Naïve Bayes (Udemy)

12 July 2024 07:53

Naïve Bayes Algorithm (Classification)

1. Probability
2. Bayes Theorem

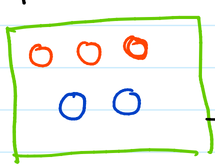
Independent events

Rolling a dice = {1, 2, 3, 4, 5, 6}

$P(1) = \frac{1}{6}$ $P(2) = \frac{1}{6}$ $P(3) = \frac{1}{6}$ {One outcome is not changing the outcome of other}

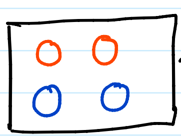
Dependent events

1) What is the probability of removing a orange marble and then a yellow marble?



$\rightarrow P(\text{orange}) = \frac{3}{5} \rightarrow 1^{st}$ event

after this event



$\rightarrow P(\text{yellow}) = \frac{2}{4}$

$P(Y/O)$

∴ Count of marbles change after 1 event

∴ 2^{nd} event is dependent on 1^{st} event

Combining the events

$$P(\text{Orange and Y}) = P(\overset{1^{st}}{\text{orange}}) * \underbrace{P(\text{Yellow} \mid \overset{2^{nd}}{\text{Orange}} \text{ is taken out})}_{\text{Conditional probability}}$$

$$= \frac{3}{5} * \frac{2}{4} = \frac{6}{20} = \frac{3}{10}$$

$$\boxed{P(A \text{ and } B) = P(A) * P(B/A)}$$

Bayes Theorem

$$P(A \text{ and } B) = P(B \text{ and } A)$$

$$P(A) * P(B/A) = P(B) * P(A/B)$$

$P(A|B)$ = Probability of event A given that B has already occurred.

$P(A)$ = Probability of event A

B has already occurred

$$P(A) * P(B|A) = P(B) * P(A|B)$$

$$\boxed{\therefore P(A|B) = \frac{P(A) * P(B/A)}{P(B)}}$$

P(A) = Probability of event A

P(B) = Probability of event B

P(B/A) = Probability of event B given that A has already occurred

Use in ML (Naive Bayes)

x_1	x_2	x_3	y
			Yes
			No
			Yes

$$\rightarrow \boxed{P(y/(x_1, x_2, x_3)) = \frac{P(y) * P(x_1, x_2, x_3/y)}{P(x_1, x_2, x_3)}}$$

$$P(y/(x_1, x_2, x_3)) = \frac{P(y) * P(x_1, x_2, x_3/y)}{P(x_1, x_2, x_3)}$$

$$= \frac{P(y) * P(x_1/y) * P(x_2/y) * P(x_3/y)}{P(x_1) * P(x_2) * P(x_3)}$$

Now according to our dataset

$$P(Yes|x_1, x_2, x_3) = \frac{P(Yes) * P(x_1/Yes) * P(x_2/Yes) * P(x_3/Yes)}{\cancel{P(x_1) * P(x_2) * P(x_3)} \rightarrow \text{constant}}$$

$$P(No/x_1, x_2, x_3) = \frac{P(No) * P(x_1/No) * P(x_2/No) * P(x_3/No)}{\cancel{P(x_1) * P(x_2) * P(x_3)} \text{ Constant}}$$

We can simply calculate the numerator to get the probabilities

For an new data point, the class for which the probability is max is assigned.

Practical Example

Day	Outlook	Temp	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes

Outlook

	Yes	No	P(E/Yes)	P(E/No)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rain	3	2	3/9	2/5

D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Overcast	4	0	4/9	0/5
Rain	3	2	3/9	2/5

PLAY(Y/N)

	P(Yes)	P(No)
Yes = 9	9/14	5/14
No = 5		

Temperature

	Yes	No	P(E|Yes)	P(E|No)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5

Test data => (sunny, Hot)

$$\therefore P(Yes|Sunny, Hot) = P(Yes) * P(Sunny/Yes) * P(Hot/Yes)$$

$$= \frac{\cancel{9}}{\cancel{14}_7} * \frac{\cancel{2}}{\cancel{9}} * \frac{2}{9} = \frac{2}{63} \approx 0.0317$$

$$P(No| Sunny, Hot) = P(No) * P(Sunny/No) * P(Hot/No)$$

$$= \frac{\cancel{5}}{\cancel{14}_7} * \frac{3}{\cancel{5}} * \frac{\cancel{2}}{5} = \frac{3}{35} \approx 0.085$$

$$P(Yes/(sunny, Hot)) = \frac{0.031}{(0.031 + 0.085)} = 0.27 = 27\%$$

$$P(No|(sunny, Hot)) = \frac{0.085}{0.031 + 0.085} = 0.73 = 73\%$$

$\therefore$ If for test data, we have sunny outlook and Hot temperature

o/p => 0 → person is not playing tennis ← $\therefore$ 73% they will not play tennis

27% they will play tennis

Variants of Naive Bayes

1. Bernoulli Naive Bayes
2. Multinomial Naive Bayes
3. Gaussian Naive Bayes

1. Bernoulli Naive Bayes

Whenever your features follow a Bernoulli distribution, then we have to use Bernoulli Naive Bayes.

Bernoulli → 0, 1 (2- categories in each feature)

Dataset

f_1	f_2	f_3	O/P
Yes	Pass	Male	Yes
Yes	Fail	Female	Yes
No	Pass	Male	No

⇒

f_1	f_2	f_3
1	1	1
1	0	0
0	1	1

→ Sparse matrix

The bernoulli NB is also used during NLP use case because after converting text to vectors as well we get some similar numeric value

2. Multinomial Naive Bayes

I/P ⇒ Text

Dataset: Spam Classification

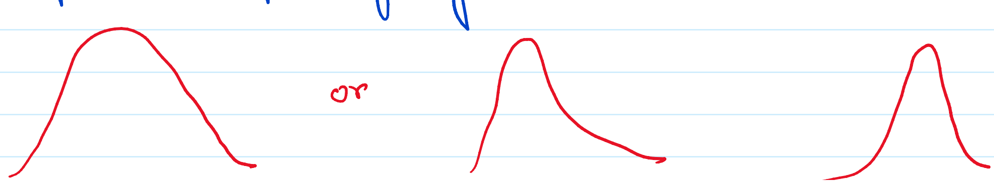
Email Body	Spam / Not Spam
You have won million $ lottery	Spam
Sakalya: you are selected for ML role	HAM

⇓

Convert text to numeric value (vector) ⇒ NLP techniques
- Bag of words
- TF-IDF
- Word2vec

3. Gaussian Naive Bayes

If the features are following a Gaussian distribution, then we use Gaussian NB



right-skewed left-skewed

Dataset → continuous features