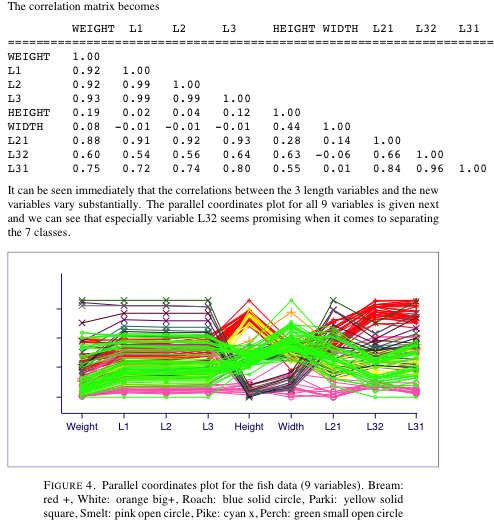
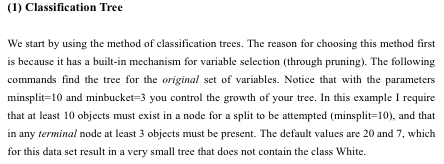




4





> fish<-read.table(“fish.dat”,h=T)

> library(rpart)

> fish.control<-rpart.control(minisplit=10,minbucket=3,xval=0)

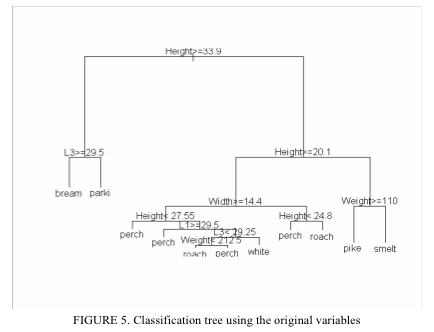
> fish.treeorig<-rpart(Species~

Weight+L1+L2+L3+Height+Width,data=fish,method=”class”,control=fish.control)

Let’s now plot the tree:

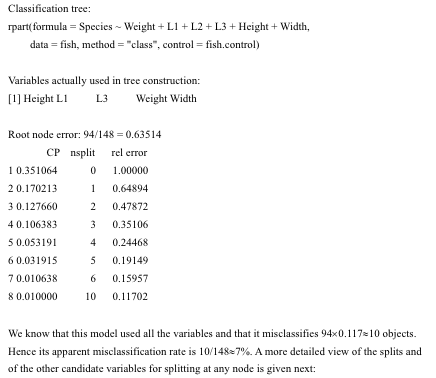
> plot(fish.treeorig)

> text(fish.treeorig)

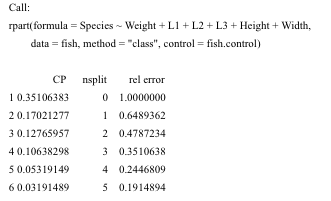


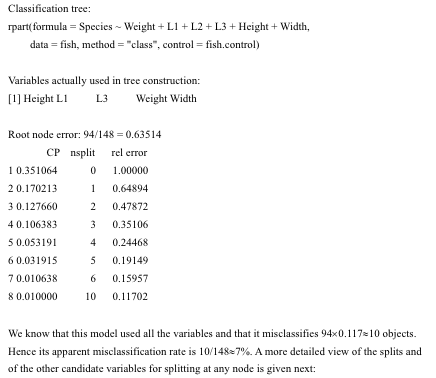
Also check out the complexity parameter (CP):

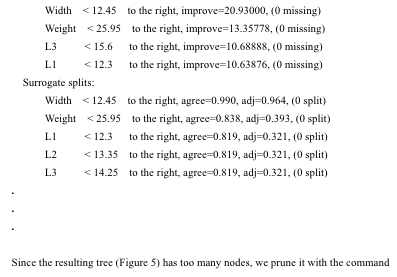
> printcp(fish.treeorig)



> summary(fish.treeorig)



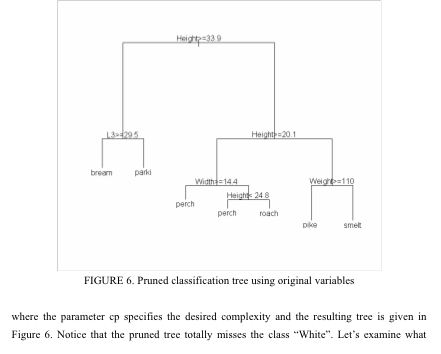




> fish.prunetree<-prune.rpart(fish.treeorig,cp=0.02)

> plot(fish.prunetree)

> text(fish.prunetree)





> L21<-fish$L2-fish$L1

> L32<-fish$L3-fish$L2

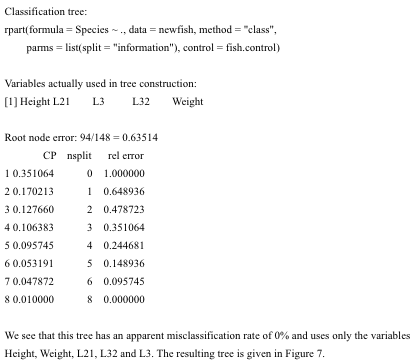
> L31<-fish$L3-fish$L1

> newfish<-cbind(fish,L21,L32,L31)

> newfish.treenew<-rpart(Species~., data=newfish,method=”class”,

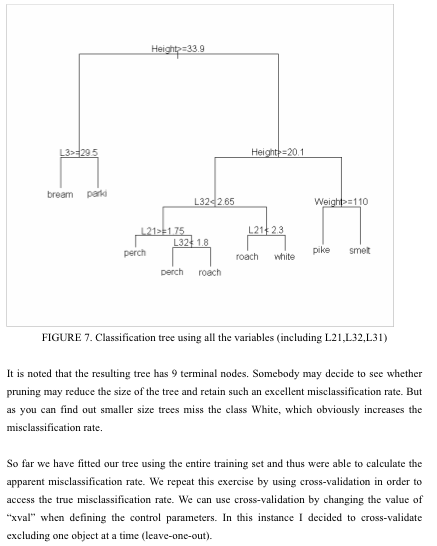
parms=list(split=’information’),control=fish.control)

> printcp(newfish.treenew)



> plot(newfish.treenew)

> text(newfish.treenew)



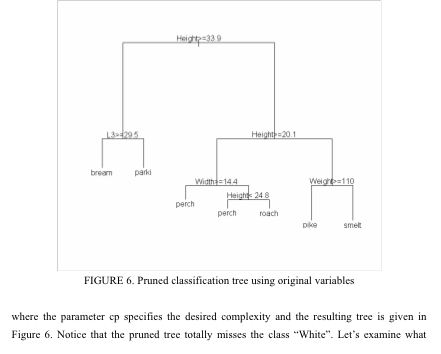
> fish.control<-rpart.control(minbucket=3,minsplit=10,xval=148)

> newfish.treenewcv<- rpart(Species~., data=newfish,method=”class”,

parms=list(split=’information’),control=fish.control)

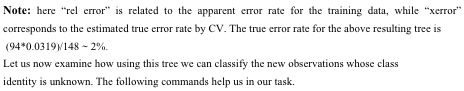
> printcp(newfish.treenewcv)





**Reference:**

Therneau, T.M., Atkinson, E.J., Foundation, M. (1997). An Introduction to Recursive Partitioning Using the RPART Routines.



> newfish.test<-read.table(“fish\_test.dat”,h=T)

> L31<-newfish.test$L3- newfish.test$L1

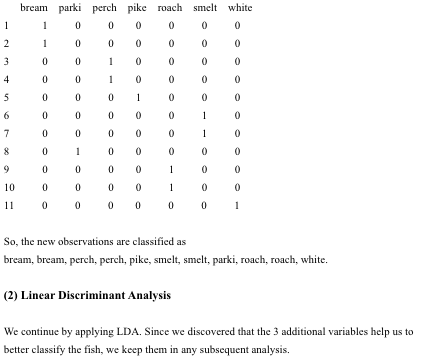
> L32<-newfish.test$L3- newfish.test$L2

> L21<-newfish.test$L2- newfish.test$L1

> newfish.test<-cbind(newfish.test,L21,L32,L31)

> newfish.tpred<-predict(newfish.treenewcv,newfish.test)

> newfish.tpred



> library(MASS)

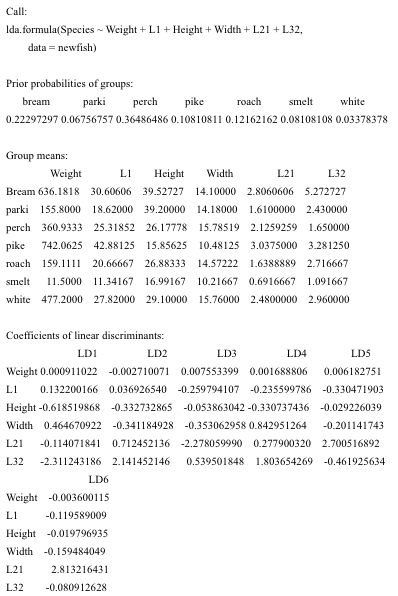
> newfish.lda<-lda(Species~.,data=newfish)

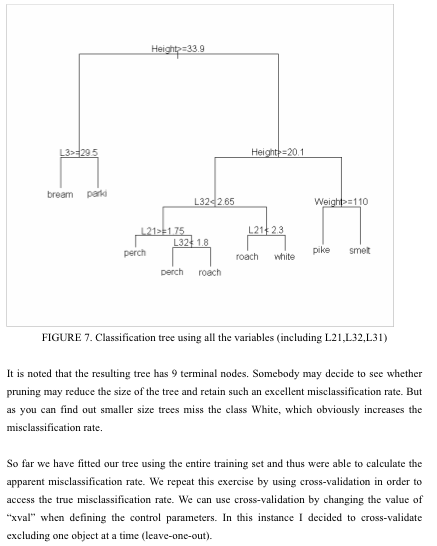




> newfish.lda<-lda(Species~Weight+L1+Height+Width+L21+L32,data=newfish)

> newfish.lda





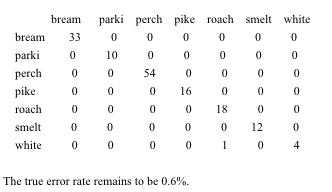
> newfish.ldapred<-predict(newfish.lda,newfish[,-1])

> table(newfish$Species,newfish.ldapred$class)



> newfish.ldacv<-lda(Species~Weight+L1+Height+Width+L21+L32,data=newfish,CV=T)

> table(newfish$Species,newfish.ldacv$class)



A plot of the various groups on the first 2 LDs can be obtained by:

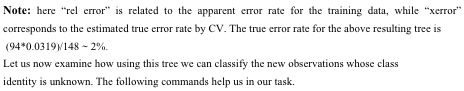
> eqscplot(newfish.ldapred$x,type=”n”,xlab=”1st LD”,ylab=”2nd LD”)

> fish.species<-c(rep(“B”,33),rep(“W”,5),rep(“R”,18),rep(“Pa”,10),rep(“S”,12),rep(“Pi”,16),

rep(“Pe”,54))

> fish.colors<-c(rep(1,33),rep(2,5),rep(3,18),rep(4,10),rep(5,12),rep(6,16),rep(7,54))

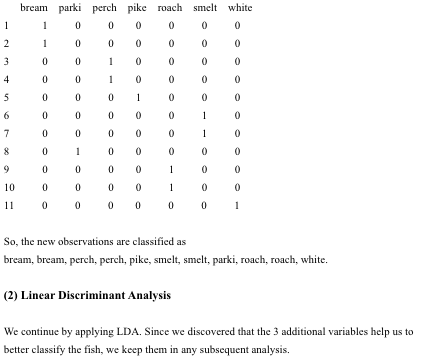
> text(newfish.ldapred$x[,1:2],fish.species,col=fish.colors)



To predict the class identities of the new data points we use:

> newfish.ldatest<-predict(newfish.lda,newfish.test)

> newfish.ldatest$class

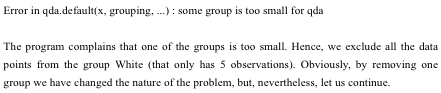


We see that the results agree with those obtained from the classification tree.

**(3) Quadratic Discriminant Analysis**

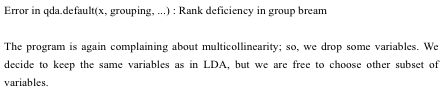
Let us examine how to apply QDA to this dataset.

> newfish.qda<-qda(Species~.,data=newfish)



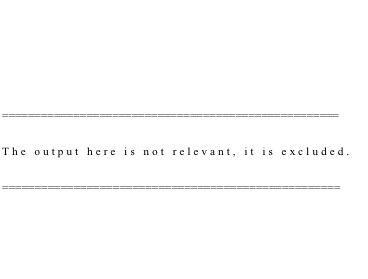
> newfish.q<-read.table(“newfish.qdat”,h=T)

> newfish.qda<-qda(Species~.,data=newfish.q)

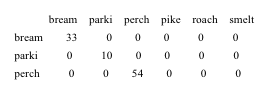


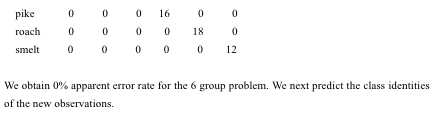
> newfish.qda<-qda(Species~Weight+L1+Height+Width+L21+L32,data=newfish.q)

> newfish.qdapred<-predict(newfish.qda,newfish.q)

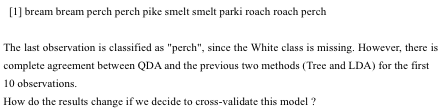


> table(newfish.q$Species,newfish.qdapred$class)



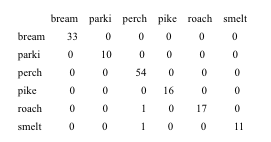


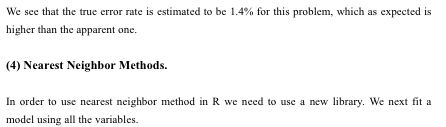
> predict(newfish.qda,newfish.test)$class



> newfish.qda<-qda(Species~Weight+L1+Height+Width+L21+L32,data=newfish.q,CV=T)

> table(newfish.q$Species,newfish.qda$class)

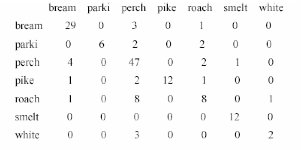




> library(class)

> newfish.knn<-knn(newfish[,2:10],newfish[,2:10],newfish[,”Species”],k=3,prob=T)

> table(newfish$Species,newfish.knn)



We see that the apparent error rate for k = 3 is about 21%. For k = 2, we have:

> newfish.knn<-knn(newfish[,2:10],newfish[,2:10],newfish[,”Species”],k=2,prob=T)

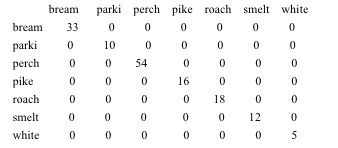
> table(newfish$Species,newfish.knn)

![](data:application/pdf;base64,)



> newfish.knn<-knn(newfish[,2:10],newfish[,2:10],newfish[,”Species”],k=1,prob=T)

> table(newfish$Species,newfish.knn)

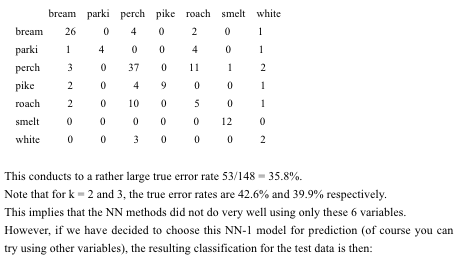




> newfish1<-newfish[,c(1,2,3,6,8,9)]

> newfish.knncv<-knn.cv(newfish1[,2:6],newfish1[,”Species”],k=1,prob=T)

> table(newfish1$Species,newfish.knncv)



> newfish1.test<-newfish.test[,c(1,2,5,7,8)]

> newfish.knntest<-knn(newfish1[,2:6],newfish1.test,newfish1[,”Species”],k=1,prob=T)

> newfish.knntest



**(5) Logistic Discrimination**

In order to use this method in R, we need another package.

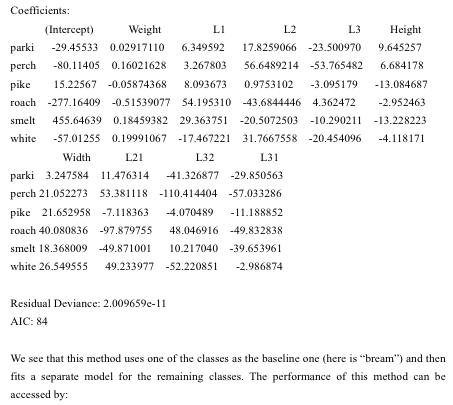
> library(nnet)

> newfish.logd<-multinom(Species~.,data=newfish,maxit=250)

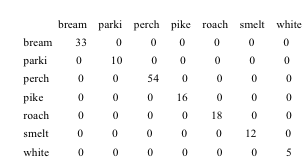
> newfish.logd

Call:

![](data:application/pdf;base64,)



> table(newfish$Species,predict(newfish.logd,newfish))







> predict(newfish.logd,newfish.test)



