Fitting Generalized Linear Models

Jie Xue

June 20, 2017

## Case Study: Prediction of Body Fat

library("mboost") ## load package  
data("bodyfat", package = "TH.data") ## load data  
head(bodyfat)

## age DEXfat waistcirc hipcirc elbowbreadth kneebreadth anthro3a anthro3b  
## 47 57 41.68 100.0 112.0 7.1 9.4 4.42 4.95  
## 48 65 43.29 99.5 116.5 6.5 8.9 4.63 5.01  
## 49 59 35.41 96.0 108.5 6.2 8.9 4.12 4.74  
## 50 58 22.79 72.0 96.5 6.1 9.2 4.03 4.48  
## 51 60 36.42 89.5 100.5 7.1 10.0 4.24 4.68  
## 52 61 24.13 83.5 97.0 6.5 8.8 3.55 4.06  
## anthro3c anthro4  
## 47 4.50 6.13  
## 48 4.48 6.37  
## 49 4.60 5.82  
## 50 3.91 5.66  
## 51 4.15 5.91  
## 52 3.64 5.14

The response variable is the body fat measured by DXA (DEXfat).

#### 1.A Reproduce formula of Garcia et al., 2005

lm1 <- lm(DEXfat ~ hipcirc + kneebreadth + anthro3a, data = bodyfat)  
coef(lm1)

## (Intercept) hipcirc kneebreadth anthro3a   
## -75.2347840 0.5115264 1.9019904 8.9096375

#### 1.B Estimate same model by glmboost

glm1 <- glmboost(DEXfat ~ hipcirc + kneebreadth + anthro3a, data = bodyfat)  
coef(glm1, off2int=TRUE) ## off2int adds the offset to the intercept

## (Intercept) hipcirc kneebreadth anthro3a   
## -75.2073365 0.5114861 1.9005386 8.9071301

Note that in this case we used the default settings in control and the default family Gaussian() leading to boosting with the loss

### We now want consider all available variables as potential predictors.

#### 2.1 Build formula

preds <- names(bodyfat[, names(bodyfat) != "DEXfat"]) ## names of predictors  
fm <- as.formula(paste("DEXfat ~", paste(preds, collapse = "+"))) ## build formula  
fm

## DEXfat ~ age + waistcirc + hipcirc + elbowbreadth + kneebreadth +   
## anthro3a + anthro3b + anthro3c + anthro4

#### 2.2 Build model in 2 ways (identical results)

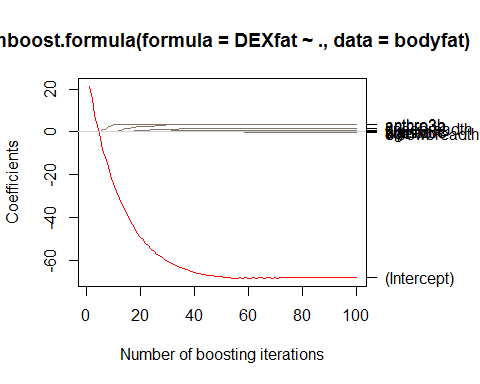
glm2a <- glmboost(fm, data = bodyfat)  
glm2b <- glmboost(DEXfat ~ ., data = bodyfat)  
identical(coef(glm2a), coef(glm2b))

## [1] TRUE

### 3. A plot of the coefficient paths

#### 3.1 All coefficients

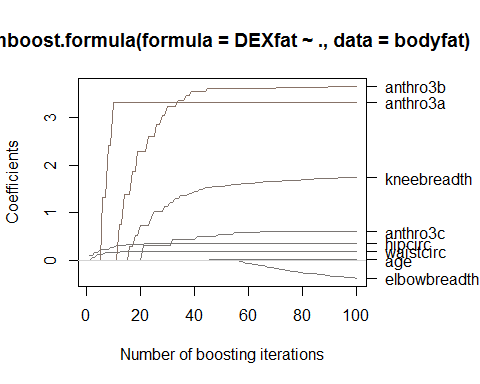
op <- par(mar = c(5,4,4,5.8) + 0.1)  
plot(glm2b, off2int = TRUE) ## default plot, offset added to intercept



par(op)

#### 3.2 Change ylim to the range of the coefficients without intercept (zoom-in)

op <- par(mar = c(5,4,4,5.8) + 0.1)  
plot(glm2b, ylim = range(coef(glm2b, which = preds)))



par(op)