28 Sep 2023

Naive Bayes

Conditional Probability

if there is a two probability given A and B

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{where } P(B) \neq 0$$

example rolling a two dice D_1 and D_2

sample space. $\{(1,1), (1,2), (1,3) \ldots\ldots (6,6)\}$

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

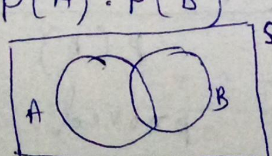
(i) what is probability D_1 has 5: $P(A=5) = \frac{1}{36}$

(ii) what is probability that $D_1 + D_2 <= 10$: $\frac{33}{36} = \frac{11}{12}$

(iii) what is probability that $\underset{A}{\underset{\uparrow}{D_1 = 5}}$ given $\underset{B}{\underset{\uparrow}{D_1 + D_2 <= 10}}$

$$P(D_1 = 5 \mid D_1 + D_2 <= 10) = P(A|B) = \frac{5}{33}$$

independent Events : A and B is called independent if $P(A \cap B) = P(A) \cdot P(B)$



P(A) already occur. $P(A|B) = \frac{n(A \cap B)}{n(B)}$ — (i)

$$P(A) = \frac{P(A \cap B)}{P(B)} = \frac{n(A \cap B)}{\cancel{n(S)} \times \frac{n(B)}{\cancel{n(S)}}}$$

$$P(A) = \frac{n(A \cap B)}{n(B)} \text{ ——— (ii)}$$

from (i) and (ii) $P(A|B) = P(A)$

Mutually Exclusive Events:

A and B are mutually exclusive events when

$P(A \cap B) = 0$

Hence $P(A|B) = \frac{P(A \cap B)}{P(B)} = 0$

Bayes Theorem: for two event A, B.

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)} \quad \text{where } P(B) \neq 0$$

postirior   liklihood   evidence   prior

proof:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ ——— (i)}$$

$\because A \cap B = B \cap A$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)}$$

$$P(A \cap B) = P(B|A) \cdot P(A) \text{ ——— (ii)}$$

from (i) and (ii)

$$P(A|B) = \frac{P(B|A)(A)}{P(B)}$$

Hence proved.

example:

Factory

M_1 M_2 M_2

M_1, M_2, M_3 - marker

D - Defective.

	M_1	M_2	M_2
produced -	20%	30	50
defective.	5%	3%	1%

Ques÷ randomly pick marker from bag. find if it is already defective and it it come from M_3

$P(M_1) = \frac{1}{5}$ $P(M_2) = \frac{3}{10}$ $P(M_3) = \frac{1}{2}$

$P(D|M_1) = \frac{1}{20}$ $P(D|M_2) = \frac{3}{100}$ $P(D|M_3) = \frac{1}{100}$

find = $P(M_3|D) = ?$

$$P(M_3|D) = \frac{P(D|M_3)\,P(M_3)}{P(D)}$$

$$P(D) = P(D\cap M_1) + P(D\cap M_2) + P(D\cap M_3)$$

$\because P(A|B) = \frac{P(A\cap B)}{P(B)}$

$$P(D) = P(D|M_1)P(M_1) + P(D|M_2)P(M_2) + P(D|M_3)P(M_3)$$

$$= \frac{1}{20}\times\frac{1}{5} + \frac{3}{100}\times\frac{3}{10} + \frac{1}{100}\times\frac{1}{2}$$

$$= \frac{1}{100}\left(1 + \frac{9}{10} + \frac{1}{2}\right)$$

$$= \frac{1}{100}\left(\frac{10+9+5}{10}\right)$$

$$= \frac{2.4}{100} = 0.024$$

Hence $P(M_3|D) = \frac{1/100 \times 1/2}{0.024} = \frac{1}{4.800} = 0.20$

Naive Bayes:

eSK

toss	venue	outlook	Result
won	Mumbai	Overcast	Won
lost	Chennai	sunny	won.
won.	kolkata	sunny	won
won	chennai	sunny	won.
lost	Mumbai	Sunny	lost
won.	chennai	overcast	lost
won	Kolkata	Overcost	Lost
won	Mumbai	sunny	won

Naive Bayes. Question example

(i) $P(W \mid (lost \cap Mumbai \cap Sunny)) = ?$ = 0.56

(ii) $P(L \mid (lost \cap Mumbai \cap Sunny)) = ?$ 0.27

Naive Bayes. (i) and (ii) में जो value बड़ा होगा वो output देगा।

we know that the $P(A|B) = \dfrac{P(B|A)\,P(A)}{P(B)}$

$P(W \mid lost \cap Mumbai \cap Sunny)$

$$= \frac{P(lost \cap Mumbai \cap Sunny \mid W)\;P(W)}{P(lost \cap Mumbai \cap Sunny)}$$

=

for $P(L \mid lost \cap Mumbai \cap Sunny)$ = similar

The denominator are same in both the cases hence don't calculate denominator

Hence

$$P(W \mid lost \cap Mumbai \cap Sunny) = P(lost \cap Mumbai \cap Sunny \mid W)\, P(W)$$

$$P(L \mid lost \cap Mumbai \cap Sunny) = P(lost \cap Mumbai \cap Sunny \mid L)\, P(L)$$

$$P(W) = \frac{5}{8} \qquad P(L) = \frac{3}{8}$$

$$P(lost \cap Mumbai \cap Sunny \mid W) = P(lost \mid W)\, P(Mumbai \mid W)\, P(Sunny \mid W)\, P(W)$$

$$= \frac{1}{5} \times \frac{2}{5} \times \cancel{\frac{4}{5}} \times \cancel{\frac{5}{8}}$$

$$X = \{x_1, x_2, x_3, \ldots x_n\}$$

$$C_k = \{C_1, C_2, C_3 \ldots C_k\}$$

toss	venue	outlook	Result
X_1	X_2	X_3	WON / Lost
			C_1 = Won
			C_2 = Lost

$$P(C_k \mid x) = \frac{P(x \mid C_k) \cdot P(C_k)}{P(X)}$$

for $k = 1 \ldots K$ we can remove $P(X)$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B) = P(A \mid B)\, P(B)$$

$$p(C_k|x) = p(x, C_k)$$

$$= p(\underset{A}{x_1}, \underbrace{x_2, x_3 \ldots x_n, C_k}_{B})$$

$$= \underbrace{p(x_1 | x_2, x_3 \ldots x_n, C_k)}_{a} \, p(x_2, x_3 \ldots x_n, C_k)$$

$$p(C_k|x) = a \times p(\underset{A}{x_2}, \underbrace{x_3, \ldots x_n, C_k}_{B})$$

$$= a \cdot \underbrace{p(x_2 | x_3, x_4, \ldots x_n, C_k)}_{b} \cdot p(x_3, x_4, \ldots x_n, C_k)$$

$$= a \cdot b \cdot p(x_3, x_4, \ldots x_n, C_k)$$

∴ the above method is chain rule for conditional probability

$$= \cancel{p(\underset{A}{x_n} | \underset{B}{C_k}) \, p(C_k)}$$

Hence we also write

$$= p(x_1 | x_2, x_3 \ldots x_n, C_k) * p(x_2 | x_3, x_4, \ldots x_n, C_k) \ldots$$

$$\ldots \; p(x_{n-1} | x_{n-1}, C_k) \cdot p(x_n | C_k) \, p(C_k)$$

conditional independence.

x_1 depend on C_k, x_2 depend on C_k and so on.

or $\quad p(A|B) = p(A)$

$\quad\quad p(A|B, C) = p(A|C)$

$$p(C_k|x) = p(x_1|C_k) \cdot p(x_2|C_k) \cdot p(x_3|C_k) \ldots p(x_n|C_k) \cdot p(C_k)$$

$$p(C_k|x) = p(C_k) \prod_{i=1}^{n} p(x_i|C_k)$$

$$p(C_k|x) = \frac{1}{Z} \times p(C_k) \prod_{i=1}^{n} p(x_i|C_k)$$

$z = p(x)$

$$\hat{y} = \underset{k \in \{1, 2, \ldots K\}}{\operatorname{argmax}} \; p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)$$

Maximum a postiorior Rule (MAP)

Handling Numerical Data in Naive Bayes

Height	Weight	Gender
172	150	M
180	170	M
165	140	M
190	200	M
139	100	F
145	120	F
160	140	F
172	150	F

$\{H = 185, W = 170, G = ?\}$

$p(M \mid H = 185, W = 170) = ?$

$= p(H = 185 \mid M) * p(W = 170 \mid M)\, p(M)$

$=$

$p(F \mid H = 185, W = 170) =$

$= p(H = 185 \mid F) * p(W = 170 \mid F)\, p(F)$

$=$

* Height is Gaussian distributed random variable

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- the above method we used is Gaussian Naive Bayes
- the above method is only used when our data is Normally distributed.
- if the data is not normally distributed we used other Naive Bayes
  - Binomial Naive Bayes
  - Multinomial Naive Bayes
  - Posylon Naive Bayes

Sentiment Analysis

there is three steps is involved

(i) Text preprocessing

(ii) Vectorize (Bag of words)

(iii) Data pass into Algo then Train

(iv) Deployement

Bag of words example.

review	great	awesome	poor	pathetic	label
1st	3	1	0	0	1
2nd	0	1	2	4	0

the above trick is done by sklearn (CountVectorizer)

example:

input review = ["The movie was great but slightly boring"]

= [0, 1, 1]

label = ?

f_1 awesome	f_2 great	f_3 boring.	label
3	1	0	positive
0	0	2	negative
3	3	1	positive

$$\left.\begin{array}{l} P(l = +ve \mid f_1 = 0,\ f_2 = 1,\ f_3 = 1) \\ P(l = -ve \mid f_1 = 0,\ f_2 = 1,\ f_3 = 1) \end{array}\right\}$$

we have to calculate these two for finding the sentiment of given review.

(i) $P(l = +ve \mid f_1 = 0, f_2 = 1, f_3 = 0)$

$= P(f_1 = 0 \mid +ve) . P(f_2 = 1 \mid +ve) . P(f_3 = 1 \mid +ve) . P(+ve)$

(ii) $P(l = -ve \mid f_1 = 0, f_2 = 1, f_3 = 1)$

$= P(f_1 = 0 \mid -ve) . P(f_2 = 1 \mid -ve) . P(f_3 = 1 \mid -ve) . P(-ve)$

- in the above (i) and (ii) which one is the big is consider as the output.
- The above all analysis is behind How the Naive Bayes is work on Sentiment Analysis.