

# DA Assignment 1

GAURAV PANCHAL

2019120046

BE EXTC

## Problem 1 - Air Traffic Data

- There are four attributes:

$A = [\text{Day, Season, Fog, Rain}]$

- There categories of classes are:

$C = [\text{On time, Late, V Late, Cancelled}]$

	Attribute	On time	Late	Very late	Cancelled
Days	Weekday	$9/14 = 0.64$	$1/2 = 0.5$	$3/3 = 1$	0
	Saturday	$2/14 = 0.14$	0	0	$1/1 = 1$
	Sunday	$1/14 = 0.07$	0	0	0
	Holiday	$2/14 = 0.14$	$1/2 = 0.5$	0	0
Season	Spring	$4/14 = 0.28$	0	0	$1/1 = 1$
	Summer	$6/14 = 0.42$	0	0	0
	Autumn	$2/14 = 0.14$	0	$1/3 = 0.33$	0
	Winter	$2/14 = 0.14$	$2/2 = 1$	$2/3 = 0.667$	0
FOG	None	$5/14 = 0.35$	$1/2 = 0.5$	$1/3 = 0.33$	0
	High	$4/14 = 0.28$	$1/2 = 0.5$	0	0
	Normal	$5/14 = 0.35$	0	$2/3 = 0.67$	$1/1 = 1$
Rain	None	$6/14 = 0.42$	$1/2 = 0.5$	$1/3 = 0.33$	0
	Slight	$6/14 = 0.42$	$1/2 = 0.5$	0	0
	Rainy	$2/14 = 0.14$	0	$2/3 = 0.67$	$1/1 = 1$
Prior Probability		$14/20 = 0.7$	$2/20 = 0.1$	$3/20 = 0.15$	$1/20 = 0.05$



Instance

Weekday Winter High None ???

CASE I: Class - On Time

$$P_{\text{on time}} = 0.7 \times 0.64 \times 0.14 \times 0.28 \times 0.42 \\ = 0.00737$$

CASE II: Class - Late

$$P_{\text{LATE}} = 0.1 \times 0.5 \times 1 \times 0.5 \times 0.5 \\ = 0.0125$$

CASE III: Class - Very Late

$$P_{\text{VLATE}} = 0.15 \times 1 \times 0.67 \times 0.33 \times 0.33 \\ = 0.0109$$

CASE IV - Class - Cancelled

$$P_{\text{cancel}} = 0.05 \times 0 \times 0 \times 1 \times 0 \\ = 0$$

Using Formula

$$P(\text{yes}) = P(\text{yes}) P(\text{weekday/yes}) P(\text{winter/yes}) \\ P(\text{High/yes}) P(\text{None/yes})$$

As the probability of Class - Late is 0.0125 the greatest, Instance - weekday, Winter, High, None will fall under Late category.

Q.2  $H_0$ : Preferred reading and Gender are not correlated in the group.

$H_1$ : Both are correlated.

Computing the  $\chi^2$  value.

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

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

$$= 507.98$$

For  $2 \times 2$  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 (look from  $\chi^2$  table)

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

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