SDSC3006 Lab
7 Tree-based Methods

Langming LIU langmiliu2-c@my.cityu.edu.hk

School of Data Science
City University of Hong Kong

Contents

- Classification Tree

- Regression Tree

Classification Tree

Introduction

- Datasets: Carseats data set in the ISLR2 library
- Target: predict Sales based on all predictors
- Key: create a binary response variable using ifelse() function. Specifically, it takes on Yes if Sales>8, and No otherwise. This is just for demonstrating classification tree!

Initializing

```
install.packages("tree")
library (ISLR2)
library (tree)
attach(Carseats)
##First create a binary response to do classification
High=factor(ifelse(Sales <=8,"No","Yes"))
##Add this column to the table
Carseats=data.frame(Carseats,High)
##head(Carseats)
```

Creating Tree

```
##Use the tree() function to build tree
tree.carseats=tree(High~.-Sales,Carseats)
summary(tree.carseats)
##deviance is related to the entropy

##plot the tree
plot(tree.carseats) #display tree structure
text(tree.carseats,pretty=0) #display node labels
```

Pruning

```
##use cv.tree() to perform cross validation for tree pruning
set.seed(3)
cv.carseats=cv.tree(tree.carseats,FUN=prune.misclass)
##FUN=prune.misclass indicates that misclassification error
##rate is used to guide cross validation
names(cv.carseats)
cv.carseats
##dev corresponds to the number of cross-validation errors
##visualize results
plot(cv.carseats$size,cv.carseats$dev,type="b")
```

Pruning

```
##prune.misclass() based on cv results
prune.carseats=prune.misclass(tree.carseats,best=12)
plot(prune.carseats)
text(prune.carseats,pretty=0)

##test the pruned tree
tree.pred=predict(prune.carseats,Carseats,type="class")
table(tree.pred,High)
```

Regression Tree

Creating Tree

```
set.seed (5)
attach(Boston)
##use one-third of the data to be training set
train = sample (1: nrow (Boston), nrow (Boston) / 3)
tree.boston = tree (medv ~ . , Boston , subset = train)
summary (tree.boston)

plot (tree.boston)
text (tree.boston , pretty = 0)
```

Pruning Tree

```
##use cv.tree() to perform cross validation
cv.boston = cv.tree (tree.boston)
plot (cv.boston$size , cv.boston$dev, type = "b")
cv.boston

prune.boston = prune.tree (tree.boston , best = 7)
plot (prune.boston)
text (prune.boston , pretty = 0)
```