No.
Date : :

Homework 6 B11410G5 電機四乙 陳昱安

(1) Given a data sample with the values "systems", "junior", and "26...30" for the attributes department, status, and age, respectively, what would a naive Bayesian classification of the salary for the sample be?

Let salary be the class label attribute.

department	status	age	salary	count
sales	senior	31~35	46k~50k	30
sales	junior	26~30	26k~30k	40
sales	junior	31~35	31k~35k	40
systems	junior	21~25	46k~50k	20
systems	senior	31~35	66k~70k	5
systems	junior	26~30	46k~50k	3
systems	senior	41~45	66k~70k	13
marketing	senior	36~40	46k~50k	10
marketing	junior	31~35	41k~45k	4
secretary	senior	46~50	36k~40k	4
secretary	junior	26~30	26k~30k	6

解:(已知): systems, junior, 26~30 ⇒預測 salary

1° 計算先驗機率 P(salary): ① Total = 30+40+20+5+3+3+10+4+4+6 = 165 (筆)

⇒(1) $P(26k\sim30k) = \frac{40+6}{165} = \frac{46}{165}$ (2) $P(31k\sim35k) = \frac{40}{165}$ (3) $P(36k\sim40k) = P(41k\sim45k) = \frac{4}{165}$

(4) $P(46k\sim50k) = \frac{63}{165}$ (5) $P(66k\sim70k) = \frac{8}{165}$

2° 條件機率 P(X|Salary): <key> 令 X 為 systems, junior, 26~30 下之事件

(1) $P(X|26k\text{-}30k) = P(X|31k\text{-}35k) = P(X|36k\text{-}40k) = P(X|41k\text{-}45k) = P(X|66k\text{-}70k) = 0$

(2) $P(X|46k\text{-}50k) = \frac{P(X\cap 46k\text{-}50k)}{P(46k\text{-}50k)} = \frac{6}{63}$

3° 後驗機率 $P(Salary|X) \propto P(Salary) \times P(X|Salary)$

(1) $P(26k\text{-}30k|X) = P(31k\text{-}35k|X) = P(36k\text{-}40k|X) = P(41k\text{-}45k|X) = P(66k\text{-}70k|X) = 0$

(2) $P(46k\text{-}50k|X) = \frac{63}{165} \times \frac{3}{63} = \frac{3}{165} \div 0.0182$ (最大)

A: 因為 Salary 中在 46k-50k 的後驗機率最大，因此在此樣本下最有可能的薪水落在 46k-50k ＃

No.
Date : :

(2) According to the table, given a data sample with the values "Urban", "Below 21", "Married", and "Female" for the attributes Location, Age, Marriage Status, and Gender, respectively, what would a naive Bayesian classification of the Loyalty for the sample be?

	Attributes				
No	Location	Age	Marriage status	Gender	Loyalty
1	Urban	Below 21	Married	Female	Low
2	Urban	Below 21	Married	Male	Low
3	Suburban	Below 21	Married	Female	High
4	Rural	21~30	Married	Female	High
5	Rural	Above 30	Single	Female	High
6	Rural	Above 30	Single	Male	Low
7	Suburban	Above 30	Single	Male	High
8	Urban	21~30	Married	Female	Low
9	Urban	Above 30	Single	Female	High
10	Rural	21~30	Single	Female	High
11	Urban	21~30	single	Male	High
12	Suburban	21~30	Married	Male	High
13	Suburban	Below 21	Single	Female	High
14	Rural	21~30	Married	Male	Low

解: (已知) "Urban", "Below 21", "Married", "Female" ⇒ 以來預測 Loyalty

1° 計算先驗機率 $P(Loyalty)$

$P(Low) = \frac{5}{14}$, $P(High) = \frac{9}{14}$

2° 計算條件機率 $P(X|Loyalty)$: <key>令X: 分別"Urban", "Below 21", "Married", "Female" 下述事件

(一) $P(High)$

① $P(Urban|High) = \frac{2}{9}$, $P(Below\ 21|High) = \frac{2}{9}$, $P(Married|High) = \frac{3}{9}$, $P(Female|High) = \frac{6}{9}$

$\Rightarrow P(X|High) = \frac{2}{9} \times \frac{2}{9} \times \frac{3}{9} \times \frac{6}{9} = \frac{72}{6561}$

② $P(Urban|Low) = \frac{3}{5}$, $P(Below\ 21|Low) = \frac{2}{5}$, $P(Married|Low) = \frac{3}{5}$, $P(Female|Low) = \frac{2}{5}$

$\Rightarrow P(X|Low) = \frac{3}{5} \times \frac{2}{5} \times \frac{3}{5} \times \frac{2}{5} = \frac{36}{625}$

3° 計算後驗機率 $P(Loyalty|X) \propto P(Loyalty) \times P(X|Loyalty)$

① High: $P(High|X) = \frac{9}{14} \times \frac{72}{6561} \doteq 0.0071$

② Low: $P(Low|X) = \frac{5}{14} \times \frac{36}{625} \doteq 0.0123$

$\because P(Low|X) > P(High|X)$ ⇒ loyalty = low #

A: 應用貝式分類法後, 在 Location = Urban, Age = Below 21, Marriage = Married, Gender = Female 之條件下, 所預測之 loyalty 為 Low。