云南大学数学与统计学院
《数据挖掘与决策支持实验》上机实践报告

课程名称： 运筹学数据挖掘与决策支持实验	**年级：** 2015 级	**上机实践成绩：**
指导教师： 彭程	**姓名：** 刘鹏	**专业：** 信息与计算科学
上机实践名称： 对乳腺癌数据进行决策树分析	**学号：** 20151910042	**上机实践日期：** 2018-07-04
上机实践编号： 05	**组号：**	

一、实验目的

学习使用 R 语言进行变量选择。

二、实验内容

三、实验平台

Windows 10 Pro 1803；

Microsoft© Visual Studio 2017 Enterprise。

Version 1.1.442 – © 2009-2018 RStudio, Inc.

四、算法设计

数据预先分析：

近年来，疾病早期诊断越来越受到医学专家的重视，乳腺癌是其中较为常见的一种。

该病的特点：乳腺癌最早的表现是患乳出现单发的、无痛的并呈进行性生长的小肿块。肿块位于外上象限最多见，其次是乳头、乳晕区和内上象限。因多无自觉症状，肿块常是病人在无意中（如洗澡、更衣）发现的。少数病人可有不同程度的触痛或刺痛和乳头溢液。肿块的生长速度较快，侵及周围组织可引起乳房外形的改变，出现一系列体征。如：肿瘤表面皮肤凹陷；邻近乳头的癌肿可将乳头牵向癌肿方向；乳头内陷等。癌肿较大者，可使整个乳房组织收缩，肿块明显凸出。癌肿继续增大，形成所谓“桔皮样”。这些都是乳腺癌的重要体征。

乳癌发展到晚期，表面皮肤受侵犯，可出现皮肤硬结，甚至皮肤破溃形成溃烂。癌肿向深层侵犯，可侵入胸筋膜、胸肌，致使肿块固定于胸壁而不易推动。

乳癌的淋巴转移多表现为同侧窝淋巴结肿大，初为散在、无痛、质硬，数目较少，可被推动；以后肿大的淋巴结数目增多，互相粘连成团，与皮肤或腋窝深部组织粘连而固定。少数病人可出现对侧腋窝淋巴结转移。

乳腺癌的远处转移，至肺时，可出现胸痛、气促、胸水等；椎骨转移时，出现患处剧痛甚至截肢；肝转移时，可出现黄疸，肝肿大等。

医学上常用乳癌诊断 9 个医学指标分别为

clump thickness	丛厚度
cell size	均匀细胞大小
cell shape	细胞的均匀形状
marginal adhesion	细胞边缘粘附程度
SECell Size	胆上皮细胞大小
Bare Nuclei	裸核大小
Bland Chromatin BLAND	染色体
Normal Nucleoli	正常核仁大小
Mitoses	有丝分裂程度

决策树算法：

五、程序代码

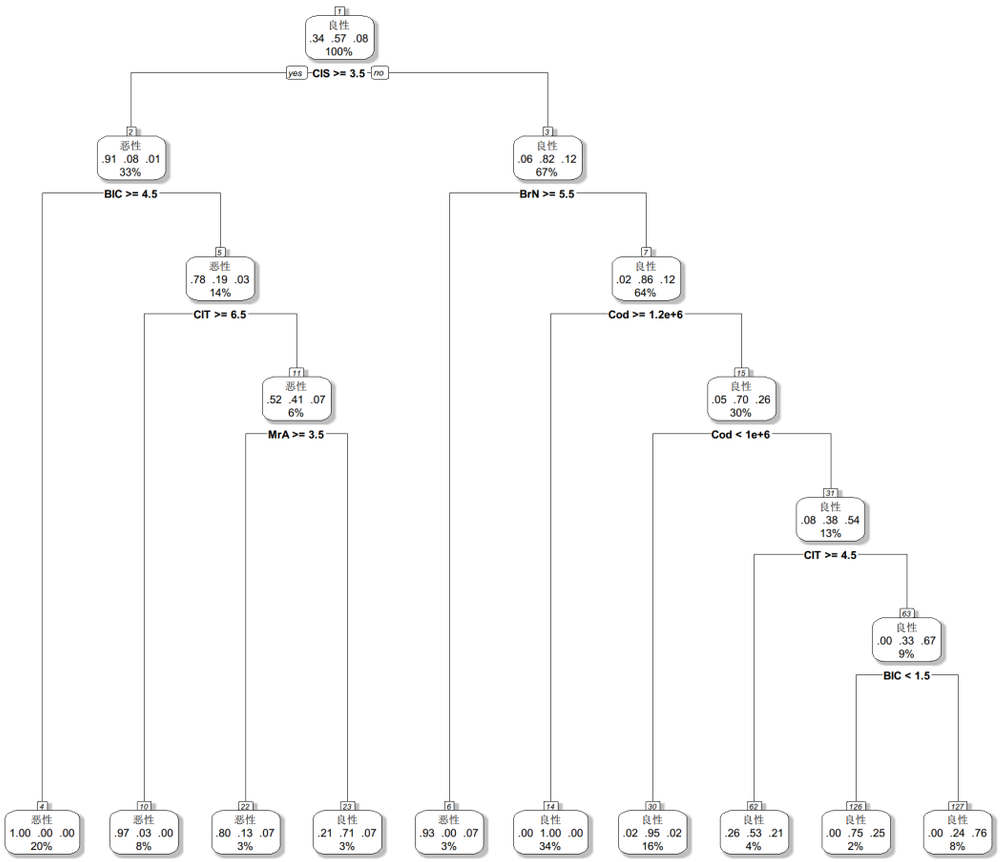
5.1 程序描述

```
library(rpart)
library(rpart.plot)
library(caret)

mydata <- read.csv("Data_Breast_Cancer.CSV")
head(mydata)

set.seed(1000)
train.idx <- createDataPartition(mydata$Class,p=0.7,list=FALSE)

mod <- rpart(Class~.,data=mydata[train.idx,], method="class", control=rpart.control(minsplit=20,cp=0.01))
mod

prp(mod, type=2, extra=104, nn=TRUE, fallen.leaves=TRUE, faclen=4, varlen=3, shadow.col="gray")
mod$cptable

mod.pruned=prune(mod, mod$captable[5, "CP"])

```

```
pred.pruned <- predict (mod, mydata[-train, idx, ], type = "class")
table(mydata[-train.idx,]$Class, pred.pruned, dnn=c("Actualn", "Predicted"))
```

5.2 程序代码

程序代码 1



七、实验体会

八、参考文献

http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.names
https://blog.csdn.net/ruoyunliufeng/article/details/79369142
https://zhuanlan.zhihu.com/p/33984536