

正文

<< [上一篇](#) [下一篇](#) >>

学者尤著宏 发表于2010-7-16 15:19:25

[查看评论](#) : 2 | [浏览](#) : 397 [打印](#)  [推荐给朋友](#)

机器学习相关

[转载]R语言中的机器学习(转)

机器学习是计算机科学和统计学的边缘交叉领域，R关于机器学习的包主要包括以下几个方面：

1) 神经网络 (Neural Networks) :

nnet包执行单隐层前馈神经网络，nnet是VR包的一部分 (<http://cran.r-project.org/web/packages/VR/index.html>) 。

2) 递归拆分 (Recursive Partitioning) :

递归拆分利用树形结构模型，来做回归、分类和生存分析，主要在**rpart**包 (<http://cran.r-project.org/web/packages/rpart/index.html>) 和**tree**包 (<http://cran.r-project.org/web/packages/tree/index.html>) 里执行，尤其推荐rpart包。Weka里也有这样的递归拆分法，如：J4.8, C4.5, M5，包**Rweka**提供了R与Weka的函数的接口 (<http://cran.r-project.org/web/packages/RWeka/index.html>) 。

party包提供两类递归拆分算法，能做到无偏的变量选择和停止标准：函数**ctree()**用非参条件推断法检测自变量和因变量的关系；而函数**mob()**能用来建立参数模型 (<http://cran.r-project.org/web/packages/party>)

/index.html)。另外，**party**包里也提供二分支树和节点分布的可视化展示。

mvpart包是**rpart**的改进包，处理多元因变量的问题 (<http://cran.r-project.org/web/packages/mvpart/index.html>)。

rpart.permutation包用置换法 (permutation) 评估树的有效性 (<http://cran.r-project.org/web/packages/rpart.permutation/index.html>)。**knnTree**包建立一个分类树，每个叶子节点是一个knn分类器 (<http://cran.r-project.org/web/packages/knnTree/index.html>)。**LogicReg**包做逻辑回归分析，针对大多数自变量是二元变量的情况 (<http://cran.r-project.org/web/packages/LogicReg/index.html>)。**maptree**包 (<http://cran.r-project.org/web/packages/maptree/index.html>) 和**pinktoe**包 (<http://cran.r-project.org/web/packages/pinktoe/index.html>) 提供树结构的可视化函数。

3) 随机森林 (Random Forests) :

randomForest包提供了用随机森林做回归和分类的函数 (<http://cran.r-project.org/web/packages/randomForest/index.html>)。**ipred**包用bagging的思想做回归，分类和生存分析，组合多个模型 (<http://cran.r-project.org/web/packages/ipred/index.html>)。**party**包也提供了基于条件推断树的随机森林法 (<http://cran.r-project.org/web/packages/party/index.html>)。**varSelRF**包用随机森林法做变量选择 (<http://cran.r-project.org/web/packages/varSelRF/index.html>)。

4) Regularized and Shrinkage Methods :

lasso2包 (<http://cran.r-project.org/web/packages/lasso2/index.html>) 和**lars**包 (<http://cran.r-project.org/web/packages/lars/index.html>) 可以执行参数受到某些限制的回归模型。**elasticnet**包可计算所有的收缩参数 (<http://cran.r-project.org/web/packages/elasticnet/index.html>)。**glmPath**包可以得到广义线性模型和COX模型的L1 regularization path (<http://cran.r-project.org/web/packages/glmPath/index.html>)。**penalized**包执行lasso (L1) 和ridge (L2)惩罚回归模型 (penalized regression models) (<http://cran.r-project.org/web/packages/penalized/index.html>)。**pamr**包执行缩小重心分类法(shrunk centroids classifier) (<http://cran.r-project.org/web/packages/pamr/index.html>)。**earth**包可做多元自适应样条回归 (multivariate adaptive regression splines) (<http://cran.r-project.org/web/packages/earth/index.html>)。

5) Boosting :

gbm包 (<http://cran.r-project.org/web/packages/gbm/index.html>) 和**boost**包 (<http://cran.r-project.org/web/packages/boost/index.html>) 执行多种多样的梯度boosting算法，**gbm**包做基于树的梯度下降boosting，**boost**包包括LogitBoost和L2Boost。**GAMMoost**包提供基于boosting的广义相加模型(generalized additive models)的程序 (<http://cran.r-project.org/web/packages/GAMMoost/index.html>)。**mboost**包做基于模型的boosting (<http://cran.r-project.org/web/packages/mboost/index.html>)。

6) 支持向量机 (Support Vector Machines) :

e1071包的svm()函数提供R和LIBSVM的接口 (<http://cran.r-project.org/web/packages/e1071/index.html>) 。
kernlab包为基于核函数的学习方法提供了一个灵活的框架, 包括SVM、RVM.....(<http://cran.r-project.org/web/packages/kernlab/index.html>) 。**klaR** 包提供了R和SVMlight的接口 (<http://cran.r-project.org/web/packages/klaR/index.html>) 。

7) 贝叶斯方法 (Bayesian Methods) :

BayesTree包执行Bayesian Additive Regression Trees (BART)算法 (<http://cran.r-project.org/web/packages/BayesTree/index.html>, http://www-stat.wharton.upenn.edu/~edgeorge/Research_papers/BART%206--06.pdf) 。**tgp**包做Bayesian半参数非线性回归 (Bayesian nonstationary, semiparametric nonlinear regression) (<http://cran.r-project.org/web/packages/tgp/index.html>) 。

8) 基于遗传算法的最优化 (Optimization using Genetic Algorithms) :

gafit包 (<http://cran.r-project.org/web/packages/gafit/index.html>) 和**rgenoud**包 (<http://cran.r-project.org/web/packages/rgenoud/index.html>) 提供基于遗传算法的最优化程序。

9) 关联规则 (Association Rules) :

arules包提供了有效处理稀疏二元数据的数据结构, 而且提供函数执行Apriori和Eclat算法挖掘频繁项集、最大频繁项集、闭频繁项集和关联规则 (<http://cran.r-project.org/web/packages/arules/index.html>) 。

10) 模型选择和确认 (Model selection and validation) :

e1071包的tune()函数在指定的范围内选取合适的参数 (<http://cran.r-project.org/web/packages/e1071/index.html>) 。**ipred**包的errorest()函数用重抽样的方法 (交叉验证, bootstrap) 估计分类错误率 (<http://cran.r-project.org/web/packages/ipred/index.html>) 。**svmpath**包里的函数可用来选取支持向量机的cost参数C (<http://cran.r-project.org/web/packages/svmpath/index.html>) 。**ROCR**包提供了可视化分类器执行效果的函数, 如画ROC曲线 (<http://cran.r-project.org/web/packages/ROCR/index.html>) 。**caret**包供了各种建立预测模型的函数, 包括参数选择和重要性量度 (<http://cran.r-project.org/web/packages/caret/index.html>) 。**caretLSF**包 (<http://cran.r-project.org/web/packages/caretLSF/index.html>) 和**caretNWS** (<http://cran.r-project.org/web/packages/caretNWS/index.html>) 包提供了与caret包类似的功能。

11) 统计学习基础 (Elements of Statistical Learning) :

书《The Elements of Statistical Learning: Data Mining, Inference, and Prediction》 (<http://www-stat.stanford.edu/~tibs/ElemStatLearn/>) 里的数据集、函数、例子都被打包放在**ElemStatLearn**包里 (<http://cran.r-project.org/web/packages/ElemStatLearn/index.html>) 。