Abhinav Pallayil
2019120045 BE EXTC

Assignment - 1

Q.1) Attributes: (Day, Season, Fog, Rain)

Categories: [On time, late, very late, Cancelled]

P[On time] = 14/20, P[Very late] = 3/20,
P[late] = 2/20 ; P[cancelled] = 1/20

Tabulation of all posterior probabilities:

For the attribute 'Day'

Attribute Day	class On time	Late	Very late	Cancelled
Weekday	9/14	1/2	3/3	0/1 = 0
Saturday	2/14	0/2	0/3	1/1 = 1
Sunday	1/14	0/2	0/3	0/1 = 0
Holiday	2/14	1/2	0/3	0/1 = 0

For the attribute 'Season'

Season	On-time	Late	very late	Cancelled
Spring	4/14	0/2	0/3	1/1
Summer	6/14	0/2	0/3	0/1
Autumn	2/14	0/2	1/3	0/1
Winter	2/14	2/2	2/3	0/1

Abhinav Pallayil
2019120045

For attribute 'Fog'

Fog	On time	Late	very late	Cancelled
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

For attribute 'Rain'

Rain	On-time	Late	very late	Cancelled
None	6/14	1/2	1/3	0/1
Slight	6/14	1/2	0/3	0/1
Heavy	2/14	0/2	2/3	1/1

For the given instance in the question
<weekday, winter, High, None>

P_{NB} [On time] = P(On time) x P(weekday | On time) x P(winter / on time) x P(High | On time) x P(None | On time)

$$P_{NB}(\text{On time}) = \frac{14}{20} \times \frac{9}{14} \times \frac{2}{14} \times \frac{4}{14} \times \frac{6}{14} = 0.0079$$

Abhinav Pallayil
2019120045

classmate

Date

Page

Similarly,

$$P_{NB}[\text{Late}] = \frac{2}{20} \times \frac{1}{2} \times \frac{2}{2} \times \frac{1}{2} \times \frac{1}{2} = \boxed{0.0125}$$

$$P_{NB}[\text{very late}] = \frac{3}{20} \times \frac{3}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} = 0.0111$$

$$P_{NB}[\text{cancelled}] = \frac{1}{20} \times \frac{0}{1} \times \frac{0}{1} \times \frac{1}{1} \times \frac{0}{1} = 0$$

$\therefore P_{NB}[\text{Late}]$ is highest, hence correct classification is 'late'.

The correct classificate is 'LATE'

Abhinav Pallayil
2019120045

classmate
Date
Page

Q.2) In this problem, we have to test the hypothesis that ~~given~~ gender and preferred reading have no correlation between them and they are independent.

Ho (Null hypothesis) : Preferred reading and gender are not correlated in the group.

H1 (Alternate hypothesis) : Preferred reading and gender are correlated in the group.

$$\boxed{\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}}$$

As per the table,

$o_{11} = 250$	$e_{11} = 90$
$o_{12} = 200$	$e_{12} = 360$
$o_{21} = 50$	$e_{21} = 210$
$o_{22} = 1000$	$e_{22} = 840$

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

$$= 284.44 + 121.90 + 71.11 + 30.48$$

$$\boxed{\chi^2 = 507.93}$$

For 2x2 table, degree of freedom ~~are~~ is $(2-1)(2-1) = 1$

∴ Degr

Abhinav Pallayil
2019120045.

classmate
Date
Page

Degree of freedom value : 1
significance level : 0.001 } 10.828 (from χ^2 distribution table).

∴ Since, the computed value is greater than permitted value (507.93 > 10.828), we reject the null hypothesis

Ho is rejected!.

∴ We conclude that preferred reading and gender are correlated.