

Classification and Regression Tree Example

Packages used

Data

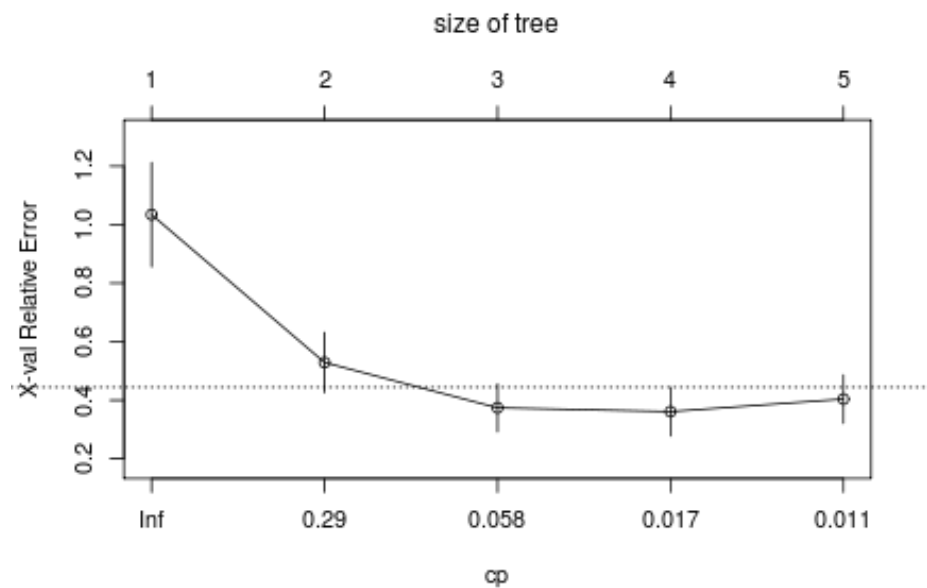
```
library(rpart)
summary(cu.summary)
```

Grow tree

```
(fit <- rpart(Mileage~Price + Country + Reliability + Type,
              method="anova", xval =100, data=cu.summary))
```

(if you know you want to use all the variables other than the response variable as input variables, `rpart(response~.,...)` will work) Note: If your data are binary and you are interested in performing a classification tree, `method = "class"`

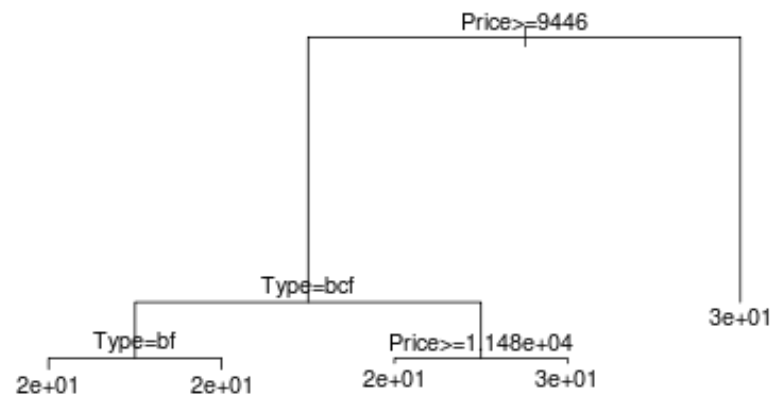
```
printcp(fit) # display the results
plotcp(fit) # visualize cross-validation results
```



```
summary(fit) # detailed summary of splits
```

plot

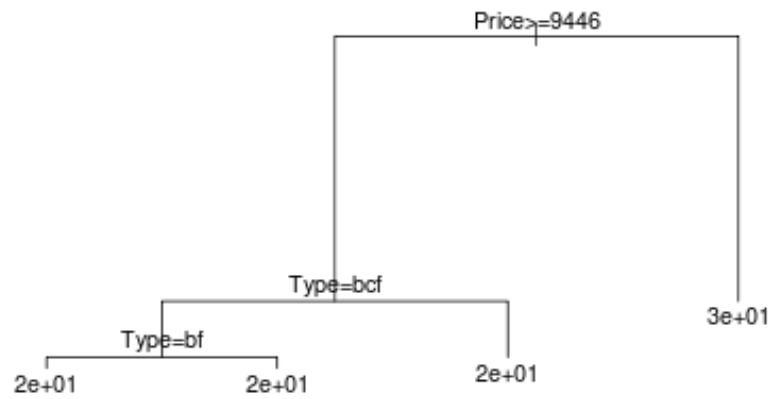
```
plot(fit)
text(fit)
```



prune

```
cpval <- fit$cptable[which.min(fit$cptable[, "xerror"]), "CP"]
pfit <- prune(fit, cp=cpval)
```

```
plot(pfit)
text(pfit)
```



```
summary(pfit)
```

Pick your own tree size

```
dfit <- rpart(Mileage~., method="anova",
              maxdepth=2, data=cu.summary)
```

```
plot(dfit)
text(dfit)
```



```
summary(dfit)
```

(Code modified from [Quick-R: Accessing the Power of R](#))

Multivariate Regression Tree Example

Data

Data Information

```
library(ade4)
data(doubs)
env=doubs$env
spe = doubs$fish
```

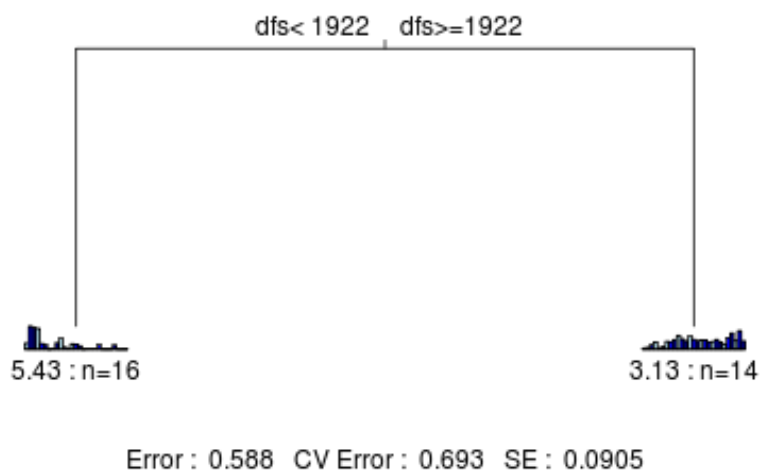
Transform response variables

The transformation consists of expressing each fish density as a proportion of the sum of all densities in the analytical unit and taking the square root of the resulting value (Legendre and Gallagher 2001). The square-root portion of the transformation decreases the importance of the most abundant species.

```
library(vegan)
spe.norm=decostand(spe, "hellinger")
```

Multivariate Regression Tree

```
library(mvpart)
spe.ch.mvpart<-mvpart(data.matrix(spe.norm) ~., env,
                      xv="1se", xval=nrow(spe),
                      xvmult=100, which=4)
```



```
summary(spe.ch.mvpart)
printcp(spe.ch.mvpart)
```

Or we can use pick

```
spe.ch.mvpart<-mvpart(data.matrix(spe.norm) ~., env,
                      xv="pick", xval=nrow(spe),
                      xvmult=100, which=4)
summary(spe.ch.mvpart)
printcp(spe.ch.mvpart)
```

- xv = Selection of tree by cross-validation:
 - "1se" - gives best tree within one SE of the overall best,
 - "min" - the best tree
 - "pick" - pick the tree size interactively,
 - "none" - no cross-validation.

- `xval` = Number of cross-validations or vector defining cross-validation groups (here we use as many rows there are in the dataset because it is a small dataset)
- `xvmult` = Number of multiple cross-validations.
- `which` = Which split labels and where to plot them, 1=centered, 2 = left, 3 = right and 4 = both.

(Modified R Code from “Numerical Ecology with R” by Borcard et al. 2012)