HW2

Kungang Zhang

February 10, 2017

# If document rendering becomes time consuming due to long computations or plots that are expensive to generate you can use knitr caching to improve performance. The documentation knitr chunk and package options describe how caching works and the cache examples provide additional details.

# If you want to enable caching globally for a document you can include a code chunk like this at the top of the document:

# ```{r setup, include=FALSE}  
# knitr::opts\_chunk$set(cache=TRUE)  
# ```  
#Set working directory and latter, I don't need to type the full directory  
#The results between chunk will remain  
#The working directory will also remain  
setwd("/Users/kungangzhang/Documents/OneDrive/Northwestern/Study/Courses/MSiA-420-0/HW2")  
rm(list = ls())  
require(gdata)

## Loading required package: gdata

## gdata: read.xls support for 'XLS' (Excel 97-2004) files ENABLED.

##

## gdata: read.xls support for 'XLSX' (Excel 2007+) files ENABLED.

##   
## Attaching package: 'gdata'

## The following object is masked from 'package:stats':  
##   
## nobs

## The following object is masked from 'package:utils':  
##   
## object.size

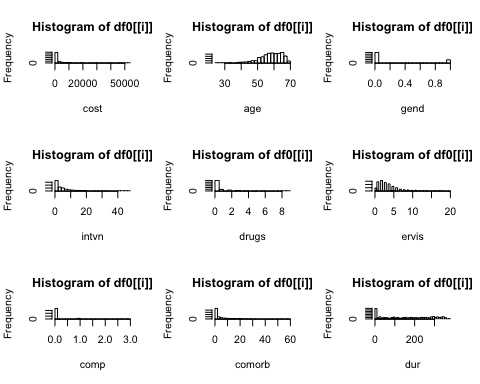
set.seed(111)

# Prob 1)

## (a)Fit a linear model and discuss the predictive power.

## Answer: First I take log transform to the cost, the response variable, and then fit the model with all predictors unchanged. The is . Then, I tried to standardize everything and the is . I saw some of predictors also have skewed distribution or long tail problem, so that I try log transform (or some special log transform depending on whether it is left-skewed and right-skewed), and the histograms look more symmetric. For the rest of predictors, I just let them be. The increases to . Generally, those predictors significant before transform are also significant afterwards.

##The histogram of each columns  
df0<-read.xls("./HW2\_data.xls",sheet=1,header=TRUE)  
par(mfrow=c(3,3))  
df<-df0  
df$gend <- as.factor(df$gend)  
par(mfrow=c(3,3))  
for (i in seq(2,10)) hist(df0[[i]],breaks=30,xlab=names(df0)[i])



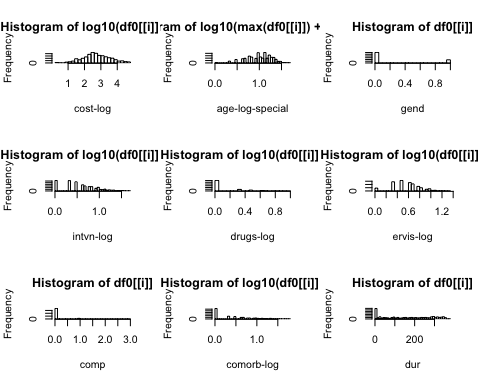
df$cost <- log10(df$cost)  
mod1<-lm(cost~.,data = df[-1])  
summary(mod1)

##   
## Call:  
## lm(formula = cost ~ ., data = df[-1])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.44852 -0.30093 0.01049 0.28276 1.72581   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.2228290 0.1698975 13.083 < 2e-16 \*\*\*  
## age -0.0044135 0.0028817 -1.532 0.1260   
## gend1 -0.0669173 0.0460024 -1.455 0.1462   
## intvn 0.0878065 0.0038090 23.053 < 2e-16 \*\*\*  
## drugs -0.0257198 0.0213709 -1.203 0.2291   
## ervis 0.0224358 0.0090588 2.477 0.0135 \*   
## comp 0.3270883 0.0794497 4.117 4.25e-05 \*\*\*  
## comorb 0.0228849 0.0037393 6.120 1.48e-09 \*\*\*  
## dur 0.0012181 0.0001874 6.501 1.43e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5373 on 779 degrees of freedom  
## Multiple R-squared: 0.5831, Adjusted R-squared: 0.5789   
## F-statistic: 136.2 on 8 and 779 DF, p-value: < 2.2e-16

##Also, I tried to standardize each variable to see effect.  
df\_std<-df  
df\_std[c(2,3,5:10)]<-sapply(df\_std[c(2,3,5:10)], function(x) (x-mean(x))/sd(x))  
mod2<-lm(cost~.,data = df\_std[-1])  
summary(mod2)

##   
## Call:  
## lm(formula = cost ~ ., data = df\_std[-1])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.95741 -0.36347 0.01268 0.34153 2.08450   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.01846 0.02637 0.700 0.4841   
## age -0.03600 0.02351 -1.532 0.1260   
## gend1 -0.08083 0.05556 -1.455 0.1462   
## intvn 0.59335 0.02574 23.053 < 2e-16 \*\*\*  
## drugs -0.03305 0.02746 -1.203 0.2291   
## ervis 0.07147 0.02886 2.477 0.0135 \*   
## comp 0.09800 0.02381 4.117 4.25e-05 \*\*\*  
## comorb 0.16449 0.02688 6.120 1.48e-09 \*\*\*  
## dur 0.17790 0.02737 6.501 1.43e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.649 on 779 degrees of freedom  
## Multiple R-squared: 0.5831, Adjusted R-squared: 0.5789   
## F-statistic: 136.2 on 8 and 779 DF, p-value: < 2.2e-16

##Except the intercept, the other aspects of the linear model would not be changed by standardization, given the full-rank design matrix.   
  
##As shown in the histogram plots the columns of cost, num of interventions, num of drugs, num of emergency, num of complication, num of other diseaes are left-skewed, while the age is right-skewed. So I need to do a log transform to those columns left-skewed and do a special log transform to those columns right-skewed.  
par(mfrow=c(3,3))  
for (i in seq(2,10)) {  
 if (i %in% c(2,5,6,7,9)){  
 hist(log10(df0[[i]]+1),breaks=30,xlab=paste(names(df0)[i],'log',sep='-'))   
 }  
 if (i==3){  
 hist(log10(max(df0[[i]])+1-df0[[i]]),breaks=30,xlab = paste(names(df0)[i],'log','special',sep='-'))  
 }  
 if (i %in% c(4,8,10)){  
 hist(df0[[i]],breaks=30,xlab=names(df0)[i])  
 }  
   
}



df\_trans\_std <- df  
df\_trans\_std$age <- log10(max(df\_trans\_std$age)+1-df\_trans\_std$age)  
df\_trans\_std$intvn <- log10(df\_trans\_std$intvn+1)  
df\_trans\_std$drugs <- log10(df\_trans\_std$drugs+1)  
df\_trans\_std$ervis <- log10(df\_trans\_std$ervis+1)  
df\_trans\_std$comorb <- log10(df\_trans\_std$comorb+1)  
df\_trans\_std[c(2,3,5:10)]<-sapply(df\_trans\_std[c(2,3,5:10)], function(x) (x-mean(x))/sd(x))  
mod3<-lm(cost~.,data = df\_trans\_std[-1])#no matter use log10(age+1) or log10(I(age+1)), the result is the same.  
summary(mod3)

##   
## Call:  
## lm(formula = cost ~ ., data = df\_trans\_std[-1])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.41656 -0.35471 0.00511 0.32302 1.90417   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.016807 0.023884 0.704 0.48184   
## age 0.036743 0.021232 1.731 0.08393 .   
## gend1 -0.073577 0.050303 -1.463 0.14396   
## intvn 0.642449 0.022936 28.010 < 2e-16 \*\*\*  
## drugs -0.008038 0.023799 -0.338 0.73563   
## ervis 0.075082 0.024177 3.105 0.00197 \*\*   
## comp 0.090160 0.021528 4.188 3.13e-05 \*\*\*  
## comorb 0.269192 0.026498 10.159 < 2e-16 \*\*\*  
## dur 0.062558 0.026960 2.320 0.02058 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5878 on 779 degrees of freedom  
## Multiple R-squared: 0.658, Adjusted R-squared: 0.6545   
## F-statistic: 187.4 on 8 and 779 DF, p-value: < 2.2e-16

##After the log transform, we see the $R^2$ increases from $0.5831$ to $0.658$ and the influential predictors don't change much.  
  
##(From website by searching 'long tail distribution log transform') You don't need to assume a lognormal distribution; there's no requirement that an independent variable in linear regression itself has a normal distribution. The hope is that, with log transformation of the independent variable, the other requirements for interpreting linear regression results will better be met, such as having normally distributed residual errors independent of fitted values. If the regression against the log-transformed independent variable meets those requirements, there are no problems with interpreting p-values, etc. Regression coefficients will now mean the change in the dependent variable per log change in the independent variable. So if you use log10, the regression coefficient will be "change per 10-fold change in GDP" for your example; for log2, "change per doubling of GDP."

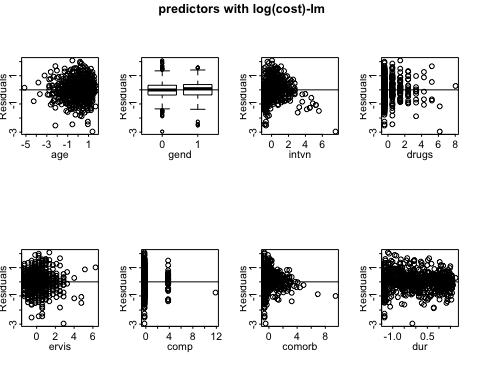
## (b)Which variables appear to have the most influence on the cost.

## Answer: From the mod2 (the standardized model without log tranforming predictors), we have the biggest coefficient of number of interventions (), so that this predictor would have the most influence on the cost. Similarily, in the mod3 (standardized model with log tranforming predictors) the number of intervention also has the biggest influence ().

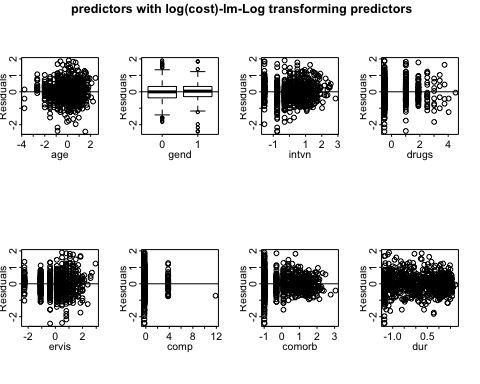
## (c)Construct appropriate diagnostics and residual plots to assess (related to nonlinearity in the relation b/w the response and the predictors.)

## Answer: From both of the plots below, we saw the residuals have little correlation with predictors, so that we don't need to change the model. If there were any nonlinear correlation, we probably need to design better predictors to capture this nonlinearity.

##For linear model without log transforming predictors  
par(mfrow=c(2,4),pin=c(0.8,0.8),tcl=-0.15,mgp=c(1,0.2,0))  
for (i in seq(3:10)) {  
 plot(df\_std[[i+2]],resid(mod2),ylab="Residuals",xlab=names(df\_std)[i+2],main="")  
 abline(0, 0)}  
title(main="Ischemic heart disease-standardized \n predictors with log(cost)-lm",outer = T)



##For linear model with log transforming skewed predictors  
par(mfrow=c(2,4),pin=c(0.8,0.8),tcl=-0.15,mgp=c(1,0.2,0))  
for (i in seq(3:10)) {  
 plot(df\_trans\_std[[i+2]],resid(mod3),ylab="Residuals",xlab=names(df\_trans\_std)[i+2],main="")  
 abline(0, 0)}  
title(main="Ischemic heart disease-standardized \n predictors with log(cost)-lm-Log transforming predictors",outer = T)



# Prob 2)Find the best neural network model for the ischemic heart disease data set, using linear output activation func, and do not rescale the response.

## (a)Use 10-fold CV to find the best combination of shrinkage param and the number of hidden nodes.

## Answer: The neural network with the smallest MSE has $= $ and number of hidden nodes , and the MSE is . It has .

##CV index random generator  
CVInd <- function(n,K) { #n is sample size; K is number of parts; returns K-length list of indices for each part  
 m<-floor(n/K) #approximate size of each part  
 r<-n-m\*K   
 I<-sample(n,n) #random reordering of the indices  
 Ind<-list() #will be list of indices for all K parts  
 length(Ind)<-K  
 for (k in 1:K) {  
 if (k <= r) kpart <- ((m+1)\*(k-1)+1):((m+1)\*k)   
 else kpart<-((m+1)\*r+m\*(k-r-1)+1):((m+1)\*r+m\*(k-r))  
 Ind[[k]] <- I[kpart] #indices for kth part of data  
 }  
 Ind  
}

## Now use multiple reps of CV to compare Neural Nets and linear reg models

library(nnet)  
CVfunc\_nnet <- function(data, lam\_seq, num\_hidnode\_seq,Nrep,K,y) {  
 n=nrow(data)  
 n.models = n.lam\*n.num\_hidnode #number of different models to fit  
 yhat=matrix(0,n,n.models)  
   
 ##Each column of mod\_par corresponds to a set of lambda and number of hidden nodes of a trail model  
 mod\_par=matrix(c(rep(lam\_seq,times=1,each=n.num\_hidnode),rep(num\_hidnode\_seq,times=n.lam,each=1)),2,n.models,byrow = T)#Store the model parameters: lambda and the number of nodes in hidden layer  
 MSE<-matrix(0,Nrep,n.models)  
 for (j in 1:Nrep) {  
 print(c(0,0,0,j))#Print out the index of replicates of CV  
 Ind<-CVInd(n,K)  
 for (k in 1:K) {  
 print(k)#Print out the index of different fold of CV  
 for (m in 1:n.models) {  
 out<-nnet(cost~.,data[-Ind[[k]],],linout = T, skip=F,size=as.integer(mod\_par[2,m]),decay=mod\_par[1,m],maxit=1000,trace=F)  
 yhat[Ind[[k]],m]<-as.numeric(predict(out,data[Ind[[k]],]))  
 }  
 } #end of k loop  
 MSE[j,]=apply(yhat,2,function(x) sum((y-x)^2))/n  
 } #end of j loop  
 MSE  
 MSEAve<- apply(MSE,2,mean); MSEAve #averaged mean square CV error  
 MSEsd <- apply(MSE,2,sd); MSEsd #SD of mean square CV error  
 r2<-1-MSEAve/var(y); r2 #CV r^2  
 ##The best model in terms of the minimum MSEAve or the maximum r2.  
 min(MSEAve)  
 max(r2)  
 ##Return the index of the minimum MSEAve or the maximum r2.  
 which(MSEAve==min(MSEAve))  
 which(r2==max(r2))  
 ##The optimal lambda and number of hidden nodes  
 mod\_par[,which(MSEAve==min(MSEAve))]  
}

## Do a CV in crude interval of lambda and number of hidden nodes.

ptm <- proc.time()  
Nrep<-2 #number of replicates of CV  
K<-10 #K-fold CV on each replicate  
n.lam = 4 #number of lambda  
n.num\_hidnode = 2 #number of different numbers of hidden nodes  
y<-df\_std$cost #observed responses  
lam\_seq = 10^seq(-as.integer(n.lam/2),as.integer(n.lam/2)-1) #seq of penalty parameters  
num\_hidnode\_seq = 5\*seq(1,n.num\_hidnode) #seq of number of hidden nodes  
  
par\_best\_crude <- CVfunc\_nnet(df\_std, lam\_seq, num\_hidnode\_seq,Nrep,K,y)

## [1] 0 0 0 1  
## [1] 1  
## [1] 2  
## [1] 3  
## [1] 4  
## [1] 5  
## [1] 6  
## [1] 7  
## [1] 8  
## [1] 9  
## [1] 10  
## [1] 0 0 0 2  
## [1] 1  
## [1] 2  
## [1] 3  
## [1] 4  
## [1] 5  
## [1] 6  
## [1] 7  
## [1] 8  
## [1] 9  
## [1] 10

proc.time() - ptm

## user system elapsed   
## 25.085 0.108 25.325

## Do a CV in smaller interval of lambda and number of hidden nodes again.

ptm <- proc.time()  
Nrep <- 2 #number of replicates of CV  
K<-10 #K-fold CV on each replicate  
n.lam = 2 #number of lambda  
n.num\_hidnode = 2 #number of different numbers of hidden nodes  
y<-df\_std$cost #observed responses  
lam\_seq = c(seq(4,4),seq(10,10,10))  
num\_hidnode\_seq = seq(15,17,2)  
  
par\_best <- CVfunc\_nnet(df\_std, lam\_seq, num\_hidnode\_seq,Nrep,K,y) #Best parameter

## [1] 0 0 0 1  
## [1] 1  
## [1] 2  
## [1] 3  
## [1] 4  
## [1] 5  
## [1] 6  
## [1] 7  
## [1] 8  
## [1] 9  
## [1] 10  
## [1] 0 0 0 2  
## [1] 1  
## [1] 2  
## [1] 3  
## [1] 4  
## [1] 5  
## [1] 6  
## [1] 7  
## [1] 8  
## [1] 9  
## [1] 10

proc.time() - ptm

## user system elapsed   
## 31.810 0.109 32.018

## (b)Fit the best model and discuss how good the predictive power is.

## Answer: The cross-validation of the best model is , with the penalization and number of hidden nodes as .

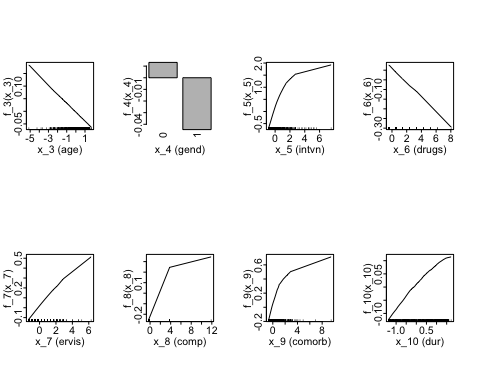
nnet\_mod<-nnet(cost~.,df\_std,linout = T, skip=F,size=as.integer(par\_best[2]),decay=par\_best[1],maxit=1000,trace=F)  
summary(nnet\_mod)

## a 9-15-1 network with 166 weights  
## options were - linear output units decay=4  
## b->h1 i1->h1 i2->h1 i3->h1 i4->h1 i5->h1 i6->h1 i7->h1 i8->h1 i9->h1   
## -0.04 0.00 0.04 0.13 -0.56 -0.01 -0.13 0.08 0.00 0.06   
## b->h2 i1->h2 i2->h2 i3->h2 i4->h2 i5->h2 i6->h2 i7->h2 i8->h2 i9->h2   
## 0.00 -0.04 -0.03 0.00 -0.02 0.00 0.01 0.00 -0.02 -0.03   
## b->h3 i1->h3 i2->h3 i3->h3 i4->h3 i5->h3 i6->h3 i7->h3 i8->h3 i9->h3   
## 0.00 -0.04 -0.04 0.00 -0.02 0.00 0.01 0.00 -0.02 -0.03   
## b->h4 i1->h4 i2->h4 i3->h4 i4->h4 i5->h4 i6->h4 i7->h4 i8->h4 i9->h4   
## -0.08 0.00 0.01 0.12 -0.62 -0.04 -0.14 -0.21 -0.06 0.04   
## b->h5 i1->h5 i2->h5 i3->h5 i4->h5 i5->h5 i6->h5 i7->h5 i8->h5 i9->h5   
## -0.01 0.04 0.07 0.00 0.03 0.00 -0.02 0.01 0.03 0.06   
## b->h6 i1->h6 i2->h6 i3->h6 i4->h6 i5->h6 i6->h6 i7->h6 i8->h6 i9->h6   
## -0.08 0.00 0.01 0.12 -0.62 -0.04 -0.14 -0.21 -0.06 0.04   
## b->h7 i1->h7 i2->h7 i3->h7 i4->h7 i5->h7 i6->h7 i7->h7 i8->h7 i9->h7   
## 0.00 -0.04 -0.03 0.00 -0.02 0.00 0.01 0.00 -0.02 -0.03   
## b->h8 i1->h8 i2->h8 i3->h8 i4->h8 i5->h8 i6->h8 i7->h8 i8->h8 i9->h8   
## 0.00 -0.04 -0.03 0.00 -0.02 0.00 0.01 0.00 -0.02 -0.03   
## b->h9 i1->h9 i2->h9 i3->h9 i4->h9 i5->h9 i6->h9 i7->h9 i8->h9 i9->h9   
## -0.01 0.04 0.07 0.00 0.03 0.00 -0.02 0.01 0.03 0.06   
## b->h10 i1->h10 i2->h10 i3->h10 i4->h10 i5->h10 i6->h10 i7->h10 i8->h10   
## 0.02 -0.02 0.11 0.02 0.08 0.09 -0.04 0.05 0.07   
## i9->h10   
## 0.12   
## b->h11 i1->h11 i2->h11 i3->h11 i4->h11 i5->h11 i6->h11 i7->h11 i8->h11   
## 0.92 0.00 -0.06 0.04 1.51 -0.17 0.05 0.15 0.84   
## i9->h11   
## 0.31   
## b->h12 i1->h12 i2->h12 i3->h12 i4->h12 i5->h12 i6->h12 i7->h12 i8->h12   
## 0.04 0.00 -0.03 -0.09 0.47 0.04 0.13 0.11 0.05   
## i9->h12   
## 0.00   
## b->h13 i1->h13 i2->h13 i3->h13 i4->h13 i5->h13 i6->h13 i7->h13 i8->h13   
## 0.00 -0.04 -0.03 0.00 -0.02 0.00 0.01 0.00 -0.02   
## i9->h13   
## -0.03   
## b->h14 i1->h14 i2->h14 i3->h14 i4->h14 i5->h14 i6->h14 i7->h14 i8->h14   
## 0.00 -0.04 -0.03 0.00 -0.02 0.00 0.01 0.00 -0.02   
## i9->h14   
## -0.03   
## b->h15 i1->h15 i2->h15 i3->h15 i4->h15 i5->h15 i6->h15 i7->h15 i8->h15   
## 0.00 -0.04 -0.03 0.00 -0.02 0.00 0.01 0.00 -0.02   
## i9->h15   
## -0.03   
## b->o h1->o h2->o h3->o h4->o h5->o h6->o h7->o h8->o h9->o   
## -0.04 -0.66 0.08 0.08 -0.75 -0.16 -0.75 0.08 0.08 -0.16   
## h10->o h11->o h12->o h13->o h14->o h15->o   
## -0.26 1.98 0.55 0.08 0.08 0.08

##(c)The variables having the most influence on cost (Use the ALEPlot package for this).  
##Answer:   
library(ALEPlot)

## Loading required package: yaImpute

yhat <- function(X.model, newdata) as.numeric(predict(X.model, newdata))  
par(mfrow=c(2,4),pin=c(0.7,0.7),tcl=-0.2,mgp = c(1,0.15,0))  
for (j in 3:10) {ALEPlot(df\_std, nnet\_mod, pred.fun=yhat, J=j, K=50, NA.plot = TRUE)  
 rug(df\_std[,j]) } ## This creates main effect ALE plots for all 8 predictors



par(mfrow=c(1,1))  
  
par(mfrow=c(2,2),pin=c(1.3,1.3),mgp = c(1,0.15,0),tcl=-0.15)   
## This creates 2nd-order interaction ALE plots for x3, x7, x6, x8, x5, x10  
ALEPlot(df\_std, nnet\_mod, pred.fun=yhat, J=c(3,7), K=50, NA.plot = TRUE)

## $K  
## [1] 25 11  
##   
## $x.values  
## $x.values[[1]]  
## [1] -5.14031174 -2.47527110 -2.03109766 -1.73498204 -1.43886641  
## [6] -1.29080860 -1.14275079 -0.99469297 -0.84663516 -0.69857735  
## [11] -0.55051953 -0.40246172 -0.25440391 -0.10634609 0.04171172  
## [16] 0.18976953 0.33782734 0.48588516 0.63394297 0.78200078  
## [21] 0.93005860 1.07811641 1.22617422 1.37423204 1.52228985  
## [26] 1.67034766  
##   
## $x.values[[2]]  
## [1] -1.2986393 -0.9194886 -0.5403379 -0.1611872 0.2179635 0.5971142  
## [7] 0.9762649 1.3554156 1.7345663 2.1137170 2.8720185 6.2843748  
##   
##   
## $f.values  
## 1 2 3 4  
## -6.428006e-04 -3.606228e-04 -2.808563e-04 -1.959205e-04 8.293719e-04  
## 1 -2.079620e-04 -8.030519e-05 2.693103e-04 2.666655e-04 3.237156e-04  
## 2 -2.816257e-04 -6.448674e-05 1.701832e-04 2.640947e-04 4.177013e-04  
## 3 -1.831639e-04 -2.728806e-05 1.290957e-04 1.779766e-04 3.326541e-04  
## 4 -1.334922e-04 -4.449034e-05 1.005730e-04 1.925952e-04 2.815117e-04  
## 5 -1.286524e-04 -3.969405e-05 8.362988e-05 2.128789e-04 2.477966e-04  
## 6 -1.045551e-04 -1.143455e-05 8.422791e-05 1.852492e-04 2.098194e-04  
## 7 -1.171003e-04 -4.356791e-05 4.748723e-05 1.739079e-04 1.986741e-04  
## 8 -1.207789e-04 -5.610210e-05 3.958052e-05 1.449167e-04 1.733634e-04  
## 9 -8.551478e-05 -3.982606e-05 3.802173e-05 1.284807e-04 1.229496e-04  
## 10 -8.500270e-05 -4.645550e-05 2.506367e-05 1.195629e-04 8.782678e-05  
## 11 -4.428142e-05 -3.253012e-05 3.095382e-06 6.367518e-05 6.671889e-05  
## 12 -7.743107e-06 -7.091140e-06 -8.193537e-06 2.718712e-05 3.540246e-05  
## 13 -2.746901e-05 -2.142038e-05 -2.421052e-05 -2.783379e-06 1.599812e-05  
## 14 -9.730431e-06 1.271554e-05 -2.511871e-05 -1.617943e-05 -1.410498e-07  
## 15 9.149961e-06 2.106970e-05 -2.316685e-05 -2.230074e-05 -3.264521e-05  
## 16 3.251068e-05 2.205871e-05 -2.601808e-05 -4.006971e-05 -5.512289e-05  
## 17 6.816875e-05 2.648832e-05 -3.958990e-05 -6.491691e-05 -8.475713e-05  
## 18 1.212779e-04 4.021486e-05 -5.953896e-05 -1.046071e-04 -1.249307e-04  
## 19 1.450489e-04 5.166729e-05 -6.397043e-05 -1.230440e-04 -1.463457e-04  
## 20 1.757585e-04 6.040052e-05 -5.857725e-05 -1.270226e-04 -1.611756e-04  
## 21 2.386077e-04 7.737904e-05 -5.161344e-05 -1.297448e-04 -2.035358e-04  
## 22 2.577888e-04 8.231070e-05 -4.875997e-05 -1.523500e-04 -2.478824e-04  
## 23 2.744127e-04 7.695069e-05 -7.691828e-05 -1.861956e-04 -2.802545e-04  
## 24 3.078836e-04 7.836064e-05 -1.041371e-04 -2.042126e-04 -2.888288e-04  
## 25 3.545254e-04 9.294138e-05 -1.539403e-04 -2.448139e-04 -3.199873e-04  
## 5 6 7 8 9  
## 1.565893e-03 1.975827e-03 2.580409e-03 1.808616e-03 8.224781e-04  
## 1 9.199410e-05 -3.247911e-04 -5.469283e-04 -7.910327e-04 -1.249482e-03  
## 2 2.560193e-04 -9.072627e-05 -2.775029e-04 -4.862468e-04 -8.667267e-04  
## 3 2.505884e-04 -2.671985e-05 -1.781360e-04 -3.089102e-04 -6.114203e-04  
## 4 2.261409e-04 1.826991e-05 -1.464804e-04 -1.992849e-04 -4.238254e-04  
## 5 2.191206e-04 3.697184e-05 -1.411128e-04 -1.303895e-04 -2.914022e-04  
## 6 1.782470e-04 1.892703e-05 -1.495365e-04 -1.863372e-04 -3.201830e-04  
## 7 1.642053e-04 5.018128e-05 -1.086612e-04 -1.929857e-04 -2.996646e-04  
## 8 1.442822e-04 4.591053e-05 -8.846626e-05 -1.673557e-04 -2.468677e-04  
## 9 9.925599e-05 1.860550e-05 -9.130557e-05 -1.647599e-04 -2.171050e-04  
## 10 7.660010e-05 4.098294e-05 -5.393400e-05 -7.854988e-05 -1.373820e-04  
## 11 6.129291e-05 3.589354e-05 -3.649948e-05 -1.102037e-04 -1.755228e-04  
## 12 6.119694e-06 -1.573806e-06 -2.995061e-05 -1.527431e-04 -2.245493e-04  
## 13 2.004281e-05 5.216376e-05 6.780316e-05 -1.157543e-05 -5.744539e-05  
## 14 -2.651528e-06 3.733679e-05 7.103979e-05 -1.642114e-05 -3.635484e-05  
## 15 -3.198373e-05 6.910740e-06 5.867733e-05 -1.046195e-05 -4.459397e-06  
## 16 -4.377688e-05 1.783602e-05 9.992484e-05 5.880064e-05 9.213478e-05  
## 17 -5.135651e-05 3.150863e-05 1.580733e-04 1.449642e-04 2.056299e-04  
## 18 -8.432969e-05 4.676030e-05 2.178008e-04 2.327067e-04 3.001699e-04  
## 19 -1.112081e-04 -2.740847e-05 1.743502e-04 2.299552e-04 3.360252e-04  
## 20 -1.544709e-04 -5.631510e-05 1.723396e-04 2.686437e-04 4.133206e-04  
## 21 -1.828350e-04 -9.278495e-05 1.140089e-04 2.510121e-04 4.170078e-04  
## 22 -1.828922e-04 -9.899498e-05 8.593804e-05 2.728752e-04 4.858017e-04  
## 23 -1.789903e-04 -7.382244e-05 1.440275e-04 3.808986e-04 6.140711e-04  
## 24 -1.512906e-04 -7.887020e-05 1.718966e-04 4.587017e-04 7.121203e-04  
## 25 -1.461751e-04 -1.065022e-04 1.771815e-04 4.710280e-04 7.446927e-04  
## 10 11  
## -5.282915e-04 -0.0043862199  
## 1 -2.072563e-03 -0.0054028033  
## 2 -1.611105e-03 -0.0048230012  
## 3 -1.277095e-03 -0.0043706483  
## 4 -1.010797e-03 -0.0039860069  
## 5 -7.996707e-04 -0.0036565372  
## 6 -7.497482e-04 -0.0034882715  
## 7 -6.505267e-04 -0.0029219352  
## 8 -5.022854e-04 -0.0023065791  
## 9 -3.770783e-04 -0.0017142572  
## 10 -2.440444e-04 -0.0012193239  
## 11 -1.767625e-04 -0.0007901426  
## 12 -2.153926e-04 -0.0004668733  
## 13 3.892341e-05 -0.0002798484  
## 14 1.472260e-04 0.0003240055  
## 15 1.406914e-04 0.0008130221  
## 16 1.988556e-04 0.0013667375  
## 17 2.553819e-04 0.0016762002  
## 18 5.060452e-04 0.0021798000  
## 19 4.664262e-04 0.0023931175  
## 20 4.682473e-04 0.0025319886  
## 21 5.421931e-04 0.0027429845  
## 22 6.812457e-04 0.0030190871  
## 23 8.060005e-04 0.0032808921  
## 24 9.005351e-04 0.0035124767  
## 25 9.295929e-04 0.0036785846

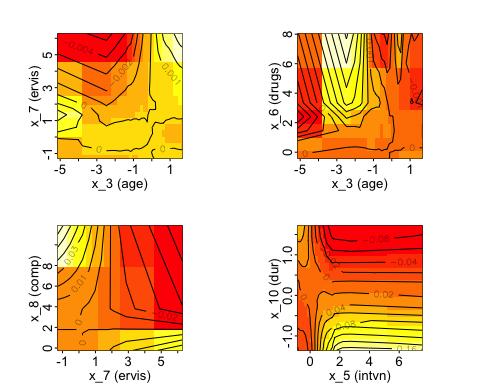
ALEPlot(df\_std, nnet\_mod, pred.fun=yhat, J=c(3,6), K=50, NA.plot = TRUE)

## $K  
## [1] 25 5  
##   
## $x.values  
## $x.values[[1]]  
## [1] -5.14031174 -2.47527110 -2.03109766 -1.73498204 -1.43886641  
## [6] -1.29080860 -1.14275079 -0.99469297 -0.84663516 -0.69857735  
## [11] -0.55051953 -0.40246172 -0.25440391 -0.10634609 0.04171172  
## [16] 0.18976953 0.33782734 0.48588516 0.63394297 0.78200078  
## [21] 0.93005860 1.07811641 1.22617422 1.37423204 1.52228985  
## [26] 1.67034766  
##   
## $x.values[[2]]  
## [1] -0.4198780 0.5200762 1.4600303 2.3999844 3.3399386 8.0397093  
##   
##   
## $f.values  
## 1 2 3 4  
## -3.639280e-04 9.985034e-04 -0.0026170348 -4.926082e-03 -2.872470e-03  
## 1 -4.620949e-04 1.209531e-04 0.0004998031 2.185144e-03 4.364209e-03  
## 2 -4.266677e-04 8.552589e-05 0.0006388952 1.827607e-03 3.831363e-03  
## 3 -4.595096e-04 1.063577e-04 0.0008342463 1.526329e-03 3.354776e-03  
## 4 -3.920794e-04 1.313829e-04 0.0005529551 7.484094e-04 2.401546e-03  
## 5 -3.350426e-04 1.565054e-04 0.0003651097 5.066294e-04 1.984457e-03  
## 6 -3.163938e-04 1.677642e-04 0.0002084721 2.960572e-04 1.598576e-03  
## 7 -2.337119e-04 1.384865e-04 0.0002694259 3.030764e-04 1.430285e-03  
## 8 -1.674770e-04 9.901176e-05 0.0002623118 3.402749e-04 1.292175e-03  
## 9 -1.392074e-04 6.517640e-05 0.0002608372 5.142901e-04 1.211220e-03  
## 10 -7.589645e-05 6.080335e-05 0.0001291570 3.075561e-04 7.495165e-04  
## 11 -8.213419e-05 4.652966e-05 0.0002724145 3.757598e-04 5.627504e-04  
## 12 -5.995769e-05 2.964687e-05 0.0002724186 1.448278e-04 7.684874e-05  
## 13 1.180742e-05 -1.491919e-05 0.0001687455 2.548575e-05 -2.974630e-04  
## 14 2.783640e-05 -2.927637e-05 0.0001179025 1.879473e-05 -3.046589e-04  
## 15 7.495605e-05 -5.156078e-05 -0.0000684681 -1.234239e-04 -4.473824e-04  
## 16 1.169198e-04 -8.498917e-05 -0.0001463389 -1.571427e-04 -4.816061e-04  
## 17 1.371257e-04 -9.154163e-05 -0.0002033082 -3.443998e-04 -8.980528e-04  
## 18 1.737095e-04 -5.266686e-05 -0.0002148502 -4.862296e-04 -1.269072e-03  
## 19 2.309107e-04 -6.929822e-05 -0.0002184896 -5.719602e-04 -1.583993e-03  
## 20 2.557385e-04 -8.465250e-05 -0.0001872419 -5.862408e-04 -1.874407e-03  
## 21 2.691201e-04 -8.388379e-05 -0.0002567787 -5.097338e-04 -2.074034e-03  
## 22 2.828073e-04 -1.054191e-04 -0.0002808844 -3.877958e-04 -1.959233e-03  
## 23 3.469433e-04 -1.066604e-04 -0.0004905700 -4.514377e-04 -2.030011e-03  
## 24 3.711086e-04 -1.136807e-04 -0.0007235797 -5.384037e-04 -2.124114e-03  
## 25 3.743743e-04 -1.169464e-04 -0.0009528348 -6.216151e-04 -2.214462e-03  
## 5  
## -1.387169e-04  
## 1 7.223415e-03  
## 2 5.820579e-03  
## 3 4.474002e-03  
## 4 2.650783e-03  
## 5 1.363703e-03  
## 6 1.078315e-04  
## 7 -9.304487e-04  
## 8 -1.241011e-03  
## 9 -1.494417e-03  
## 10 -2.002710e-03  
## 11 -2.046777e-03  
## 12 -2.389980e-03  
## 13 -1.701537e-03  
## 14 -6.459785e-04  
## 15 2.740525e-04  
## 16 1.302583e-03  
## 17 1.715958e-03  
## 18 -2.042986e-05  
## 19 -5.635107e-04  
## 20 -8.303671e-04  
## 21 -1.006436e-03  
## 22 -8.831540e-04  
## 23 -9.454520e-04  
## 24 -1.031074e-03  
## 25 -1.112942e-03

ALEPlot(df\_std, nnet\_mod, pred.fun=yhat, J=c(7,8), K=50, NA.plot = TRUE)

## $K  
## [1] 11 2  
##   
## $x.values  
## $x.values[[1]]  
## [1] -1.2986393 -0.9194886 -0.5403379 -0.1611872 0.2179635 0.5971142  
## [7] 0.9762649 1.3554156 1.7345663 2.1137170 2.8720185 6.2843748  
##   
## $x.values[[2]]  
## [1] -0.2302054 3.8009470 11.8632519  
##   
##   
## $f.values  
## 1 2  
## -2.849681e-03 2.839745e-03 0.066742957  
## 1 -2.516128e-03 2.506191e-03 0.058891379  
## 2 -2.030454e-03 2.020518e-03 0.050887681  
## 3 -1.138223e-03 1.128287e-03 0.042477425  
## 4 2.476283e-05 -3.469892e-05 0.033796414  
## 5 1.964085e-03 -1.974022e-03 0.024339066  
## 6 4.075781e-03 -4.085717e-03 0.014709346  
## 7 6.591437e-03 -6.601373e-03 0.004675664  
## 8 9.058872e-03 -9.068808e-03 -0.005309795  
## 9 1.213011e-02 -1.128626e-02 -0.015045274  
## 10 1.731731e-02 -1.647346e-02 -0.027750499  
## 11 4.006290e-02 -3.921906e-02 -0.058014121

ALEPlot(df\_std, nnet\_mod, pred.fun=yhat, J=c(5,10), K=50, NA.plot = TRUE)

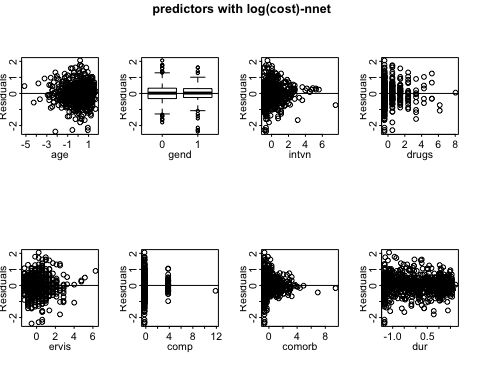


## $K  
## [1] 16 44  
##   
## $x.values  
## $x.values[[1]]  
## [1] -0.84131152 -0.66256968 -0.48382784 -0.30508600 -0.12634417  
## [6] 0.05239767 0.23113951 0.40988135 0.58862318 0.76736502  
## [11] 0.94610686 1.12484870 1.48233237 1.66107421 2.19729972  
## [16] 2.73352523 7.55955484  
##   
## $x.values[[2]]  
## [1] -1.356566133 -1.331755507 -1.298674671 -1.265593836 -1.182891748  
## [6] -1.125000286 -1.050568406 -0.992676945 -0.909974856 -0.810732350  
## [11] -0.678409009 -0.620517547 -0.554355876 -0.488194206 -0.405492117  
## [16] -0.322790029 -0.215277314 -0.132575226 -0.074683764 0.008018324  
## [21] 0.123801248 0.189962919 0.256124589 0.330556469 0.413258557  
## [26] 0.504230854 0.570392525 0.653094613 0.710986075 0.768877537  
## [31] 0.835039208 0.934281714 0.983902967 1.033524220 1.074875264  
## [36] 1.124496517 1.182387979 1.232009232 1.281630485 1.347792155  
## [41] 1.413953826 1.455304870 1.496655914 1.546277167 1.719951553  
##   
##   
## $f.values  
## 1 2 3 4  
## -0.018819989 -0.019096034 -0.019635929 -0.020096658 -0.021686585  
## 1 -0.023578049 -0.023516288 -0.023621621 -0.023602117 -0.024006082  
## 2 -0.027333823 -0.026996365 -0.026710640 -0.026388319 -0.025980839  
## 3 -0.027165223 -0.026669835 -0.026114257 -0.025585399 -0.025128850  
## 4 -0.019638280 -0.019069311 -0.018390708 -0.017835254 -0.016966771  
## 5 -0.007737575 -0.007101212 -0.006628715 -0.006282985 -0.005467176  
## 6 0.008926179 0.009391184 0.009692321 0.009754224 0.010184360  
## 7 0.028899386 0.028949978 0.028836704 0.028484195 0.028528659  
## 8 0.051409780 0.051092558 0.050611468 0.049745091 0.049275688  
## 9 0.070972082 0.070135360 0.069134771 0.067748895 0.066765624  
## 10 0.089029722 0.087673502 0.086153414 0.084437418 0.082655331  
## 11 0.102978539 0.101292198 0.099441991 0.097395875 0.094814972  
## 12 0.126674323 0.124657863 0.120988797 0.117123824 0.112724062  
## 13 0.156030635 0.152195316 0.146707392 0.141023561 0.134804941  
## 14 0.174590164 0.168935987 0.161629205 0.154126515 0.146089037  
## 15 0.179515787 0.172042752 0.162917112 0.155286199 0.147120497  
## 16 0.166921031 0.159853524 0.151133412 0.143908027 0.136147853  
## 5 6 7 8 9  
## -0.021791979 -0.021532630 -0.021751992 -0.0211411811 -0.0200682488  
## 1 -0.023463302 -0.022351263 -0.021853990 -0.0208744607 -0.0190794724  
## 2 -0.024777028 -0.023005606 -0.022106848 -0.0207270120 -0.0188155698  
## 3 -0.023432401 -0.021373064 -0.020107355 -0.0190151510 -0.0161990811  
## 4 -0.015426013 -0.013078762 -0.011694472 -0.0108776888 -0.0094044820  
## 5 -0.004135427 -0.002501838 -0.001339134 -0.0009788537 0.0003244427  
## 6 0.010943853 0.011863778 0.012185627 0.0119858672 0.0117391598  
## 7 0.028492270 0.028698532 0.028179526 0.0273132384 0.0255165272  
## 8 0.048443416 0.047473322 0.046092205 0.0445593903 0.0413862643  
## 9 0.065068494 0.062922044 0.060803630 0.0580206433 0.0534711027  
## 10 0.080093342 0.076830874 0.074321611 0.0702884524 0.0644887402  
## 11 0.091451764 0.087073278 0.084173165 0.0788898356 0.0718399520  
## 12 0.107780549 0.102038366 0.097952740 0.0906471713 0.0815750481  
## 13 0.128281121 0.121175243 0.115904104 0.1065762953 0.0967259837  
## 14 0.137984912 0.130100845 0.124051518 0.1139455206 0.1033170206  
## 15 0.138888148 0.130875858 0.124698307 0.1144640861 0.1037073626  
## 16 0.128321032 0.120714270 0.114942247 0.1051135547 0.0947623593  
## 10 11 12 13 14  
## -0.019654580 -0.019060600 -0.018612328 -0.018913414 -0.018689916  
## 1 -0.017622123 -0.017137835 -0.016347187 -0.016030962 -0.015354052  
## 2 -0.016704387 -0.015949564 -0.014888380 -0.014355682 -0.013260762  
## 3 -0.013622558 -0.012329124 -0.011207671 -0.009931540 -0.008857519  
## 4 -0.006362618 -0.005973028 -0.005394063 -0.004221326 -0.003168202  
## 5 0.001973683 0.002066251 0.002348194 0.002983260 0.003738355  
## 6 0.012105419 0.011686869 0.011186254 0.011283650 0.010933592  
## 7 0.023814851 0.022885183 0.021602009 0.020594253 0.019139044  
## 8 0.038111654 0.035609053 0.033543321 0.031729182 0.029467591  
## 9 0.048623559 0.044548025 0.041675910 0.039055389 0.035987415  
## 10 0.058068264 0.053510812 0.050156780 0.046729877 0.042855520  
## 11 0.064937558 0.059898189 0.056062240 0.051430694 0.046015091  
## 12 0.072065784 0.065821772 0.060781180 0.054944990 0.047988141  
## 13 0.084609848 0.077161193 0.070915959 0.063875126 0.055377030  
## 14 0.090422697 0.082195854 0.074745976 0.066500500 0.056461157  
## 15 0.090684815 0.082329748 0.074751647 0.066746435 0.056947355  
## 16 0.082145340 0.074195801 0.067023228 0.059509840 0.050202583  
## 15 16 17 18 19  
## -0.017971638 -0.016406136 -0.012014932 -0.0095455524 -0.008999603  
## 1 -0.014388687 -0.011957765 -0.006990853 -0.0044591971 -0.003890073  
## 2 -0.011790896 -0.008180650 -0.002834595 -0.0009188969 -0.000213883  
## 3 -0.007497682 -0.003596220 0.001436953 0.0029790959 0.003310166  
## 4 -0.001918394 0.001952187 0.006672477 0.0073983278 0.007355455  
## 5 0.003744967 0.006095044 0.010293892 0.0104983006 0.009280258  
## 6 0.009666362 0.010415977 0.013370926 0.0126696713 0.011225784  
## 7 0.016643036 0.016419858 0.018130910 0.0165239912 0.014854259  
## 8 0.025742806 0.023883603 0.024373743 0.0221822173 0.020286640  
## 9 0.031456248 0.027957254 0.027226482 0.0249220136 0.022913493  
## 10 0.037131083 0.032438819 0.030737312 0.0277792509 0.025221978  
## 11 0.038965419 0.032947920 0.030121740 0.0265100856 0.023404060  
## 12 0.039434142 0.031912317 0.028142024 0.0235862576 0.019931479  
## 13 0.045318705 0.036292553 0.031578148 0.0260782691 0.022388517  
## 14 0.046643095 0.037581970 0.032832592 0.0272977400 0.023573015  
## 15 0.047369556 0.038548694 0.033542385 0.0277440103 0.023755762  
## 16 0.041116608 0.032453127 0.027604200 0.0219632058 0.018132338  
## 20 21 22 23 24  
## -0.0057420220 -0.004348121 -0.003590824 -0.0033568648 0.0007831891  
## 1 -0.0003960318 0.001286212 0.002032487 0.0026220307 0.0066072856  
## 2 0.0035166176 0.005093963 0.005447629 0.0061748164 0.0100052723  
## 3 0.0057151244 0.006860549 0.007940934 0.0082556661 0.0118084078  
## 4 0.0084348702 0.009148374 0.010955478 0.0114397372 0.0141828072  
## 5 0.0098163358 0.010097919 0.010875061 0.0106015709 0.0120522063  
## 6 0.0105894633 0.010053855 0.009870820 0.0085825621 0.0091987081  
## 7 0.0130455401 0.011731224 0.010588011 0.0082849862 0.0080666427  
## 8 0.0168634807 0.014770456 0.012848535 0.0095307429 0.0079974772  
## 9 0.0178758939 0.014168429 0.012219555 0.0078953590 0.0051172386  
## 10 0.0185699379 0.014835520 0.012859693 0.0075290931 0.0035061181  
## 11 0.0162032670 0.012441895 0.010439115 0.0041021123 -0.0017537333  
## 12 0.0126034177 0.008608823 0.006372819 -0.0001974077 -0.0078861240  
## 13 0.0149331875 0.010811324 0.008448052 0.0015044474 -0.0065576465  
## 14 0.0158699393 0.011527393 0.008943439 0.0017791517 -0.0062209163  
## 15 0.0161268010 0.011858369 0.009348529 0.0022583569 -0.0061760078  
## 16 0.0112964866 0.007821164 0.006104433 -0.0001926302 -0.0082869735  
## 25 26 27 28 29  
## 0.0030563558 0.007697841 0.0099618761 0.013695611 0.0165922179  
## 1 0.0081957588 0.012154340 0.0143777923 0.017662026 0.0200343630  
## 2 0.0116017530 0.015077510 0.0173476598 0.020227948 0.0221819921  
## 3 0.0134249342 0.016417868 0.0178346778 0.020025029 0.0212274167  
## 4 0.0147693325 0.016799985 0.0174953315 0.018792376 0.0191014574  
## 5 0.0117168997 0.012785270 0.0128265434 0.013524301 0.0130912793  
## 6 0.0079415695 0.008284488 0.0072805378 0.007037192 0.0056630671  
## 7 0.0060840522 0.005701519 0.0036523452 0.002870831 0.0009585361  
## 8 0.0048115204 0.003428249 0.0003338513 -0.001364005 -0.0041699684  
## 9 0.0007279155 -0.001656095 -0.0054828143 -0.007912993 -0.0111542538  
## 10 -0.0021286477 -0.005758101 -0.0108221400 -0.013773658 -0.0172973841  
## 11 -0.0086339418 -0.013500715 -0.0198020737 -0.023274932 -0.0271855992  
## 12 -0.0156394967 -0.021379434 -0.0285539571 -0.032548155 -0.0368457641  
## 13 -0.0144565683 -0.020342055 -0.0276621270 -0.031920460 -0.0364822036  
## 14 -0.0140578123 -0.020088848 -0.0275473430 -0.031944099 -0.0366442656  
## 15 -0.0144472004 -0.020351439 -0.0278253290 -0.032137756 -0.0367535944  
## 16 -0.0158112695 -0.020968612 -0.0279094242 -0.031641424 -0.0356768345  
## 30 31 32 33 34  
## 0.018527403 0.023990102 0.025279527 0.0271946042 0.0291291937  
## 1 0.021551385 0.026986147 0.028047620 0.0296092990 0.0311766369  
## 2 0.023395466 0.028414884 0.028962548 0.0299226087 0.0310879140  
## 3 0.022065658 0.025829437 0.026466338 0.0267173912 0.0272187212  
## 4 0.019046392 0.021214029 0.021178420 0.0208775857 0.0208111179  
## 5 0.012294111 0.012865606 0.012149913 0.0111526924 0.0104624713  
## 6 0.003799927 0.002931229 0.001824280 0.0003297069 -0.0008578668  
## 7 -0.001970576 -0.004144480 -0.005642684 -0.0073762230 -0.0088027621  
## 8 -0.007992749 -0.011471859 -0.013708024 -0.0156455442 -0.0172227587  
## 9 -0.015412333 -0.019536268 -0.022510393 -0.0246518942 -0.0263797841  
## 10 -0.021837928 -0.026606688 -0.030225638 -0.0326845332 -0.0347298171  
## 11 -0.032113084 -0.037526669 -0.041463014 -0.0442393029 -0.0466019808  
## 12 -0.041993953 -0.047628241 -0.051785289 -0.0548789717 -0.0575590437  
## 13 -0.041894527 -0.047669526 -0.051967284 -0.0552016780 -0.0577922935  
## 14 -0.042195012 -0.048377370 -0.052599094 -0.0557440318 -0.0582451908  
## 15 -0.042261761 -0.048703916 -0.053185438 -0.0565901726 -0.0593511288  
## 16 -0.040604574 -0.046354763 -0.050144319 -0.0528570872 -0.0549260773  
## 35 36 37 38 39  
## 0.029998963 0.033414967 0.035392575 0.037775579 0.040875545  
## 1 0.031902002 0.035188106 0.036660503 0.038599064 0.041121190  
## 2 0.031654134 0.034004742 0.034862553 0.035981015 0.037469562  
## 3 0.027495697 0.029210512 0.029278698 0.029605061 0.029998663  
## 4 0.020609649 0.021460774 0.020999525 0.020535009 0.019968701  
## 5 0.009637249 0.009624683 0.008392398 0.007156845 0.005816142  
## 6 -0.002240885 -0.003101812 -0.004663740 -0.005886104 -0.007758126  
## 7 -0.010424746 -0.012154465 -0.014203724 -0.015913421 -0.018139835  
## 8 -0.019014283 -0.021280615 -0.024033939 -0.026447701 -0.028958323  
## 9 -0.028340849 -0.031143794 -0.034601183 -0.037299153 -0.040093985  
## 10 -0.037008276 -0.039974181 -0.043594529 -0.046455460 -0.049852537  
## 11 -0.049197834 -0.052597528 -0.056476316 -0.059595686 -0.063251203  
## 12 -0.060588685 -0.064422169 -0.068734746 -0.072112556 -0.076026513  
## 13 -0.061255724 -0.065522997 -0.070269363 -0.073790200 -0.077847184  
## 14 -0.061752412 -0.066063474 -0.070806859 -0.074583971 -0.078913548  
## 15 -0.062902140 -0.067256993 -0.071997396 -0.076030784 -0.080632953  
## 16 -0.057785123 -0.061448009 -0.065723227 -0.069291428 -0.073428411  
## 40 41 42 43 44  
## 0.043893407 0.045516856 0.048327220 0.050463104 0.060225642  
## 1 0.043496630 0.044744420 0.046964228 0.048990260 0.056022193  
## 2 0.039327041 0.040056868 0.041758714 0.043078350 0.047871222  
## 3 0.031226704 0.031282145 0.032695854 0.033300573 0.036537329  
## 4 0.020634242 0.020127183 0.020897234 0.020772533 0.023248031  
## 5 0.005595491 0.004789316 0.005260252 0.004984859 0.004916723  
## 6 -0.009013979 -0.010227874 -0.010098607 -0.010360948 -0.011157823  
## 7 -0.020249365 -0.022109582 -0.022539037 -0.023360101 -0.026050227  
## 8 -0.031923939 -0.033818467 -0.034282233 -0.035137608 -0.039720986  
## 9 -0.043787915 -0.045938538 -0.046772260 -0.048038615 -0.053032974  
## 10 -0.053540736 -0.055685629 -0.056879397 -0.058150051 -0.063148709  
## 11 -0.066933672 -0.069072833 -0.070259501 -0.071523056 -0.076514613  
## 12 -0.079747269 -0.081924718 -0.083339445 -0.085003001 -0.090221020  
## 13 -0.081710967 -0.084031443 -0.085589197 -0.087652755 -0.093097235  
## 14 -0.083049923 -0.085513305 -0.087213965 -0.089420429 -0.095360511  
## 15 -0.084770135 -0.087234323 -0.088935790 -0.091143061 -0.097083950  
## 16 -0.076920988 -0.078740572 -0.080043563 -0.081852358 -0.087394771

## (d)Construct appropriate residual plots to assess the nonlinearity not captured by the nnet.

## Answer: From the residual plts, there is no nonlinearity not captured by neural network.

par(mfrow=c(2,4),pin=c(0.8,0.8),tcl=-0.15,mgp=c(1,0.2,0))  
for (i in seq(3:10)) {  
 plot(df\_std[[i+2]],resid(nnet\_mod),ylab="Residuals",xlab=names(df)[i+2],main="")  
 abline(0, 0)}  
title(main="Ischemic heart disease-standardized \n predictors with log(cost)-nnet",outer = T)

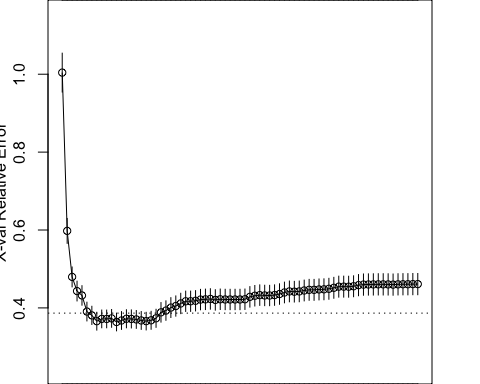


# Prob 3)Repeat Prob 2) but for a regression tree.

## (a)Use 10-fold CV to find the best tree size or complexity parameter value

## Answer: The cross-validation of the best model is , with the penalization and number of hidden nodes as .

#do not have to standardize or transform predictors to fit trees  
# the CV shell is not correct in tree?  
# cp is \lambda, the complex parameter; with small cp we will grow a big tree(overfit)  
# xval: fold of cross validation  
library(rpart)  
control <- rpart.control(minbucket = 5, cp = 0.0001, maxsurrogate = 0, usesurrogate = 0, xval = 10)  
par(mfrow=c(1,1),pin=c(4,4),mgp=c(2,1,0))  
df\_std.tr <- rpart(cost ~ .,df\_std, method = "anova", control = control)  
plotcp(df\_std.tr) #plot of CV r^2 vs. size



printcp(df\_std.tr) #same info is in df\_std.tr$cptable

##   
## Regression tree:  
## rpart(formula = cost ~ ., data = df\_std, method = "anova", control = control)  
##   
## Variables actually used in tree construction:  
## [1] age comorb comp dur ervis gend intvn X   
##   
## Root node error: 787/788 = 0.99873  
##   
## n= 788   
##   
## CP nsplit rel error xerror xstd  
## 1 0.43938070 0 1.00000 1.00418 0.049875  
## 2 0.09581897 1 0.56062 0.59773 0.031942  
## 3 0.05828524 2 0.46480 0.47955 0.025505  
## 4 0.03070183 3 0.40652 0.44327 0.025352  
## 5 0.02475265 4 0.37581 0.43202 0.025008  
## 6 0.01166647 5 0.35106 0.39063 0.024038  
## 7 0.00991235 6 0.33939 0.38080 0.023348  
## 8 0.00664712 7 0.32948 0.36589 0.022869  
## 9 0.00622922 8 0.32283 0.37176 0.022742  
## 10 0.00607735 9 0.31661 0.37168 0.022716  
## 11 0.00577605 10 0.31053 0.37310 0.022817  
## 12 0.00543282 11 0.30475 0.36386 0.022735  
## 13 0.00468591 12 0.29932 0.36769 0.022781  
## 14 0.00409089 14 0.28995 0.37211 0.023163  
## 15 0.00399151 15 0.28586 0.37170 0.023141  
## 16 0.00367634 16 0.28187 0.37038 0.022880  
## 17 0.00351925 17 0.27819 0.36778 0.022885  
## 18 0.00325939 18 0.27467 0.36605 0.022590  
## 19 0.00314084 19 0.27141 0.36817 0.023029  
## 20 0.00239211 21 0.26513 0.37331 0.023182  
## 21 0.00218520 22 0.26274 0.38859 0.025057  
## 22 0.00201994 23 0.26055 0.39313 0.024958  
## 23 0.00198681 24 0.25853 0.40083 0.025465  
## 24 0.00191879 25 0.25654 0.40479 0.025943  
## 25 0.00172807 26 0.25463 0.41144 0.026129  
## 26 0.00171146 27 0.25290 0.41702 0.026068  
## 27 0.00169081 28 0.25119 0.41707 0.026067  
## 28 0.00165318 29 0.24950 0.41831 0.026033  
## 29 0.00159386 32 0.24454 0.42138 0.026206  
## 30 0.00154465 33 0.24294 0.42194 0.026227  
## 31 0.00148975 34 0.24140 0.42296 0.026236  
## 32 0.00145187 35 0.23991 0.42048 0.026212  
## 33 0.00137539 36 0.23846 0.42216 0.026248  
## 34 0.00137445 37 0.23708 0.42164 0.026154  
## 35 0.00137418 38 0.23571 0.42164 0.026154  
## 36 0.00136847 40 0.23296 0.42164 0.026154  
## 37 0.00135649 41 0.23159 0.42164 0.026154  
## 38 0.00131296 42 0.23023 0.42222 0.026174  
## 39 0.00121457 43 0.22892 0.42837 0.026276  
## 40 0.00120267 44 0.22770 0.43123 0.026306  
## 41 0.00118913 45 0.22650 0.43280 0.026317  
## 42 0.00116990 47 0.22412 0.43197 0.026192  
## 43 0.00114827 50 0.22061 0.43183 0.026117  
## 44 0.00110054 51 0.21947 0.43291 0.026239  
## 45 0.00103407 53 0.21726 0.43561 0.026276  
## 46 0.00095212 54 0.21623 0.43960 0.026549  
## 47 0.00091540 56 0.21433 0.44232 0.026658  
## 48 0.00091359 57 0.21341 0.44162 0.026508  
## 49 0.00089511 58 0.21250 0.44235 0.026512  
## 50 0.00084125 59 0.21160 0.44471 0.026650  
## 51 0.00083257 60 0.21076 0.44640 0.026837  
## 52 0.00079518 62 0.20910 0.44635 0.026837  
## 53 0.00075828 65 0.20671 0.44742 0.026840  
## 54 0.00075577 66 0.20595 0.44776 0.026834  
## 55 0.00065863 67 0.20520 0.44908 0.026845  
## 56 0.00058355 68 0.20454 0.45163 0.026801  
## 57 0.00054158 69 0.20395 0.45457 0.026906  
## 58 0.00049403 71 0.20287 0.45461 0.026908  
## 59 0.00047174 72 0.20238 0.45442 0.026882  
## 60 0.00044834 73 0.20191 0.45557 0.026836  
## 61 0.00037554 74 0.20146 0.45843 0.027161  
## 62 0.00036051 75 0.20108 0.45986 0.027268  
## 63 0.00034685 76 0.20072 0.46013 0.027259  
## 64 0.00029060 77 0.20037 0.46031 0.027271  
## 65 0.00027881 78 0.20008 0.46050 0.027273  
## 66 0.00027063 79 0.19980 0.46048 0.027274  
## 67 0.00025437 80 0.19953 0.46034 0.027272  
## 68 0.00024858 82 0.19903 0.45980 0.027219  
## 69 0.00021032 83 0.19878 0.46051 0.027217  
## 70 0.00020138 84 0.19857 0.46078 0.027213  
## 71 0.00019722 85 0.19837 0.46078 0.027213  
## 72 0.00013895 86 0.19817 0.46112 0.027218  
## 73 0.00010000 87 0.19803 0.46113 0.027217

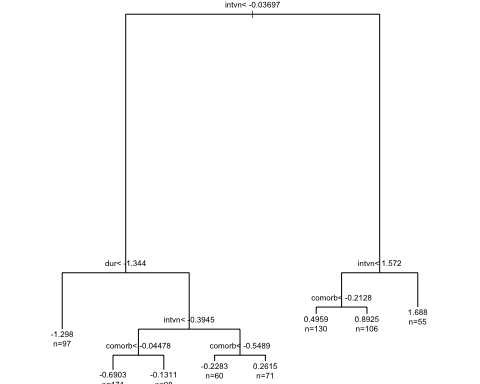
#prune back to optimal size, according to plot of CV 1-r^2  
df\_std.tr1 <- prune(df\_std.tr, cp=0.00991235) #approximately the best size pruned tree  
df\_std.tr1$variable.importance#The importance of each predictors

## intvn dur comorb   
## 415.82543 75.40953 36.46287

df\_std.tr1$cptable[nrow(df\_std.tr1$cptable),] #shows training and CV 1-r^2, and other things

## CP nsplit rel error xerror xstd   
## 0.00991235 7.00000000 0.32948180 0.36589363 0.02286890

# #prune and plot a little smaller tree than the optimal one, just for display  
# df\_std.tr2 <- prune(df\_std.tr, cp=0.00631770) #bigger cp gives smaller size tree  
# df\_std.tr2  
par(cex=.5); plot(df\_std.tr1, uniform=F); text(df\_std.tr1, use.n = T); par(cex=1)



##  
yhat<-predict(df\_std.tr1); e<-df\_std$cost-yhat  
c(1-var(e)/var(df\_std$cost), 1-df\_std.tr1$cptable[nrow(df\_std.tr1$cptable),3]) #check to see training r^2 agrees with what is in cptable

## [1] 0.6705182 0.6705182

## (b)Fit the best model and discuss how good the predictive power of model is.

## Answer: The of the best model is .

control\_best <- rpart.control(minbucket = 5, cp = 0.00991235, maxsurrogate = 0, usesurrogate = 0)  
df\_std.tr\_best <- rpart(cost ~ .,df\_std, method = "anova", control = control)  
summary(df\_std.tr\_best)

## Call:  
## rpart(formula = cost ~ ., data = df\_std, method = "anova", control = control)  
## n= 788   
##   
## CP nsplit rel error xerror xstd  
## 1 0.4393806980 0 1.0000000 1.0012859 0.04969062  
## 2 0.0958189687 1 0.5606193 0.5786511 0.03041456  
## 3 0.0582852400 2 0.4648003 0.4876924 0.02518666  
## 4 0.0307018259 3 0.4065151 0.4234470 0.02370008  
## 5 0.0247526454 4 0.3758133 0.4103806 0.02318566  
## 6 0.0116664748 5 0.3510606 0.3711441 0.02212699  
## 7 0.0099123516 6 0.3393941 0.3651516 0.02175201  
## 8 0.0066471212 7 0.3294818 0.3557923 0.02119254  
## 9 0.0062292206 8 0.3228347 0.3524980 0.02084895  
## 10 0.0060773511 9 0.3166055 0.3538237 0.02115414  
## 11 0.0057760514 10 0.3105281 0.3488309 0.02113143  
## 12 0.0054328169 11 0.3047521 0.3569755 0.02194587  
## 13 0.0046859091 12 0.2993192 0.3551443 0.02162683  
## 14 0.0040908932 14 0.2899474 0.3519707 0.02132498  
## 15 0.0039915125 15 0.2858565 0.3531667 0.02120588  
## 16 0.0036763410 16 0.2818650 0.3526794 0.02115299  
## 17 0.0035192504 17 0.2781887 0.3521567 0.02102946  
## 18 0.0032593908 18 0.2746694 0.3495872 0.02075046  
## 19 0.0031408351 19 0.2714100 0.3496215 0.02096043  
## 20 0.0023921101 21 0.2651284 0.3555114 0.02146275  
## 21 0.0021851965 22 0.2627362 0.3541340 0.02196686  
## 22 0.0020199429 23 0.2605511 0.3542153 0.02183583  
## 23 0.0019868052 24 0.2585311 0.3555504 0.02206639  
## 24 0.0019187941 25 0.2565443 0.3637513 0.02279507  
## 25 0.0017280675 26 0.2546255 0.3692774 0.02294902  
## 26 0.0017114612 27 0.2528974 0.3722436 0.02310999  
## 27 0.0016908144 28 0.2511860 0.3751710 0.02313302  
## 28 0.0016531753 29 0.2494952 0.3774764 0.02316394  
## 29 0.0015938622 32 0.2445356 0.3775323 0.02316220  
## 30 0.0015446491 33 0.2429418 0.3815622 0.02368391  
## 31 0.0014897502 34 0.2413971 0.3823339 0.02371365  
## 32 0.0014518749 35 0.2399074 0.3859007 0.02394141  
## 33 0.0013753869 36 0.2384555 0.3891178 0.02393746  
## 34 0.0013744505 37 0.2370801 0.3923620 0.02402686  
## 35 0.0013741820 38 0.2357057 0.3923620 0.02402686  
## 36 0.0013684667 40 0.2329573 0.3923620 0.02402686  
## 37 0.0013564880 41 0.2315888 0.3941048 0.02405453  
## 38 0.0013129566 42 0.2302323 0.3950325 0.02410118  
## 39 0.0012145663 43 0.2289194 0.3961811 0.02416805  
## 40 0.0012026706 44 0.2277048 0.3968755 0.02424335  
## 41 0.0011891268 45 0.2265022 0.3964759 0.02423941  
## 42 0.0011699030 47 0.2241239 0.3969063 0.02424304  
## 43 0.0011482663 50 0.2206142 0.3975923 0.02428895  
## 44 0.0011005356 51 0.2194659 0.4031947 0.02453942  
## 45 0.0010340717 53 0.2172649 0.4068800 0.02473075  
## 46 0.0009521186 54 0.2162308 0.4100927 0.02476712  
## 47 0.0009153973 56 0.2143265 0.4132673 0.02475891  
## 48 0.0009135854 57 0.2134111 0.4140330 0.02478324  
## 49 0.0008951054 58 0.2124976 0.4140330 0.02478324  
## 50 0.0008412472 59 0.2116025 0.4160754 0.02487156  
## 51 0.0008325706 60 0.2107612 0.4166382 0.02494894  
## 52 0.0007951798 62 0.2090961 0.4165106 0.02494249  
## 53 0.0007582752 65 0.2067105 0.4181090 0.02493675  
## 54 0.0007557695 66 0.2059523 0.4169767 0.02488438  
## 55 0.0006586300 67 0.2051965 0.4162883 0.02475721  
## 56 0.0005835501 68 0.2045379 0.4153512 0.02474075  
## 57 0.0005415793 69 0.2039543 0.4174531 0.02480829  
## 58 0.0004940284 71 0.2028711 0.4183426 0.02475713  
## 59 0.0004717407 72 0.2023771 0.4176996 0.02472919  
## 60 0.0004483399 73 0.2019054 0.4186287 0.02476807  
## 61 0.0003755365 74 0.2014570 0.4199849 0.02484124  
## 62 0.0003605072 75 0.2010815 0.4198727 0.02483015  
## 63 0.0003468544 76 0.2007210 0.4191001 0.02478548  
## 64 0.0002906009 77 0.2003741 0.4193551 0.02475935  
## 65 0.0002788063 78 0.2000835 0.4206516 0.02475723  
## 66 0.0002706308 79 0.1998047 0.4209884 0.02478058  
## 67 0.0002543666 80 0.1995341 0.4209884 0.02478058  
## 68 0.0002485771 82 0.1990254 0.4209884 0.02478058  
## 69 0.0002103247 83 0.1987768 0.4215253 0.02479106  
## 70 0.0002013780 84 0.1985665 0.4214592 0.02479099  
## 71 0.0001972181 85 0.1983651 0.4215459 0.02478939  
## 72 0.0001389504 86 0.1981679 0.4219656 0.02479055  
## 73 0.0001000000 87 0.1980289 0.4228664 0.02484757  
##   
## Variable importance  
## intvn dur comorb ervis age X comp   
## 70 13 8 3 3 2 1   
##   
## Node number 1: 788 observations, complexity param=0.4393807  
## mean=4.990192e-17, MSE=0.998731   
## left son=2 (497 obs) right son=3 (291 obs)  
## Primary splits:  
## intvn < -0.03697325 to the left, improve=0.43938070, (0 missing)  
## dur < -1.344161 to the left, improve=0.22953210, (0 missing)  
## comorb < -0.2128213 to the left, improve=0.14947230, (0 missing)  
## ervis < 0.4075389 to the left, improve=0.08468142, (0 missing)  
## comp < 1.785371 to the left, improve=0.06102652, (0 missing)  
##   
## Node number 2: 497 observations, complexity param=0.09581897  
## mean=-0.5068892, MSE=0.5936514   
## left son=4 (97 obs) right son=5 (400 obs)  
## Primary splits:  
## dur < -1.344161 to the left, improve=0.25558670, (0 missing)  
## comorb < -0.3808606 to the left, improve=0.19056300, (0 missing)  
## intvn < -0.3944569 to the left, improve=0.18556480, (0 missing)  
## comp < 1.785371 to the left, improve=0.04626715, (0 missing)  
## X < 81.5 to the right, improve=0.00946294, (0 missing)  
##   
## Node number 3: 291 observations, complexity param=0.05828524  
## mean=0.865718, MSE=0.5022771   
## left son=6 (236 obs) right son=7 (55 obs)  
## Primary splits:  
## intvn < 1.571703 to the left, improve=0.31383180, (0 missing)  
## ervis < -0.3507625 to the left, improve=0.12641140, (0 missing)  
## comorb < -0.2128213 to the left, improve=0.09684730, (0 missing)  
## comp < 1.785371 to the left, improve=0.04126791, (0 missing)  
## dur < -0.4137623 to the left, improve=0.03181452, (0 missing)  
##   
## Node number 4: 97 observations, complexity param=0.005776051  
## mean=-1.297894, MSE=0.4423941   
## left son=8 (80 obs) right son=9 (17 obs)  
## Primary splits:  
## intvn < -0.5731988 to the left, improve=0.10593140, (0 missing)  
## gend splits as RL, improve=0.06391250, (0 missing)  
## age < -1.809011 to the left, improve=0.03172130, (0 missing)  
## ervis < 0.4075389 to the right, improve=0.02278604, (0 missing)  
## X < 42.5 to the right, improve=0.01857077, (0 missing)  
##   
## Node number 5: 400 observations, complexity param=0.03070183  
## mean=-0.3150706, MSE=0.4418075   
## left son=10 (269 obs) right son=11 (131 obs)  
## Primary splits:  
## intvn < -0.3944569 to the left, improve=0.13672430, (0 missing)  
## comorb < 0.4593358 to the left, improve=0.12631820, (0 missing)  
## dur < 1.0542 to the left, improve=0.07286293, (0 missing)  
## comp < 1.785371 to the left, improve=0.04958728, (0 missing)  
## X < 81.5 to the right, improve=0.01470129, (0 missing)  
##   
## Node number 6: 236 observations, complexity param=0.01166647  
## mean=0.6740518, MSE=0.3684249   
## left son=12 (130 obs) right son=13 (106 obs)  
## Primary splits:  
## comorb < -0.2128213 to the left, improve=0.10559740, (0 missing)  
## intvn < 0.4992523 to the left, improve=0.08040409, (0 missing)  
## ervis < -0.3507625 to the left, improve=0.07371652, (0 missing)  
## comp < 1.785371 to the left, improve=0.02288521, (0 missing)  
## dur < 1.492521 to the left, improve=0.01794721, (0 missing)  
##   
## Node number 7: 55 observations, complexity param=0.00351925  
## mean=1.68814, MSE=0.2426159   
## left son=14 (45 obs) right son=15 (10 obs)  
## Primary splits:  
## intvn < 3.805976 to the left, improve=0.20755970, (0 missing)  
## ervis < 1.16584 to the left, improve=0.10447630, (0 missing)  
## dur < 0.194098 to the left, improve=0.08316169, (0 missing)  
## X < 64.5 to the left, improve=0.06211012, (0 missing)  
## age < -1.809011 to the right, improve=0.06004955, (0 missing)  
##   
## Node number 8: 80 observations, complexity param=0.001544649  
## mean=-1.397686, MSE=0.3949815   
## left son=16 (20 obs) right son=17 (60 obs)  
## Primary splits:  
## gend splits as RL, improve=0.038471390, (0 missing)  
## age < -1.809011 to the left, improve=0.024535370, (0 missing)  
## X < 489.5 to the left, improve=0.022957860, (0 missing)  
## ervis < 0.7866896 to the left, improve=0.019549280, (0 missing)  
## drugs < 0.05009908 to the left, improve=0.001351071, (0 missing)  
##   
## Node number 9: 17 observations, complexity param=0.00148975  
## mean=-0.8282836, MSE=0.3981151   
## left son=18 (12 obs) right son=19 (5 obs)  
## Primary splits:  
## age < 0.1897695 to the left, improve=0.17323300, (0 missing)  
## X < 404.5 to the right, improve=0.17294750, (0 missing)  
## ervis < -0.7299132 to the right, improve=0.02013875, (0 missing)  
##   
## Node number 10: 269 observations, complexity param=0.02475265  
## mean=-0.4865843, MSE=0.3824584   
## left son=20 (171 obs) right son=21 (98 obs)  
## Primary splits:  
## comorb < -0.04478204 to the left, improve=0.18934760, (0 missing)  
## dur < 0.2023682 to the left, improve=0.07974283, (0 missing)  
## ervis < 0.7866896 to the left, improve=0.03083300, (0 missing)  
## drugs < 1.930007 to the right, improve=0.02464214, (0 missing)  
## X < 762 to the right, improve=0.01123997, (0 missing)  
##   
## Node number 11: 131 observations, complexity param=0.009912352  
## mean=0.03712181, MSE=0.3792317   
## left son=22 (60 obs) right son=23 (71 obs)  
## Primary splits:  
## comorb < -0.5488999 to the left, improve=0.15702740, (0 missing)  
## dur < 1.083145 to the left, improve=0.09381864, (0 missing)  
## comp < 1.785371 to the left, improve=0.05937084, (0 missing)  
## ervis < 2.303292 to the right, improve=0.03666887, (0 missing)  
## X < 58 to the right, improve=0.03383980, (0 missing)  
##   
## Node number 12: 130 observations, complexity param=0.006229221  
## mean=0.4959442, MSE=0.3883622   
## left son=24 (96 obs) right son=25 (34 obs)  
## Primary splits:  
## intvn < 0.8567359 to the left, improve=0.09710199, (0 missing)  
## ervis < -0.3507625 to the left, improve=0.06682977, (0 missing)  
## comp < 1.785371 to the left, improve=0.06246507, (0 missing)  
## age < 0.8560297 to the right, improve=0.02531820, (0 missing)  
## drugs < 0.05009908 to the left, improve=0.02265002, (0 missing)  
##   
## Node number 13: 106 observations, complexity param=0.006647121  
## mean=0.8924857, MSE=0.2573555   
## left son=26 (63 obs) right son=27 (43 obs)  
## Primary splits:  
## intvn < 0.4992523 to the left, improve=0.19176490, (0 missing)  
## dur < 0.02869385 to the right, improve=0.03153982, (0 missing)  
## ervis < 0.4075389 to the left, improve=0.02757362, (0 missing)  
## comorb < 0.4593358 to the right, improve=0.02016753, (0 missing)  
## X < 673.5 to the left, improve=0.02005670, (0 missing)  
##   
## Node number 14: 45 observations, complexity param=0.001374182  
## mean=1.582355, MSE=0.2172972   
## left son=28 (38 obs) right son=29 (7 obs)  
## Primary splits:  
## age < 0.7079719 to the left, improve=0.10189360, (0 missing)  
## X < 619.5 to the left, improve=0.06367864, (0 missing)  
## ervis < 1.16584 to the left, improve=0.06206714, (0 missing)  
## dur < 0.194098 to the left, improve=0.05755897, (0 missing)  
## intvn < 2.286671 to the left, improve=0.03695674, (0 missing)  
##   
## Node number 15: 10 observations  
## mean=2.164174, MSE=0.0795848   
##   
## Node number 16: 20 observations, complexity param=0.001375387  
## mean=-1.611196, MSE=0.7822615   
## left son=32 (6 obs) right son=33 (14 obs)  
## Primary splits:  
## ervis < 0.4075389 to the right, improve=0.06918591, (0 missing)  
## X < 403.5 to the left, improve=0.05426015, (0 missing)  
## intvn < -0.7519406 to the right, improve=0.04088996, (0 missing)  
## age < -1.586924 to the left, improve=0.03911049, (0 missing)  
##   
## Node number 17: 60 observations, complexity param=0.0008325706  
## mean=-1.326516, MSE=0.2456275   
## left son=34 (12 obs) right son=35 (48 obs)  
## Primary splits:  
## X < 636.5 to the right, improve=0.031237200, (0 missing)  
## drugs < 0.05009908 to the left, improve=0.028561650, (0 missing)  
## ervis < 0.02838819 to the left, improve=0.011923470, (0 missing)  
## intvn < -0.7519406 to the left, improve=0.010331530, (0 missing)  
## age < 0.5599141 to the left, improve=0.007843768, (0 missing)  
##   
## Node number 18: 12 observations  
## mean=-0.9978009, MSE=0.2800886   
##   
## Node number 19: 5 observations  
## mean=-0.4214421, MSE=0.4468919   
##   
## Node number 20: 171 observations, complexity param=0.005432817  
## mean=-0.6903059, MSE=0.4157082   
## left son=40 (154 obs) right son=41 (17 obs)  
## Primary splits:  
## ervis < 0.7866896 to the left, improve=0.06014716, (0 missing)  
## X < 84 to the right, improve=0.03594703, (0 missing)  
## intvn < -0.5731988 to the left, improve=0.03573340, (0 missing)  
## comorb < -0.3808606 to the left, improve=0.03216706, (0 missing)  
## drugs < 1.930007 to the right, improve=0.01823729, (0 missing)  
##   
## Node number 21: 98 observations, complexity param=0.003259391  
## mean=-0.131111, MSE=0.1256621   
## left son=42 (77 obs) right son=43 (21 obs)  
## Primary splits:  
## comorb < 1.299532 to the left, improve=0.20829600, (0 missing)  
## dur < -0.2276826 to the left, improve=0.09581209, (0 missing)  
## ervis < 0.7866896 to the left, improve=0.04381983, (0 missing)  
## X < 170.5 to the left, improve=0.02533905, (0 missing)  
## age < 1.448261 to the right, improve=0.02445647, (0 missing)  
##   
## Node number 22: 60 observations, complexity param=0.006077351  
## mean=-0.2283348, MSE=0.3726663   
## left son=44 (39 obs) right son=45 (21 obs)  
## Primary splits:  
## age < -0.3284328 to the right, improve=0.21390340, (0 missing)  
## X < 132.5 to the right, improve=0.07888370, (0 missing)  
## dur < 0.3181512 to the right, improve=0.03303060, (0 missing)  
## comp < 1.785371 to the left, improve=0.03015901, (0 missing)  
## ervis < 0.7866896 to the right, improve=0.02865753, (0 missing)  
##   
## Node number 23: 71 observations, complexity param=0.003676341  
## mean=0.2614514, MSE=0.2749064   
## left son=46 (66 obs) right son=47 (5 obs)  
## Primary splits:  
## comp < 1.785371 to the left, improve=0.14823380, (0 missing)  
## comorb < 0.7954144 to the left, improve=0.11798930, (0 missing)  
## ervis < 0.02838819 to the left, improve=0.07963678, (0 missing)  
## X < 66.5 to the right, improve=0.04915741, (0 missing)  
## dur < 1.07901 to the left, improve=0.03078364, (0 missing)  
##   
## Node number 24: 96 observations, complexity param=0.004090893  
## mean=0.3803765, MSE=0.3485705   
## left son=48 (87 obs) right son=49 (9 obs)  
## Primary splits:  
## comp < 1.785371 to the left, improve=0.09621239, (0 missing)  
## ervis < -0.3507625 to the left, improve=0.09599961, (0 missing)  
## age < -0.6245484 to the left, improve=0.02468846, (0 missing)  
## dur < -0.6205175 to the right, improve=0.01943866, (0 missing)  
## X < 51.5 to the right, improve=0.01645515, (0 missing)  
##   
## Node number 25: 34 observations, complexity param=0.001711461  
## mean=0.8222529, MSE=0.3565271   
## left son=50 (29 obs) right son=51 (5 obs)  
## Primary splits:  
## ervis < 0.7866896 to the left, improve=0.11111440, (0 missing)  
## age < -0.03231719 to the left, improve=0.08202016, (0 missing)  
## drugs < 0.9900532 to the left, improve=0.06417642, (0 missing)  
## gend splits as RL, improve=0.04642856, (0 missing)  
## dur < -0.2318177 to the right, improve=0.03547546, (0 missing)  
##   
## Node number 26: 63 observations, complexity param=0.001169903  
## mean=0.7089522, MSE=0.2079169   
## left son=52 (38 obs) right son=53 (25 obs)  
## Primary splits:  
## age < 0.4118563 to the left, improve=0.05560168, (0 missing)  
## ervis < -0.3507625 to the left, improve=0.03908651, (0 missing)  
## X < 673.5 to the left, improve=0.03725504, (0 missing)  
## intvn < 0.1417686 to the left, improve=0.03634097, (0 missing)  
## dur < 1.016984 to the left, improve=0.03045981, (0 missing)  
##   
## Node number 27: 43 observations, complexity param=0.002185197  
## mean=1.161383, MSE=0.2081309   
## left son=54 (16 obs) right son=55 (27 obs)  
## Primary splits:  
## comorb < 1.131493 to the right, improve=0.19215880, (0 missing)  
## dur < 0.8226339 to the right, improve=0.12748010, (0 missing)  
## age < 0.8560297 to the right, improve=0.10568630, (0 missing)  
## intvn < 1.21422 to the right, improve=0.06893980, (0 missing)  
## ervis < 0.4075389 to the left, improve=0.02044871, (0 missing)  
##   
## Node number 28: 38 observations, complexity param=0.001374182  
## mean=1.518491, MSE=0.2236022   
## left son=56 (16 obs) right son=57 (22 obs)  
## Primary splits:  
## age < -0.03231719 to the right, improve=0.13729850, (0 missing)  
## comorb < 0.1232572 to the left, improve=0.05227259, (0 missing)  
## ervis < 0.7866896 to the left, improve=0.04656738, (0 missing)  
## X < 628.5 to the left, improve=0.04380298, (0 missing)  
## intvn < 2.107929 to the left, improve=0.03738194, (0 missing)  
##   
## Node number 29: 7 observations  
## mean=1.929047, MSE=0.04073398   
##   
## Node number 32: 6 observations  
## mean=-1.96656, MSE=0.8252277   
##   
## Node number 33: 14 observations  
## mean=-1.458897, MSE=0.686531   
##   
## Node number 34: 12 observations  
## mean=-1.501704, MSE=0.232573   
##   
## Node number 35: 48 observations, complexity param=0.0008325706  
## mean=-1.282719, MSE=0.2393002   
## left son=70 (35 obs) right son=71 (13 obs)  
## Primary splits:  
## X < 496 to the left, improve=0.07400948, (0 missing)  
## age < 0.1157406 to the right, improve=0.03296519, (0 missing)  
## ervis < 0.02838819 to the left, improve=0.02996235, (0 missing)  
## drugs < 0.05009908 to the left, improve=0.02408183, (0 missing)  
## intvn < -0.7519406 to the left, improve=0.01192502, (0 missing)  
##   
## Node number 40: 154 observations, complexity param=0.004685909  
## mean=-0.742843, MSE=0.3612642   
## left son=80 (75 obs) right son=81 (79 obs)  
## Primary splits:  
## comorb < -0.5488999 to the left, improve=0.05660563, (0 missing)  
## intvn < -0.5731988 to the left, improve=0.04917607, (0 missing)  
## X < 84 to the right, improve=0.04221382, (0 missing)  
## dur < 1.050065 to the left, improve=0.02818956, (0 missing)  
## ervis < 0.4075389 to the right, improve=0.01774486, (0 missing)  
##   
## Node number 41: 17 observations, complexity param=0.001593862  
## mean=-0.2143818, MSE=0.6573995   
## left son=82 (7 obs) right son=83 (10 obs)  
## Primary splits:  
## age < 0.1157406 to the right, improve=0.11223990, (0 missing)  
## dur < -0.07468376 to the left, improve=0.07571241, (0 missing)  
## X < 259.5 to the right, improve=0.06922836, (0 missing)  
## ervis < 1.544991 to the right, improve=0.04754902, (0 missing)  
## drugs < 0.05009908 to the right, improve=0.03055514, (0 missing)  
##   
## Node number 42: 77 observations, complexity param=0.001202671  
## mean=-0.2156013, MSE=0.1054947   
## left son=84 (40 obs) right son=85 (37 obs)  
## Primary splits:  
## comorb < 0.4593358 to the left, improve=0.11651990, (0 missing)  
## ervis < 0.4075389 to the left, improve=0.11160690, (0 missing)  
## dur < -0.2276826 to the left, improve=0.06680239, (0 missing)  
## X < 472.5 to the right, improve=0.06022436, (0 missing)  
## age < 0.2637984 to the right, improve=0.02358988, (0 missing)  
##   
## Node number 43: 21 observations, complexity param=0.0003605072  
## mean=0.1786868, MSE=0.07745963   
## left son=86 (12 obs) right son=87 (9 obs)  
## Primary splits:  
## intvn < -0.7519406 to the left, improve=0.1744191, (0 missing)  
## comorb < 2.811886 to the left, improve=0.1620935, (0 missing)  
## age < 0.4118563 to the left, improve=0.1566595, (0 missing)  
## ervis < -0.3507625 to the left, improve=0.1046579, (0 missing)  
## X < 581 to the left, improve=0.1038960, (0 missing)  
##   
## Node number 44: 39 observations, complexity param=0.002019943  
## mean=-0.4355141, MSE=0.3048099   
## left son=88 (13 obs) right son=89 (26 obs)  
## Primary splits:  
## age < 0.7079719 to the right, improve=0.13372730, (0 missing)  
## ervis < -0.7299132 to the right, improve=0.11413300, (0 missing)  
## X < 579 to the left, improve=0.07164840, (0 missing)  
## dur < 0.272665 to the right, improve=0.05499074, (0 missing)  
## drugs < 0.05009908 to the right, improve=0.04125141, (0 missing)  
##   
## Node number 45: 21 observations, complexity param=0.001034072  
## mean=0.1564266, MSE=0.2709294   
## left son=90 (5 obs) right son=91 (16 obs)  
## Primary splits:  
## gend splits as RL, improve=0.14303750, (0 missing)  
## ervis < 0.02838819 to the left, improve=0.13015920, (0 missing)  
## X < 266 to the right, improve=0.09795595, (0 missing)  
## dur < -0.6660037 to the left, improve=0.02535338, (0 missing)  
## drugs < 0.05009908 to the right, improve=0.01466328, (0 missing)  
##   
## Node number 46: 66 observations, complexity param=0.00239211  
## mean=0.2058892, MSE=0.2484325   
## left son=92 (39 obs) right son=93 (27 obs)  
## Primary splits:  
## comorb < 0.2912965 to the left, improve=0.11481630, (0 missing)  
## X < 66.5 to the right, improve=0.07728612, (0 missing)  
## ervis < 0.02838819 to the left, improve=0.04453639, (0 missing)  
## intvn < -0.2157151 to the left, improve=0.03452760, (0 missing)  
## age < 0.1157406 to the left, improve=0.03088927, (0 missing)  
##   
## Node number 47: 5 observations  
## mean=0.9948719, MSE=0.04570561   
##   
## Node number 48: 87 observations, complexity param=0.003140835  
## mean=0.3214756, MSE=0.3140433   
## left son=96 (41 obs) right son=97 (46 obs)  
## Primary splits:  
## ervis < -0.3507625 to the left, improve=0.07399046, (0 missing)  
## dur < 0.9714977 to the right, improve=0.04592374, (0 missing)  
## age < 1.300203 to the left, improve=0.04057243, (0 missing)  
## X < 539 to the right, improve=0.03506547, (0 missing)  
## drugs < 0.05009908 to the left, improve=0.03019103, (0 missing)  
##   
## Node number 49: 9 observations  
## mean=0.9497526, MSE=0.3246076   
##   
## Node number 50: 29 observations, complexity param=0.001100536  
## mean=0.7396078, MSE=0.292288   
## left son=100 (13 obs) right son=101 (16 obs)  
## Primary splits:  
## dur < -0.2318177 to the right, improve=0.09456307, (0 missing)  
## X < 447 to the left, improve=0.07842554, (0 missing)  
## age < -0.03231719 to the left, improve=0.04664487, (0 missing)  
## ervis < 0.02838819 to the left, improve=0.04512998, (0 missing)  
## comorb < -0.5488999 to the left, improve=0.02787724, (0 missing)  
##   
## Node number 51: 5 observations  
## mean=1.301595, MSE=0.4597301   
##   
## Node number 52: 38 observations, complexity param=0.0005415793  
## mean=0.6217421, MSE=0.1853497   
## left son=104 (32 obs) right son=105 (6 obs)  
## Primary splits:  
## X < 672.5 to the left, improve=0.05435871, (0 missing)  
## age < -1.068722 to the right, improve=0.04400341, (0 missing)  
## ervis < 1.16584 to the right, improve=0.04115982, (0 missing)  
## dur < 0.946687 to the left, improve=0.03659770, (0 missing)  
## comorb < -0.04478204 to the right, improve=0.02155456, (0 missing)  
##   
## Node number 53: 25 observations, complexity param=0.001169903  
## mean=0.8415117, MSE=0.2130866   
## left son=106 (5 obs) right son=107 (20 obs)  
## Primary splits:  
## X < 160 to the left, improve=0.17230680, (0 missing)  
## ervis < 0.4075389 to the left, improve=0.16451140, (0 missing)  
## intvn < 0.1417686 to the left, improve=0.08331400, (0 missing)  
## drugs < 0.05009908 to the left, improve=0.06418527, (0 missing)  
## comorb < 0.4593358 to the right, improve=0.05714859, (0 missing)  
##   
## Node number 54: 16 observations, complexity param=0.0003468544  
## mean=0.9015948, MSE=0.1440944   
## left son=108 (5 obs) right son=109 (11 obs)  
## Primary splits:  
## X < 256 to the left, improve=0.11840080, (0 missing)  
## comorb < 2.811886 to the left, improve=0.09034923, (0 missing)  
## intvn < 1.035478 to the right, improve=0.05590831, (0 missing)  
## age < 1.004088 to the right, improve=0.03145365, (0 missing)  
## dur < 1.294036 to the left, improve=0.01521802, (0 missing)  
##   
## Node number 55: 27 observations, complexity param=0.001451875  
## mean=1.315332, MSE=0.182384   
## left son=110 (22 obs) right son=111 (5 obs)  
## Primary splits:  
## intvn < 0.6779941 to the right, improve=0.23203490, (0 missing)  
## dur < 1.269225 to the right, improve=0.10848350, (0 missing)  
## age < 0.8560297 to the right, improve=0.08583497, (0 missing)  
## X < 230 to the right, improve=0.08369517, (0 missing)  
## gend splits as RL, improve=0.07471168, (0 missing)  
##   
## Node number 56: 16 observations, complexity param=0.0004483399  
## mean=1.313033, MSE=0.2495448   
## left son=112 (5 obs) right son=113 (11 obs)  
## Primary splits:  
## X < 467 to the right, improve=0.08837178, (0 missing)  
## dur < 0.8019584 to the right, improve=0.05701744, (0 missing)  
## ervis < 0.02838819 to the right, improve=0.05490333, (0 missing)  
## intvn < 2.286671 to the left, improve=0.04005110, (0 missing)  
## comorb < 0.1232572 to the left, improve=0.02891695, (0 missing)  
##   
## Node number 57: 22 observations, complexity param=0.0004717407  
## mean=1.667914, MSE=0.1517071   
## left son=114 (11 obs) right son=115 (11 obs)  
## Primary splits:  
## intvn < 2.107929 to the left, improve=0.11123700, (0 missing)  
## dur < 0.194098 to the left, improve=0.10343920, (0 missing)  
## ervis < 0.7866896 to the left, improve=0.09097616, (0 missing)  
## gend splits as RL, improve=0.08430905, (0 missing)  
## X < 332.5 to the right, improve=0.07350113, (0 missing)  
##   
## Node number 70: 35 observations, complexity param=0.0002906009  
## mean=-1.363825, MSE=0.1490721   
## left son=140 (29 obs) right son=141 (6 obs)  
## Primary splits:  
## age < 1.004088 to the left, improve=0.043833620, (0 missing)  
## intvn < -0.7519406 to the left, improve=0.041505760, (0 missing)  
## X < 422.5 to the right, improve=0.018340160, (0 missing)  
## comorb < -0.5488999 to the right, improve=0.013175120, (0 missing)  
## ervis < -0.7299132 to the left, improve=0.001690954, (0 missing)  
##   
## Node number 71: 13 observations  
## mean=-1.064357, MSE=0.4168296   
##   
## Node number 80: 75 observations, complexity param=0.004685909  
## mean=-0.8896089, MSE=0.4504312   
## left son=160 (7 obs) right son=161 (68 obs)  
## Primary splits:  
## ervis < 0.4075389 to the right, improve=0.12510630, (0 missing)  
## intvn < -0.5731988 to the left, improve=0.10673890, (0 missing)  
## X < 84 to the right, improve=0.10671580, (0 missing)  
## age < -1.364838 to the right, improve=0.04757302, (0 missing)  
## dur < 1.004578 to the left, improve=0.03090843, (0 missing)  
##   
## Node number 81: 79 observations, complexity param=0.001368467  
## mean=-0.6035083, MSE=0.2367482   
## left son=162 (27 obs) right son=163 (52 obs)  
## Primary splits:  
## intvn < -0.7519406 to the left, improve=0.05758313, (0 missing)  
## comorb < -0.3808606 to the left, improve=0.02918951, (0 missing)  
## drugs < 0.05009908 to the right, improve=0.02637457, (0 missing)  
## ervis < 0.4075389 to the left, improve=0.02383657, (0 missing)  
## dur < 1.050065 to the left, improve=0.01701144, (0 missing)  
##   
## Node number 82: 7 observations  
## mean=-0.5390496, MSE=0.673527   
##   
## Node number 83: 10 observations  
## mean=0.01288572, MSE=0.5206734   
##   
## Node number 84: 40 observations, complexity param=0.0009135854  
## mean=-0.322233, MSE=0.1264226   
## left son=168 (33 obs) right son=169 (7 obs)  
## Primary splits:  
## ervis < 0.02838819 to the left, improve=0.14218020, (0 missing)  
## X < 453 to the right, improve=0.08627539, (0 missing)  
## dur < -0.2276826 to the left, improve=0.07154348, (0 missing)  
## gend splits as LR, improve=0.07108591, (0 missing)  
## age < -0.03231719 to the right, improve=0.03656625, (0 missing)  
##   
## Node number 85: 37 observations, complexity param=0.0002543666  
## mean=-0.1003238, MSE=0.05728872   
## left son=170 (16 obs) right son=171 (21 obs)  
## Primary splits:  
## dur < 1.103821 to the right, improve=0.09360318, (0 missing)  
## intvn < -0.5731988 to the right, improve=0.06956834, (0 missing)  
## age < 0.7079719 to the right, improve=0.05956940, (0 missing)  
## X < 369 to the right, improve=0.05833095, (0 missing)  
## comorb < 0.7954144 to the right, improve=0.01834478, (0 missing)  
##   
## Node number 86: 12 observations  
## mean=0.07802482, MSE=0.07262581   
##   
## Node number 87: 9 observations  
## mean=0.3129027, MSE=0.05238037   
##   
## Node number 88: 13 observations  
## mean=-0.7210361, MSE=0.242639   
##   
## Node number 89: 26 observations, complexity param=0.00137445  
## mean=-0.2927531, MSE=0.2747532   
## left son=178 (16 obs) right son=179 (10 obs)  
## Primary splits:  
## intvn < -0.2157151 to the left, improve=0.15142160, (0 missing)  
## ervis < -0.7299132 to the right, improve=0.12591760, (0 missing)  
## drugs < 0.05009908 to the right, improve=0.10777980, (0 missing)  
## X < 235.5 to the left, improve=0.05835144, (0 missing)  
## age < 0.4118563 to the left, improve=0.05222621, (0 missing)  
##   
## Node number 90: 5 observations  
## mean=-0.1957237, MSE=0.09622386   
##   
## Node number 91: 16 observations, complexity param=0.0007582752  
## mean=0.2664736, MSE=0.2746615   
## left son=182 (9 obs) right son=183 (7 obs)  
## Primary splits:  
## ervis < 0.02838819 to the left, improve=0.13579500, (0 missing)  
## age < -1.364838 to the left, improve=0.06760745, (0 missing)  
## X < 299 to the right, improve=0.04721696, (0 missing)  
## drugs < 0.05009908 to the right, improve=0.04353354, (0 missing)  
## dur < -0.5088697 to the right, improve=0.01657195, (0 missing)  
##   
## Node number 92: 39 observations, complexity param=0.001728067  
## mean=0.06536358, MSE=0.2654394   
## left son=184 (29 obs) right son=185 (10 obs)  
## Primary splits:  
## dur < -0.4964644 to the right, improve=0.13137280, (0 missing)  
## X < 328.5 to the right, improve=0.08219991, (0 missing)  
## age < 0.5599141 to the right, improve=0.06824538, (0 missing)  
## comorb < -0.3808606 to the left, improve=0.02799128, (0 missing)  
## drugs < 0.05009908 to the right, improve=0.01759428, (0 missing)  
##   
## Node number 93: 27 observations, complexity param=0.001918794  
## mean=0.4088707, MSE=0.1541413   
## left son=186 (17 obs) right son=187 (10 obs)  
## Primary splits:  
## ervis < 0.02838819 to the left, improve=0.36284430, (0 missing)  
## age < 0.4118563 to the left, improve=0.07565637, (0 missing)  
## dur < 1.455305 to the right, improve=0.07312091, (0 missing)  
## comorb < 1.971689 to the left, improve=0.06339109, (0 missing)  
## X < 171 to the left, improve=0.03154854, (0 missing)  
##   
## Node number 96: 41 observations, complexity param=0.001189127  
## mean=0.1600138, MSE=0.2261795   
## left son=192 (34 obs) right son=193 (7 obs)  
## Primary splits:  
## ervis < -1.109064 to the right, improve=0.09907769, (0 missing)  
## age < 0.8560297 to the right, improve=0.09741340, (0 missing)  
## dur < 0.4339341 to the left, improve=0.06929544, (0 missing)  
## X < 682.5 to the left, improve=0.03832117, (0 missing)  
## comorb < -0.3808606 to the right, improve=0.01129648, (0 missing)  
##   
## Node number 97: 46 observations, complexity param=0.003140835  
## mean=0.4653871, MSE=0.3484099   
## left son=194 (18 obs) right son=195 (28 obs)  
## Primary splits:  
## dur < 0.115531 to the right, improve=0.18232670, (0 missing)  
## ervis < 1.924142 to the right, improve=0.11887350, (0 missing)  
## X < 612 to the right, improve=0.11615000, (0 missing)  
## age < -0.4764906 to the right, improve=0.03468014, (0 missing)  
## drugs < 0.05009908 to the left, improve=0.02667370, (0 missing)  
##   
## Node number 100: 13 observations  
## mean=0.555168, MSE=0.2953872   
##   
## Node number 101: 16 observations, complexity param=0.001100536  
## mean=0.8894651, MSE=0.239673   
## left son=202 (10 obs) right son=203 (6 obs)  
## Primary splits:  
## ervis < -0.7299132 to the left, improve=0.24269880, (0 missing)  
## dur < -0.9885418 to the left, improve=0.15550700, (0 missing)  
## age < -0.03231719 to the left, improve=0.12326820, (0 missing)  
## drugs < 0.05009908 to the left, improve=0.06255544, (0 missing)  
## X < 260 to the left, improve=0.04376938, (0 missing)  
##   
## Node number 104: 32 observations, complexity param=0.0005415793  
## mean=0.5782779, MSE=0.1903667   
## left son=208 (14 obs) right son=209 (18 obs)  
## Primary splits:  
## dur < 0.6489595 to the left, improve=0.07708507, (0 missing)  
## age < -1.068722 to the right, improve=0.06921209, (0 missing)  
## intvn < 0.3205104 to the left, improve=0.06653473, (0 missing)  
## X < 459.5 to the right, improve=0.06273523, (0 missing)  
## gend splits as LR, improve=0.04047772, (0 missing)  
##   
## Node number 105: 6 observations  
## mean=0.8535508, MSE=0.09478151   
##   
## Node number 106: 5 observations  
## mean=0.4582819, MSE=0.05979883   
##   
## Node number 107: 20 observations, complexity param=0.001169903  
## mean=0.9373191, MSE=0.2055132   
## left son=214 (14 obs) right son=215 (6 obs)  
## Primary splits:  
## X < 332 to the right, improve=0.27149620, (0 missing)  
## dur < 0.2850703 to the left, improve=0.15245640, (0 missing)  
## ervis < 0.4075389 to the left, improve=0.06899738, (0 missing)  
## comorb < 0.4593358 to the right, improve=0.05221294, (0 missing)  
## age < 1.152145 to the right, improve=0.01866991, (0 missing)  
##   
## Node number 108: 5 observations  
## mean=0.7078579, MSE=0.1191009   
##   
## Node number 109: 11 observations  
## mean=0.989657, MSE=0.1306393   
##   
## Node number 110: 22 observations, complexity param=0.0009153973  
## mean=1.217261, MSE=0.1641226   
## left son=220 (8 obs) right son=221 (14 obs)  
## Primary splits:  
## age < 0.4118563 to the right, improve=0.19952310, (0 missing)  
## gend splits as RL, improve=0.10905120, (0 missing)  
## X < 545.5 to the right, improve=0.09979841, (0 missing)  
## dur < -0.3517358 to the right, improve=0.05222840, (0 missing)  
## ervis < 0.7866896 to the right, improve=0.04520298, (0 missing)  
##   
## Node number 111: 5 observations  
## mean=1.746848, MSE=0.03420911   
##   
## Node number 112: 5 observations  
## mean=1.09277, MSE=0.3877943   
##   
## Node number 113: 11 observations  
## mean=1.413153, MSE=0.1546274   
##   
## Node number 114: 11 observations  
## mean=1.538009, MSE=0.2035506   
##   
## Node number 115: 11 observations  
## mean=1.79782, MSE=0.0661127   
##   
## Node number 140: 29 observations, complexity param=0.0002706308  
## mean=-1.400594, MSE=0.169683   
## left son=280 (18 obs) right son=281 (11 obs)  
## Primary splits:  
## intvn < -0.7519406 to the left, improve=0.043282830, (0 missing)  
## age < 0.1157406 to the right, improve=0.037674640, (0 missing)  
## X < 168.5 to the right, improve=0.013645510, (0 missing)  
## comorb < -0.5488999 to the right, improve=0.006327138, (0 missing)  
## ervis < -0.3507625 to the right, improve=0.002756803, (0 missing)  
##   
## Node number 141: 6 observations  
## mean=-1.18611, MSE=0.01133558   
##   
## Node number 160: 7 observations  
## mean=-1.629486, MSE=0.3270255   
##   
## Node number 161: 68 observations, complexity param=0.003991512  
## mean=-0.8134451, MSE=0.400982   
## left son=322 (35 obs) right son=323 (33 obs)  
## Primary splits:  
## intvn < -0.5731988 to the left, improve=0.11520690, (0 missing)  
## X < 84 to the right, improve=0.10550340, (0 missing)  
## age < -1.364838 to the right, improve=0.03793938, (0 missing)  
## dur < 1.004578 to the left, improve=0.02641534, (0 missing)  
## ervis < 0.02838819 to the left, improve=0.02109461, (0 missing)  
##   
## Node number 162: 27 observations, complexity param=0.001312957  
## mean=-0.765544, MSE=0.2803853   
## left son=324 (9 obs) right son=325 (18 obs)  
## Primary splits:  
## dur < -0.5708963 to the left, improve=0.13649170, (0 missing)  
## ervis < 0.02838819 to the left, improve=0.09776928, (0 missing)  
## age < 0.7079719 to the left, improve=0.09283472, (0 missing)  
## X < 540.5 to the left, improve=0.07563882, (0 missing)  
## comorb < -0.3808606 to the left, improve=0.02736853, (0 missing)  
##   
## Node number 163: 52 observations, complexity param=0.0007951798  
## mean=-0.5193743, MSE=0.1933792   
## left son=326 (20 obs) right son=327 (32 obs)  
## Primary splits:  
## comorb < -0.3808606 to the left, improve=0.04181459, (0 missing)  
## dur < 0.6448244 to the left, improve=0.04107682, (0 missing)  
## age < -0.180375 to the right, improve=0.04004286, (0 missing)  
## X < 144.5 to the left, improve=0.02530745, (0 missing)  
## drugs < 0.05009908 to the right, improve=0.02027467, (0 missing)  
##   
## Node number 168: 33 observations, complexity param=0.0002788063  
## mean=-0.3839811, MSE=0.05606277   
## left son=336 (22 obs) right son=337 (11 obs)  
## Primary splits:  
## dur < 0.9714977 to the left, improve=0.11860110, (0 missing)  
## comorb < 0.2912965 to the right, improve=0.09280048, (0 missing)  
## intvn < -0.7519406 to the left, improve=0.08430351, (0 missing)  
## age < -0.6245484 to the left, improve=0.08375077, (0 missing)  
## X < 598 to the right, improve=0.04967185, (0 missing)  
##   
## Node number 169: 7 observations  
## mean=-0.0311345, MSE=0.3554058   
##   
## Node number 170: 16 observations, complexity param=0.0002543666  
## mean=-0.1842175, MSE=0.03494047   
## left son=340 (9 obs) right son=341 (7 obs)  
## Primary splits:  
## age < -0.03231719 to the right, improve=0.361264600, (0 missing)  
## dur < 1.28163 to the left, improve=0.134869400, (0 missing)  
## X < 540.5 to the left, improve=0.058258510, (0 missing)  
## intvn < -0.5731988 to the right, improve=0.049449400, (0 missing)  
## comorb < 0.7954144 to the right, improve=0.008783127, (0 missing)  
##   
## Node number 171: 21 observations, complexity param=0.0001389504  
## mean=-0.03640476, MSE=0.06486791   
## left son=342 (14 obs) right son=343 (7 obs)  
## Primary splits:  
## comorb < 0.6273751 to the right, improve=0.08027590, (0 missing)  
## X < 369 to the right, improve=0.07685477, (0 missing)  
## dur < 0.2809352 to the left, improve=0.07326069, (0 missing)  
## ervis < -0.3507625 to the right, improve=0.01700568, (0 missing)  
## intvn < -0.7519406 to the left, improve=0.01069074, (0 missing)  
##   
## Node number 178: 16 observations, complexity param=0.0008412472  
## mean=-0.4540051, MSE=0.2903912   
## left son=356 (10 obs) right son=357 (6 obs)  
## Primary splits:  
## age < 0.2637984 to the left, improve=0.14249350, (0 missing)  
## X < 576.5 to the left, improve=0.11077350, (0 missing)  
## dur < -0.6866792 to the right, improve=0.03943537, (0 missing)  
## ervis < 0.2179635 to the right, improve=0.03368417, (0 missing)  
##   
## Node number 179: 10 observations  
## mean=-0.0347498, MSE=0.1415631   
##   
## Node number 182: 9 observations  
## mean=0.09615243, MSE=0.2326087   
##   
## Node number 183: 7 observations  
## mean=0.4854579, MSE=0.2434776   
##   
## Node number 184: 29 observations, complexity param=0.0007557695  
## mean=-0.04429343, MSE=0.1669351   
## left son=368 (17 obs) right son=369 (12 obs)  
## Primary splits:  
## intvn < -0.2157151 to the left, improve=0.12286220, (0 missing)  
## comorb < -0.3808606 to the left, improve=0.10830150, (0 missing)  
## dur < 0.2602597 to the left, improve=0.08909159, (0 missing)  
## age < 0.4858852 to the right, improve=0.04147986, (0 missing)  
## ervis < 0.4075389 to the left, improve=0.03593895, (0 missing)  
##   
## Node number 185: 10 observations  
## mean=0.3833689, MSE=0.4151029   
##   
## Node number 186: 17 observations, complexity param=0.000201378  
## mean=0.2274883, MSE=0.06970879   
## left son=372 (8 obs) right son=373 (9 obs)  
## Primary splits:  
## X < 199.5 to the left, improve=0.13373660, (0 missing)  
## ervis < -0.3507625 to the right, improve=0.11718990, (0 missing)  
## dur < 1.24855 to the right, improve=0.07143137, (0 missing)  
## comorb < 0.9634536 to the right, improve=0.06828797, (0 missing)  
## age < 0.633943 to the right, improve=0.05964198, (0 missing)  
##   
## Node number 187: 10 observations  
## mean=0.7172208, MSE=0.1466675   
##   
## Node number 192: 34 observations, complexity param=0.001189127  
## mean=0.09208969, MSE=0.1940952   
## left son=384 (12 obs) right son=385 (22 obs)  
## Primary splits:  
## ervis < -0.7299132 to the left, improve=0.14439590, (0 missing)  
## dur < 0.4339341 to the left, improve=0.06743822, (0 missing)  
## X < 76.5 to the left, improve=0.04214739, (0 missing)  
## age < -0.6245484 to the left, improve=0.03346771, (0 missing)  
## intvn < 0.1417686 to the left, improve=0.01853122, (0 missing)  
##   
## Node number 193: 7 observations  
## mean=0.489931, MSE=0.2507628   
##   
## Node number 194: 18 observations, complexity param=0.001690814  
## mean=0.1510373, MSE=0.2937364   
## left son=388 (6 obs) right son=389 (12 obs)  
## Primary splits:  
## X < 620 to the right, improve=0.25167520, (0 missing)  
## age < -0.180375 to the left, improve=0.12371510, (0 missing)  
## intvn < 0.3205104 to the left, improve=0.07330974, (0 missing)  
## dur < 0.6158787 to the right, improve=0.06547892, (0 missing)  
## ervis < 1.16584 to the right, improve=0.06466393, (0 missing)  
##   
## Node number 195: 28 observations, complexity param=0.001986805  
## mean=0.6674691, MSE=0.2791956   
## left son=390 (20 obs) right son=391 (8 obs)  
## Primary splits:  
## X < 218 to the right, improve=0.20001540, (0 missing)  
## intvn < 0.1417686 to the right, improve=0.12949880, (0 missing)  
## ervis < 0.7866896 to the right, improve=0.04820343, (0 missing)  
## drugs < 0.05009908 to the left, improve=0.04398700, (0 missing)  
## dur < -0.876894 to the left, improve=0.03914327, (0 missing)  
##   
## Node number 202: 10 observations  
## mean=0.702647, MSE=0.203822   
##   
## Node number 203: 6 observations  
## mean=1.200829, MSE=0.144309   
##   
## Node number 208: 14 observations  
## mean=0.4409203, MSE=0.2612511   
##   
## Node number 209: 18 observations, complexity param=0.0004940284  
## mean=0.6851117, MSE=0.1091465   
## left son=418 (12 obs) right son=419 (6 obs)  
## Primary splits:  
## X < 316.5 to the left, improve=0.19789940, (0 missing)  
## age < 0.1157406 to the right, improve=0.09921979, (0 missing)  
## dur < 1.041794 to the right, improve=0.09102804, (0 missing)  
## comorb < 0.1232572 to the right, improve=0.06168906, (0 missing)  
## gend splits as LR, improve=0.03357727, (0 missing)  
##   
## Node number 214: 14 observations  
## mean=0.7826821, MSE=0.1831875   
##   
## Node number 215: 6 observations  
## mean=1.298139, MSE=0.07161962   
##   
## Node number 220: 8 observations  
## mean=0.977874, MSE=0.08837788   
##   
## Node number 221: 14 observations  
## mean=1.354053, MSE=0.1559469   
##   
## Node number 280: 18 observations, complexity param=0.0002103247  
## mean=-1.467588, MSE=0.1142321   
## left son=560 (5 obs) right son=561 (13 obs)  
## Primary splits:  
## X < 146.5 to the left, improve=0.08050155, (0 missing)  
## age < -0.03231719 to the right, improve=0.07428325, (0 missing)  
## ervis < -0.7299132 to the left, improve=0.01437987, (0 missing)  
##   
## Node number 281: 11 observations  
## mean=-1.290967, MSE=0.2410583   
##   
## Node number 322: 35 observations, complexity param=0.001653175  
## mean=-1.022146, MSE=0.3323596   
## left son=644 (5 obs) right son=645 (30 obs)  
## Primary splits:  
## dur < -1.162216 to the left, improve=0.10790040, (0 missing)  
## X < 697.5 to the right, improve=0.08666097, (0 missing)  
## age < -0.4764906 to the right, improve=0.08663404, (0 missing)  
## ervis < -0.3507625 to the left, improve=0.03008663, (0 missing)  
## gend splits as LR, improve=0.02191865, (0 missing)  
##   
## Node number 323: 33 observations, complexity param=0.001356488  
## mean=-0.5920955, MSE=0.3785719   
## left son=646 (28 obs) right son=647 (5 obs)  
## Primary splits:  
## age < -1.290809 to the right, improve=0.08545322, (0 missing)  
## X < 97.5 to the right, improve=0.07253317, (0 missing)  
## dur < -1.112595 to the right, improve=0.05412884, (0 missing)  
## ervis < -0.3507625 to the left, improve=0.01654424, (0 missing)  
## gend splits as RL, improve=0.01395576, (0 missing)  
##   
## Node number 324: 9 observations  
## mean=-1.042204, MSE=0.0367601   
##   
## Node number 325: 18 observations, complexity param=0.001148266  
## mean=-0.6272142, MSE=0.3447925   
## left son=650 (13 obs) right son=651 (5 obs)  
## Primary splits:  
## ervis < 0.02838819 to the left, improve=0.14560860, (0 missing)  
## X < 540.5 to the left, improve=0.12550060, (0 missing)  
## comorb < -0.3808606 to the left, improve=0.06693848, (0 missing)  
## dur < 0.2850703 to the right, improve=0.06055252, (0 missing)  
## age < 0.7079719 to the left, improve=0.05271006, (0 missing)  
##   
## Node number 326: 20 observations, complexity param=0.00065863  
## mean=-0.6331184, MSE=0.1391463   
## left son=652 (9 obs) right son=653 (11 obs)  
## Primary splits:  
## X < 493 to the right, improve=0.186257800, (0 missing)  
## age < 0.1897695 to the right, improve=0.176115500, (0 missing)  
## dur < -0.7445707 to the right, improve=0.142289000, (0 missing)  
## ervis < -0.3507625 to the right, improve=0.073518900, (0 missing)  
## intvn < -0.5731988 to the right, improve=0.004969151, (0 missing)  
##   
## Node number 327: 32 observations, complexity param=0.0007951798  
## mean=-0.4482843, MSE=0.2141349   
## left son=654 (6 obs) right son=655 (26 obs)  
## Primary splits:  
## X < 144.5 to the left, improve=0.10157680, (0 missing)  
## intvn < -0.5731988 to the left, improve=0.05353456, (0 missing)  
## age < 0.7079719 to the right, improve=0.04496460, (0 missing)  
## dur < 0.7027159 to the left, improve=0.04150807, (0 missing)  
## ervis < -0.3507625 to the left, improve=0.02518043, (0 missing)  
##   
## Node number 336: 22 observations, complexity param=0.0001972181  
## mean=-0.4416401, MSE=0.0578959   
## left son=672 (8 obs) right son=673 (14 obs)  
## Primary splits:  
## comorb < 0.2912965 to the right, improve=0.12185720, (0 missing)  
## age < -0.6245484 to the left, improve=0.07749122, (0 missing)  
## dur < -0.6039771 to the right, improve=0.05257631, (0 missing)  
## intvn < -0.5731988 to the left, improve=0.03168148, (0 missing)  
## X < 465.5 to the left, improve=0.02660603, (0 missing)  
##   
## Node number 337: 11 observations  
## mean=-0.2686632, MSE=0.03244919   
##   
## Node number 340: 9 observations  
## mean=-0.2833018, MSE=0.02125358   
##   
## Node number 341: 7 observations  
## mean=-0.05682344, MSE=0.02368588   
##   
## Node number 342: 14 observations  
## mean=-0.08743088, MSE=0.03787926   
##   
## Node number 343: 7 observations  
## mean=0.06564748, MSE=0.1032232   
##   
## Node number 356: 10 observations  
## mean=-0.611572, MSE=0.288099   
##   
## Node number 357: 6 observations  
## mean=-0.1913937, MSE=0.183868   
##   
## Node number 368: 17 observations, complexity param=0.0002485771  
## mean=-0.1646166, MSE=0.114006   
## left son=736 (11 obs) right son=737 (6 obs)  
## Primary splits:  
## ervis < -0.3507625 to the right, improve=0.10093910, (0 missing)  
## age < 0.1157406 to the right, improve=0.09514170, (0 missing)  
## X < 318 to the right, improve=0.06422201, (0 missing)  
## gend splits as LR, improve=0.05556417, (0 missing)  
## comorb < -0.3808606 to the left, improve=0.04236504, (0 missing)  
##   
## Node number 369: 12 observations  
## mean=0.1261644, MSE=0.1923522   
##   
## Node number 372: 8 observations  
## mean=0.1250776, MSE=0.0543313   
##   
## Node number 373: 9 observations  
## mean=0.31852, MSE=0.06576827   
##   
## Node number 384: 12 observations  
## mean=-0.1345863, MSE=0.1931774   
##   
## Node number 385: 22 observations, complexity param=0.0003755365  
## mean=0.2157311, MSE=0.151282   
## left son=770 (11 obs) right son=771 (11 obs)  
## Primary splits:  
## dur < -0.4840591 to the right, improve=0.08880081, (0 missing)  
## X < 279.5 to the right, improve=0.08548585, (0 missing)  
## intvn < 0.1417686 to the left, improve=0.07477387, (0 missing)  
## comorb < -0.5488999 to the left, improve=0.04664731, (0 missing)  
## age < -0.6985773 to the left, improve=0.02705166, (0 missing)  
##   
## Node number 388: 6 observations  
## mean=-0.2334784, MSE=0.03025144   
##   
## Node number 389: 12 observations  
## mean=0.3432951, MSE=0.3145896   
##   
## Node number 390: 20 observations, complexity param=0.001214566  
## mean=0.5180122, MSE=0.2010572   
## left son=780 (6 obs) right son=781 (14 obs)  
## Primary splits:  
## X < 349.5 to the left, improve=0.2377094, (0 missing)  
## dur < -0.876894 to the left, improve=0.1978829, (0 missing)  
## drugs < 0.9900532 to the right, improve=0.1368225, (0 missing)  
## ervis < 0.7866896 to the right, improve=0.1278937, (0 missing)  
## intvn < 0.1417686 to the right, improve=0.1093300, (0 missing)  
##   
## Node number 391: 8 observations  
## mean=1.041111, MSE=0.2790898   
##   
## Node number 418: 12 observations  
## mean=0.5811886, MSE=0.05397554   
##   
## Node number 419: 6 observations  
## mean=0.8929578, MSE=0.1546883   
##   
## Node number 560: 5 observations  
## mean=-1.622214, MSE=0.07395438   
##   
## Node number 561: 13 observations  
## mean=-1.408116, MSE=0.1169908   
##   
## Node number 644: 5 observations  
## mean=-1.486011, MSE=0.3040113   
##   
## Node number 645: 30 observations, complexity param=0.001653175  
## mean=-0.9448354, MSE=0.2952456   
## left son=1290 (24 obs) right son=1291 (6 obs)  
## Primary splits:  
## ervis < 0.02838819 to the left, improve=0.14227880, (0 missing)  
## X < 477 to the right, improve=0.11686980, (0 missing)  
## age < -0.4764906 to the right, improve=0.10835110, (0 missing)  
## dur < 0.7688775 to the left, improve=0.01947408, (0 missing)  
## intvn < -0.7519406 to the left, improve=0.01794481, (0 missing)  
##   
## Node number 646: 28 observations, complexity param=0.0009521186  
## mean=-0.6681009, MSE=0.4005724   
## left son=1292 (15 obs) right son=1293 (13 obs)  
## Primary splits:  
## ervis < -0.3507625 to the left, improve=0.06562993, (0 missing)  
## X < 204 to the right, improve=0.06191728, (0 missing)  
## dur < 0.5662574 to the left, improve=0.05468035, (0 missing)  
## age < -0.4764906 to the left, improve=0.04802425, (0 missing)  
## gend splits as RL, improve=0.02818121, (0 missing)  
##   
## Node number 647: 5 observations  
## mean=-0.1664654, MSE=0.04185773   
##   
## Node number 650: 13 observations  
## mean=-0.7661729, MSE=0.2936071   
##   
## Node number 651: 5 observations  
## mean=-0.2659216, MSE=0.2971375   
##   
## Node number 652: 9 observations  
## mean=-0.8110972, MSE=0.1171051   
##   
## Node number 653: 11 observations  
## mean=-0.4874994, MSE=0.1100581   
##   
## Node number 654: 6 observations  
## mean=-0.7552939, MSE=0.3628807   
##   
## Node number 655: 26 observations, complexity param=0.0007951798  
## mean=-0.3774359, MSE=0.1530383   
## left son=1310 (10 obs) right son=1311 (16 obs)  
## Primary splits:  
## age < 0.7079719 to the right, improve=0.19123110, (0 missing)  
## X < 288.5 to the right, improve=0.11622730, (0 missing)  
## dur < -0.8024621 to the left, improve=0.08812371, (0 missing)  
## intvn < -0.5731988 to the left, improve=0.07110992, (0 missing)  
## ervis < -0.3507625 to the left, improve=0.01443268, (0 missing)  
##   
## Node number 672: 8 observations  
## mean=-0.552754, MSE=0.02526336   
##   
## Node number 673: 14 observations  
## mean=-0.3781464, MSE=0.06545659   
##   
## Node number 736: 11 observations  
## mean=-0.2438435, MSE=0.1418592   
##   
## Node number 737: 6 observations  
## mean=-0.01936723, MSE=0.03033672   
##   
## Node number 770: 11 observations  
## mean=0.09982617, MSE=0.2099931   
##   
## Node number 771: 11 observations  
## mean=0.3316361, MSE=0.06570294   
##   
## Node number 780: 6 observations  
## mean=0.18407, MSE=0.150911   
##   
## Node number 781: 14 observations  
## mean=0.6611303, MSE=0.1542724   
##   
## Node number 1290: 24 observations, complexity param=0.001653175  
## mean=-1.047314, MSE=0.2395569   
## left son=2580 (16 obs) right son=2581 (8 obs)  
## Primary splits:  
## age < -0.4764906 to the right, improve=0.24137810, (0 missing)  
## X < 498 to the right, improve=0.14073170, (0 missing)  
## dur < -0.6825441 to the left, improve=0.08028015, (0 missing)  
## ervis < -0.7299132 to the right, improve=0.02643623, (0 missing)  
## gend splits as LR, improve=0.00522808, (0 missing)  
##   
## Node number 1291: 6 observations  
## mean=-0.5349223, MSE=0.3079646   
##   
## Node number 1292: 15 observations, complexity param=0.0009521186  
## mean=-0.8190455, MSE=0.2388639   
## left son=2584 (8 obs) right son=2585 (7 obs)  
## Primary splits:  
## dur < -0.4096272 to the left, improve=0.21282070, (0 missing)  
## age < -0.5505195 to the left, improve=0.06477352, (0 missing)  
## X < 375.5 to the left, improve=0.05065016, (0 missing)  
## ervis < -0.7299132 to the right, improve=0.02239685, (0 missing)  
##   
## Node number 1293: 13 observations  
## mean=-0.4939341, MSE=0.5305355   
##   
## Node number 1310: 10 observations  
## mean=-0.593827, MSE=0.1754965   
##   
## Node number 1311: 16 observations, complexity param=0.0005835501  
## mean=-0.2421915, MSE=0.09144514   
## left son=2622 (10 obs) right son=2623 (6 obs)  
## Primary splits:  
## age < 0.2637984 to the left, improve=0.31388620, (0 missing)  
## ervis < -0.3507625 to the right, improve=0.13055160, (0 missing)  
## dur < -0.5667612 to the right, improve=0.07526464, (0 missing)  
## X < 404 to the right, improve=0.04887992, (0 missing)  
## intvn < -0.5731988 to the left, improve=0.02278307, (0 missing)  
##   
## Node number 2580: 16 observations, complexity param=0.0008951054  
## mean=-1.217349, MSE=0.1809823   
## left son=5160 (7 obs) right son=5161 (9 obs)  
## Primary splits:  
## X < 498 to the right, improve=0.243272400, (0 missing)  
## ervis < -0.3507625 to the right, improve=0.171675600, (0 missing)  
## age < 0.2637984 to the left, improve=0.068778840, (0 missing)  
## dur < 0.2354491 to the right, improve=0.028218950, (0 missing)  
## gend splits as LR, improve=0.002970729, (0 missing)  
##   
## Node number 2581: 8 observations  
## mean=-0.7072437, MSE=0.1832347   
##   
## Node number 2584: 8 observations  
## mean=-1.02995, MSE=0.2783213   
##   
## Node number 2585: 7 observations  
## mean=-0.5780116, MSE=0.08483716   
##   
## Node number 2622: 10 observations  
## mean=-0.3734242, MSE=0.0444054   
##   
## Node number 2623: 6 observations  
## mean=-0.02347035, MSE=0.09330239   
##   
## Node number 5160: 7 observations  
## mean=-1.455272, MSE=0.1378365   
##   
## Node number 5161: 9 observations  
## mean=-1.032297, MSE=0.1362682

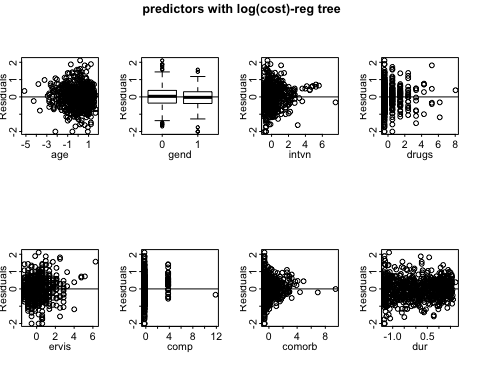
## (c)The most influencing variable on the cost and the effect.

## Answer: From the above result, the intvn has the most influence on the cost, and the effect is the larger intvn, the more of final cost.

## (d)Construct appropriate residual plots to assess whether there remains any linearity not captured by the regression tree model.

## Answer: From the residual plts, there is no nonlinearity not captured by regression tree.

par(mfrow=c(2,4),pin=c(0.8,0.8),tcl=-0.15,mgp=c(1,0.2,0))  
for (i in seq(3:10)) {  
 plot(df\_std[[i+2]],resid(df\_std.tr1),ylab="Residuals",xlab=names(df\_std)[i+2],main="")  
 abline(0, 0)}  
title(main="Ischemic heart disease-standardized \n predictors with log(cost)-reg tree",outer = T)



## (e)Linear reg, nnet,reg tree, which you recommand for this data set and why?

## Answer:

# Prob 4)Forensic example, keep all 6-category to do classification

##Prepare dataset  
FGL<-read.table("../Data\_for\_Lecture\_Examples/fgl.txt",sep="\t")  
FGL1<-FGL  
k<-ncol(FGL1)-1;  
FGL1[1:k]<-sapply(FGL1[1:k], function(x) (x-mean(x))/sd(x))  
FGL1<-data.frame(FGL1,"type\_ind"=as.numeric(factor(FGL1$type)))#add a column of categories with index, instead of strings  
##Or use: as.numeric(factor(FGL1$type, levels=levels(FGL1$type)))

## (a)10-fold CV to find the best nnet for classifying the class type

## Answer: The neural network with the smallest misclassification rate has $= $ and number of hidden nodes as . The misclassification rate is .

##CV function for classification  
CVfunc\_nnet\_clf <- function(data, lam\_seq, num\_hidnode\_seq,Nrep,K,y) {  
 n=nrow(data)  
 n.models = n.lam\*n.num\_hidnode #number of different models to fit  
 yhat=matrix(0,n,n.models)  
   
 ##Each column of mod\_par corresponds to a set of lambda and number of hidden nodes of a trail model  
 mod\_par=matrix(c(rep(lam\_seq,times=1,each=n.num\_hidnode),rep(num\_hidnode\_seq,times=n.lam,each=1)),2,n.models,byrow = T)#Store the model parameters: lambda and the number of nodes in hidden layer  
 MSE<-matrix(0,Nrep,n.models)  
 for (j in 1:Nrep) {  
 print(c(0,0,0,j))#Print out the index of replicates of CV  
 Ind<-CVInd(n,K)  
 for (k in 1:K) {  
 print(k)#Print out the index of different fold of CV  
 for (m in 1:n.models) {  
 out<-nnet(type~.,data[-Ind[[k]],],linout = F, skip=F,size=as.integer(mod\_par[2,m]),decay=mod\_par[1,m],maxit=1000,trace=F)  
 phat<-predict(out,data[Ind[[k]],])  
 yhat[Ind[[k]],m]<-apply(phat,1,function(x) which(x==max(x)))  
 }  
 } #end of k loop  
 MSE[j,]=apply(yhat,2,function(x) sum(y != x)/n)  
 } #end of j loopE  
 MSEAve<- apply(MSE,2,mean); MSEAve #averaged mean square CV error  
 MSEsd <- apply(MSE,2,sd); MSEsd #SD of mean square CV error  
 r2<-1-MSEAve/var(y); r2 #CV r^2  
 ##The best model in terms of the minimum MSEAve or the maximum r2.  
 min(MSEAve)  
 max(r2)  
 ##Return the index of the minimum MSEAve or the maximum r2.  
 which(MSEAve==min(MSEAve))  
 which(r2==max(r2))  
 ##The optimal lambda and number of hidden nodes  
 mod\_par[,which(MSEAve==min(MSEAve))]  
}  
  
##Do a CV on crude interval of lambda and number of hidden nodes again.  
library(nnet)  
ptm <- proc.time()  
Nrep<-2 #number of replicates of CV  
K<-10 #K-fold CV on each replicate  
n.lam = 4 #number of lambda  
n.num\_hidnode = 2 #number of different numbers of hidden nodes  
y<-FGL1$type\_ind  
lam\_seq = 10^seq(-as.integer(n.lam/2),as.integer(n.lam/2)-1)  
num\_hidnode\_seq = 5\*seq(1,n.num\_hidnode)   
  
par\_best\_crude <- CVfunc\_nnet\_clf(FGL1[,c(1:10)], lam\_seq, num\_hidnode\_seq,Nrep,K,y)

## [1] 0 0 0 1  
## [1] 1  
## [1] 2  
## [1] 3  
## [1] 4  
## [1] 5  
## [1] 6  
## [1] 7  
## [1] 8  
## [1] 9  
## [1] 10  
## [1] 0 0 0 2  
## [1] 1  
## [1] 2  
## [1] 3  
## [1] 4  
## [1] 5  
## [1] 6  
## [1] 7  
## [1] 8  
## [1] 9  
## [1] 10

proc.time() - ptm

## user system elapsed   
## 6.538 0.026 6.585

##Do a CV in smaller interval of lambda and number of hidden nodes again.  
ptm <- proc.time()  
Nrep<-2 #number of replicates of CV  
K<-10 #K-fold CV on each replicate  
n.lam = 2 #number of lambda  
n.num\_hidnode = 2 #number of different numbers of hidden nodes  
y<-FGL1$type\_ind  
lam\_seq = c(seq(0.05,0.05,0.01),seq(0.1,0.1,0.1))  
num\_hidnode\_seq = seq(24,26,2)   
  
par\_best <- CVfunc\_nnet\_clf(FGL1[,c(1:10)], lam\_seq, num\_hidnode\_seq,Nrep,K,y)

## [1] 0 0 0 1  
## [1] 1  
## [1] 2  
## [1] 3  
## [1] 4  
## [1] 5  
## [1] 6  
## [1] 7  
## [1] 8  
## [1] 9  
## [1] 10  
## [1] 0 0 0 2  
## [1] 1  
## [1] 2  
## [1] 3  
## [1] 4  
## [1] 5  
## [1] 6  
## [1] 7  
## [1] 8  
## [1] 9  
## [1] 10

proc.time() - ptm

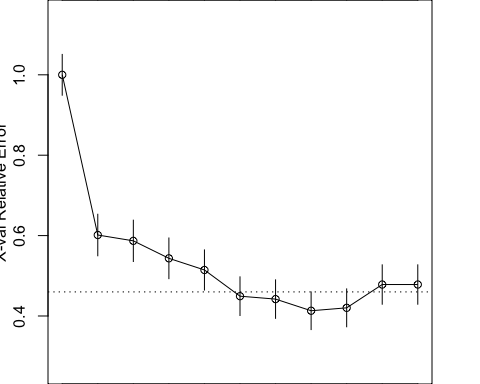
## user system elapsed   
## 27.869 0.099 28.125

##Fit the best nnet model  
out<-nnet(type~.,FGL1[,c(1:10)],linout = F, skip=F,size=as.integer(par\_best[2]),decay=par\_best[1],maxit=1000,trace=F)##type is a factor  
phat<-predict(out,FGL1)  
yhat<-apply(phat,1,function(x) which(x==max(x)))  
e.nnet<-sum(yhat!=y)\*1.0/length(y)

## (b)10-fold CV to find the best tree model for classifying the class type.

## Answer: The classification tree with the smallest misclassification rate has complexity parameter , and the misclassification rate is .

library(rpart)  
control <- rpart.control(minbucket = 1, cp = 0.0001, maxsurrogate = 0, usesurrogate = 0, xval = 10)  
par(mfrow=c(1,1),pin=c(4,4),mgp=c(2,1,0))  
FGL1.tr <- rpart(type ~ .,FGL1[,c(1:10)], method = "class", control = control)  
plotcp(FGL1.tr) #plot of CV r^2 vs. size



printcp(FGL1.tr) #same info is in df\_std.tr$cptable

##   
## Classification tree:  
## rpart(formula = type ~ ., data = FGL1[, c(1:10)], method = "class",   
## control = control)  
##   
## Variables actually used in tree construction:  
## [1] Al Ba Ca Fe K Mg Na RI Si  
##   
## Root node error: 138/214 = 0.64486  
##   
## n= 214   
##   
## CP nsplit rel error xerror xstd  
## 1 0.2065217 0 1.000000 1.00000 0.050729  
## 2 0.0724638 2 0.586957 0.60145 0.051652  
## 3 0.0579710 3 0.514493 0.58696 0.051414  
## 4 0.0362319 4 0.456522 0.54348 0.050577  
## 5 0.0326087 5 0.420290 0.51449 0.049913  
## 6 0.0217391 7 0.355072 0.44928 0.048087  
## 7 0.0144928 8 0.333333 0.44203 0.047855  
## 8 0.0108696 15 0.231884 0.41304 0.046860  
## 9 0.0072464 18 0.195652 0.42029 0.047118  
## 10 0.0036232 38 0.050725 0.47826 0.048957  
## 11 0.0001000 44 0.028986 0.47826 0.048957

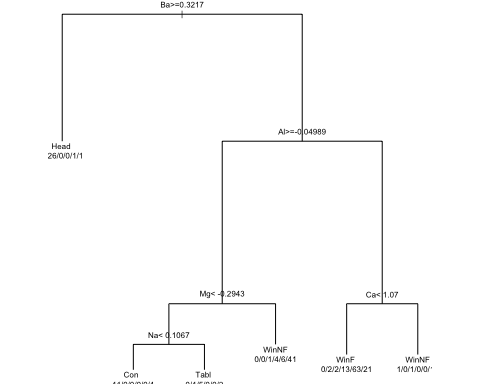
#prune back to optimal size, according to plot of CV 1-r^2  
FGL1.tr1 <- prune(FGL1.tr, cp=0.0326087) #approximately the best size pruned tree  
FGL1.tr1$variable.importance#The importance of each predictors

## Ba Al Mg Ca Na   
## 26.044912 16.085776 11.340598 8.668054 6.116667

FGL1.tr1$cptable[nrow(FGL1.tr1$cptable),] #shows training and CV 1-r^2, and other things

## CP nsplit rel error xerror xstd   
## 0.03260870 5.00000000 0.42028986 0.51449275 0.04991272

# #prune and plot a little smaller tree than the optimal one, just for display  
# FGL1.tr2 <- prune(FGL1.tr, cp=0.0108696) #bigger cp gives smaller size tree  
# FGL1.tr2  
par(cex=.5); plot(FGL1.tr1, uniform=F); text(FGL1.tr1, use.n = T); par(cex=1)



##  
yhat<-apply(predict(FGL1.tr1),1,function(x) which(x==max(x)))  
e.tr<-sum(FGL1$type\_ind!=yhat)/length(yhat)

## (c)Fit multinomial results and discuss it

## Answer: The misclassification rate is .

FGL1.multinom<-multinom(type~.,FGL1[,c(1:10)])

## # weights: 66 (50 variable)  
## initial value 383.436526   
## iter 10 value 177.590797  
## iter 20 value 138.457855  
## iter 30 value 131.091430  
## iter 40 value 126.200258  
## iter 50 value 124.021003  
## iter 60 value 122.318924  
## iter 70 value 121.792280  
## iter 80 value 121.490672  
## iter 90 value 121.385524  
## iter 100 value 121.347733  
## final value 121.347733   
## stopped after 100 iterations

yhat<-predict(FGL1.multinom,FGL1[,c(1:10)])  
e.multi<-sum(FGL1$type!=yhat)/length(yhat)

## (d)Compare the three models from parts (a)-(c).

## Answer: The neural network has the best predictive ability but not very interpretable. Classification tree has very good interpretability, but the predictive ability is not as good as that of neural network. The multinomial regression has the worse predictive ability and the interpretability is better than neural network, but it can only capture the linear relation between predictors and response. For simple predicting purpose, I think neural network is the best for this problem.