

Jeenal Mehta

2019140039

# DA Assignment - I

BE - IT

1. Calculating all the posterior & prior probabilities

For prior probabilities for the following:

Attribute	On time	Late	Very late	Cancelled
Day:				
weekday	$9/14 = 0.64$	$1/2 = 0.5$	$2/3 = 1$	$0/1 = 0$
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	$6/14 = 0.43$	$0/2 = 0$	$0/3 = 0$	$0/1 = 0$
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:

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: Holiday, Autumn, Normal, Heavy

Case I:

Class = on time

$$= 0.14 \times 0.14 \times 0.36 \times 0.67 = 0.0004939$$

Case II:

Class = late

$$= 0 \times 0 \times 0.5 \times 0.5 = 0$$

Case III:

Class = Very late

$$= 0 \times 0.33 \times 0.67 \times 0.67 = 0$$

Case IV:

Class = cancelled

$$= 0 \times 0 \times 0 \times 1 = 0$$

Visibly, Case 1 is strong.

∴ The instance will be categorized under on time class.

2.

$H_0$ : Preferred reading in genders not correlated

$H_1$ : Preferred reading in genders are correlated

Calculating the  $\chi^2$  value

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

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

$$= 254.44 + 121.90 + 71.11 + 30.48 = \underline{507.93}$$

For a  $2 \times 2$  table, degree of freedom are  $(2-1)(2-1) = 1$

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

Since the computed value is above this, we can reject the null hypothesis that genders and preferred reading are independent

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

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