

Case-study-hypertension

Hanchao Zhang, M.S.

```
rhc <- read.csv("http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/rhc.csv")  
rhc <- rhc[,-1]
```

Data preparation

Data preparation

Cox Regression

```

data <- rhc[,c("t3d30", "dth30", "swang1", "age", "cat1",
"ca", "sex", "race", "edu", "income", "ninsclas", "resp",
"card", "neuro", "gastr", "renal", "meta", "hema",
"seps", "trauma", "ortho", "temp1", "meanbp1",
"resp1", "hrt1", "pafil", "paco21", "ph1", "wblc1", "hemal",
"sod1", "pot1", "creal", "bili1", "alb1", "scomal",
"aps1", "dnr1", "das2d3pc", "cardiohx", "chfhx",
"dementhx", "chrpulhx", "renalhx", "liverhx", "gibledhx",
"malighx", "immunhx", "transhx", "amihx")
]

dd <- rhc[,c("t3d30", "dth30", "swang1", "age", "cat1", "ca",
"sex", "race", "edu", "income", "ninsclas", "resp",
"card", "neuro", "gastr", "renal", "meta", "hema",
"seps", "trauma", "ortho", "temp1", "meanbp1",
"resp1", "hrt1", "pafil", "paco21", "ph1", "wblc1", "hemal",
"sod1", "pot1", "creal", "bili1", "alb1", "scomal",
"aps1", "dnr1", "das2d3pc", "cardiohx", "chfhx",
"dementhx", "chrpulhx", "renalhx", "liverhx", "gibledhx",
"malighx", "immunhx", "transhx", "amihx")
]

for (i in names(data)) {
  data[,i] <- as.numeric((data[,i]))
  if(mean(data[,i], na.rm = T) < 5){
    data[,i] <- as.factor((dd[,i]))
  }else{
    data[,i] <- ((dd[,i]))
  }
}

data$dth30 <- as.numeric(data$dth30) - 1

index <- sample(nrow(data), 0.6*nrow(data), replace =
T)
train <- data[index,]
valid <- data[-index,]

```

Cox Regression

```
fit3 <- coxph(Surv(t3d30,dth30) ~ swangl + age + sex +
  race + edu + income + ninsclas + resp + card + neuro +
  gastr + renal + meta + hema + seps + trauma + ortho +
  templ + meanbp1 + resp1 + hrt1 + pafil + paco21 + ph1
  + wblcl + hema1 + sod1 + pot1 + creal + bili1 + alb1 +
  scomal + apsl + dnrl + das2d3pc + cardiohx + chfhx + d
  ementhx + chrpulhx + renalhx + liverhx + gibledhx + ma
  lighx + immunhx + transhx + amihx, data = train, metho
  d="breslow")
```

```
## Warning in fitter(X, Y, strats, offset, init, contr
ol,
## weights = weights, : Loglik converged before variab
le
## 36,38,43,88,94,98,102,103,106,109,111,113,114,115,1
16,117,118,119,120,121,122,123,124,125,126,127,128,129
,130,131,132,133,134,135,136,137,138,139,140,141,142,1
43,144,145,146,147,148,149,150,151,152,153,154,155,156
,157,158,159,160,161,162,163,164,165,166,167,168,169,1
70,171,172,173,174,175,176,178,179,180,181,183,184,185
,186,187,188,189,190,191,192,193,194,196,197,199,200,2
02,203,204,205,207,208,209,210,212,213,214,215,216,217
,218,221,222,223,224,225,226,229,230,242,243,246,252,2
53,254,260,301,309,312,320,321,326,327,329,330,333,335
,337,338,339,341,355,356,361,381,383,386,389,391,394,3
98,407,423,433,438,440,451,452,465,475,481,505,514,530
,531,532,533,534,535,536,537,538,539,540,541,542,543,5
44,545,546,547,548,550,551,552,553,554,555,556,557,558
,559,560,561,562,563,564,565,566,567,568,569,570,571,5
73,576,578 ;
## beta may be infinite.
```

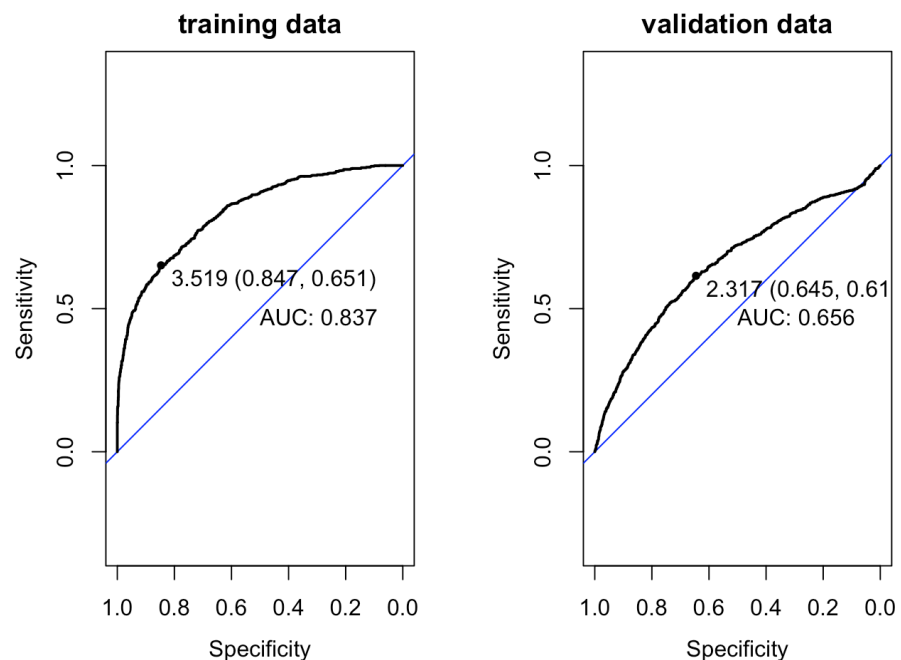
```

pred1.1 <- predict(fit3, newdata = valid, type = "risk")
pred2.2 <- predict(fit3, newdata = train, type = "risk")

r1.1 <- roc(valid$dth30~pred1.1)
r2.1 <- roc(train$dth30~pred2.2)

par(mfrow = c(1,2))
plot(r2.1, print.auc = T, print.thres = T, identity.col="blue", main = "training data")
plot(r1.1, print.auc = T, print.thres = T, identity.col="blue", main = "validation data")

```



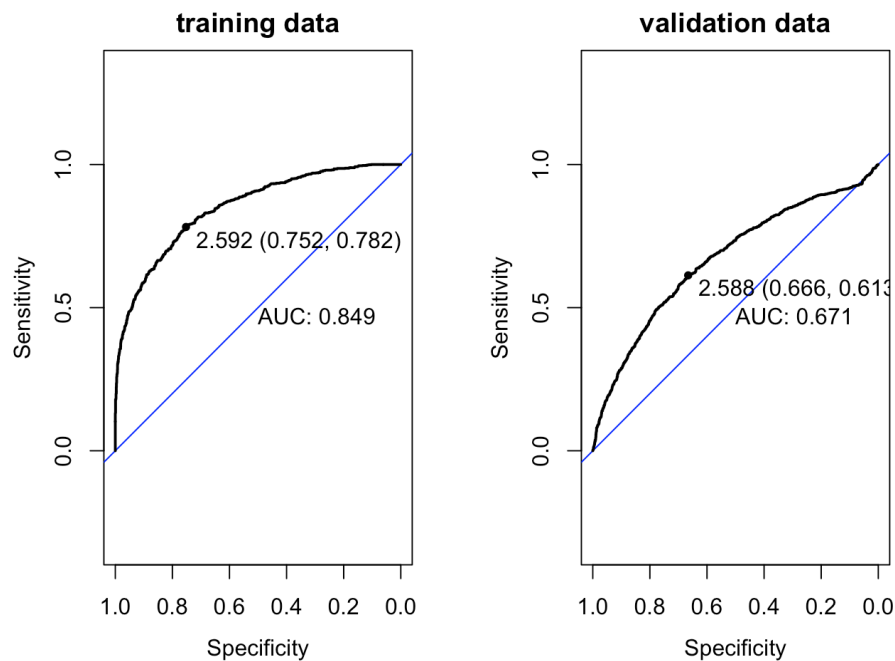
```

fit1 <- coxph(Surv(t3d30,dth30) ~ swang1 + age*cat1 +
age*ca + age + sex + race + edu + income + ninsclas +
resp + card + neuro + gastr + renal + meta + hema + se
ps + trauma + ortho + templ + meanbp1 + resp1 + hrt1 +
pafil + paco21 + ph1 + wblc1 + hema1 + sod1 + pot1 + c
real + bili1 + alb1 + scomal + aps1 + dnr1 + das2d3pc
+ cardiohx + chfhx + dementhx + chrpulhx + renalhx + l
iverhx + gibledhx + malighx + immunhx + transhx + amih
x, data = train, method="breslow")

```

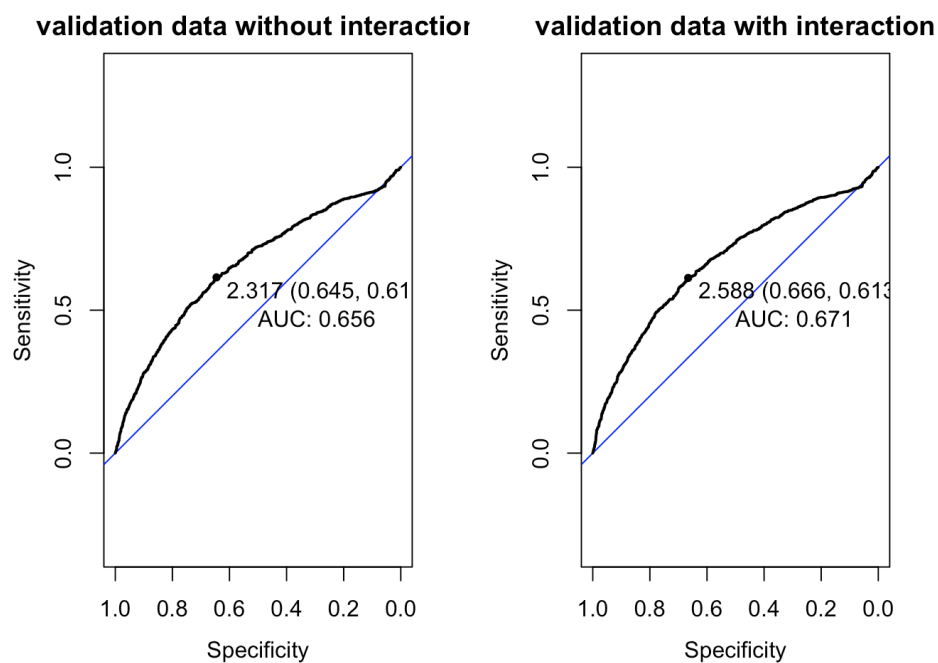
```
## Warning in fitter(X, Y, strats, offset, init, contr  
ol,  
## weights = weights, : Loglik converged before variab  
le  
## 5,46,48,53,98,104,108,112,113,119,121,123,124,125,1  
26,127,128,129,130,131,132,133,134,135,136,137,138,139  
,140,141,142,143,144,145,146,147,148,149,150,151,152,1  
53,154,155,156,157,158,159,160,161,162,163,164,165,166  
,167,168,169,170,171,172,173,174,175,176,177,178,179,1  
80,181,182,183,184,185,186,188,189,190,191,193,194,195  
,196,197,198,199,200,201,202,203,204,206,207,209,210,2  
12,213,214,215,217,218,219,220,222,223,224,225,226,227  
,228,231,232,233,234,235,236,239,240,252,253,256,263,2  
64,269,270,271,311,319,322,330,331,336,337,339,340,343  
,345,347,348,349,351,365,366,371,391,393,396,399,401,4  
04,408,417,433,443,448,450,461,462,475,485,491,515,524  
,540,541,542,543,544,545,546,547,548,549,550,551,552,5  
53,554,555,556,557,558,560,561,562,563,564,565,566,567  
,568,569,570,571,572,573,574,575,576,577,579,580,581,5  
83,586,588 ;  
## beta may be infinite.
```

```
pred1 <- predict(fit1, newdata = valid, type = "risk")  
pred2 <- predict(fit1, newdata = train, type = "risk")  
  
r1 <- roc(valid$dth30~pred1)  
r2 <- roc(train$dth30~pred2)  
  
par(mfrow = c(1,2))  
plot(r2, print.auc = T, print.thres = T, identity.col=  
"blue", main = "training data")  
plot(r1, print.auc = T, print.thres = T, identity.col=  
"blue", main = "validation data")
```



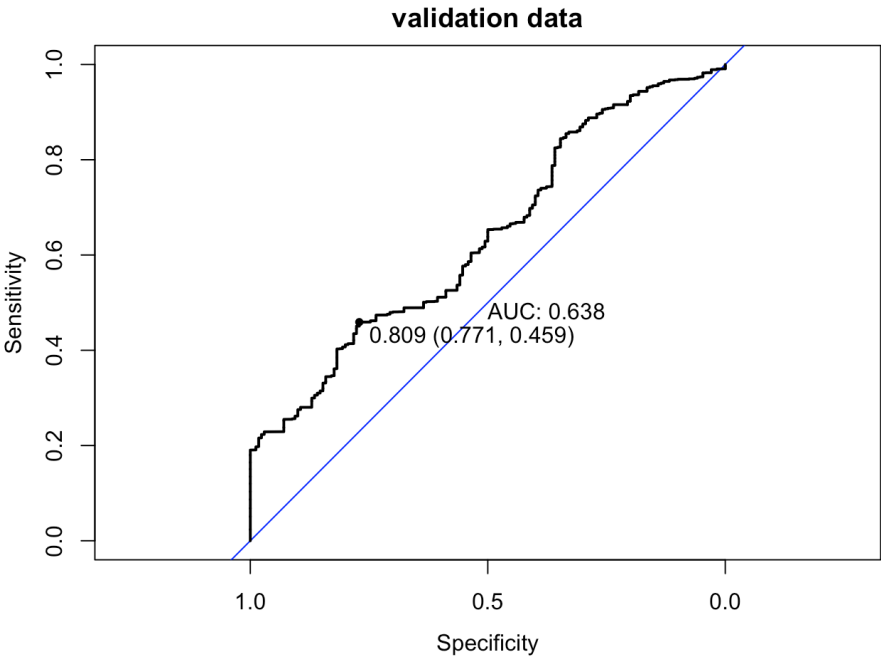
```
par(mfrow = c(1,2))
plot(r1.1, print.auc = T, print.thres = T, identity.col="blue", main = "validation data without interaction"
)

plot(r1, print.auc = T, print.thres = "best", identity.col="blue", main = "validation data with interaction"
)
```



```
fit2 <- rfsrc(Surv(t3d30,dth30) ~ ., data = train, nsp  
lit = 10, importance = TRUE, ntree = 100)
```

```
p1 <- predict(fit2, newdata = valid)  
Fsurv <- data.frame(p1$survival)  
  
Rsurv <- matrix(NA,nrow = nrow(Fsurv), ncol = 30)  
  
for (i in 1:30) {  
  Rsurv[i,valid$t3d30[i]:30] <- 0  
}  
  
Rsurv <- data.frame(Rsurv)  
  
Rsurv[is.na(Rsurv)] <- 1  
  
Fsurv <- data.frame(t(Fsurv))  
Rsurv <- data.frame(t(Rsurv[,1:29]))  
  
fsurv <- gather(Fsurv, key = "time", value = "status",  
index = 1:ncol(Fsurv))  
rsurv <- gather(Rsurv, key = "time", value = "status",  
index = 1:ncol(Rsurv))  
  
r2 <- roc(rsurv$status ~fsurv$status)  
plot(r2, print.auc = T, print.thres = T, identity.col=  
"blue", main = "validation data")
```




```
p2 <- predict(fit2, newdata = train)
Fsurv2 <- data.frame(p2$survival)

Rsurv2 <- matrix(NA,nrow = nrow(Fsurv2), ncol = 30)

for (i in 1:30) {
  Rsurv2[i,train$t3d30[i]:30] <- 0
}

Rsurv2 <- data.frame(Rsurv2)

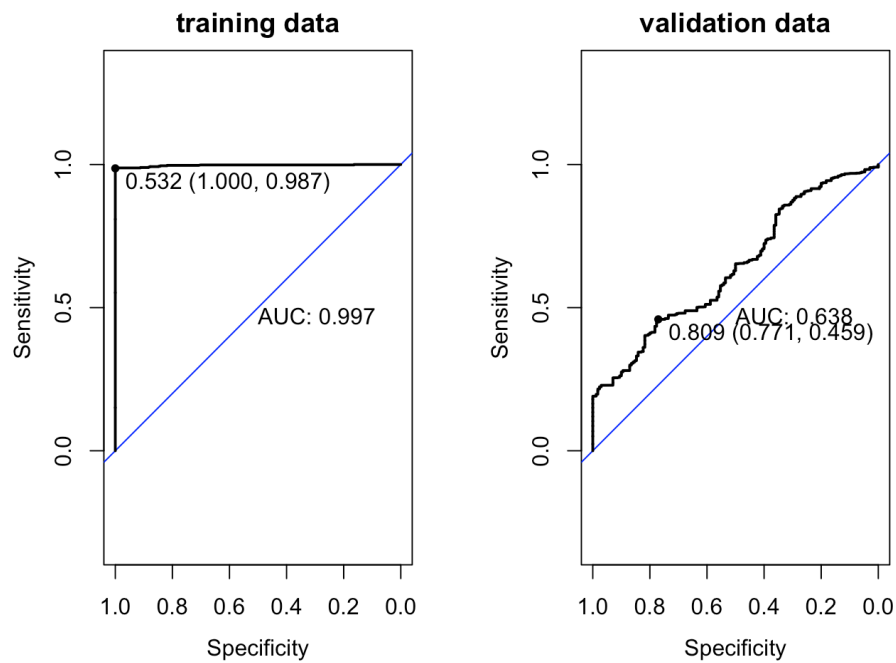
Rsurv2[is.na(Rsurv2)] <- 1

Fsurv2 <- data.frame(t(Fsurv2))
Rsurv2 <- data.frame(t(Rsurv2[,1:29]))

fsurv2 <- gather(Fsurv2, key = "time", value = "status",
  index = 1:nrow(Fsurv2))
rsurv2 <- gather(Rsurv2, key = "time", value = "status",
  index = 1:nrow(Rsurv2))

r3 <- roc(rsurv2$status ~fsurv2$status)

par(mfrow = c(1,2))
plot(r3, print.auc = T, print.thres = T, identity.col=
"blue", main = "training data")
plot(r2, print.auc = T, print.thres = T, identity.col=
"blue", main = "validation data")
```



```
data1 <- data0 <- data
data1$swang1 <- data$swang1[2]
data0$swang1 <- data$swang1[1]

fit1pred1 <- survfit(fit1, newdata = data1)
fit1pred2 <- survfit(fit1, newdata = data0)
```