

DSC520Week9KalaivaniKalyan

Kalaivani Kalyanasundaram

8/8/2021

Week 9 Assignment

Github link

<https://github.com/kalaikalyan/hello-world>

Fit a binary logistic regression model to the data set that predicts whether or not the patient survived for one year (the Risk1Y variable) after the surgery. Use the glm() function to perform the logistic regression. See Generalized Linear Models for an example. Include a summary using the summary() function in your results.

```
##
## Call:
## glm(formula = surgery_df$Risk1Yr ~ surgery_df$DGN + surgery_df$PRE4 +
##      surgery_df$PRE5 + surgery_df$PRE6 + surgery_df$PRE7 + surgery_df$PRE8
##      +
##      surgery_df$PRE9 + surgery_df$PRE10 + surgery_df$PRE11 +
surgery_df$PRE14 +
##      surgery_df$PRE17 + surgery_df$PRE19 + surgery_df$PRE25 +
##      surgery_df$PRE30 + surgery_df$PRE32 + surgery_df$AGE, family =
binomial(),
##      data = surgery_df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6084  -0.5439  -0.4199  -0.2762   2.4929
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.655e+01  2.400e+03  -0.007  0.99450
## surgery_df$DGNDGN2  1.474e+01  2.400e+03   0.006  0.99510
## surgery_df$DGNDGN3  1.418e+01  2.400e+03   0.006  0.99528
## surgery_df$DGNDGN4  1.461e+01  2.400e+03   0.006  0.99514
## surgery_df$DGNDGN5  1.638e+01  2.400e+03   0.007  0.99455
## surgery_df$DGNDGN6  4.089e-01  2.673e+03   0.000  0.99988
## surgery_df$DGNDGN8  1.803e+01  2.400e+03   0.008  0.99400
## surgery_df$PRE4    -2.272e-01  1.849e-01  -1.229  0.21909
## surgery_df$PRE5    -3.030e-02  1.786e-02  -1.697  0.08971 .
## surgery_df$PRE6PRZ1 -4.427e-01  5.199e-01  -0.852  0.39448
```

```
## surgery_df$PRE6PRZ2 -2.937e-01 7.907e-01 -0.371 0.71030
## surgery_df$PRE7T 7.153e-01 5.556e-01 1.288 0.19788
## surgery_df$PRE8T 1.743e-01 3.892e-01 0.448 0.65419
## surgery_df$PRE9T 1.368e+00 4.868e-01 2.811 0.00494 **
## surgery_df$PRE10T 5.770e-01 4.826e-01 1.196 0.23185
## surgery_df$PRE11T 5.162e-01 3.965e-01 1.302 0.19295
## surgery_df$PRE14OC12 4.394e-01 3.301e-01 1.331 0.18318
## surgery_df$PRE14OC13 1.179e+00 6.165e-01 1.913 0.05580 .
## surgery_df$PRE14OC14 1.653e+00 6.094e-01 2.713 0.00668 **
## surgery_df$PRE17T 9.266e-01 4.445e-01 2.085 0.03709 *
## surgery_df$PRE19T -1.466e+01 1.654e+03 -0.009 0.99293
## surgery_df$PRE25T -9.789e-02 1.003e+00 -0.098 0.92227
## surgery_df$PRE30T 1.084e+00 4.990e-01 2.172 0.02984 *
## surgery_df$PRE32T -1.398e+01 1.645e+03 -0.008 0.99322
## surgery_df$AGE -9.506e-03 1.810e-02 -0.525 0.59944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 395.61 on 469 degrees of freedom
## Residual deviance: 341.19 on 445 degrees of freedom
## AIC: 391.19
##
## Number of Fisher Scoring iterations: 15
```

According to the summary, which variables had the greatest effect on the survival rate?

Based on the summary, the variables that had the greatest effect are PRE9T , PRE14OC14, PRE17T, and PRE32T on the survival rate based on the p values.

To compute the accuracy of your model, use the dataset to predict the outcome variable. The percent of correct predictions is the accuracy of your model. What is the accuracy of your model?

```
## Warning: 'newdata' had 111 rows but variables found have 359 rows

##           Predicted_Value
## Actual_Value FALSE TRUE
##           F    302     6
##           T     44     7

## [1] 0.8607242
```

The accuracy is around 83%

Fit a logistic regression model to the binary-classifier-data.csv dataset

The dataset (found in binary-classifier-data.csv) contains three variables; label, x, and y. The label variable is either 0 or 1 and is the output we want to predict using the x and y variables.

```
##
## Call:
## glm(formula = binary_df$label ~ binary_df$x + binary_df$y, family =
binomial(),
##     data = binary_df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3728  -1.1697  -0.9575   1.1646   1.3989
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.424809   0.117224   3.624  0.00029 ***
## binary_df$x -0.002571   0.001823  -1.411  0.15836
## binary_df$y -0.007956   0.001869  -4.257 2.07e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2075.8  on 1497  degrees of freedom
## Residual deviance: 2052.1  on 1495  degrees of freedom
## AIC: 2058.1
##
## Number of Fisher Scoring iterations: 4
```

What is the accuracy of the logistic regression classifier? Keep this assignment handy, as you will be comparing your results from this week to next week.

```
## Warning: 'newdata' had 499 rows but variables found have 1498 rows

## Warning: 'newdata' had 999 rows but variables found have 1498 rows

##              Predicted_Value
## Actual_Value FALSE TRUE
##              0    429   338
##              1    286   445

## [1] 0.5834446
```

The accuracy is around 58.3% which is relatively lower.