DA - assignment

Mihir Nikam
2019140042 (I.T.)

Q1) Calculating all posterior and prior probabilities

- Prior probabilities for the following:

Day

Day/class	On time	Late	Very Late	Cancelled
Weekday	9/14 = 0.64	1/2 = 0.5	3/3 = 1	0/1 = 0
Saturday	2/14 = 0.14	1/2 = 0.5	0/3 = 0	1/1 = 1
Sunday	1/14 = 0.07	0/2 = 0	0/3 = 0	0/1 = 0
Holiday	2/14 = 0.14	0/2 = 0	0/3 = 0	0/1 = 0

Season

Season/class	on time	Late	Very Late	Cancelled
Spring	4/14 = 0.29	0/2 = 0	0/3 = 0	0/1 = 0
Summer	6/14 = 0.43	0/2 = 0	0/3 = 0	0/1 = 0
Autumn	2/14 = 0.14	0/2 = 0	1/3 = 0.33	0/1 = 0
Winter	2/14 = 0.14	2/2 = 1	2/3 = 0.67	0/1 = 0

Fog

Fog/class	on time	Late	Very Late	Cancelled
None	5/14 = 0.36	0/2 = 0	0/3 = 0	0/1 = 0
High	4/14 = 0.29	1/2 = 0.5	1/3 = 0.33	1/1 = 1
Normal	5/14 = 0.36	1/2 = 0.5	2/3 = 0.67	0/1 = 0

Rain

Rain/class	on time	Late	Very Late	Cancelled
None	5/14 = 0.36	1/2 = 0.5	1/3 = 0.33	0/1 = 0
Slight	8/14 = 0.57	0/2 = 0	0/3 = 0	0/1 = 0
Heavy	1/14 = 0.07	1/2 = 0.5	2/3 = 0.67	1/1 = 1

∴ Prior probability

on time	Late	Very Late	Cancelled
$\frac{14}{20} = 0.7$	$\frac{2}{20} = 0.1$	$\frac{3}{20} = 0.15$	$\frac{1}{20} = 0.05$

Now, we take a particular instance
Weekday, Winter, High, None, ??

Case 1: Class = on time
$= 0.7 \times 0.64 \times 0.14 \times 0.29 \times 0.36$
$= 6.547 \times 10^{-3}$

Case 2: Class = Late
$= 0.1 \times 0.5 \times 1 \times 0.5 \times 0.5$
$= 0.0125$

Class 3: Class = very late
$= 0.15 \times 1 \times 0.67 \times 0.33 \times 0.33$
$= 0.0109$

Class 4: class = cancelled
$= 0.05 \times 0.0 \times 0.0 \times 1.0 \times 0$
$= 0$

As we can see, case 2 has the highest probability.
Hence, we can conclude that the person will be "LATE"

201914042 - Mihir Nikam

Q2] H_0: preferred reading and gender are not correlated in the group

H_a: Both are co-related in the group.

- We compute the χ^2 value;

$$\chi^2 = \frac{count(A = a_i) \times count(B = b_j)}{n}$$

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

$$= 264.44 + 121.9 + 71.11 + 30.48$$

$$= 507.93$$

D.F for 2x2 table : $(2-1)(2-1) = 1$

From the χ^2 table, $\alpha = 0.001$, to reject null hypothesis value is 10.828

Since, the calculated value is greater than the value in the table, we can reject H_0.

∴ Both (gender and preferred reading) are co-related.

∴ We can conclude that two attributes are corelated for the given group