INT104 Artificial Intelligence

Review

Fang Kang
Sichen Liu
Shengchen Li

Xi'an Jiaotong-Liverpool University

7th May 2024

Naïve Bayes

Outlook	Temprature	Humidity	Windy	Play
Overcast	Hot	High	False	Yes
Overcast	Cool	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Rainy	Mild	Normal	False	Yes
Rainy	Mild	High	True	No
Sunny	Hot	High	True	No
Sunny	Hot	High	False	No
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Sunny	Mild	Normal	True	Yes

Will you play on the day of Mild?

Solution

Naïve Bayes

Temp.	Yes	No	p
Hot	2	2	0.28
Mild	4	2	0.43
Cool	3	1	0.28
p	0.64	0.36	

By this table we have $p(\text{Mild}|\text{Yes}) = \frac{4}{9} = 0.44$ and $p(\text{Mild}|\text{No}) = \frac{2}{5} = 0.4$

Posterior $p(\text{Yes}|\text{Mild}) = \dfrac{p(\text{Mild}|\text{Yes})p(\text{Yes})}{p(\text{Mild})} = \dfrac{0.44 \times 0.64}{0.43} = 0.65$

Posterior $p(\text{No}|\text{Mild}) = \dfrac{p(\text{Mild}|\text{No})p(\text{No})}{p(\text{Mild})} = \dfrac{0.4 \times 0.36}{0.43} = 0.33$

As $p(\text{Yes}|\text{Mild}) > p(\text{No}|\text{Mild})$, it is likely to play.

Exercise

Will the following condition be considered as a proper day for play?

- `Sunny, Windy`

- `Overcast, Normal Humidity & Cool`

CCP Tables

Outlook	Yes	No	p
Overcast	4	0	0.28
Rainy	3	2	0.36
Sunny	2	3	0.36
p	0.64	0.36	

Wind	Yes	No	p
True	3	3	0.43
False	6	2	0.57
p	0.64	0.36	

Humid	Yes	No	p
Normal	6	1	0.5
High	3	4	0.5
p	0.64	0.36	

Solution

`Sunny, Windy`

$$p(\text{Yes}|\text{Sunny, Windy}) = \frac{p(\text{Sunny, Windy}|\text{Yes})p(\text{Yes})}{p(\text{Sunny, Windy})} \propto$$

$p(\text{Sunny, Windy}|\text{Yes})p(\text{Yes}) = p(\text{Sunny}|\text{Yes})p(\text{Windy}|\text{Yes})p(\text{Yes}) =$
$0.22 \times 0.33 \times 0.64 = 0.05$

$$p(\text{No}|\text{Sunny, Windy}) = \frac{p(\text{Sunny, Windy}|\text{No})p(\text{No})}{p(\text{Sunny, Windy})} \propto$$

$p(\text{Sunny, Windy}|\text{No})p(\text{No}) = p(\text{Sunny}|\text{No})p(\text{Windy}|\text{No})p(\text{No}) =$
$0.6 \times 0.6 \times 0.36 = 0.13$

As $p(\text{Yes}|\text{Sunny, Windy}) < p(\text{No}|\text{Sunny, Windy})$, so the combination of weather is unlikely to be suitable for playing

`Rainy, Normal Humidity & Cool`

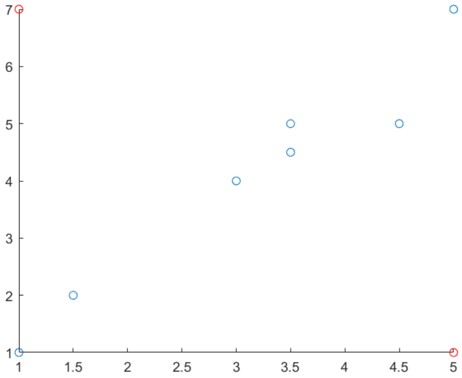
$p(\text{Yes}|\text{Rainy, Normal, Cool}) = \dfrac{p(\text{Rainy, Normal, Cool}|\text{Yes})p(\text{Yes})}{p(\text{Rainy, Normal, Cool})} \propto$
$p(\text{Rainy, Normal, Cool}|\text{Yes})p(\text{Yes}) =$
$p(\text{Rainy}|\text{Yes})p(\text{Normal}|\text{Yes})p(\text{Cool}|\text{Yes})p(\text{Yes}) = 0.33 \times 0.67 \times 0.33 \times$
$0.64 = 0.047$

$p(\text{No}|\text{Rainy, Normal, Cool}) = \dfrac{p(\text{Rainy, Normal, Cool}|\text{No})p(\text{No})}{p(\text{Rainy, Normal, Cool})} \propto$
$p(\text{Rainy, Normal, Cool}|\text{No})p(\text{No}) =$
$p(\text{Rainy}|\text{No})p(\text{Normal}|\text{No})p(\text{Cool}|\text{No})p(\text{No}) = 0.4 \times 0.2 \times 0.2 \times 0.36 =$
0.0058

As $p(\text{Yes}|\text{Rainy, Normal, Cool}) > p(\text{No}|\text{Rainy, Normal, Cool})$, it is likely to play

k-means

Given the following 7 points: A (1.0, 1.0), B (1.5, 2.0), C (3.0, 4.0), D (5.0, 7.0), E (3.5, 5.0), F (4.5, 5.0), G (3.5, 4.5), use *k*-means algorithm to divide all points into two clusters with two initial points A and F.



Solution

1st iteration

The following table shows the distance between samples and centroids

	A	F
A	0	7.5
B	1.5	6
C	5	2.5
D	10	2.5
E	6.5	1
F	7.5	0
G	6	1.5

So A, B belong to cluster Z1, whose centroid is (1.25, 1.5). C, D, E, F, G belong to cluster Z2, whose centroid is (3.9, 5.1).

Solution

3rd iteration

The following table shows the distance between samples and centroids

	Z1'	Z2'
A	0.75	7.3
B	0.75	5.8
C	4.25	2.3
D	9.25	2.7
E	5.75	0.8
F	6.75	1
G	5.25	1.3

So A, B belong to cluster $Z1'$. C, D, E, F, G belong to cluster $Z2'$. There are no points that changes the cluster hence the algorithm is converged.

Examples

Given the following table that shows the distance between samples ("city block distance"), using agglomerative clustering method, draw the final dendrogram obtained.

	A	B	C	D	E
A	0				
B	8	0			
C	3	6	0		
D	5	5	8	0	
E	13	10	2	7	0

Single Linkage

	A	B	D	CE
A	0			
B	8	0		
D	5	5	0	
CE	3	6	7	0

Table: 1st round

	B	D	ACE
B	0		
D	5	0	
ACE	6	5	0

Table: 2nd round

The resulting dendrograms are:

Single Linkage

	BD	ACE
BD	0	
ACE	5	0

Table: 3rd round

The resulting dendrograms are:

kNN

kNN could also be tested though it is very easy