Classification in R Programming

**1. Objective**

In this tutorial, we will study classification in [R](https://data-flair.training/blogs/r-programming-tutorial/) in detail. We will also cover decision tree, Naïve Bayes classification. Along with this, we will also learn Support vector machine. To understand it in the best manner we will use images and real-time examples.

[](https://d2h0cx97tjks2p.cloudfront.net/blogs/wp-content/uploads/Classification-in-R-01.jpg)

**2. Introduction to Classification in R**

We use it to predict a categorical class label, such as weather: rainy, sunny, cloudy or snowy.

**2.1 Important points of Classification**

**a**. Many different Classifiers available.

**b**. Strengths and Weakness.

**c**. Decision Tree Classifiers:Good in explaining the classification result.

**d**. Naive Bayes Classifiers: Strong theory.

**e.**K – nn Classifiers: Lazy Classifiers.

**f**. Support Vector Machines: It states the art. It also performs very well in across in different domains in practice.

**Keywords**

classif

**Usage**

classification(trExemplObj,classLabels,valExemplObj=NULL,kf=5,kernel=”linear”)

**Arguments**

**a. trExemplObj**

It is an exemplars train eSet object.

**b. classLabels**

It is being stored in eSet object as variable name e.g “type”.

**c. valExemplObj**

It is known as exemplars validation eSet object.

**d. kf**

It is termed as the k-folds value of the cross-validation parameter. Also, the default value is 5-folds. By setting “Loo” or “LOO” a Leave-One-Out Cross-Validation which we have to perform.

**e. kernel**

In classification analysis, we use a type of Kernel.The default kernel is “linear”.

**f. classL**

The labels of the train set.

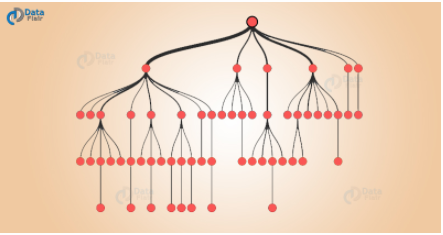
**g. valClassL**

It is termed as the labels of the validation set if not NULL.

**h. predLbls**

it is defined as the predicted labels according to the classification analysis.

**3. Decision Tree in R**

[](https://d2h0cx97tjks2p.cloudfront.net/blogs/wp-content/uploads/class-3.png)

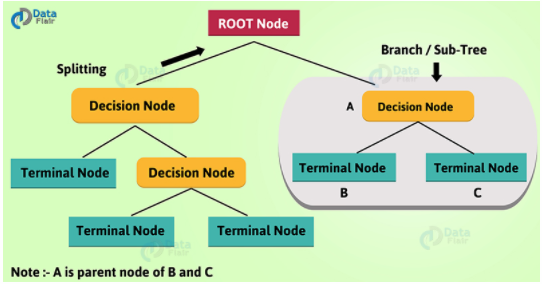
It is a type of supervised learning algorithm. We use it for classification problems. It works for both types input and output variables. In this technique, we split the population into two or more homogeneous sets. Moreover, it is based on most significant splitter/differentiator in input variables.

The decision tree is powerful non-linear classifiers. It uses a tree structure to model the relationships among the features. Also, potential outcomes. A decision tree classifier uses a structure of branching decisions.

**In classifying data, the decision tree follows the steps mentioned below:**

* It puts all training examples to a root.
* Decision tree divides training examples based on selected attributes.
* Then it will select attributes by using some statistical measures.
* Recursive partitioning continues until no training example remains.

**3.1 Important terminologies related to Decision Tree**

[](https://d2h0cx97tjks2p.cloudfront.net/blogs/wp-content/uploads/Class-3.2.png)

* **Root Node**: It represents entire population or sample. Moreover, it gets divided into two or more homogeneous sets.
* **Splitting**: Process of dividing a node into two or more sub-nodes.
* **Decision Tree**: it is produced when a sub-node splits into further sub-nodes.
* **Leaf/Terminal Node**: Nodes do not split is being called Leaf or Terminal node.
* **Pruning:** When we remove sub-nodes of a decision node, this process is being called pruning. You can say opposite process of splitting.
* **Branch / Sub-Tree**: A subsection of the entire tree is being called branch or sub-tree.
* **Parent and Child Node**:

A node, which is being divided into sub-nodes is being called parent node of sub-nodes. Whereas sub-nodes are the child of a parent node.

**3.2 Types of Decision Tree**

**a. Categorical(classification) Variable Decision Tree**: Decision Tree which has categorical target variable.

**b. Continuous(Regression) Variable Decision Tree:** Decision Tree has continuous target variable.

**3.3 Categorical(classification)Trees vs Continuous(regression)Trees**

**a**. Regression trees are used when the dependent variable is continuous. While classification trees are used when the dependent variable is categorical.

**b**. In continuous, a value obtained is mean response of observation falls in that region.

**c**. In classification, a value obtained by a terminal node is a mode of observations falls in that region.

**d**. There is one similarity in both the cases. Splitting process continuos results into grown trees until it reaches to stopping criteria. But, the grown tree is likely to overfit data, leading to poor accuracy on unseen data. This brings ‘pruning’. Pruning is one of the techniques which uses tackle overfitting.

**3.4 Advantages of Decision Tree in R**

**a. Easy to Understand:**

It does not need any statistical knowledge to read and interpret them. Its graphical representation is very intuitive and users can relate their hypothesis.

**b. Less data cleaning required**:

Compared to some other modeling techniques, it requires fewer data.

**c**. The data type is not a constraint: It can handle both numerical and categorical variables.

**d**. It is Simple to understand and interpret.

**e**. It requires little data preparation.

**f.**It works with both numerical and categorical data.

**g**. It handles nonlinearity.

**h**. It is possible to confirm a model using statistical tests. Gives you confidence it will work on new data sets.

**i**. It is Robust. It performs well even if you deviate from assumptions.

**j**. It scales to big data.

**3.5 Disadvantages of R Decision Tree**

**a. Overfitting**: It is one of the most practical difficulties for decision tree models. By setting constraints on model parameters and pruning we can solve this problem.

**b. Not fit for continuous variables**: While using continuous numerical variables. Whenever it categorizes variables in different categories, the decision tree loses information.

**3.6 Limitations of Decision Tree**

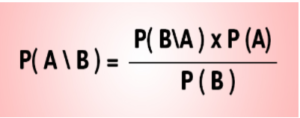
* To learn globally optimal tree is NP-hard, algos rely on greedy search.
* Easy to overfit the tree.
* Complex “if-then” relationships between features inflate tree size. eg XOR gate, multiplexor

**4. Introduction to Naïve Bayes classification**

[](https://data-flair.training/blogs/r-interview-questions-and-answers/)

We use Bayes’ theorem to make the prediction. It is based on prior knowledge and current evidence.

**Bayes’ theorem is expressed by the following equation:**

[](https://d2h0cx97tjks2p.cloudfront.net/blogs/wp-content/uploads/Class-4-formula.png)

where P(A) and P(B) are the probability of events A and B without regarding each other. P(A|B) is the probability of A conditional on B and P(B|A) is the probability of B conditional on A.

**4.1 Naive Bayes Classifier**

**Usage**

## S3 method for class ‘formula’:

naiveBayes(formula, data, …, subset, na.action = na.pass)

## Default S3 method:

naiveBayes(x, y, …)

**5. Introduction to Support Vector Machines**

**5.1 What is Support Vector Machine**

We use it to find the optimal hyperplane (line in 2D, a plane in 3D and hyperplane in more than 3 dimensions). Which helps in maximizes the margin between two classes. Support Vectors are observations that support hyperplane on either side.

It helps in solving a linear optimization problem. It also helps out in finding the hyperplane with the largest margin. We use the “Kernel Trick” to separate instances that are inseparable.

**Usage**

## S3 method for class ‘formula’:

svm(formula, data = NULL, …, subset, na.action =

na.omit, scale = TRUE)

## Default S3 method:

svm(x, y = NULL, scale = TRUE, type = NULL, kernel =

“radial”, degree = 3, gamma = 1 / ncol(as.matrix(x)), coef0 = 0, cost = 1, nu = 0.5,

class.weights = NULL, cachesize = 40, tolerance = 0.001, epsilon = 0.1,

shrinking = TRUE, cross = 0, probability = FALSE, fitted = TRUE,

…, subset, na.action = na.omit)

**5.2 Terminologies related to R SVM**

**Why Hyper plane?**

It is a line in 2D and plane in 3D. In higher dimensions (more than 3D), it’s called as a hyperplane. Moreover, SVM helps us to find a hyperplane that can separate two classes.

**What is Margin?**

A distance between the hyperplane and the closest data point is called a margin. But if we want to double it, then it would be equal to the margin.

**How to find the optimal hyperplane?**

First, we have to select two hyperplanes. They must separate the data with no points between them. Then maximize the distance between these two hyperplanes. The distance here is ‘margin’.

**What is Kernel?**

It is a method which helps to make SVM run in case of non-linear separable data points. We use a kernel function to transforms the data into a higher dimensional feature space. And also with the help of it to make it possible to perform the linear separation.

**Different Kernels**

**1**. linear: u’\*v

**2**. polynomial: (gamma\*u’\*v + coef0)^degree

**3**. radial basis (RBF) : exp(-gamma\*|u-v|^2)sigmoid : tanh(gamma\*u’\*v + coef0)

RBF is generally the most popular one.

**How SVM works?**

**a**. Choose an optimal hyperplane which maximizes margin.

**b**. Applies penalty for misclassifications (cost ‘c’ tuning parameter).

**c**. If the non-linearly separable the data points. Then transform data to high dimensional space. Where it is easier to classify with linear decision surfaces (Kernel trick).

**5.3 Advantages of SVM in R**

* If we are using Kernel trick in case of non-linear separable data then it performs very well.
* SVM works well in high dimensional space and in case of text or image classification.
* It does not suffer multicollinearity problem.

**5.4 Disadvantages of R SVM**

* It takes more time on large-sized data sets.
* SVM does not return probability estimates.
* In case of linearly separable data, this is almost like logistic regression.

**[Test Your R Knowledge](https://data-flair.training/blogs/category/quiz/r-quiz/" \t "_blank)**

**5.5 Support Vector Machine – Regression**

* Yes, we can use it for regression problem. Wherein dependent or target variable is continuous.
* The aim of SVM regression is same as classification problem i.e. to find the largest margin.

**6. Applications of Classification in R**

* An emergency room in a hospital measures 17 variables of newly admitted patients. Variables, like blood pressure, age and many more. A decision has to be taken whether to put the patient in an intensive care unit. Due to a high cost of I.C.U, those patients who may survive more a month are given high priority. Also, the problem is to predict high-risk patients. And to discriminate them from low-risk patients.
* A credit company receives hundreds of thousands of applications for new cards. The application contains information about several different attributes. Moreover, the problem is to categorize those who have good credit, bad credit or fall into a gray area
* Astronomers have been cataloging distant objects in the sky using a long exposure C.C.D images. Thus, the object needs to be labeled a star, galaxy etc. The data is noisy, and the images are very faint. hence, the cataloging can decades to complete. How can physicists automate the cataloging process and improve its effectiveness?