

Naïve Bayes

25 May 2023 20:13

- * It comes under Supervised Learning.
- * It is a classification algorithm based on Baye's theorem.
- * It is a classification algorithm based on Baye's theorem with an assumption of independence among the independent variables [Correlation among independent variables should be low]
- * It works well with huge data and mostly used to solve text kind of data.
- Eg: * Gmail Classification
Twitter Sentiment analysis.

Baye's Equation

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

(posterior probability) (prior probability)

$P(c|x)$ → posterior probability of class given predictors
 Posterior probability means if input data is given then to which target/class is decided by $P(c|x)$.
 Input / output

Hey, you have $\overset{(x)}{\text{Ham}}$ $\overset{?}{\text{Spam}}$
 won prize worth of \$10000.

$P(x|c)$ → probability of predictor given class.
 $P(x)$ → predictor prior probability
 $P(c)$ → class prior probability.

How Naïve Bayes works?

- Step1:- Convert the data into frequency table.
- Step2:- Create likelihood table by finding the probability.
- Step3:- Use naïve bayes equation to calculate the posterior probability for each class. $P(c|x)$
- Step4:- The class with highest posterior probability is considered as outcome of the prediction.

Problem Statement:- Predict whether children can play or not if weather is sunny?
 $x = \text{Sunny}$ $y = ?$ $\frac{P(Yes|\text{Sunny})}{P(No|\text{Sunny})} = ?$

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes

Step1: Create frequency table.

weather	Yes	No
Overcast	4	0
Sunny	3	2
Rainy	2	3
Total	9	5

Step2: Create likelihood table by finding probabilities

weather	Yes	No	
Overcast	4	0	$P(\text{Overcast}) = \frac{4}{14}$
Sunny	3	2	$P(\text{Sunny}) = \frac{5}{14}$
Rainy	2	3	$P(\text{Rainy}) = \frac{5}{14}$
	$P(\text{Yes}) = \frac{9}{14}$	$P(\text{No}) = \frac{5}{14}$	

Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No ✓
Rainy	No ✓
Sunny	Yes
Rainy	Yes
Rainy	No ✓
Overcast	Yes
Overcast	Yes
Rainy	No ✓

Rainy	2	3
Total	9	5

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(Yes|Sunny) = \frac{P(Sunny|Yes)P(Yes)}{P(Sunny)}$$

$$P(Sunny|Yes) = \frac{3}{9}$$

$$P(Yes|Sunny) = \frac{\left(\frac{3}{9}\right) \times \left(\frac{9}{14}\right)}{\frac{5}{14}}$$

$$P(Yes|Sunny) = \frac{3}{5}$$

$$P(Yes|Sunny) = 0.6 \approx 60\%$$

Rainy	2	3	$P(Rainy) = \frac{5}{14}$
Total	9	5	

$$P(Yes) = \frac{9}{14} \quad P(No) = \frac{5}{14}$$

$$P(Go|Sunny) = \frac{P(Sunny|Go)P(Go)}{P(Sunny)}$$

$$P(Sunny|No) = \frac{2}{5}$$

$$P(No|Sunny) = \frac{\frac{2}{5} \times \frac{5}{14}}{\frac{5}{14}} = \frac{2}{5}$$

$$P(No|Sunny) = \frac{2}{5}$$

$$P(No|Sunny) = 0.4 \approx 40\%$$

weather	prediction
Sunny	Yes

Rainy ?