PAGE No.
DATE / /

Answer:

Q. 17.

Applying Naive Bayle's Classifier to map input tuples into accurate to class.

Table class has 4 diff. values, we need to find all the prior & posterior probabilities

$P(\text{on time}) = \frac{14}{20}$

$P(\text{Late}) = \frac{2}{20}$

$P(\text{very Late}) = \frac{3}{20}$

$P(\text{cancelled}) = \frac{1}{20}$

Calculating posterior probalitities.

for attribute 'Day'

$P(\text{weekday} / \text{on Time}) = 9/14$

$P(\text{weekday} / \text{Late}) = 1/2$

$P(\text{weekday} / \text{very Late}) = 3/3$

$P(\text{weekday} / \text{cancelled}) = 0/1$

Similarly calculating posterior probabilities for all other ~~attributes~~ value of attribute.

Day	on Time	Late	very Late	cancelled
Week day	9/14	1/2	3/3	0/1
Saturday	2/14	1/2	0/3	1/1
Sunday	1/14	0/2	0/3	0/1
Holiday	2/14	1/2	0/3	0/1

Attribute 'season'

season	on Time	Late	very late	cancelled
spring	4/14	0/2	0/3	1/1
Summer	6/14	0/2	0/3	0/1
Autumn	2/14	0/2	1/3	0/1
winter	2/14	1/2	2/3	0/1

Attribute 'Fog'

Fog	on Time	Late	very Late	Cancelled
None	5/14	0/2	0/3	0/1
High	4/14	1/2	1/3	1/1
Normal	5/14	1/2	2/3	0/1

Attribute 'Rain'

Rain	on Time	Late	very Late	Cancelled
None	6/14	1/2	1/3	0/1
Slight	6/14	1/2	0/3	0/1
Heavy	2/14	0/2	2/3	1/1

Apply Naives Bayes formula.

P_{NB}(on Time) = P(on Time) × P(we-day/ontime) × P(Winter/ontime) × P(High/ontime) × P(None/onTime)

$= \frac{14}{20} \times \frac{9}{14} \times \frac{2}{14} \times \frac{4}{14} \times \frac{5}{14} = 0.0079$

Similarly

P_{NB} (late) $= \frac{2}{20} \times \frac{1}{2} \times \frac{2}{2} \times \frac{1}{2} \times \frac{1}{2}$

$= 0.0125$

P_{NB} (very late) $= \frac{3}{20} \times \frac{3}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3}$

$= 0.0041$

P_{NB} (cancelled) $= \frac{1}{20} \times \frac{0}{1} \times \frac{-}{1} \times \frac{0}{1} \times \frac{1}{1} = 0$

~~P_{NB} (late) > P~~

P_{NB} (late) is highest

The correct classification is late.

Any other unseen instances prediction can be found out by this method.

Q2.

To test hypothesis that genders & preferred reading are independent, i.e. there is no correlation bet^n them.

Using chi Square Test,

The contingency table size is 2×2, which is given as

	Male	female
fiction	250 (90)	200 (360)
non fiction	50 (210)	1000 (840)

Degree of freedom $= (2-1) \times (2-1) = 1$

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$

O_{ij} → observed frequency

e_{ij} → Expected frequency.

$$\chi^2 = \frac{(250-70)^2}{90} + \frac{(50-210)^2}{210} + \frac{(200-360)^2}{360} + \frac{(1000-840)^2}{840} = 507.9365$$

Referring the table, for degree of freedom 1. & significant 0.01,

PAGE NO.
DATE / /

χ^2 value needed to χ^2:

χ^2 value needed to reject hypothesis is 6.635.

Our received value is above this value, therefore we can reject the hypothesis that gender & preferred reading are independent and we can conclude that two attributes are correlated.