

# R Language学习Day1

2020年2月22日 17:58

## 包的安装

- 第一次安装一个包，使用命令`install.packages("package_name","dir")`即可。
- `dir`:包安装的路径。默认情况下是安装在`..\\library` 文件夹中的。可以通过本参数来进行修改，来选择安装的文件夹。
- 一个包仅需要安装一次。但和其他软件类似，包经常被其作者更新。使用命令`update.packages()`可以更新已经安装的包。
- 要查看已安装包的描述，可以使用`installed.packages()`命令，这将列出安装的包，以及它们的版本号、依赖关系等信息。

## 查看包的相关信息

- 查看包帮助：`library(help="package_name")`
- 主要内容包括：例如：包名、作者、版本、更新时间、功能描述、开源协议、存储位置、主要的函数
- 查看当前环境哪些包加载：`find.package()`
- 加载包：`library(package_name)` 或者 `require(package_name)`
- 移除包出内存：`detach(package:package_name)`
- 把包的数据加载到内存中：`attach(iris)`
- 列出所有安装的包：`library()`

## 常用 R 包

机器学习与统计学习类别就包含有实现分类、聚类、关联规则、时间序列分析等功能的R包，通过加载不同的R包就能够实现相应的数据挖掘功能

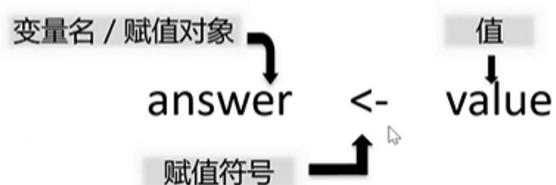
功能	函数及加载包
分类与预测	nnet函数需要加载BP神经网络nnet包； randomForest函数需要加载随机森林randomForest包； svm函数需要加载e1071包； tree函数需要加载CRAT决策树tree包等
聚类分析	hclust函数、kmeans()函数在stats包中
关联规则	apriori函数需要加载arules包
时间序列	arima函数需要加载forecast、tseries包

## 2.1数据对象与数据读写

## 数据对象类型

### 基本赋值语句

- `x <- 8`
- `a <- 'city'`



- R 语言的对象常见的数据类型有：字符型、数值型、逻辑型、复数型。此外，也可能是缺省值(NA)。

#### 判别和转换数据对象类型的函数

类型	判别	转换
character	<code>is.character()</code>	<code>as.character()</code>
complex	<code>is.complex()</code>	<code>as.complex()</code>
integer	<code>is.integer()</code>	<code>as.integer()</code>
logical	<code>is.logical()</code>	<code>as.logical()</code>
NA	<code>is.na()</code>	<code>as.na()</code>
numeric	<code>is.numeric()</code>	<code>as.numeric()</code>

```

1 data <- data.frame(c1=c(1,2,-4),c2=1:3,c3=TRUE,c4='123')
2 # 一、数据类型
3 ##1、查看数据类型: class(x);mode(x);typeof(X);
4 sapply(data,class)
5 sapply(data,mode)
6 sapply(data,typeof)
7 #####精度上, typeof>class>mode
8
9 ##2、判断数据类型
10 c4 <- '123'
11 is.numeric(c4)
12 is.character(c4)
13
14 ##3、数据类型转换
15 c4 <- as.numeric(c4)
16 is.numeric(c4)
17
18

```

## 数据结构

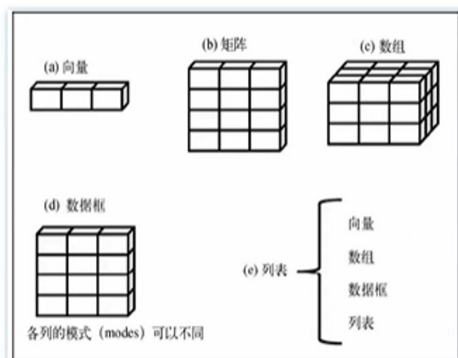
➤ 一个向量的所有元素都必须属于相同的类型。如果不是，R将强制执行类型转换。

```

• > w4 <- c(w,w1)
• > w4
• [1] "1"      "3"      "4"      "5"      "6"      "7"      "张三" "李四" "王五"
• > mode(w4)
• [1] "character"
• > w5 <- c(w1,w2)
• > w5
• [1] "张三" "李四" "王五" "TRUE" "FALSE" "TRUE"
• > mode(w5)
• [1] "character"

```

- R拥有许多用于存储数据的对象类型，包括向量、矩阵、数组、数据框和列表。



- 数据框(data frame)是R中用于存储数据的一种结构：列表示变量，行表示观测。在同一个数据框中可以存储不同类型(如数值型、字符型)的变量。数据框将是你用来存储数据集的主要数据结构。

## 向量创建

- 向量是以一维数组的方法管理数据的一种对象类型。可以说向量是R语言中最基本的数据类型，很多算法函数都是以向量的形式输入的。
- 向量可以是数值型、字符型、逻辑值型(T、F) 和复数型。
- 实例：创建不同类型的向量
 

```
> x1 <- c(1, 2, 3, 4) # 创建数值型向量,可写成x1=c(1:4)
> x2 <- c("a", "b", "c", "d") # 创建字符型变量
> x3 <- c(TRUE, FALSE, FALSE, TRUE) # 创建逻辑型变量
```
- 一个向量的所有元素都必须属于相同的类型。如果不是，R将强制执行类型转换。

## 向量

- 向量是以一维数组的方法管理数据的一种对象类型。可以说向量是R语言中最基本的数据类型，很多算法函数都是以向量的形式输入的；
- 字符型、逻辑值型(T、F)、数值型和复数型；
- Length函数测长度；
- c函数创建：

• > w <- c(1,3,4,5,6,7)	• > w2 <- c(T,F,T)	• > w1 <- c('张三','李四','王五')
• > length(w)	• > length(w2)	• > length(w1)
• [1] 6	• [1] 3	• [1] 3
• > mode(w)	• > mode(w2)	• > mode(w1)
• [1] "numeric"	• [1] "logical"	• [1] "character"