

CI603 Data Mining

SOLUTIONS

Tutorial 6

Naïve Bayes Classifier

1. Using the data table below predict the class for the data record

X=(Refund = yes, Marital Status = Divorced, Income = 110K)

Tid	Refund	Marital Status	Taxable income	Defaulted
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	No	Single	85K	Yes
8	No	Married	75K	No
9	No	Single	90K	Yes

Class No

$$P(\text{Refund}=\text{yes} | \text{No}) = (2+1)/(6+2) = 3/8$$

$$P(\text{Marital Status}=\text{Divorced} | \text{No}) = (0+1)/(6+3) = 1/9$$

$$P(\text{income}=110 | \text{No}) =$$

$$= \frac{1}{\sqrt{2\pi}(26.17)} e^{-\frac{(110-91)^2}{2(685)}} = 0.01171$$

Class Yes

$$P(\text{Refund}=\text{yes} | \text{Yes}) = (0+1)/(3+2) = 1/5$$

$$P(\text{Marital Status}=\text{Divorced} | \text{Yes}) = (1+1)/(3+3) = 1/3$$

$$P(\text{income}=110 | \text{Yes}) =$$

$$= \frac{1}{\sqrt{2\pi}(5)} e^{-\frac{(110-90)^2}{2(25)}} = 0.0002677$$

$$P(\text{No}) P(X | \text{No}) = 6/9 \times 3/8 \times 1/9 \times 0.01171 = 0.000325$$

$$P(\text{Yes}) P(X | \text{Yes}) = 3/9 \times 1/5 \times 1/3 \times 0.0002677 = 5.94 \times 10^{-6}$$