LAB - 11 : Naive Bayes classifiers

Example	Day	Outlook	Temp	Humidity	Wind	Play Tennis
	1	Sunny	Hot	High	weak	No
	2	Sunny	Hot	High	strong	No
	3	Overcast	Hot	High	weak	yes
	4	Rain	mild	High	weak	yes
	5	Rain	Cool	Normal	weak	yes
	6	Rain	cool	Normal	strong	No
	7	Overcast	cool	Normal	strong	yes
	8	Sunny	mild	High	weak	No
	9	Sunny	Cool	Normal	Weak	yes
	10	Rain	mild	Normal	weak	yes
	11	Sunny	mild	Normal	strong	yes
	12	Overcast	mild	High	strong	yes
	13	Overcast	Hot	Normal	weak	yes
	14	Rain	mild	High	strong	No

→ Given a new instance

$x' = ($ outlook = Sunny , Temp = Cool , Humidity = High , Wind = Strong $)$

Soln $P($ play = yes $) = 9/14$

$P($ play = No $) = 5/14$

Outlook	yes	No	
Sunny	2/9	3/5	
Overcast	4/9	0/5	
Rain	3/9	2/5	

Temp.	yes	No	
Hot	2/9	2/5	
mild	4/9	2/5	
Cool	3/9	1/5	

Humidity	yes	No	
High	3/9	4/5	
Normal	6/9	1/5	

Wind	yes	No	
Strong	yes 3/9	No 3/5	
Weak	6/9	2/5	

$$P(c_i \mid x) = P(x_1 \mid c_i) * P(x_2 \mid c_i) * \ldots * P(x_n \mid c_i) * P(c_i)$$

→ $P(yes \mid x')$ = [$P(sunny \mid yes)$ * $P(cool \mid yes)$
  * $P(High \mid yes)$ * $P(strong \mid yes)$]
  * $P(play = yes)$

  = [$2/9$ * $3/9$ * $3/9$ * $3/9$] * $9/14$

  = 0.0053

→ $P(No \mid x')$ = [$P(sunny \mid No)$ * $P(cool \mid No)$ *
  $P(High \mid No)$ * $P(strong \mid No)$]
  * $P(play = No)$

  = [$3/5$ * $1/5$ * $4/5$ * $3/5$] * $5/14$

  = 0.0206 ✓

→ we label x' to be "No".

| Example | Apply Gini Index For the given Dataset. |

RID	Age	income	Student	credit rating	class: buyes_comp
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle	medium	no	excellent	yes
13	middle	high	yes	fair	yes
14	senior	medium	no	excellent	no

Sol^n   consider the Data

$\rightarrow$ buys_computer = yes = 9 tuple

$\rightarrow$ buys_computer = No = 5 tuple

* Gini index of Data

$$Gini(D) = 1 - \left(\frac{9}{14}\right)^2 - \left(\frac{5}{14}\right)^2$$

$$= 0.4592$$

* Consider splitting subset of income

income	yes	No	Total
Low	3	1	4
medium	4	2	6
High	2	2	4

i) Gini income c- {high, Low} (D)

$$= \frac{8}{14}\left[1 - \left(\frac{5}{8}\right)^2 - \left(\frac{3}{8}\right)^2\right]$$

$$+ \frac{6}{14}\left[1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2\right]$$

$$= 0.4583$$

ii) Gini income c- {high, medium} (D)

$$= \frac{10}{14}\left[1 - \left(\frac{6}{10}\right)^2 - \left(\frac{4}{10}\right)^2\right]$$

$$+ \frac{4}{14}\left[1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2\right]$$

$$= 0.4500$$

iii) Gini income $\leftarrow$ { low, medium }

$$= \frac{10}{14} \left[1 - \left(\frac{7}{10}\right)^2 - \left(\frac{3}{10}\right)^2 \right]$$

$$+ \frac{4}{14} \left[1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 \right]$$

$$= 0.4429$$

* Consider Splitting Subset For Age

Age	yeg	No	Total	
youth	2	3	5	
middle	4	0	4	
senior	3	2	5	

1) Gini Age $\leftarrow$ { youth, middle } (D)

$$= \frac{9}{14} \left[1 - \left(\frac{6}{9}\right)^2 - \left(\frac{3}{9}\right)^2 \right]$$

$$+ \frac{5}{14} \left[1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2 \right]$$

$$= 0.4571$$

ii) Gini Age c- { youth, senior } (0)

$$= \frac{10}{14}\left[1 - \left(\frac{5}{10}\right)^2 - \left(\frac{5}{10}\right)^2\right]$$

$$+ \frac{4}{14}\left[1 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{4}\right)^2\right]$$

$$= 0.3571$$

iii) Gini Age c- { middle, senior }

$$= \frac{9}{14}\left[1 - \left(\frac{7}{9}\right)^2 - \left(\frac{2}{9}\right)^2\right]$$

$$+ \frac{5}{14}\left[1 - \left(\frac{2}{5}\right)^2 - \left(\frac{3}{5}\right)^2\right]$$

$$= 0.3937$$

* Consider splitting Subset o for Student

Student	yes	No	Total
yes	6	1	7
No	3	4	7

i) Gini Student (D)

$$= \frac{7}{14}\left[1 - \left(\frac{6}{7}\right)^2 - \left(\frac{1}{7}\right)^2\right]$$

$$+ \frac{7}{14}\left[1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2\right]$$

$$= 0.3673$$

* Consider for credit rating

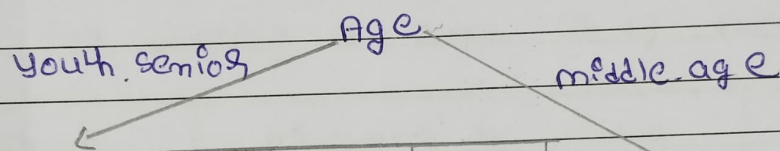| credit rating	yes	No	Total
fair	6	2	8
excellent	3	3	6

→ Gini credit rating (D)

$$= \frac{8}{14}\left[1 - \left(\frac{6}{8}\right)^2 - \left(\frac{2}{8}\right)^2\right] + \frac{6}{14}\left[1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2\right]$$

$$= 0.4286$$

Darshan UNIVERSITY

Attribute	Gini-index	Δgini
income	0.4429	0.4592 - 0.4429 =
Age	0.3571	0.4592 - 0.3571 =
Student	0.3673	0.4592 - 0.3673 =
Credit rating	0.4286	0.4592 - 0.4286 =

Age

youth, senior   middle age

Income	student	CR	class
high	no	fair	no
high	no	excellent	no
medium	no	fair	no
low	yes	fair	yes
medium	yes	excellent	yes
medium	no	fair	yes
low	yes	fair	yes
low	yes	excellent	no
medium	yes	fair	yes
medium	no	excellent	no

income	student	CR	class
high	no	fair	yes
low	yes	excellent	yes
medium	no	excellent	yes
high	yes	fair	yes