2019110067

Page No.
Date.

DA Assignment.

Vinayak Tripathi (BE ETRX)

Q)

Days	Season	Fog
Weekday - 13	Spring - 5	None - 5
Holiday - 3	Winter - 6	High - 7
Satuday - 3	Summer - 6	Normal - 8
Sunday - 1.	Autumn - 3	

Rain	Class
None - 8	On time - 14
Slight - 7	Late - 2
Heavy - 5.	Very Late - 3
	Cancelled - 1.

Above are the unique values of each attribute & their count. Now we will consider prior probabilities. We will be using Naive Bayes Classification technique since the attributes & final classe are categorical.

Prior Probabilities.

Total cases - 20.

P(On Time) = 14/20

P(Late) = 2/20

P(Verylate) = 3/20

P(cancelled) = 1/20

We will calculate posterior probalities for each ~~class~~ attribute & class

Attribute Day : 20.

Counts

	Ontime	late	very late	cancelled
Weekday	9	1	3	0
Holiday	2	1	0	0
Saturday	2	0	0	1
Sunday	1	0	0	0

P(on time / Weekday) = 9/14

P(ontime / Holiday) = 2/14

P(ontime / Saturday) = 2/14

P(Ontime / ~~saturday~~) = 1/14

similarly

P(late / weekday) = 1/2

P(late / Holiday) = 1/2

P(late / Saturday) = 0

P(late / Sunday) = 0

creating a table of posterior probabilities

	very late	Cancelled
Weekday	1	0
~~late~~ Holiday	0	0
Saturday	0	1
Sunday	0	0

Attribute : Season

	Count				Probabilies			
	Ontime	late	Verylate	Cancel	Ontime	late	VL	Canc
Spring	4	0	0	1	0.29	0.0	0	1
Winter	2	2	2	0	0.14	1	0.67	0
Summer	6	0	0	0	0.43	0	0	0
Autum	2	0	1	0	0.14	0	0.33	0

Attribute : Fog

	Count				Probabilities			
	Ontime	Late	Verylate	Cancel	Ontime	late	Very late	Can
None	5	0	0	0	0.36	0	0.1	0
High	4	1	1	1	0.29	0.5	0.33	1
Normal	5	1	2	0	0.36	0.5	0.67	0

Attribute : Rain

	Count				Probabilities			
	Ontime	late	Verylate	Cancel	Ontime	late	Very late	Cancel
None	6	1	1	0	0.43	0.5	0.33	0
Slight	6	1	0	0	0.43	0.5	0.0	0
Heavy	2	0	2	1	0.14	0	0.67	1

Using above prior & posterior probabilities we can devise a classification technique

eg:

X = { WeekDay, Winter, High, None }

These are the attribute value.

We have to predict its class.

$P(X/\text{Ontime}) = P(\text{ontime}) \times P(\text{Weekday}/\text{Ontime}) \times P(\text{Winter}/\text{Ontime}) \times P(\text{high}/\text{ontime}) \times P(\text{None}/\text{ontime})$

$= 0.7 \times 0.64 \times 0.14 \times 0.29 \times 0.43$

$= 0.0078$

$P(X/\text{late}) = 0.1 \times 0.5 \times 1 \times 0.5 \times 0.5$

$= 0.0125$

$P(X/\text{verylate}) = 0.15 \times 1 \times 0.67 \times 0.33 \times 0.33$

$= 0.0111$

$P(X/\text{cancelled}) = 0.05 \times 0 \times 0 \times 1 \times 0 = 0$

We can see here for given X
P(X/ late) ~~has hig~~ is highest &
hence given tuple has class
~~ver~~ late

Page No.
Date.

Q.2) For this question we will be using χ^2 Correlation Test for Nominal Data.

The given problem is
hyp test hypothesis that gender & preferred reading have ~~non~~ correlation.
Let us assume that above is not the case.
Let us find the χ^2 value using formula.

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

For 2x2 table DF are $(2-1)(2-1) = 1$

$$e_{11} = \frac{\text{count(male)} \times \text{count(fiction)}}{n} = \frac{300 \times 450}{1500}$$

$e_{11} = 90$

similarly
$e_{12} = 360$. $e_{21} = 210$. $e_{22} = 840$.

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

χ^2 = ~~284~~ 507.93.

From χ^2 distribution for 0.001 significance level the χ^2 ~~value~~

critical χ^2 value is 10.828.
above which the hypothesis will be rejected.

Since our χ^2. is above critical value we can reject. the hypothesis that gender & preferred reding are independent.
We can say that two attributes are ~~str~~ correlated