

ASSIGNMENT No: 11

Title: Perform the data classification algorithm using any Classification algorithm

Problem Statement: Perform the data classification algorithm using any Classification algorithm

Prerequisite:

Basics of Python

Software Requirements: Jupyter

Hardware Requirements:

PIV, 2GB RAM, 500 GB HDD

Learning Objectives:

Learn to Perform the data classification algorithm using any Classification algorithm

Outcomes:

After completion of this assignment students are able to understand Perform the data classification algorithm using any Classification algorithm

Theory:

What is Classification?

We use the training dataset to get better boundary conditions which could be used to determine each target class. Once the boundary conditions are determined, the next task is to predict the target class. The whole process is known as classification.

Target class examples:

- Analysis of the customer data to predict whether he will buy computer accessories (Target class: Yes or No)
- Classifying fruits from features like color, taste, size, weight (Target classes: Apple, Orange, Cherry, Banana)
- Gender classification from hair length (Target classes: Male or Female)

Let's understand the concept of classification algorithms with gender classification using hair length (by no means am I trying to stereotype by gender, this is only an example). To classify gender (target class) using hair length as feature parameter we could train a model using any classification algorithms to come up with some set of boundary conditions which can be used to differentiate the male and female genders using hair length as the training feature. In gender classification case the boundary condition could be the proper hair length value. Suppose the differentiated boundary hair length value is 15.0 cm then we can say that if hair length is less than 15.0 cm then gender could be male or else female.

Classification Algorithms vs Clustering Algorithms

In clustering, the idea is not to predict the target class as in classification, it's more ever trying to group the similar kind of things by considering the most satisfied condition, all the items in the same group should be similar and no two different group items should not be similar.

Group items Examples:

- While grouping similar language type documents (Same language documents are one group.)
- While categorizing the news articles (Same news category(Sport) articles are one group)

Let's understand the concept with clustering genders based on hair length example. To determine gender, different similarity measure could be used to categorize male and female genders. This could be done by finding the similarity between two hair lengths and keep them in the same group if the similarity is less (Difference of hair length is less). The same process could continue until all the hair length properly grouped into two categories.

Basic Terminology in Classification Algorithms

- Classifier: An algorithm that maps the input data to a specific category.
- Classification model: A classification model tries to draw some conclusion from the input values given for training. It will predict the class labels/categories for the new data.
- Feature: A feature is an individual measurable property of a phenomenon being observed.
- Binary Classification: Classification task with two possible outcomes. Eg: Gender classification (Male / Female)
- Multi-class classification: Classification with more than two classes. In multi-class classification, each sample is assigned to one and only one target label. Eg: An animal can be a cat or dog but not both at the same time.
- Multi-label classification: Classification task where each sample is mapped to a set of target labels (more than one class). Eg: A news article can be about sports, a person, and location at the same time.

Applications of Classification Algorithms

- Email spam classification
- Bank customers loan pay willingness prediction.
- Cancer tumor cells identification.
- Sentiment analysis
- Drugs classification
- Facial key points detection
- Pedestrians detection in an automotive car driving.

Types of Classification Algorithms

Classification Algorithms could be broadly classified as the following:

- Linear Classifiers
 - Logistic regression
 - Naive Bayes classifier
 - Fisher's linear discriminant
- Support vector machines
 - Least squares support vector machines
- Quadratic classifiers
- Kernel estimation
 - k-nearest neighbor
- Decision trees
 - Random forests
- Neural networks
- Learning vector quantization

Consider the annual rainfall details at a place starting from January 2012. We create an R time series object for a period of 12 months and plot it.

```
# Get the data points in form of a R vector.rainfall <-  
c(799,1174.8,865.1,1334.6,635.4,918.5,685.5,998.6,784.2,985,882.8,1071)
```

```
# Convert it to a time series object.  
rainfall.timeseries <- ts(rainfall,start = c(2012,1),frequency = 12)
```

```
# Print the timeseries data.  
print(rainfall.timeseries)
```

```
# Give the chart file a name.png(file =  
"rainfall.png")
```

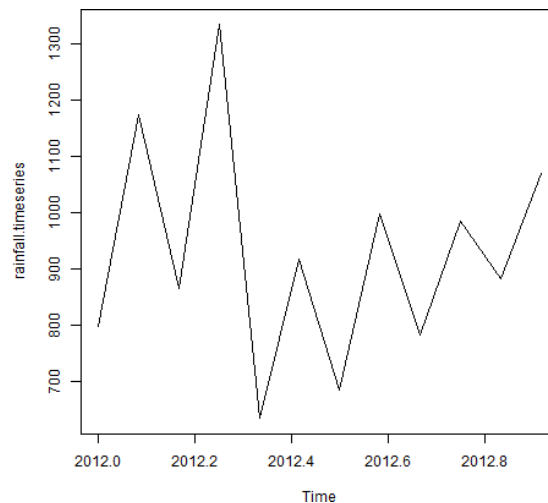
```
# Plot a graph of the time series.  
plot(rainfall.timeseries)
```

```
# Save the file.  
dev.off()
```

Output:

When we execute the above code, it produces the following result and chart –

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
2012	799.0	1174.8	865.1	1334.6	635.4	918.5	685.5	998.6
	Oct	Nov	Dec					
2012	985.0	882.8	1071.0					



Conclusion:- Thus, this way Performed data classification algorithm using any Classification algorithm