'adult.csv p9': 35603.0, 'adult.csv p10': 23808.0}

cross validation error: 36248.3

sensitivity: 0.646459474202

specificity: 0.524332591769

precision: 0.844629565904

recall: 0.646459474202

15082-knn in adult.csv data set with 30162 elements:

matrix confusion: {'FN': 176859486, 'TP': 278013634}

sample error: {'adult.csv p6': 18242594.0, 'adult.csv p7': 17182268.0, 'adult.csv p4':
18515156.0, 'adult.csv p5': 17559826.0, 'adult.csv p2': 17286028.0, 'adult.csv p3':
18650856.0, 'adult.csv p1': 17599404.0, 'adult.csv p8': 16888928.0, 'adult.csv p9':
17125090.0, 'adult.csv p10': 17809336.0}

cross validation error: 17685948.6

sensitivity: 0.611189410357

specificity: 0.5

precision: 1.0

recall: 0.611189410357

4. Conclusão

Após finalizar todas as interações de cada conjunto, pode-se perceber que para os conjuntos com multi classes, ou seja, que possuem mais do que duas classes resultando em uma acurácia ruim conforme o número de vizinhos a serem obtidos cresce, esse fato pode ser analisado no conjunto de dados Wine Quality, composto por 11 classes e 11 atributos, onde é aplicado o 799-NN, ou seja, o algoritmo neste caso deve retornar os 799 vizinhos mais próximos, entretanto, apesar de obter os valores, a acurácia no final do teste foi de apenas 35,83%, valor muito inferior quando comparado com as buscas menores, por exemplo, 1-NN, onde o resultado final foi de 42,51% de acurácia.

Além disso, é possível analisar que no caso dos conjuntos binários, quanto maior o número de vizinho a serem encontrados, maior é a precisão obtida, um exemplo desse fenômeno pode ser visto no conjunto de dados Adult, composto por 2 classes e 14 atributos, nele existem 30162 elementos e um dos algoritmos aplicados nele exige 15082-NN, ou seja, 15082 vizinhos mais próximos, praticamente metade do conjunto. Se analisarmos, o conjunto todo tem 30162 dados e aplicando o 10-fold-cross validation obtemos 10 partes iguais com 3016 elementos, onde usamos uma dessas partes para teste e os demais 27144 elementos para obter os vizinhos mais próximos, ou seja, vamos mapear 55,56% dos elementos como vizinhos próximos, dados suficientes para ter uma precisão impecável no final da interação, como podemos ver na matrix confusion {'FN': 176859486, 'TP': 278013634}.