NAIVE BAYES CLASSIFIER

Contents

Independent Event

- An event that does not affect the occurrence of another subsequent event in a random experiment is an independent event.
- Ex. Tossing a coin.

Here, Sample Space S = {H, T} and both H and T are independent events.

If A and B are independent events then the probability of both occurring is

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

Dependent Event

- The occurance\ non-occurance of one event affect the occurance\ non-occurance of the other.

If A and B are dependent events then the probability of both occurring is

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

Probability of B given A

Marbles in a Bag

- 2 blue and 3 red marbles are in a bag.
- What are the chances of getting a blue marble?
- Answer: - The chance is 2 in 5

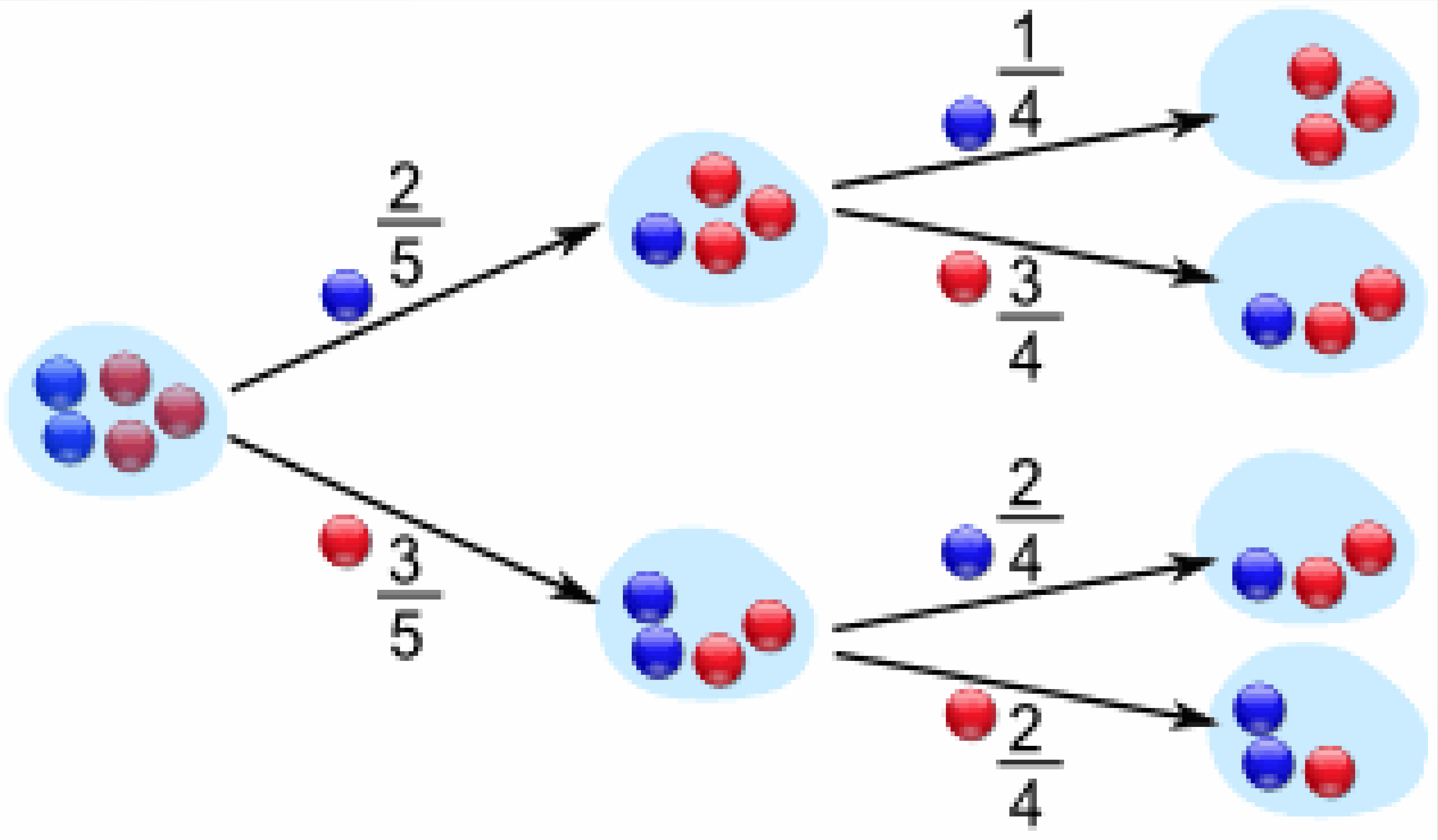
Cont...

- But after taking one marble
- **Situation may change!**
- So the next time:

  if we got a red marble before, then the chance of a blue marble next is 2 in 4

- if we got a blue marble before, then the chance of a blue marble next is 1 in 4

Cont..

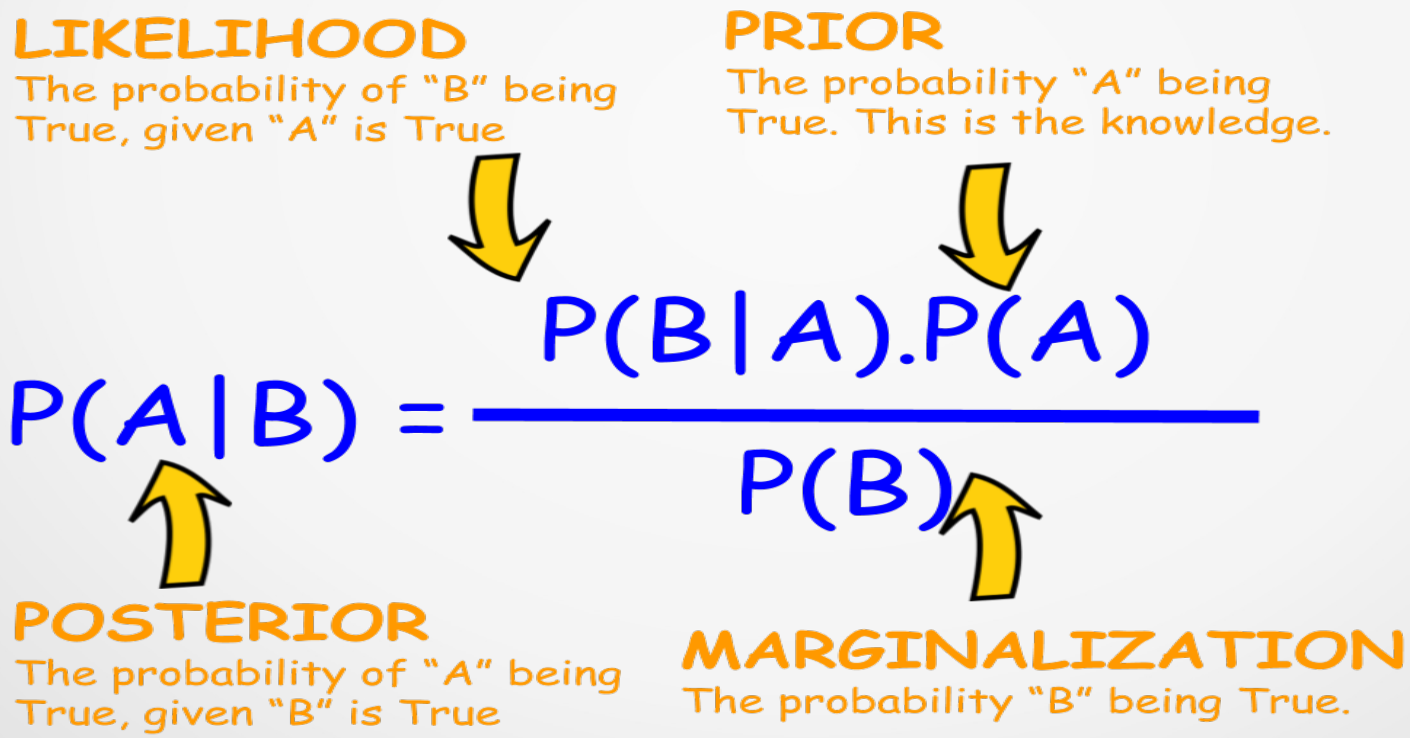


Conditional Probability

- In probability theory, conditional probability is a measure of the probability of an event given that another event has already occurred.
- If the event of interest is A and the event B is assumed to have occurred, "the conditional probability of A given B", or "the probability of A under the condition B", is usually written as P(A|B).
- The probability that any given person has a cough on any given day maybe only 5%. But if we know or assume that the person has a cold, then they are much more likely to be coughing. The conditional probability of coughing given that person have a cold might be a much higher 75%.

Baye’s Theorem

- In probability theory and Statistics, Bayes' theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event.



Cont...

- A doctor knows that meningitis causes stiff neck 50% of the time
  $\longrightarrow$ **likelihood**
- Prior probability of any patient having meningitis is 1/50,000
  $\longrightarrow$ **prior**
- Prior probability of any patient having stiff neck is 1/20
  $\longrightarrow$ **Marginalization**

  If a patient has stiff neck,
- what's the probability he/she has meningitis ?

$$\boxed{P(M \mid S) = \frac{P(S \mid M)P(M)}{P(S)} = \frac{0.5 \times 1/50000}{1/20} = 0.0002}$$

Naive Bayes Classifiersifier

- A classifier is a machine learning model that is used to discriminate different objects based on certain features.
- Naive bayes is a supervised learning algorithm for classification so the task is to find the class of observation (data point) given the values of features. Naive bayes classifier calculates the probability of a class given a set of feature values.
- The crux of the classifier is based on the Bayes theorem.

Cont...

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

- $P(c|x)$ is the posterior probability of *class* (*target*) given *predictor* (*attribute*).
- $P(c)$ is the prior probability of *class*.
- $P(x|c)$ is the likelihood which is the probability of *predictor* given *class*.
- $P(x)$ is the prior probability of *predictor*.

Example using weather dataset

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Cont...

- The posterior probability can be calculated by:
  - Constructing a frequency table for each attribute against the target.
  - Transforming the frequency tables to likelihood tables
  - Use the Naive Bayesian equation to calculate the posterior probability for each class
  - The class with the highest posterior probability is the outcome of prediction.

Cont..

$$P(x \mid c) = P(Sunny \mid Yes) = 3/9 = 0.33$$

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

Likelihood Table		Play Golf		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	2/9	3/5	5/14
		9/14	5/14	

$$P(x) = P(Sunny) = 5/14 = 0.36$$

$$P(c) = P(Yes) = 9/14 = 0.64$$

Posterior Probability: $P(c \mid x) = P(Yes \mid Sunny) = 0.33 \times 0.64 \div 0.36 = 0.60$

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
		9	5	14

$$P(x \mid c) = P(Sunny \mid No) = 2/5 = 0.4$$

$$P(x) = P(Sunny) = 5/14 = 0.36$$

$$P(c) = P(No) = 5/14 = 0.36$$

Posterior Probability: $P(c \mid x) = P(No \mid Sunny) = 0.40 \times 0.36 \div 0.36 = 0.40$

Cont...

The likelihood tables for all four predictors.

Frequency Table

		Play Golf	
		Yes	No
Outlook	Sunny	**3**	**2**
	Overcast	**4**	**0**
	Rainy	**2**	**3**

		Play Golf	
		Yes	No
Humidity	High	**3**	**4**
	Normal	**6**	**1**

		Play Golf	
		Yes	No
Temp.	Hot	**2**	**2**
	Mild	**4**	**2**
	Cool	**3**	**1**

		Play Golf	
		Yes	No
Windy	False	**6**	**2**
	True	**3**	**3**

Likelihood Table

		Play Golf	
		Yes	No
Outlook	Sunny	**3/9**	**2/5**
	Overcast	**4/9**	**0/5**
	Rainy	**2/9**	**3/5**

		Play Golf	
		Yes	No
Humidity	High	**3/9**	**4/5**
	Normal	**6/9**	**1/5**

		Play Golf	
		Yes	No
Temp.	Hot	**2/9**	**2/5**
	Mild	**4/9**	**2/5**
	Cool	**3/9**	**1/5**

		Play Golf	
		Yes	No
Windy	False	**6/9**	**2/5**
	True	**3/9**	**3/5**

Cont...

- In this example we have 4 inputs (predictors). The final posterior probabilities can be standardized between 0 and 1.

Outlook	Temp	Humidity	Windy	Play
Rainy	Cool	High	True	?

Cont...

$$P(Yes \mid X) = P(Rainy \mid Yes) \times P(Cool \mid Yes) \times P(High \mid Yes) \times P(True \mid Yes) \times P(Yes)$$

$$P(Yes \mid X) = 2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.00529 \qquad 0.2 = \frac{0.00529}{0.02057 + 0.00529}$$

$$P(No \mid X) = P(Rainy \mid No) \times P(Cool \mid No) \times P(High \mid No) \times P(True \mid No) \times P(No)$$

$$P(No \mid X) = 3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.02057 \qquad 0.8 = \frac{0.02057}{0.02057 + 0.00529}$$

Applications of Naive Baye's Algorithms

- **Real time Prediction:** Naive Bayes is an eager learning classifier and it is sure fast. Thus, it could be used for making predictions in real time.
- **Multi class Prediction:** This algorithm is also well known for multi class prediction feature. Here we can predict the probability of multiple classes of target variable.
- **Text classification/ Spam Filtering/ Sentiment Analysis:** Naive Bayes classifiers mostly used in text classification (due to better result in multi class problems and independence rule) have higher success rate as compared to other algorithms. As a result, it is widely used in Spam filtering (identify spam e-mail) and Sentiment Analysis (in social media analysis, to identify positive and negative customer sentiments)
- **Recommendation System:** Naive Bayes Classifier and Collaborative Filtering together builds a Recommendation System that uses machine learning and data mining techniques to filter unseen information and predict whether a user would like a given resource or not

Applications of Naive Baye's Algorithms

- **Real time Prediction:** Naive Bayes is an eager learning classifier and it is sure fast. Thus, it could be used for making predictions in real time.
- **Multi class Prediction:** This algorithm is also well known for multi class prediction feature. Here we can predict the probability of multiple classes of target variable.
- **Text classification/ Spam Filtering/ Sentiment Analysis:** Naive Bayes classifiers mostly used in text classification (due to better result in multi class problems and independence rule) have higher success rate as compared to other algorithms. As a result, it is widely used in Spam filtering (identify spam e-mail) and Sentiment Analysis (in social media analysis, to identify positive and negative customer sentiments)
- **Recommendation System:** Naive Bayes Classifier and Collaborative Filtering together builds a Recommendation System that uses machine learning and data mining techniques to filter unseen information and predict whether a user would like a given resource or not

Pros and Cons of Naive Bayes Algorithm

- Pros
  - The assumption that all features are independent makes naive bayes algorithm very fast compared to complicated algorithms. In some cases, speed is preferred over higher accuracy.
  - It works well with high-dimensional data such as text classification, email spam detection.
- Cons
  - The assumption that all features are independent is not usually the case in real life so it makes naive bayes algorithm less accurate than complicated algorithms. Speed comes at a cost!