Algoritmos de Machine Learning

Árvore de Decisão
kNN
kNN improve

vamos lá

Waveform Database Generator

https://archive.ics.uci.edu/ml/datasets/Waveform+Database+Generator+%28Version+2%29

Nessa base de dados temos, 3 classes de onda, 40 atributos que incluem ruído e possui 5000 instâncias.

Usos do passado:
No livro temos destacado:
-- Optimal Bayes classification rate: 86% accuracy
-- CART decision tree algorithm: 72%
-- Nearest Neighbor Algorithm: 38%
-- 300 training and 5000 test instances

vamos lá

Waveform Database Generator

https://archive.ics.uci.edu/ml/datasets/Waveform+Database+Generator+%28Version+2%29

Data Set Characteristics: Multivariate, Data-Generator

Attribute Characteristics: Real

Associated Tasks: Classification

Number of Instances: 5000

Number of Attributes: 40

Missing Values? No

Date Donated: 1988-11-10

vamos lá

Waveform Database Generator

https://archive.ics.uci.edu/ml/datasets/Waveform+Database+Generator+%28Version+2%29

Exemplo dos dados do dataset:

-1.23,-1.56,-1.75,-0.28,0.60,2.22,0.85,0.21,-0.20,0.89,1.08,4.20,2.89,7.75,4.59,3.15,5.12,3.32,1.20,0.24,-0.56,2

-0.69,2.43,0.61,2.08,2.30,3.25,5.52,4.55,2.97,2.22,2.81,1.61,1.24,1.89,1.88,-1.34,0.83,1.41,1.78,0.60,2.42,1

-0.12,-0.94,1.29,2.59,2.42,3.55,4.94,3.25,1.90,2.07,0.51,1.45,2.50,0.12,1.41,2.78,0.64,0.62,-0.01,-0.79,-0.12,0

0.86,0.29,2.19,-0.02,1.13,2.51,2.37,5.45,5.45,4.84,4.65,4.05,2.58,1.40,1.24,1.41,1.07,-1.43,2.84,-1.18,1.12,1

1.16,0.37,0.40,-0.59,2.66,1.00,2.69,4.06,5.34,3.53,4.82,4.79,4.30,1.84,1.73,0.21,-0.18,0.13,-0.21,-0.80,-0.68,1

vamos lá

Image Segmentation Dataset

https://archive.ics.uci.edu/ml/datasets/Image+Segmentation

As instâncias foram desenhadas aleatoriamente a partir de uma base de dados de 7 imagens de outdoor.
As imagens foram segmentadas à mão para criar uma classificação de cada pixel.
Cada instância é uma região de 3x3.

vamos lá

Image Segmentation Dataset

https://archive.ics.uci.edu/ml/datasets/Image+Segmentation

Data Set Characteristics: Multivariate
Attribute Characteristics: Real
Associated Tasks: Classification
Number of Instances: 2310
Number of Attributes: 19
Missing Values? No
Date Donated: 1990-11-01

vamos lá

Image Segmentation Dataset

https://archive.ics.uci.edu/ml/datasets/Image+Segmentation

Informações dos atributos:

1. region-centroid-col: the column of the center pixel of the region.
2. region-centroid-row: the row of the center pixel of the region.

...

17. value-mean: 3-d nonlinear transformation of RGB. (Algorithm can be found in Foley and VanDam, Fundamentals of Interactive Computer Graphics)
18. saturatoin-mean: (see 17)
19. hue-mean: (see 17)

vamos lá

Image Segmentation Dataset

https://archive.ics.uci.edu/ml/datasets/Image+Segmentation

Exemplo dos dados do dataset:

1,110.0,189.0,9,0.0,0.0,1.0000004,0.66666675,1.2222223,1.1863422,12.925926,10.888889,9.222222,18.666668,-6.111111,-11.111111,17.222221,18.666668,0.50813884,1.9108642

1,86.0,187.0,9,0.0,0.0,1.1111108,0.7200825,1.4444445,0.750309,13.740741,11.666667,10.333334,19.222221,-6.2222223,-10.222222,16.444445,19.222221,0.46332908,1.9414649

1,225.0,244.0,9,0.0,0.0,3.3888886,2.1951127,2.9999998,1.5202343,12.259259,10.333334,9.333334,17.11111,-5.7777777,-8.777778,14.555555,17.11111,0.48014903,1.9879022

1,47.0,232.0,9,0.0,0.0,1.2777778,1.254621,1.0,0.8944273,12.703704,11.0,9.0,18.11111,-5.111111,-11.111111,16.222221,18.11111,0.50096595,1.875362

1,97.0,186.0,9,0.0,0.0,1.1666671,0.6912147,1.1666671,1.0055404,15.592592,13.888889,11.777778,21.11111,-5.111111,-11.444445,16.555555,21.11111,0.44266057,1.8636538

Vamos?

Resultados

Decision Tree

WAVEFORM

ENTROPY 77%, 76%, 75%, 76%, 75%, 76%, 76%, 76%, 75%, 75%, 75%, 76%, 76% (75.69%)

GINI 76%, 75%, 74%, 75%, 76%, 75%, 75%, 75%, 76%, 74%, 74%, 75%, 74% (74.92%)

SEGMENTATION

ENTROPY 89%, 92%, 91%, 90%, 91%, 91%, 92%, 90%, 90%, 91%, 91%, 92%, 91% (90.84%)

GINI 79%, 82%, 78%, 80%, 83%, 79%, 81%, 78%, 84%, 82%, 80%, 80%, 81% (80.54%)

kNN

WAVEFORM

MANHATTAN

K=1 = 78%, 78%, 76%, 76%, 78%, 77%, 76%, 79%, 76%, 75%, 77%, 75%

K=2 = 75%, 76%, 76%, 77%, 75%, 74%, 76%, 74%, 75%, 74%, 75%, 76%

K=3 = 81%, 80%, 79%, 80%, 80%, 80%, 80%, 78%, 81%, 82%, 81%, 82%

K=7 = 80%, 80%, 77%, 81%, 80%, 82%, 81%, 81%, 82%, 81%, 82%, 80%

K=11 = 85%, 85%, 84%, 86%, 84%, 85%, 84%, 85%, 82%, 85%, 83%, 84%

K=21 = 85%, 84%, 84%, 85%, 85%, 85%, 86%, 83%, 85%, 84%, 85%, 83%

K=41 = 83%, 84%, 84%, 88%, 87%, 85%, 85%, 84%, 85%, 83%, 85%, 85%

K=101 = 85%, 85%, 85%, 85%, 87%, 86%, 85%, 86%, 87%, 85%, 85%, 85%

kNN

WAVEFORM

EUCLIDEAN

K=1 = 78%, 78%, 76%, 78%, 77%, 77%, 76%, 79%, 79%, 79%, 78%, 77%

K=2 = 78%, 78%, 75%, 76%, 75%, 78%, 78%, 78%, 77%, 78%, 79%, 76%

K=3 = 82%, 82%, 82%, 82%, 82%, 80%, 80%, 81%, 82%, 81%, 81%, 83%

K=7 = 82%, 82%, 85%, 82%, 82%, 84%, 83%, 82%, 85%, 82%, 84%, 83%

K=11 = 82%, 84%, 85%, 86%, 85%, 86%, 84%, 86%, 84%, 84%, 85%, 83%

K=21 = 85%, 85%, 86%, 86%, 85%, 85%, 84%, 86%, 85%, 84%, 84%, 85%

K=41 = 84%, 85%, 84%, 86%, 85%, 87%, 86%, 85%, 86%, 85%, 84%, 86%

K=101 = 85%, 84%, 85%, 87%, 85%, 88%, 84%, 86%, 88%, 84%, 85%, 85%

kNN

SEGMENTATION

MANHATTAN

K=1 = 96%, 96%, 98%, 97%, 96%, 95%, 96%, 95%, 97%, 98%, 97%, 96%

K=2 = 95%, 95%, 95%, 95%, 96%, 95%, 95%, 95%, 97%, 93%, 95%, 95%

K=3 = 95%, 96%, 96%, 96%, 96%, 95%, 96%, 96%, 95%, 96%, 97%, 95%

K=7 = 94%, 93%, 95%, 94%, 97%, 96%, 95%, 94%, 95%, 96%, 96%, 94%

K=11 = 94%, 93%, 95%, 94%, 93%, 93%, 94%, 96%, 95%, 93%, 95%, 95%

K=21 = 94%, 94%, 95%, 92%, 90%, 93%, 92%, 91%, 92%, 92%, 94%, 92%

K=41 = 89%, 92%, 88%, 90%, 89%, 91%, 89%, 92%, 91%, 90%, 91%, 92%

K=101 = 86%, 84%, 82%, 86%, 88%, 86%, 85%, 83%, 85%, 85%, 83%, 89%

kNN

SEGMENTATION

EUCLIDEAN

K=1 = 96%, 96%, 95%, 94%, 97%, 97%, 95%, 98%, 95%, 96%, 95%, 95%

K=2 = 94%, 93%, 96%, 95%, 95%, 95%, 94%, 95%, 94%, 94%, 95%, 95%

K=3 = 95%, 93%, 93%, 95%, 94%, 94%, 95%, 93%, 95%, 97%, 95%, 95%

K=7 = 94%, 93%, 94%, 94%, 93%, 93%, 95%, 93%, 91%, 92%, 93%, 92%

K=11 = 92%, 92%, 89%, 88%, 94%, 89%, 93%, 92%, 90%, 92%, 92%, 92%

K=21 = 90%, 89%, 93%, 90%, 89%, 88%, 86%, 90%, 91%, 90%, 88%, 88%

K=41 = 86%, 87%, 86%, 86%, 84%, 87%, 88%, 86%, 86%, 84%, 84%, 84%

K=101 = 80%, 80%, 79%, 78%, 81%, 81%, 82%, 82%, 82%, 81%, 79%, 80%

kNN improve

WAVEFORM

#K=1 = 83%, 84%, 84%, 84%, 83%, 83%, 82%, 83%, 83%, 82%, 83%, 83%, 83%

#K=2 = 86%, 84%, 81%, 83%, 86%, 83%, 85%, 85%, 85%, 85%, 84%, 82%, 85%

#K=3 = 86%, 84%, 82%, 85%, 86%, 84%, 84%, 85%, 84%, 84%, 84%, 85%, 84%

#K=7 = 86%, 85%, 85%, 86%, 86%, 86%, 85%, 86%, 85%, 86%, 85%, 87%, 86%

#K=11 = 85%, 86%, 85%, 86%, 87%, 86%, 85%, 87%, 88%, 86%, 86%, 86%, 86%

#K=21 = 86%, 87%, 87%, 86%, 86%, 86%, 86%, 85%, 87%, 88%, 87%, 86%, 87%

#K=41 = 87%, 86%, 87%, 87%, 87%, 87%, 87%, 86%, 86%, 88%, 85%, 86%, 87%

#K=101 = 86%, 89%, 87%, 87%, 85%, 87%, 89%, 86%, 86%, 86%, 87%, 88%, 85%

kNN improve

SEGMENTATION

#K=1 = 95%, 94%, 95%, 95%, 95%, 95%, 95%, 95%, 95%, 96%, 96%, 96%, 96%

#K=2 = 94%, 97%, 95%, 96%, 95%, 95%, 95%, 95%, 97%, 95%, 96%, 93%, 97%

#K=3 = 94%, 94%, 93%, 95%, 95%, 94%, 93%, 93%, 94%, 94%, 94%, 92%, 95%

#K=7 = 93%, 93%, 92%, 93%, 91%, 92%, 91%, 90%, 93%, 93%, 89%, 92%, 92%

#K=11 = 90%, 89%, 87%, 88%, 92%, 91%, 89%, 88%, 88%, 89%, 90%, 90%, 91%

#K=21 = 82%, 83%, 83%, 85%, 84%, 85%, 86%, 81%, 84%, 85%, 85%, 85%, 83%

#K=41 = 76%, 78%, 79%, 79%, 76%, 77%, 74%, 77%, 76%, 77%, 81%, 79%, 74%

#K=101 = 62%, 56%, 61%, 62%, 63%, 61%, 59%, 63%, 63%, 62%, 60%, 64%, 63%