

ECT_HW4_109403021_Weka 題

Python 找到的最佳三個欄位為('energy', 'acousticness', 'loudness')

1. 先將 3 個欄位以外的欄位都刪除

No.	Name
1	<input type="checkbox"/> energy
2	<input checked="" type="checkbox"/> liveness
3	<input checked="" type="checkbox"/> tempo
4	<input checked="" type="checkbox"/> speechiness
5	<input type="checkbox"/> acousticness
6	<input checked="" type="checkbox"/> instrumentalness
7	<input checked="" type="checkbox"/> time_signature
8	<input checked="" type="checkbox"/> danceability
9	<input checked="" type="checkbox"/> key
10	<input checked="" type="checkbox"/> duration_ms
11	<input type="checkbox"/> loudness
12	<input checked="" type="checkbox"/> valence
13	<input checked="" type="checkbox"/> mode
14	<input checked="" type="checkbox"/> type
15	<input checked="" type="checkbox"/> uri

Remove

Status Remove selected attributes.

2. 再將資料標準化

Filter

Choose Standardize

3. 從 cluster 選 SimpleKMeans 並將 numCluster 設為 4、seed 設為 15

weka.gui.GenericObjectEditor

weka.clusterers.SimpleKMeans

About

Cluster data using the k means algorithm.

More

Capabilities

canopyMaxNumCanopiesToHoldInMemory 100

canopyMinimumCanopyDensity 2.0

canopyPeriodicPruningRate 10000

canopyT1 -1.25

canopyT2 -1.0

debug False

displayStdDevs False

distanceFunction Choose EuclideanDistance -R first-last

doNotCheckCapabilities False

dontReplaceMissingValues False

fastDistanceCalc False

initializationMethod Random

maxIterations 500

numClusters 4

numExecutionSlots 1

preserveInstancesOrder False

reduceNumberOfDistanceCalcsViaCanopies False

seed 15

Open... Save... OK Cancel

4. 得到結果

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Cluster

Choose SimpleKMeans -ini 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 1.25 -t2 1.0 -N 4 -A "weka.core.EuclideanDistance -R first-last" -J 500 -min-clusts 1 -S 15

Cluster mode

☒ Use training set
☐ Supplied test set Set...
☐ Percentage split % 66
☐ Classes to clusters evaluation
Global loadings
☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

7/26/26 - SimpleKMeans

Cluster output

Means
=====

Number of iterations: 26
Within cluster sum of squared errors: 42.10330926720498

Initial starting points (random):
Cluster 0: 0.950941,-0.504695,-2.725
Cluster 1: 0.859395,-0.501119,-4.823
Cluster 2: -0.301805,-0.504734,-12.218
Cluster 3: 1.061761,-0.504485,-5.648

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data	Cluster# 0	Cluster# 1	Cluster# 2	Cluster# 3
	(1923.0)	(1011.0)	(259.0)	(182.0)	(471.0)
energy	-0	0.7376	-1.1987	-1.9438	-0.173
acousticness	-0	-0.445	0.4572	2.7202	-0.3043
loudness	-6.0253	-4.3293	-8.3945	-11.3053	-6.3227

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

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

Clustered Instances

0	1011 (53%)
1	259 (13%)
2	182 (9%)
3	471 (24%)

Status
OK

Log 0