THADOMAL SHAHANI
OWM
THADOMAL SHAHANI TSEC ENGINEERING COLLEGE

Naive Bayes Classifier.

Car Theft example Attributes are color type original & the subject, stolen be either yes or no.

Data set:

| Car No | Color | type | orignal | stolen. |
|---|---|---|---|---|
| 1 | Red | Sports | Domestic | Yes |
| 2 | Red | Sports | Domestic | No |
| 3 | Red | Sports | Domistic | Yes |
| 4 | Yellow | Sports | Domistic | No |
| 5 | Yellow | Sports | Imported | Yes |
| 6 | Yellow | SUV | Imported | No |
| 7 | Yellow | SUV | Imported | Yes |
| 8 | Yellow | SUV | Domestic. | No |
| 9 | Red | SUV | Imported | No |
| 10 | Red. | Sports | Imported | Yes. |

We want to classify a < Red, domestic, SUV unseen sample Note there is no example of a ~~ret~~ Domestic SUV. SUVs in our data set.

P ( Yes ) = 5/10
P ( NO ) = 5/10.

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

Color:

$P(\text{Red} / \text{Yes}) = 3/5$ $\quad$ $P(\text{Red} / \text{No}) = 2/5$

$P(\text{Yellow} / \text{Yes}) = 2/5$ $\quad$ $P(\text{Yellow} / \text{No}) = 3/5$

Type.

$P(\text{SUV} / \text{Yes}) = 1/5$ $\quad$ $P(\text{SUV} / \text{No}) = 3/5$

$P(\text{Sports} / \text{Yes}) = 4/5$ $\quad$ $P(\text{Sports} / \text{No}) = 2/5$

Origin.

$P(\text{Domestic} / \text{Yes}) = 2/5$ $\quad$ $P(\text{Domestic} / \text{No}) = 3/5$

$P(\text{Imported} / \text{Yes}) = 3/5$ $\quad$ $(\text{Imported} / \text{NO}) = 2/5$

An unseen sample $x = \langle \text{Red}, \text{Domestic}, \text{SUV} \rangle$

$P(x/\text{Yes}) \cdot P(\text{Yes})$.

$= P(\text{Red}, \text{Yes}) \cdot P(\text{Domestic} . \text{Yes}) \cdot P(\text{SUV}/\text{Yes}) \cdot P(\text{Yes})$

$= \frac{3}{5} \times \frac{2}{5} \times \frac{1}{5} \times \frac{5}{10}$

$= 0.024$

$P(x / \text{NO}) \cdot P(\text{NO}$

$= P(\text{Red} / \text{No}) \cdot P(\text{Domestic} / \text{NO}) \cdot P(\text{SUV} / \text{NO}) \cdot P(\text{NO})$

$= \frac{2}{5} \times \frac{3}{5} \times \frac{3}{5} \times \frac{5}{10}$

$= 0.072$.

Since $0.072 > 0.024$, our example gets classific as 'No'.

Apriori example :

| T/D | items |
|---|---|
| T1 | I1 I2 I5 |
| T2 | I2 I4 |
| T3 | I2 I3 |
| T4 | I1 I2 I4 |
| T5 | I1 I3 |
| T6 | I2 I3 |
| T7 | I1 I3 |
| T8 | I1 I2 I3 I5 |
| T9 | I1 I2 I3 |
| ~~I9~~ | |

minimum confidence = 50 %.
minimum support cost is 2 .

i) k = 1

| | |
|---|---|
| I1 | 6 |
| I2 | 7 |
| I3 | 6 |
| I4 | 2 |
| I5 | 2 |

ii)

| Itemset | Sup_count |
|---|---|
| $I_1, I_2$ | 4 |
| $I_1, I_3$ | 4 |
| $I_1, I_4$ | 1 |
| $I_1, I_5$ | 2 |
| $I_2, I_3$ | 4 |
| $I_2, I_4$ | 2 |
| $I_2, I_5$ | 2 |
| $I_3, I_4$ | 0 |
| $I_3, I_5$ | 1 |
| $I_4, I_5$ | 0 |

$\because$ remove

$I_3, I_4$ $0 < 2$

$I_3, I_5$ $1 < 2$

$I_4, I_5$ $0 < 2$.

iii)

| | |
|---|---|
| $I_1, I_2, I_3$ | 2 |
| $I_1, I_2, I_5$ | 2 |

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

Now,

confidence (A → B) = support count (A ∪ B) / support count (A).

| | | |
|---|---|---|
| (I1, I2) → I3 | 2/4 | 50% |
| (I1, I3) → I2 | 2/4 | 50% |
| (I2, I3) → I1 | 2/4 | 50% |
| I1 → (I2, I3) | 2/6 | 33% |
| I2 → (I1, I3) | 2/7 | 28% |
| I3 → (I1, I2) | 2/6 | 33% |

∵ min confidence is 50%.

∴ first three (I1, I2) → I3
(I1, I3) → I2
(I2, I3) → I1

can be consider as strong association rules.

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

Q). Perform Agglomerative Algorithm on the following data & plot a dendogram using single link Approach. the given data indicates the distance b/w elements.

| Item | E | A | C | B | D. |
|---|---|---|---|---|---|
| E | 0 | 1 | 2 | 2 | 3 |
| A | 1 | 0 | 2 | 5 | 3 |
| C | 2 | 2 | 0 | 1 | 6 |
| B | 2 | 5 | 1 | 0 | 3 |
| D | 3 | 3 | 6 | 3 | 0. |

Pair E, A.

| | (E, A) | C | B | D. |
|---|---|---|---|---|
| (E, A) | 0 | | | |
| C | 2 | 0 | | |
| B | 2 | 1 | 0 | |
| D | 3 | 6 | 3 | 0 |

Pair (B, C).

| | (E, A) | (B, C) | D. |
|---|---|---|---|
| (E, A) | 0 | | |
| (B, C) | 2 | 0 | |
| D | 3 | 3 | 0 |

Pair (E,A) & (B,C).

| | ((E,A), (B,C)) | D |
|---|---|---|
| (E,A), (B,C) | 0 | |
| D | 2 | 0 |

Pair (E,A,)(B,C), D.

| | (E, A,B,C,D) |
|---|---|
| (E,A,B,C,D) | 0 |