Naïve Bayes Classifier

Naïve Bayes Classifier Algorithm

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.
- It is mainly used in text classification that includes a high-dimensional training dataset.
- Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
- Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

- The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as:
- Naïve: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.
- Bayes: It is called Bayes because it depends on the principle of Bayes' Theorem.

Bayes' Theorem:

- Bayes' theorem is also known as Bayes' Rule or Bayes' law, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
- The formula for Bayes' theorem is given as:
- Naïve Bayes Classifier Algorithm

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

Where,

- P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

- P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.
- P(A) is Prior Probability: Probability of hypothesis before observing the evidence.
- P(B) is Marginal Probability: Probability of Evidence.

STEPS

- find the P(A)- prior probability
- Convert the given dataset into frequency tables.
- Generate Likelihood table by finding the probabilities of given features. (Conditional probabilities)
- Now, use Bayes theorem to calculate the posterior probability.

Advantages of Naïve Bayes Classifier

- Naïve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
- It can be used for Binary as well as Multi-class Classifications.
- It performs well in Multi-class predictions as compared to the other Algorithms.
- It is the most popular choice for text classification problems.

- **Disadvantages of Naïve Bayes Classifier:**

- Naive Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between features.

Applications of Naïve Bayes Classifier:

- It is used for Credit Scoring.
- It is used in medical data classification.
- It can be used in real-time predictions because Naïve Bayes Classifier is an eager learner.
- It is used in Text classification such as Spam filtering and Sentiment analysis.

NAIVE BAYES CLASSIFIER – Example -1

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
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
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

$\langle Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong \rangle$

Calculate prior probability(P)-> Y & N

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
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
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

$$P(PlayTennis = yes) = 9/14 = .64$$

$$P(PlayTennis = no) = 5/14 = .36$$

Calculate the conditional probability->each attribute

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
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
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

Outlook	Y	N
sunny	2/9	3/5
overcast	4/9	0
rain	3/9	2/5

Calculate the conditional probability->each attribute

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
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
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

Outlook	Y	N
sunny	2/9	3/5
overcast	4/9	0
rain	3/9	2/5
Tempreature		
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humidity	Y	N
high	3/9	4/5
normal	6/9	1/5
Windy		
Strong	3/9	3/5
Weak	6/9	2/5

Classify the new example ->Y or N

$$\langle Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong\rangle$$

$$v_{NB} = \underset{v_j \in \{yes,no\}}{\text{argmax}} \; P(v_j) \prod_i P(a_i|v_j)$$

$$= \underset{v_j \in \{yes,no\}}{\text{argmax}} \; P(v_j) \quad P(Outlook = sunny|v_j) P(Temperature = cool|v_j)$$

$$\cdot \; P(Humidity = high|v_j) P(Wind = strong|v_j)$$

$$v_{NB}(yes) = P(yes)\; P(sunny|yes)\; P(cool|yes)\; P(high|yes)\; P(strong|yes) = .0053$$

$$v_{NB}(no) = P(no)\; P(sunny|no)\; P(cool|no)\; P(high|no)\; P(strong|no) = .0206$$

$$v_{NB}(yes) = \frac{v_{NB}(yes)}{v_{NB}(yes)+v_{NB}(no)} = 0.205 \qquad v_{NB}(no) = \frac{v_{NB}(no)}{v_{NB}(yes)+v_{NB}(no)} = 0.795$$

Example 2

- Estimate conditional probabilities of each attributes {color, legs, height, smelly} for the species classes: {M, H} using the data given in the table.
- Using these probabilities estimate the probability values for the new instance – (Color=Green, legs=2, Height=Tall, and Smelly=No).

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

Calculate Prior probability (M) & P(H)

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

New Instance

(Color=Green, legs=2, Height=Tall, and Smelly=No)

$$P(M) = \frac{4}{8} = 0.5 \quad P(H) = \frac{4}{8} = 0.5$$

Color	M	H
White	2/4	3/4
Green	2/4	1/4

Calculate for all attributes(conditional probabilities) ->possible outcome, possible values

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

Color	M	H
White	2/4	3/4
Green	2/4	1/4

Legs	M	H
2	1/4	4/4
3	3/4	0/4

Height	M	H
Tall	3/4	2/4
Short	1/4	2/4

Smelly	M	H
Yes	3/4	1/4
No	1/4	3/4

Calculate the probability of new instance->classify whether M or H(Using prior & Conditional)

$$P(M) = \frac{4}{8} = 0.5 \quad P(H) = \frac{4}{8} = 0.5$$

Color	M	H
White	2/4	3/4
Green	2/4	1/4

Legs	M	H
2	1/4	4/4
3	3/4	0/4

Height	M	H
Tall	3/4	2/4
Short	1/4	2/4

Smelly	M	H
Yes	3/4	1/4
No	1/4	3/4

$$p(M|New\ Instance) = p(M) * p(Color = Green|M) * p(Legs = 2|M) * p(Height = tall|M) * p(Smelly = no\ |M)$$

$$p(M|New\ Instance) = 0.5 * \frac{2}{4} * \frac{1}{4} * \frac{3}{4} * \frac{1}{4} = 0.0117$$

$$p(H|New\ Instance) = p(H) * p(Color = Green|H) * p(Legs = 2|H) * p(Height = tall|H) * p(Smelly = no\ |H)$$

$$p(H|New\ Instance) = 0.5 * \frac{1}{4} * \frac{4}{4} * \frac{2}{4} * \frac{3}{4} = 0.047$$

$$p(H|New\ Instance) > p(M|New\ Instance)$$

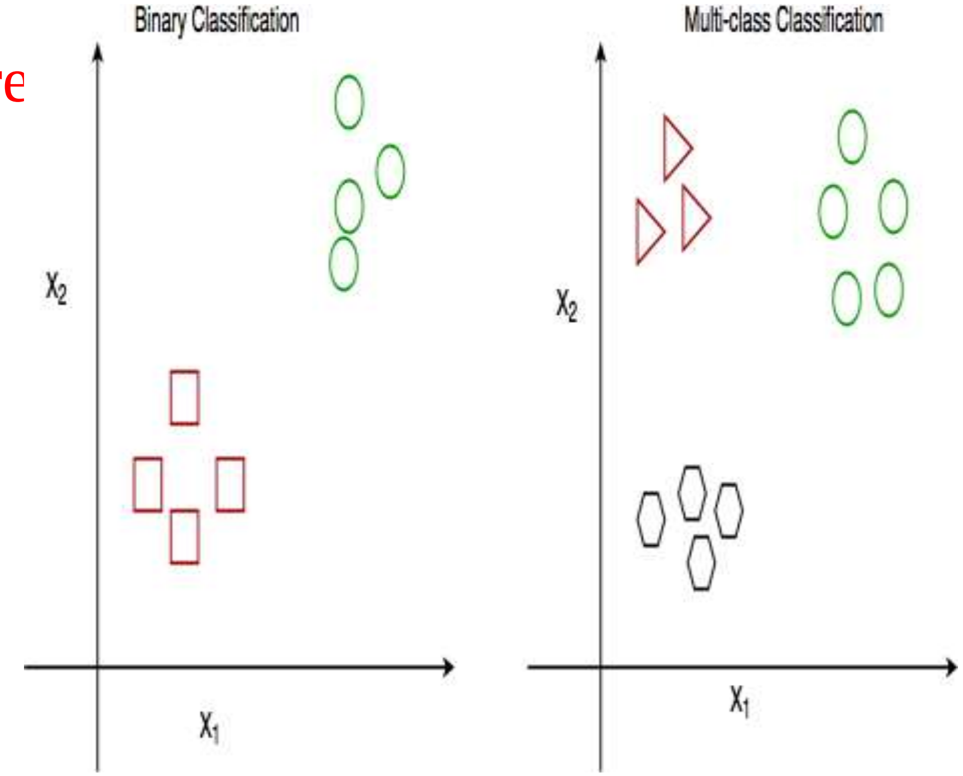
Hence the new instance belongs to Speices H

Text Classification

- **Text Classification** is an example of supervised **machine learning** task since a labelled dataset containing **text** documents and their labels is used for train a **classifier.**
- The **Naive Bayes classifier** is a simple **classifier** that classifies based on probabilities of events. It is the applied commonly to **text classification**. **With** the training set, we can train a **Naive Bayes classifier** which we can **use** to automatically categorize a new sentence.
- TYPES
- **Binary Classification** : When we have to categorize given data into 2 distinct classes. Example – On the basis of given health conditions of a person, we have to determine whether the person has a certain disease or not.
- **Multiclass Classification** : The number of classes is more than 2. For Example – On the basis of data about different species of flowers, we have to determine which specie does our observation belong to.

Example Fig : Binary and Multiclass Classification. Here x_1, x_2 are our variables upon which the class is predicted.

How does classification works?



- Which means there are two possible outcomes:

- Suppose we have to predict whether a given patient has a certain disease or not, on the basis of 3 variables, called features.
- The patient has the said disease. Basically a result labelled “Yes” or “True”.
- The patient is disease free. A result labelled “No” or “False”.
- This is a binary classification problem.
- We have a set of observations called training data set, which comprises of sample data with actual classification results. We train a model, called Classifier on this data set, and use that model to predict whether a certain patient will have the disease or not.

EXAMPLE 3

Naïve Bayes Model – Text Classification Example

- Dataset for the text classification with training and test data is given below.
- The goal is to classify the test data into the right class as h or —h (read as not h).

	Document ID	Keywords in the document	Class h
Training Set	1	Love Happy Joy Joy Happy	Yes
	2	Happy Love Kick Joy Happy	Yes
	3	Love Move Joy Good	Yes
	4	Love Happy Joy Love Pain	Yes
	5	Joy Love Pain Kick Pain	No
	6	Pain Pain Love kick	No
Testing Set	7	Love Pain Joy Love Kick	?

Priori Probability

$$p(c|d) \propto p(c) \prod_{1 \le k \le n_d} p(t_k|c)$$

- The prior probabilities of a document being classified using the six documents are,

$$P(h) = \frac{4}{6} = \frac{2}{3}$$

$$and$$

$$p(-h) = \frac{2}{6} = \frac{1}{3}$$

- That is there is $\frac{2}{3}$ prior probability that a document will be classified as h and $\frac{1}{3}$ probability of not h.

Conditional Probability

	Document ID	Keywords in the document	Class h
Training Set	1	Love Happy Joy Joy Happy	Yes
	2	Happy Love Kick Joy Happy	Yes
	3	Love Move Joy Good	Yes
	4	Love Happy Joy Love Pain	Yes
	5	Joy Love Pain Kick Pain	No
	6	Pain Pain Love kick	No
Testing Set	7	Love Pain Joy Love Kick	?

- The conditional probability for each term is the relative frequency of the term occurring in each class of the documents 'h class' and 'not h class'.

Testing Example:

Love Pain Joy Love Kick = ?

Class h	Class -h
$P(Love \mid h) = 5/19$	$P(Love \mid -h) = 2/9$
$P(Pain \mid h) = 1/19$	$P(Pain \mid -h) = 4/9$
$P(Joy \mid h) = 5/19$	$P(Joy \mid -h) = 1/9$
$P(Kick \mid h) = 1/19$	$P(Kick \mid -h) = 2/9$

Testing Example:

Love Pain Joy Love Kick = ?

$$p(c|d) \propto p(c) \prod_{1 \le k \le n_d} p(t_k|c)$$

Class h	Class -h
$P(Love \mid h) = 5/19$	$P(Love \mid -h) = 2/9$
$P(Pain \mid h) = 1/19$	$P(Pain \mid -h) = 4/9$
$P(Joy \mid h) = 5/19$	$P(Joy \mid -h) = 1/9$
$P(Kick \mid h) = 1/19$	$P(Kick \mid -h) = 2/9$

$$P(h|d_7) = P(h) * (P(Love|h) * (P(Love|h) * P(Pain|h) * P(Joy|h) * P(Kick|h)$$
$$= (2/3) * (5/19) * (5/19) * (1/19) * (5/19) * (1/19) = \mathbf{-0.0000067}$$

$$p(-h|d7) = p(-h) * P(Love| - h) * P(Love| - h) * P(Pain| - h) * P(Joy| - h) * P(Kick| - h)$$
$$= (1/3) * (2/9) * (2/9) * (4/9) * (1/9) * (2/9) = \mathbf{0.00018}$$

Class label for testing example is: **No**