DA - Assignment I

Nachiket Sankhe
BE EXTC
2019120053

Q.1 Calculating all the posterior and prior probabilities.

for prior probabilities for the following.

attribute	ontime	late	very late	cancelled
Day :-				
Weekday	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	0/2	0/3	0/1
Season :-				
Spring	4/14	0/2	0/3	0/1
Summer	8/14	0/2	0/3	0/1
Autumn	2/14	0/2	1/3	0/1
Winter	2/14	2/2	2/3	0/1
Fog :-				
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
Rain :-				
None	5/14	0+1/2	0+1/3	0/1
Slight	8/14	0/2	0/3	0/1
Heavy	1/14	1/2	2/3	1/1

Prior Probability :-

	14/20	2/20	3/20	1/20

Instance : weekday, winter, High, None

Case 1 :-
    class :- On time
$$= 0.7 \times 0.64 \times 0.14 \times 0.29 \times 0.86$$
$$= 6.547 \times 10^{-3}$$
$$= 0.0065$$

Case 2 :-
    class :- late
$$= 0.1 \times 0.5 \times 1 \times 0.5 \times 0.5$$
$$= 0.0109$$

Case 3 :-
    class = Cancelled
$$0.05 \times 0 \times 0 \times 1 \times 0$$
$$= 0$$

Case 2 is strong

∴ The instance will be categorized under class late.

Q.3

Ho :

The preffered reading and gender are not correlated in the group

H₁ :-

Both are correlated

Computing the χ^2 value.

e_{ij} : $\dfrac{count\ (A = a_i)\ \times\ count\ (B = b_j)}{n}$

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

$$+ \frac{(1000 - 840)^2}{840}$$

$$= \underline{507.93}$$

For 2x2 table, degree of freedom = 1

∴ χ^2 value needed to reject the hypothesis at 0.001 significance level is 10.828 (from table).

Since, computed value value is above this, we reject the null hypothesis that gender and preffered reading are independent.

∴ we conclude that 2 attributes are correlated for the given group

used formula :- $\chi^2 = \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n} \left(\frac{b_{ij} - e_{ij}}{e_{ij}}\right)^2$