Advanced Statistical Modeling

Non-parametric models - Iteratively Re-Weighted Least Squares

Haoran Mo, Alexandra Yamaui

November 22, 2017

In this task we are going to implement the Iteratively Re-Weighted Least Squares algorithm (IRWLS), which is the most frequently used method to solve the maximization problem of the log-likelihood function. This function is at the same time used to estimate the coefficients of the Logistic Regression model. Later we are going to use the glm() R function and compare the results.

We will used zero as the initial value of the coefficients β_0 (beta_0) and β_1 (beta_1) and we will build a new response variable z, which is a linear combination of the points x. The formula is presented below

$$z_i = \beta_0 + \beta_1 x_i + \frac{y_i - p_i}{p_i(1-p_i)}, \ i = 1, ..., n$$

where y_i is the original response variable and p_i is defined as below, which comes from the logistic function for the conditional distribution of the response variable y:

$$p_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

```
IRWLS <- function(x,y) {
  n <- length(x)
  beta_0 <- 0
  beta_1 <- 0
  s <- 0
  p <- c()
  v <- c()
  z <- c()
  convergence = 1
  #convergence != TRUE

  while (convergence > 0.0001) {  # we set 0.0001 instead of 0 due to computing cost concerned.
    for (i in 1:n) {
      p[i] <- exp(beta_0 + beta_1*x[i])/(1 + exp(beta_0 + beta_1*x[i]))
      z[i] <- beta_0 + beta_1*x[i] + ((y[i]-p[i])/p[i]*(1-p[i]))
      v[i] <- p[i]*(1-p[i])
    }

    lr <- lm(z ~ x, weights = v)


    beta_0_pre <- beta_0
    beta_1_pre <- beta_1
    beta_0 <- lr$coefficients[1]
    beta_1 <- lr$coefficients[2]

    convergence <- (abs(beta_0-beta_0_pre) + abs(beta_1-beta_1_pre))/2
    s <- s + 1
  }
  return(c(beta_0,beta_1))
}


x <- burn.injuries$e
y <- burn.injuries$superv
betas <- IRWLS(x,y)

glm.model <- glm(y~x, family = 'binomial')
summary(glm.model)
```

```
##
## Call:
## glm(formula = y ~ x, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6256  -1.5274   0.8141   0.8422   0.8725
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.8912     0.1059   8.412   <2e-16 ***
## x             0.1221     0.1872   0.652    0.514
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 525.39  on 434  degrees of freedom
## Residual deviance: 524.96  on 433  degrees of freedom
## AIC: 528.96
##
## Number of Fisher Scoring iterations: 4
```

The results obtained with our IRWLS function and glm are similar, the values of β_0 are 0.8929646 and 0.8911719, respectively, and the values for β_1 are 0.1495506 and 0.1221222.

Comparing the output of the IRWLS algorithm and glm() function we can see that from both methods the coefficients and their significance can be obtained. However, the glm function calculates de deviance and the Akaike information criterion (AIC), which are statistics to measure the goodness of fit for model comparisson (nested models in the case of the deviance and non-nested models in the case of the AIC). The IRWLS method does not adjust the coefficients for other models and therefore the deviance and AIC information cannot be obtained.