Chapter 12 –
Discriminant Analysis 判别分析

Instructor: Zach Zhizhong ZHOU,
Shanghai Jiao Tong University
主讲教师：周志中，上海交通大学

Data Mining for Business Intelligence

Shmueli, Patel & Bruce

RidingMowers - 1

```
library(MASS) ## needed for Discriminant Analysis
library(caret)
setwd("C:/BA/DA")
RidingMowers <- read.csv("RidingMowers.csv",header =
TRUE)
NumData = nrow(RidingMowers)
set.seed(100)
SampleIdx <- sample(1:NumData,floor(0.7*NumData))
TrainSet <- RidingMowers[SampleIdx,]
TestSet <- RidingMowers[-SampleIdx,]

RM_LDA <- lda(Ownership~.,data=TrainSet)
table(TestSet$Ownership,predict(RM_LDA,TestSet)$class)
```

RidingMowers - 2

```
Predictors <- RidingMowers[,c("Income","Lot_Size")]
myControl=trainControl(method='repeatedcv', number=4,
repeats=20)

model <- train( Predictors, RidingMowers[,"Ownership"],
                method='lda',  metric='Accuracy',
trControl=myControl)
model
confusionMatrix(model)
#train函数可以处理的方法在:
http://topepo.github.io/caret/bytag.html
#判别分析所用的方法在:
http://topepo.github.io/caret/Discriminant_Analysis.html
```

RidingMowers - 3

```
model <- train( Predictors, RidingMowers[,"Ownership"],
                method='qda',  metric='Accuracy',
trControl=myControl)
model
#使用Quandratic Discriminant Analysis
confusionMatrix(model)
model <- train( Predictors, RidingMowers[,"Ownership"],
                method='pda',  metric='Accuracy',
trControl=myControl)
model
#使用Penalized Discriminant Analysis
confusionMatrix(model)
PredictTest <-
predict(model,RidingMowers[,c("Income","Lot_Size")])
table(TestSet[,"Ownership"],PredictTest)
```