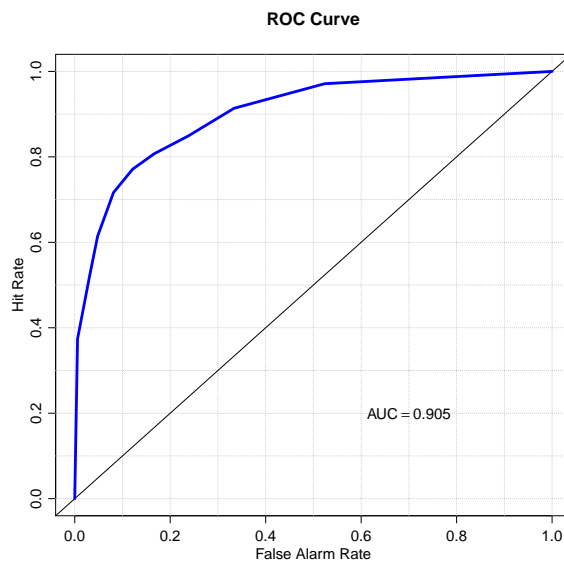


ROC curve for binomial signal regression (Phoneme data)



The receiver operating characteristic (ROC) curve for the logit penalized signal regression for the *aa-ao* phoneme data, optimally tuned with AIC. R code in `f-vowel-roc.R`

```
# ROC curve for binomial signal regression (Phoneme data)
# A graph in the book 'Practical Smoothing. The Joys of P-splines'
# Paul Eilers and Brian Marx, 2019

library(ggplot2)
library(gridExtra)
library(JOPS)
library(verification)

# Get the data
library(ElemStatLearn)
data(phoneme)
X = as.matrix(phoneme[, 1:150])
y = as.vector(phoneme[, 257])

# Select the 'aa' and 'ao' only
inn = which(y == "aa" | y == "ao")
X = X[inn, ]
y = y[inn]

# Change y to 0/1 ('aa' is success)
y[y == "aa"] = 1
y[y == "ao"] = 0
y = as.numeric(y)

iindex = 1:ncol(X)
pord = 3
nseg = 20

# Search for min AIC
llamb = seq(from = -2.5, to = 0.5, by = 0.25)
lambin = 10^llamb
r = length(llamb)
cv_ = rep(-999, r)
```

```
for (i in 1:r) {
  psrloop = psSignal(y, X, iindex, nseg = nseg, lambda = lambin[i],
                    pord = pord, family = "binomial")
  cv_[i] = psrloop$aic
}
lamopt = lambin[which(cv_ == min(cv_))]
fit2 = psSignal(y, X, iindex, nseg = nseg, lambda = lamopt,
               pord = pord, family = "binomial")

# Plot and save pdf
par(mfrow=c(1,1))
roc_aaa0 = verify(y, fit2$mu, bins = FALSE)
roc.plot(roc_aaa0, plot = FALSE, thresholds = seq(0.005, 0.995, by = 0.005))
lines.roc(roc_aaa0, col='blue', lwd = 3)
area1=roc.area(y, fit2$mu)$A
print(area1)
text(.7, .2, bquote("AUC"==.(round(area1, 3))))
```
