

Assignment 1

Smriti Agrawal

2019140001

Page No.

Date

Calculating all the position & prior probabilities

For prior probabilities for the following

Attribute on time Late Very Late Cancelled

Day:

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:

Spring	$4/14 = 0.29$	$0/2 = 0$	$0/3 = 0$	$0/1 = 0$
Summer	$8/14 = 0.57$	$0/2 = 0$	$0/3 = 0$	$1/1 = 1$
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:

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:

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	$14/20 = 0.70$	$2/20 = 0.10$	$3/20 = 0.15$	$1/20 = 0.05$
-------------	----------------	---------------	---------------	---------------

Instance: Weekday, Winter, High, None

Case 1 : Class = on time

$$= 0.7 \times 0.64 \times 0.14 \times 0.29 \times 0.36 \\ = 0.006547$$

Case 2 : Class = Late = $0.10 \times 0.5 \times 1.0 \times 0.5 = 0.5 = 0.0125$

Case 3 : Class = Very Late = $0.15 \times 1.0 \times 0.67 \times 0.33 \times 0.33 \\ = 0.0109$

Case 4 : Class = Cancelled
 $= 0.05 \times 0.0 \times 0.0 \times 1.0 \times 0 = 0$

\therefore Case 2 is very strong.

\therefore The instance will be categorized under Class Late.

Q2) H_0 : Preferred heading and gender are not correlated in the group.

H_1 : Both are correlated

\therefore Computing the χ^2 value.

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

$$\begin{aligned} \chi^2 &= \frac{(250-90)^2}{90} + \frac{(50-210)^2}{210} + \frac{(200-360)^2}{360} + \frac{(1000-240)^2}{240} \\ &= 284.44 + 121.90 + 71.11 + 30.43 = 507.93 \end{aligned}$$

For 2×2 table, degree of freedom are $(2-1)(2-1)=1$

For 1 degree of freedom, χ^2 value needed to reject the hypothesis at 0.001 significance level is 10.828
(took from χ^2 distribution table)

Since the computed value is above this, we can reject the null hypothesis that gender & preferred reading are independent.

\therefore We conclude that 2 attributes are correlated for the given group.

$$\text{Used the formula } \rightarrow \chi^2 = \sum_{i=1}^m \sum_{j=1}^n \frac{n}{e_{ij}} (O_{ij} - e_{ij})^2$$