R语言常用包

1、聚类

* 常用的包： fpc，cluster，pvclust，mclust
* 基于划分的方法: kmeans, pam, pamk, clara
* 基于层次的方法: hclust, pvclust, agnes, diana
* 基于模型的方法: mclust
* 基于密度的方法: dbscan
* 基于画图的方法: plotcluster, plot.hclust
* 基于验证的方法: cluster.stats

2、分类

* 常用的包：

rpart，party，randomForest，rpartOrdinal，tree，marginTree，

maptree，survival

* 决策树: rpart, ctree
* 随机森林: cforest, randomForest
* 回归, Logistic回归, Poisson回归: glm, predict, residuals
* 生存分析: survfit, survdiff, coxph

3、关联规则与频繁项集

* 常用的包：

arules：支持挖掘频繁项集，最大频繁项集，频繁闭项目集和关联规则

DRM：回归和分类数据的重复关联模型

* APRIORI算法，广度RST算法：apriori, drm
* ECLAT算法： 采用等价类，RST深度搜索和集合的交集： eclat

4、序列模式

* 常用的包： arulesSequences
* SPADE算法： cSPADE

5、时间序列

* 常用的包： timsac
* 时间序列构建函数： ts
* 成分分解: decomp, decompose, stl, tsr

6、统计

* 常用的包： Base R, nlme
* 方差分析: aov, anova
* 密度分析: density
* 假设检验: t.test, prop.test, anova, aov
* 线性混合模型：lme
* 主成分分析和因子分析：princomp
* 方差分析对应的是Kruskal-Wallis秩和检验（R: kruskal.test)  
  T检验对应的是Wilcoxon符号秩和检验（R: wilcox.test)

7、图表

* 条形图: barplot
* 饼图: pie
* 散点图: dotchart
* 直方图: hist
* 密度图: densityplot
* 箱形图 boxplot
* QQ (quantile-quantile) 图: qqnorm, qqplot, qqline
* Bi-variate plot: coplot
* 树: rpart
* Parallel coordinates: parallel, paracoor, parcoord
* 热图, contour: contour, filled.contour
* 其他图: stripplot, sunflowerplot, interaction.plot, matplot, fourfoldplot,  
  assocplot, mosaicplot
* 保存的图表格式: pdf, postscript, win.metafile, jpeg, bmp, png

8、数据操作

* 缺失值：na.omit
* 变量标准化：scale
* 变量转置：t
* 抽样：sample
* 堆栈：stack, unstack
* 其他：aggregate, merge, reshape

9、与数据挖掘软件Weka做接口

* RWeka: 通过这个接口，可以在R中使用Weka的所有算法。

10.人工神经网络： nnet

11.支持向量机SVM：e1071

12：核函数：kernlab

13.随机森林randomforest：randomForest

14.制作分位箱图：Hmisc

15.绘制图形：qplot，ggplot2，