Bayesian Classification

Sunday, January 24, 2021   12:31 PM

X = {age = youth, income = medium, student = yes, credit-rating = fair, buy.computer = ?}

P(buy computer = yes) = 9/14 = 0.643

RID	age	income	student	credit_rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	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_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

P(playtennis = yes) = 9/14 = 0.643

P(playtennis = no) = 5/14 = 0.357

Q. (Outlook = sunny, temperature = cool, Humidity = high, wind = strong, play tennis = ?)

Day	Outlook	Temperature	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

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

Temperature	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	3/9	3/5
Weak	6/9	2/5

P(outlook = sunny | Play tennis = Yes) = 2/9 = 0.222   P(

P(temp = cool | Play tennis = Yes) = 3/9 = 0.333

P(humidity = high | playtennis = Yes) = 3/9 = 0.333

P(wind = strong | playtennis = Yes) = 3/9 = 0.333

P(X | play tennis = yes) = 0.222 * 0.333 * 0.333 * 0.333 = 0.0082

[illegible] = 0.058

$P(X \mid playtennis=no) = \frac{[illegible]}{5} \times \frac{[illegible]}{5} \times \frac{[illegible]}{5} \times \frac{[illegible]}{5} = 0.058$

$P(yes) = P(X \mid playtennis=yes) * P(playtennis=yes) = 0.0082 \times 0.643 = 0.0053$

$P(no) = P(X \mid playtennis=no) * P(playtennis=no) = 0.058 \times 0.357 = 0.0207$

$\frac{P_{AB}(yes)}{P_{AB}(yes) + P_{AB}(no)} = 0.205 \quad \frac{P_{AB}(no)}{P_{AB}(yes) + P_{AB}(no)} = 0.795$

Since $P(C_i \mid X)$ of playtennis=no has greater value, naive bayesian classifier predicts playtennis = no for tuple X (outlook = sunny, temperature = cool, humidity = high, wind = strong)

$P(flu=yes) = 5/8 =$

$P(flu=no) = 3/8 =$

$P(X \mid flu=yes) = P(chill=yes) \times P(running\ nose = ~~yes~~\ no) \times P(headache = ~~yes~~\ mild) \times P(fever=yes)$

$= 3/5 * 4/5 * 3/5 * 4/5 =$

$P(X \mid flu=no) = P(chill=no) \times P(running\ nose = yes) \times P(headache = no) \times P(fever = no)$

$P(X \mid flu=yes) * P(flu=yes) =$

$P(X \mid flu=no) * P(flu=no) =$

Q) Consider following dataset and predict the fruit = {yellow, sweet, long} using Naive Bayes Classification algorithm.

Fruit	Yellow	Sweet	Long	Total
Mango	350	450	0	650
Banana	400	300	350	400
Others	50	100	50	150
Total	800	850	400	1200

$P(fruit = Mango) = 650/1200 = 0.542$

$P(X \mid Mango) = P(Yellow \mid Mango) * P(Sweet \mid Mango) * P(Long \mid Mango)$

$= 0$

$P(fruit = banana) = 400/1200 = 0.333$

$P(Yellow|banana) = \frac{P(Banana|Y) * P(Y)}{P(B)} = \frac{400/800 * 800/1200}{400/1200} = 1$

$P(sweet|banana) = \frac{P(B|S) \times P(S)}{P(B)} = \frac{300/850 * 850/1200}{400/1200} = \frac{3}{4} = 0.75$

$P(long|banana) = \frac{350}{400} = 0.875$

$P(X|banana) = 0.656$

$P(fruit = others) = 0.417$

$P(Yellow|others) = 0.333$

$P(sweet|others) =$

$P(long|others) =$

$P(X|others) =$

$P(X|Mango) = 0$

$P(X|banana) = 0.656$

So, fruits{yellow, sweet, long} = banana

$P(X|others) = 0.074$

Q. Consider following dataset and predict the class of new instance X = (Refund = No, Marital Status = Married, Income = 120K) using Naive Bayes Classification algorithm.

Tid	Refund	Marital Status	Taxable Income Amount	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No

$P(X|evade = No)$

$= P(refund = No|evade = No) * P(married|evade = No) \times P(income = 120K|evade = No) = \frac{4}{7} * \frac{4}{7} * 0.0072$

$= 0.0023$

$P(A_i|C_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{\left\{\frac{(A_i - \mu_{ij})}{2\sigma_{ij}^2}\right\}}$

where μ_{ij} = mean, σ_{ij} = variance

$P(income = 120K | evade = NO) = 0.0072$

$n = 7$

$\mu_{ij} = \frac{125 + 100 + 70 + 120 + 60 + 220 + 75}{7}$

$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$

$\mu_{ij} = \frac{\sum x_i}{\ldots}$

8	No	Single	85k	Yes
9	No	Married	75k	No
10	No	Single	80k	Yes

$(\sigma_{ij})^2 = 110 = \frac{(125-110)^2 + (100-110)^2 \ldots}{7-1} = 2975$

$\mu_{ij} = \frac{\sum x_i}{n}$

Married|Yes = 0/3 = 0

(0.0023)

No

Q. Color = Green, legs = 2, Height = Tall, Smelly = No, Species = ?

No.	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

P(X | evade = No) = 0.0023

P(X | evade = Yes) = 0

Ans = No