

2019120060

Pratham Solanki

Q1

Calculating all the posterior & prior probabilities

Q2

database attributes are: Day, day, season, job, rain.

the categories of classes are: ontime, late, very late, cancelled.

Naive bayesian Classifier:

$$P(y|x_1, \dots, x_n) = \frac{P(x_1|y) P(x_2|y) \dots P(x_n|y) P(y)}{P(x_1) P(x_2) P(x_3) \dots P(x_n)}$$

day:

attribute	ontime	late	very late	cancelled
week day	9/14	1/2	3/3	0
saturday	2/14	1/2	0	1
sunday	1/14	0	0	0
holiday	2/14	0	0	0

Season:

spring	4/14	0	0	0
summer	6/14	0	0	0
autumn	2/14	0	1/3	0
winter	2/14	2/2	2/3	0

Fog:	5/14	0	0	0
none	4/14	0.5	0.33	1
high	5/14	0.5	0.67	0
normal				

Rain:	0.82	0.5	0.33	0
none	0.07	0	0	0
slight	0.07	0.5	0.67	1
heavy				

prior
probability

14/20

2/20

3/20

1/20

Past on case 3 week days, winter, high, none ???

Case 1 ~~Case 1~~ on fine

$$= 0.70 \times 0.64 \times 0.14 \times 0.29 \times 0.36$$

$$= 6.547 \times 10^{-3}$$

Case 2

late

$$0.10 \times 0.50 \times 1.0 \times 0.5 \times 0.5$$

$$= 0.0125$$

Case 3

very late

$$= 0.15 \times 1.0 \times 0.67 \times 0.33 \times 0.33$$

$$= 0.0109$$

Case 4: Cancelled

$$= 6.05 \times 0.0 \times 0.0 \times 1.0 \times 0.0$$
$$= 0.$$

Case 2 is strong.
The instance will be categorized under
Class Late.

2019120060 Pratham D Solanki

classmate

Date

Page

Q.2 Problem statement: to check whether the two attributes gender & reading have no correlation & they are independent.

using χ^2 test.

$$\text{dof} = (2-1) \times (2-1).$$

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^n \frac{|\text{O}_{ij} - \text{E}_{ij}|^2}{\text{E}_{ij}}$$

$$\therefore \chi^2 = \frac{[250 - 90]^2}{90} + \frac{[50 - 210]^2}{210} + \frac{[200 - 360]^2}{360} + \frac{[1000 - 840]^2}{840} = 570.93$$

Since $570.93 > 2.706$,

gender & preferred reading are strongly correlated.

So we can reject the null hypothesis of independence at confidence level of 0.1.