hw7-report

Jack McShane

2022-04-08

1. Let y_i be the default, x_{i1} be the student factor, x_{i2} be the balance, and x_{i3} be the income, assume $Y_i \sim bernoulli(p_i)$ with:

$$p_i = P(Y_i = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}, i = 1, \ldots, n$$

Write down the pmf $f(y_i)$.

The pmf of the Bernoulli distribution takes the following form:

$$f(y_i) = p_i^{y_i}(1 - p_i)^{(1-y_i)}$$

Following this equation and the way that we have modeled our probability of success, p_i, using the Sigmoid function, the pmf of Y_i can be written as shown below:

$$\begin{aligned} f(y_i) &= \left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}\right)^{y_i} \left(1 - \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}\right)^{(1-y_i)} \\ &= \left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}\right)^{y_i} \left(\frac{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}} - \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}\right) \\ &= \boxed{\left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}\right)^{y_i} \left(\frac{e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}\right)^{(1-y_i)}} \end{aligned}$$

2. Write down the joint distribution $f(y_1, \ldots, y_n)$.

$$f(y_i, \ldots, y_n) = \boxed{\prod_{i=1}^{n} \left[\left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}\right)^{y_i} \left(\frac{e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}\right)^{(1-y_i)}\right]}$$

3. What is the likelihood function $L(\beta_0, \beta_1, \beta_2, \beta_3)$?

$$L(\beta_0, \beta_1, \beta_2, \beta_3) = \boxed{\prod_{i=1}^{n} \left[\left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}\right)^{y_i} \left(\frac{e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}\right)^{(1-y_i)}\right]}$$

where all y_i have been observed.

4. Write down the log likelihood function, $l(\beta_0, \beta_1, \beta_2, \beta_3) = logL(\beta_0, \beta_1, \beta_2, \beta_3)$, and negative log likelihood function $-l(\beta_0, \beta_1, \beta_2, \beta_3)$.

$$\begin{aligned} l(\beta_0, \beta_1, \beta_, \beta_3) &= logL(\beta_0, \beta_1, \beta_2, \beta_3) \\ &= log\left[\prod_{i=1}^{n}\left[\left(\frac{1}{1+e^{-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}}\right)^{y_i}\left(\frac{e^{-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}}{1+e^{-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}}\right)^{(1-y_i)}\right]\right] \\ &= \boxed{\sum_{i=1}^{n}\left[y_i log\left(\frac{1}{1+e^{-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}}\right)+(1-y_i)log\left(\frac{e^{-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}}{1+e^{-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}}\right)\right]} \end{aligned}$$

$$-l(\beta_0, \beta_1, \beta_2, \beta_3) = \boxed{-\sum_{i=1}^{n}\left[y_i log\left(\frac{1}{1+e^{-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}}\right)+(1-y_i)log\left(\frac{e^{-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}}{1+e^{-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}}\right)\right]}$$

5. The maximum likelihood estimators of $\beta_0, \beta_1, \beta_2, \beta_3$ are:

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = argmaxL(\beta_0, \beta_1, \beta_2, \beta_3)$$

Explain that it is equivalent to the following:

$$\begin{aligned} (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) &= argmaxl(\beta_0, \beta_1, \beta_2, \beta_3) \\ &= argmin[-l(\beta_0, \beta_1, \beta_2, \beta_3)] \end{aligned}$$

The first equation, $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = argmax[l(\beta_0, \beta_1, \beta_2, \beta_3)]$, is equivalent to the original due to the fact that logarithmic functions are monotonically increasing (i.e. the value of the function is forever increasing over its range of x). This property allows us to apply a logarithmic transformation, but preserve the values of the parameters, in this case β_0, β_1, β_2 and β_3 that maximize the likelihood function as they will also be the values that maximize the log likelihood function.

The second function, the negative log likelihood, is simply a reflection across the x-axis, result of which is the previously maximum values of the log likelihood function now represent the minimum values of the *negative* log likelihood function. It therefore follow that the values of β_0, β_1, β_2, and β_3 which minimize the negative log likelihood function are the same values that maximize both the likelihood and the log likelihood functions.

6. In class, we use Newton-Raphson iteration to obtain the estimates:

$$\beta^{(t+1)} = \beta^{(t)} + (X^T D(\beta^{(t)})X)^{-1} X^T (y - p(\beta^{(t)})), t = 0, 1, \ldots,$$

Carry out this computation in R. What are your $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$?

```
# using the iteration approach
y<- as.numeric(Default$default) - 1
x1 <- as.numeric(Default$student) - 1
x2 <- Default$balance
x3 <- Default$income

X <- cbind(1, x1, x2, x3)
beta0 <- rep(0,4)
phat <- 1 / (1 + exp(-X %*% beta0))
beta1 <- beta0 + solve( t(X) %*% diag(c(phat*(1-phat))) %*% X ) %*% t(X) %*% (y - phat)

while( sum((beta1 - beta0)^2) > 1e-6 ) {
  beta0 <- beta1
  phat <- 1 / (1 + exp(-X %*% beta0))
  beta1 <- beta0 + solve( t(X) %*% diag(c(phat*(1-phat))) %*% X ) %*% t(X) %*% (y - phat)
}
```

```
## [1] "The estimated value of Beta"
```

```
##             [,1]
##    -1.086905e+01
## x1 -6.467758e-01
## x2  5.736505e-03
## x3  3.033450e-06
```

7. How do you interpret $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$?

The great thing about a Logistic Regression model is its parameter's interpretability. Given its form, we can calculate the effect of a change in one of the input features simply by taking the ratio of the odds for a positive outcome prior to and post this change.

Understanding and deriving our metric:

$$
\begin{aligned}
ratio &= \frac{P(Y=1)}{P(Y=0)} \\
&= \frac{p}{1-p} \\
&= p(1-p)^{-1} \\
&= \left(\frac{1}{1+e^{-(\beta_0+\beta 1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}}\right)\left(\frac{e^{-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}}{1+e^{-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}}\right)^{-1} \\
&= \frac{1}{e^{-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3})}} \\
&= e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3}}
\end{aligned}
$$

The above gives us the likelihood of the positive class relative to the negative class for a given input $< 1, x_{i1}, x_{i2}, x_{i3} >$. We can then use this concept and the simplified equation that it gives us to compare the effect that a change to one of the input features has on the chances for the positive class outcome as so:

Assuming two vectored inputs, $< 1, x_{i1}, x_{i2}, x_{i3} >$ and $< 1, x_{j1}, x_{i2}, x_{i3} >$, where only one feature value differs.

$$
\begin{aligned}
\frac{ratio_2}{ratio_1} &= \frac{e^{\beta_0+\beta_1 x_{j1}+\beta_2 x_{i2}+\beta_3 x_{i3}}}{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3}}} \\
&= \frac{e^{\beta_0} e^{\beta_1 x_{j1}} e^{\beta_2 x_{i2}} e^{\beta_3 x_{i3}}}{e^{\beta_0} e^{\beta_1 x_{i1}} e^{\beta_2 x_{i2}} e^{\beta_3 x_{i3}}} \\
&= e^{\beta_1 (x_{j1}-x_{i1})}
\end{aligned}
$$

The above equation gives us the difference in magnitude between positive class outcomes based on the change in a single feature; similar equations will result for changes in the other features. This is how we quantify the effect changes in a particular feature affect the outcome of our logistic model. We can now use this metric to compare the effects of our $\hat{\beta}_1$ based on a change in our x_1 feature input.

```
## [1] "Effect of being a student on probability of defaulting: 0.523731668809862"
## [1] "Effect of $100 increase in amount due: 1.77473395392338"
## [1] "Effect of $1000 increase in yearly income: 1.00303805568489"
```

We can see from the above that being a student decreases the likelihood of defaulting by a factor of roughly 2 while increasing the amount due on the loan increases the probability of default by a factor of ~ 1.75 (nearly 2). As well, an increase in yearly income seems to have little effect on the outcome.

8. Given your estimates, what is your default prediction of a person who is not a student, and has a balance of 900 and income of 20,000?

```
p <- 1 / (1 + exp( -(beta1[1] + beta1[2]*0 + beta1[3]*900 + beta1[4]*20000) ))
p
```

```
## [1] 0.003520865
```

Because the value of p (i.e. the probability of default) is lower than .5, the prediction of the model is that the debtor will not default on their payments.

9. Use the glm command, what is your $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$? Are they the same as your answers in Question 6?

```
summary(glm(default~student+balance+income, family="binomial", data=Default))
```

```
##
## Call:
## glm(formula = default ~ student + balance + income, family = "binomial",
##     data = Default)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4691  -0.1418  -0.0557  -0.0203   3.7383
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.087e+01  4.923e-01 -22.080  < 2e-16 ***
## studentYes  -6.468e-01  2.363e-01  -2.738  0.00619 **
## balance      5.737e-03  2.319e-04  24.738  < 2e-16 ***
## income       3.033e-06  8.203e-06   0.370  0.71152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2920.6  on 9999  degrees of freedom
## Residual deviance: 1571.5  on 9996  degrees of freedom
## AIC: 1579.5
##
## Number of Fisher Scoring iterations: 8
```

Yes, the estimations for $\beta_0, \beta_1, \beta_2, \beta_3$ are roughly equivalent.