

Data warehouse & Data mining lab

Assignment 3a

Submitted by :
Shiv Kumar
2016UIT2563
Semester 8
IT2

Experiment 9 : Write a procedure for Clustering Weather data using EM Algorithm.

1. Data Loading

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose Apply Stop

Current relation
Relation: weather
Instances: 14
Attributes: 5
Sum of weights: 14

Attributes
All | None | Invert | Pattern

No.	Name
1	<input checked="" type="checkbox"/> outlook
2	<input type="checkbox"/> temperature
3	<input type="checkbox"/> humidity
4	<input type="checkbox"/> windy
5	<input type="checkbox"/> play

Remove

Selected attribute
Name: outlook
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

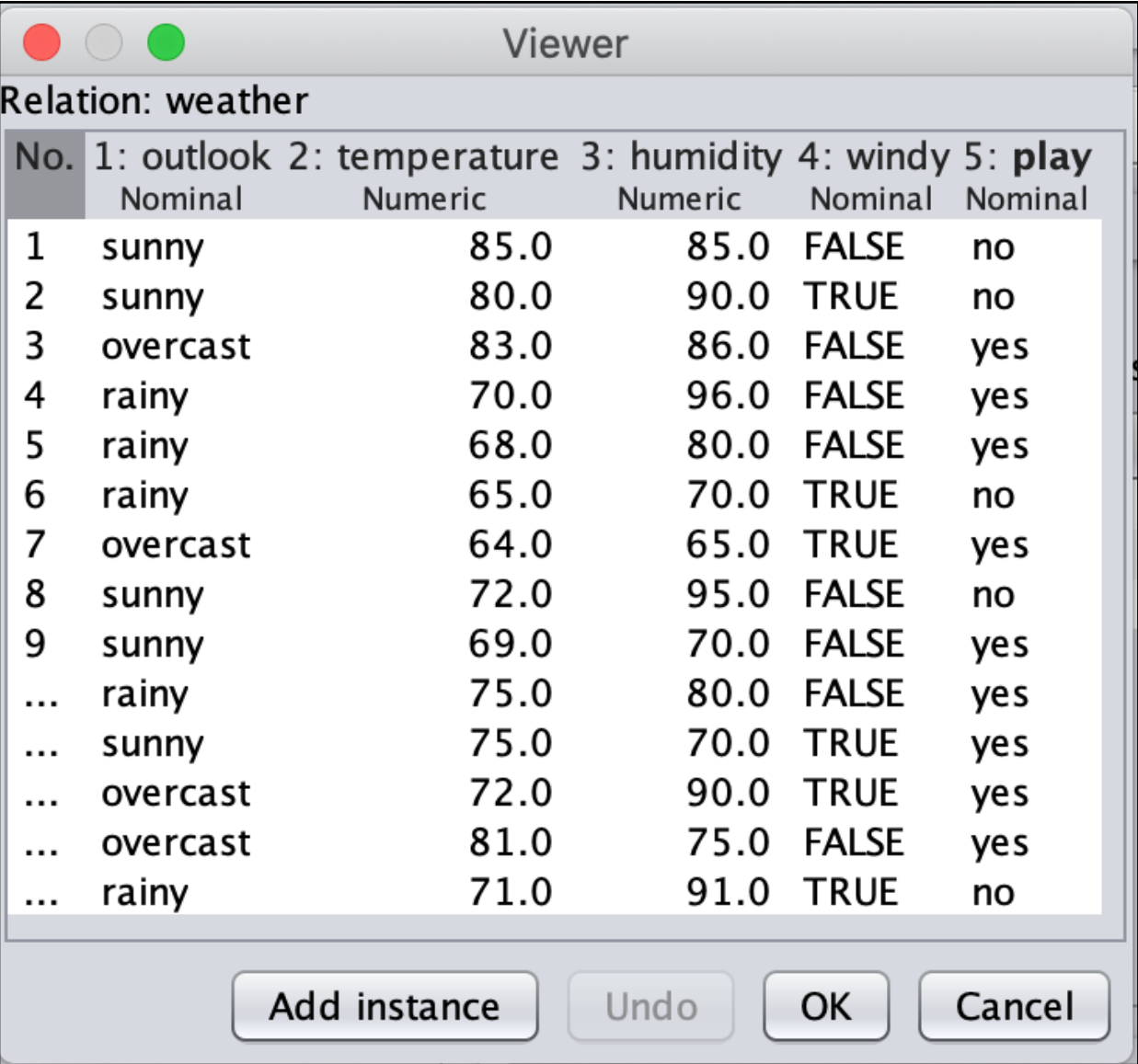
No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

Class: play (Nom) Visualize All

Outlook	Blue (play = no)	Red (play = yes)	Total
sunny	3	2	5
overcast	4	0	4
rainy	3	2	5

Status: OK Log x 0

2. Review of Loaded Data



Relation: weather					
No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Numeric	Numeric	Nominal	Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	70.0	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	70.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny	69.0	70.0	FALSE	yes
...	rainy	75.0	80.0	FALSE	yes
...	sunny	75.0	70.0	TRUE	yes
...	overcast	72.0	90.0	TRUE	yes
...	overcast	81.0	75.0	FALSE	yes
...	rainy	71.0	91.0	TRUE	no

Add instance Undo OK Cancel

3. Result after EM algorithm

The screenshot shows the Weka Explorer interface with the 'Cluster' tab selected. The 'Clusterer' dropdown is set to 'EM' with the command line: `EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100`.

Cluster mode:

- ☒ Use training set
- ☐ Supplied test set (Set...)
- ☐ Percentage split (% 66)
- ☐ Classes to clusters evaluation (Norm) play
- ☒ Store clusters for visualization

Clusterer output:

```
=== Run information ===
Scheme:      weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Relation:    weather
Instances:   14
Attributes:  5
              outlook
              temperature
              humidity
              windy
              play
Test mode:   evaluate on training data

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

EM
==
|
Number of clusters selected by cross validation: 1
Number of iterations performed: 2

Attribute      Cluster
              0
              (1)
=====
outlook
sunny          6
overcast       5
rainy          6
[total]       17
temperature
mean          73.5714
std. dev.     6.3326

humidity
mean          81.6429
std. dev.     9.9111

windy
TRUE          7
```

Result list (right-click for options):

06:21:44 - EM

Status: OK

Log x 0

Clusterer output

Number of clusters selected by cross validation: 1
Number of iterations performed: 2

Attribute	Cluster
	0
	(1)
=====	
outlook	
sunny	6
overcast	5
rainy	6
[total]	17
temperature	
mean	73.5714
std. dev.	6.3326
humidity	
mean	81.6429
std. dev.	9.9111
windy	
TRUE	7
FALSE	9
[total]	16
play	
yes	10
no	6
[total]	16

Time taken to build model (full training data) : 0.06 seconds

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

Clustered Instances

0 14 (100%)

Log likelihood: -9.4063