

José Cleison de Lima – Graduação Tecnológica em Ciência de Dados

Atividade prática – 2:

Segue imagem com a quantidade de instancias e atributos junto com o erro RMS ao realizar o agrupamento.

```
Clusterer output

=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A
Relation:    A4
Instances:   4773
Attributes:  6
             Idade
             Atraso
             Valor
             CONTATO
             EFETIVO
             Acordo
Test mode:   evaluate on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 10
Within cluster sum of squared errors: 1861.1424641830763

Initial starting points (random):
```

Abaixo o resultado do cluster em dois atributos 0 e 1. 3906 instâncias com acordo e 867 instâncias sem acordo.

```
Cluster 0: 4,1,1,4,1,1
Cluster 1: 4,2,5,10,2,1

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
              (4773.0)    (3906.0)    (867.0)
=====
Idade          1.7461         1.7087         1.9146
Atraso         2.4588         2.5507         2.045
Valor          1.4718         1.3525         2.0092
CONTATO        4.7591         4.649          5.2549
EFETIVO        1.1672         1.1792         1.113
Acordo         0.8184         1              0

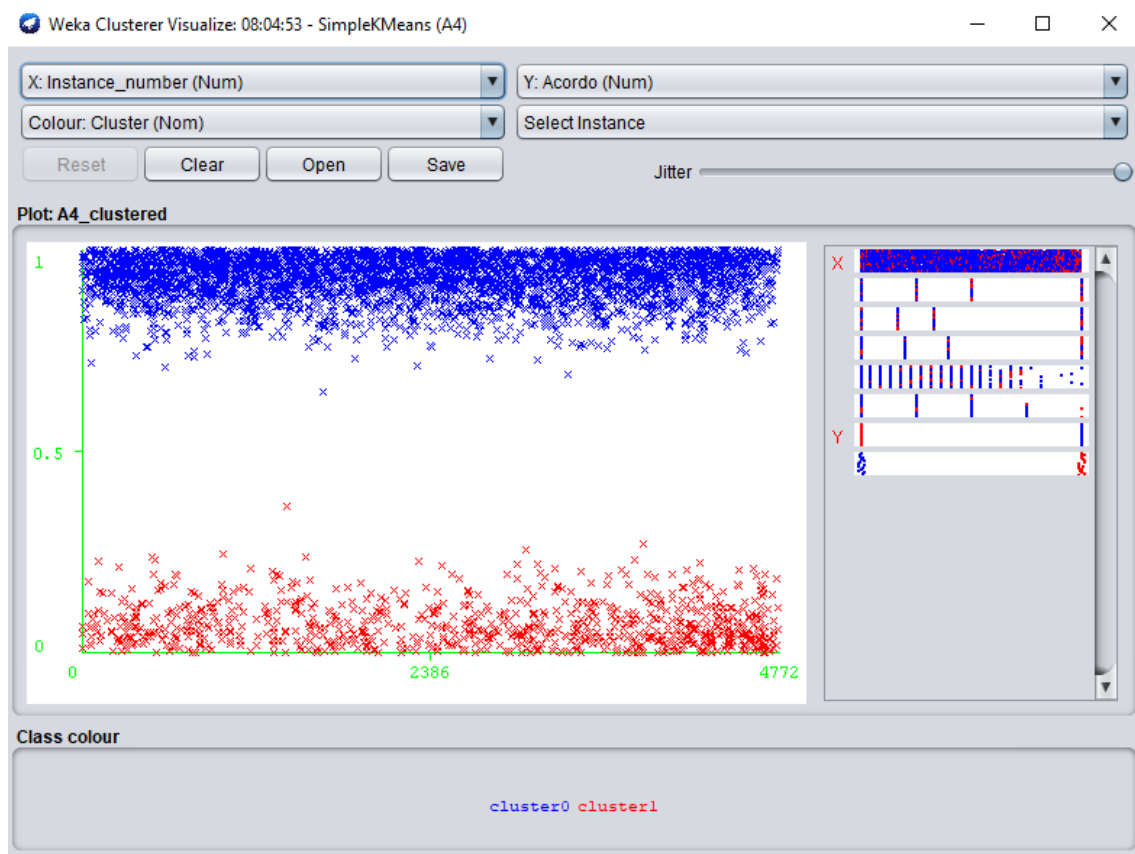
Time taken to build model (full training data) : 0.05 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      3906 ( 82%)
1       867 ( 18%)
```

Ao observarmos o gráfico abaixo conseguimos notar os clientes que **não fecham acordo em vermelho** e os clientes que **fecham acordo em azul** e as instâncias nos grupos não se misturam.



Atividade prática – 2:

```

=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 16 -A
Relation:    iris
Instances:   150
Attributes:  5
    sepal.length
    sepal.width
    petal.length
    petal.width
    variety
Test mode:   evaluate on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 6
Within cluster sum of squared errors: 1.884232737321628

Initial starting points (random):

```

Within cluster sum of squared errors: 1.884232737321628

Initial starting points (random):

Cluster 0: 6.1,2.9,4.7,1.4,Versicolor
Cluster 1: 6.2,2.9,4.3,1.3,Versicolor
Cluster 2: 6.9,3.1,5.1,2.3,Virginica
Cluster 3: 5.5,4.2,1.4,0.2,Setosa
Cluster 4: 6.9,3.1,4.9,1.5,Versicolor
Cluster 5: 6.1,3,4.6,1.4,Versicolor
Cluster 6: 4.9,3.6,1.4,0.1,Setosa
Cluster 7: 4.4,3,1.3,0.2,Setosa
Cluster 8: 5.5,2.4,3.7,1,Versicolor
Cluster 9: 4.3,3,1.1,0.1,Setosa
Cluster 10: 6,2.7,5.1,1.6,Versicolor
Cluster 11: 5.7,2.5,5,2,Virginica
Cluster 12: 4.6,3.1,1.5,0.2,Setosa
Cluster 13: 7.4,2.8,6.1,1.9,Virginica
Cluster 14: 5.9,3,5.1,1.8,Virginica
Cluster 15: 6.9,3.2,5.7,2.3,Virginica

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (150.0)	Cluster#								
		0 (8.0)	1 (12.0)	2 (18.0)	3 (7.0)	4 (9.0)	5 (6.0)	6 (24.0)	7 (1.0)	8 (11.0)
sepal.length	5.8433	6.1875	5.6417	6.5889	5.5286	6.7111	5.85	5.0958	4.5	5.3909
sepal.width	3.0573	2.875	2.7833	3.1333	4.0429	3.0444	3.15	3.5333	2.3	2.3727
petal.length	3.758	4.4625	4.1	5.5222	1.4714	4.6667	4.5333	1.5083	1.3	3.6364
petal.width	1.1993	1.3625	1.25	2.2444	0.2857	1.4556	1.5833	0.2708	0.3	1.0818
variety	Setosa	Versicolor	Versicolor	Virginica	Setosa	Versicolor	Versicolor	Setosa	Setosa	Versicolor

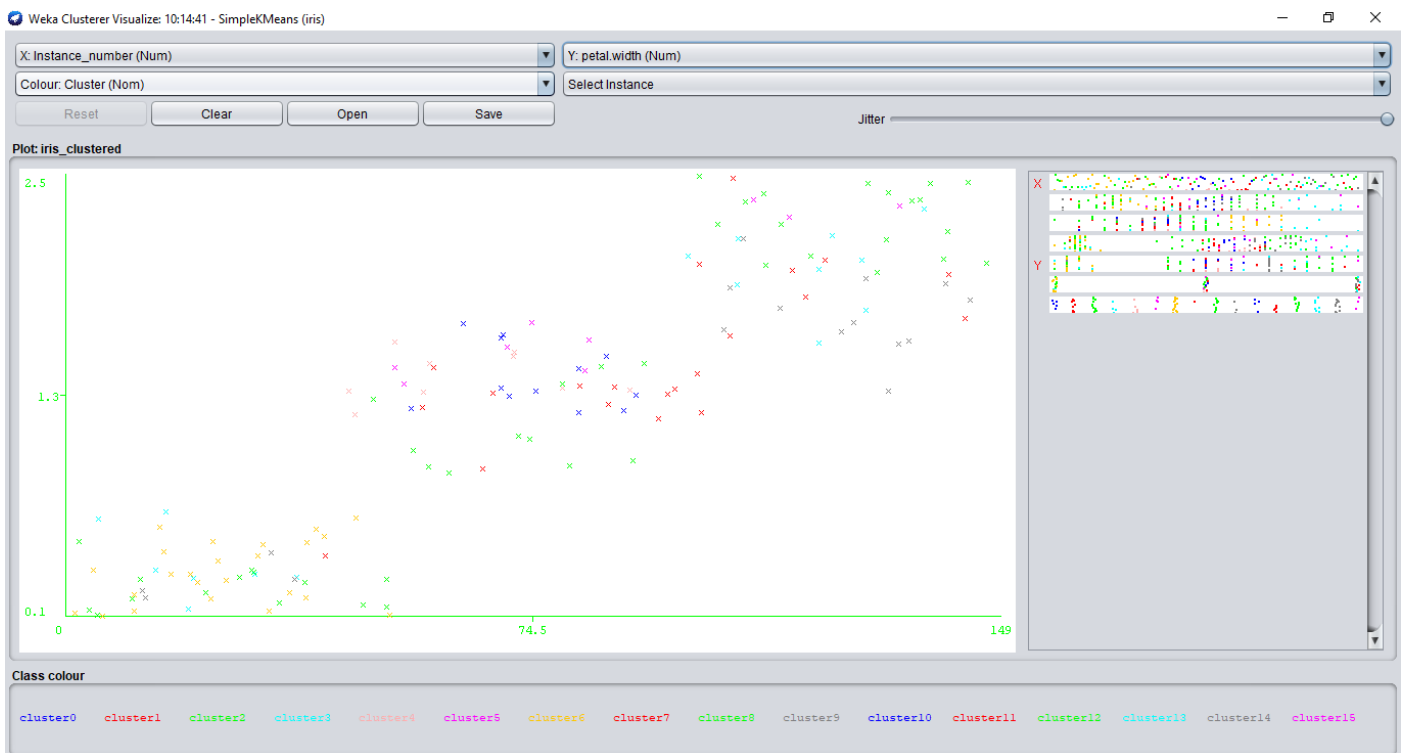
9 (4.0)	10 (4.0)	11 (8.0)	12 (14.0)	13 (9.0)	14 (12.0)	15 (3.0)
4.375	6.2	5.7375	4.8071	7.4333	6.2667	7.6
3.025	2.425	2.5875	3.1357	2.9222	2.8417	3.7333
1.275	4.725	4.9625	1.4429	6.2667	5.2417	6.4
0.175	1.475	1.9125	0.2	1.9889	1.75	2.2333
Setosa	Versicolor	Virginica	Setosa	Virginica	Virginica	Virginica

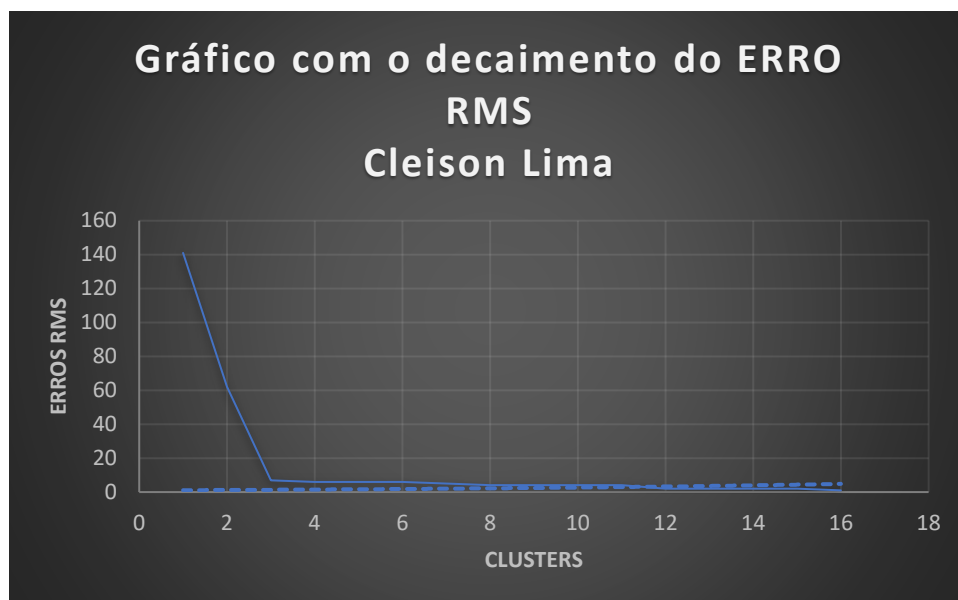
Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	8	(5%)
1	12	(8%)
2	18	(12%)
3	7	(5%)
4	9	(6%)
5	6	(4%)
6	24	(16%)
7	1	(1%)
8	11	(7%)
9	4	(3%)
10	4	(3%)
11	8	(5%)
12	14	(9%)
13	9	(6%)
14	12	(8%)
15	3	(2%)





Conforme mostra o gráfico acima a melhor quantidade de clusters é “3”, a partir desse número o decaimento varia.