

## Practical No. 4

Aim: To demonstrate classification rule process on dataset using naive bayes algorithm.

Date :

## Practical No. 4



Aim : To demonstrate classification rule process on dataset using naive bayes algorithm.

Theory :

Naive Bayes classifier algorithm :

- Naive Bayes algo. is a supervised learning algo., 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.
- Naive Bayes classifier is one of the simple and most effective classification algo.
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
- Some popular examples of Naive Bayes algo. are spam filtration, sentiment analysis and classifying articles.

Why is it called Naive Bayes ?

The Naive Bayes algo. is comprised of two words Naive and Bayes, which can be described as :

Naive : It is called Naive 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 basis of color, shape and taste then red, spherical, and sweet fruit is identified or 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's theorem.

Output :

(1)

Weka Explorer

Preprocess Classifier Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☒ Use training set

☐ Supplied test set

☒ Cross-validation Folds 10

☐ Percentage split

More options...

(Nom) play

Start Stop

Result list (right-click for options)

16.03.17: NaiveBayes

Classifier output

--- An information ---

4. Naive Bayes classifier

Relation weather.symbolic

Instances 14

Attributes 4

outlook

sunny 2.0 4.0

overcast 3.0 1.0

rainy 4.0 3.0

[total] 12.0 8.0

temperature

hot 1.0 3.0

mild 5.0 3.0

cool 4.0 2.0

[total] 12.0 8.0

humidity

high 4.0 5.0

normal 7.0 2.0

[total] 11.0 7.0

windy

TRUE 4.0 4.0

FALSE 7.0 3.0

[total] 11.0 7.0

Test mode: 10-fold cross-validation

--- Classifier model (full training set) ---

Naive Bayes classifier

Attribute Class

yes no

(0.61) (0.39)

Time taken to build model: 0 seconds

--- Stratified cross-validation ---

--- Summary ---

Correctly Classified Instances 8 57.1429 %

Incorrectly Classified Instances 6 42.8571 %

Kappa statistic -0.0244

Mean absolute error 0.4374

Root mean squared error 0.4216

Relative absolute error 91.8971 %

Root relative squared error 99.6452 %

Status OK

(2)

Weka Explorer

Preprocess Classifier Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☒ Use training set

☐ Supplied test set

☒ Cross-validation Folds 10

☐ Percentage split

More options...

(Nom) play

Start Stop

Result list (right-click for options)

16.03.17: NaiveBayes

Classifier output

--- An information ---

4. Naive Bayes classifier

Relation weather.symbolic

Instances 14

Attributes 4

outlook

sunny 2.0 4.0

overcast 3.0 1.0

rainy 4.0 3.0

[total] 12.0 8.0

temperature

hot 1.0 3.0

mild 5.0 3.0

cool 4.0 2.0

[total] 12.0 8.0

humidity

high 4.0 5.0

normal 7.0 2.0

[total] 11.0 7.0

windy

TRUE 4.0 4.0

FALSE 7.0 3.0

[total] 11.0 7.0

Test mode: 10-fold cross-validation

--- Classifier model (full training set) ---

Naive Bayes classifier

Attribute Class

yes no

(0.61) (0.39)

Time taken to build model: 0 seconds

--- Stratified cross-validation ---

--- Summary ---

Correctly Classified Instances 8 57.1429 %

Incorrectly Classified Instances 6 42.8571 %

Kappa statistic -0.0244

Mean absolute error 0.4374

Root mean squared error 0.4216

Relative absolute error 91.8971 %

Root relative squared error 99.6452 %

Total number of instances 14

--- Detailed Accuracy By Class ---

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
yes	0.778	0.0	0.639	0.778	0.7	0.578	yes
no	0.2	0.222	0.333	0.2	0.25	0.578	no
Aweighted Avg.	0.571	0.504	0.528	0.571	0.538	0.578	

--- Confusion Matrix ---

a b

a b

2 2

4 1

Status OK



Date :



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:

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

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.

Working of Naive Bayes' classifier:

Working of Naive Bayes classifier can be understood with the help of the below example:

Suppose we have a dataset of weather conditions and corresponding target variable "play". So using this dataset we need to decide that whether we should play or not on a particular day according to the weather condition. So to solve this problem, we need to follow the below steps:

- Convert the given dataset into frequency tables.
- Generate likelihood table by finding the probabilities of given frequency.
- Now, use Bayes' theorem to calculate the posterior probability.

Step 1: Initially, we have to load the required dataset in the weka tool using 'choose file' options. Here we select the weather-nominal dataset to execute.



Date :



Step 2: Now we have to go to the classify tab on the left side and click on the choose button and select the naive Bayes algorithm in it.

Step 3: Now to change the parameters click on the right side at the choose button and we are accepting the default values in this example.

Step 4: We choose the percentage split as our measurement method from the "Test" choices in the main panel. Since we don't have a separate test data collection, we'll use the percentage split of 66 percent to get a good idea of the model's accuracy. Our dataset contains 14 examples, with 9 being used for training and 5 being used for testing.

Step 5: To generate the model, we now click 'start'. When the model is done, the evaluation statistic will appear in right panel.

Result :

Thus the classification on data set is performed by Naive Bayes classification.

Viva Questions:

① What is a Naive Bayes classifier?

→ A naive Bayes classifier is an algo. that uses Bayes' theorem to classify objects. Naive Bayes classifiers assume strong or naive independence between attributes of data points.

② What are the basic assumption?

→ The basic assumption in naive Bayes is one of conditional independence between all independent variable features.



Date :

Ques.3) What are the advantages of Naive Bayes classifier?

- ① It is simple and easy to implement.
- ② It doesn't require as much training data.
- ③ It handles both discrete and continuous data.
- ④ It is fast and can be used to make real time predictions.
- ⑤ It is not sensitive to irrelevant features.

Ques.4) Name the different problem statements you can solve using Naive Bayes.

- ① A fruit may be considered to be an apple if it is red, round and about 3 inch in diameter.
- ② If the weather is sunny, then the player should play or not?