

Exercise 5.7/8

Does anyone know how to start on 5.8(b)? Steps on the other parts would be helpful as well, and this question can also be used if anyone has questions on the other problem. (Question edited to include screenshots)

7. In Sections 5.3.2 and 5.3.3, we saw that the `cv.glm()` function can be used in order to compute the LOOCV test error estimate. Alternatively, one could compute those quantities using just the `glm()` and

200 5. Resampling Methods

`predict.glm()` functions, and a for loop. You will now take this approach in order to compute the LOOCV error for a simple logistic regression model on the `Weekly` data set. Recall that in the context of classification problems, the LOOCV error is given in (5.4).

- Fit a logistic regression model that predicts `Direction` using `Lag1` and `Lag2`.
- Fit a logistic regression model that predicts `Direction` using `Lag1` and `Lag2` using all but the first observation.
- Use the model from (b) to predict the direction of the first observation. You can do this by predicting that the first observation will go up if $P(\text{Direction} = \text{"Up"} | \text{Lag1}, \text{Lag2}) > 0.5$. Was this observation correctly classified?
- Write a for loop from $i = 1$ to $i = n$, where n is the number of observations in the data set, that performs each of the following steps:
 - Fit a logistic regression model using all but the i th observation to predict `Direction` using `Lag1` and `Lag2`.
 - Compute the posterior probability of the market moving up for the i th observation.
 - Use the posterior probability for the i th observation in order to predict whether or not the market moves up.
 - Determine whether or not an error was made in predicting the direction for the i th observation. If an error was made, then indicate this as a 1, and otherwise indicate it as a 0.
- Take the average of the n numbers obtained in (d)iv in order to obtain the LOOCV estimate for the test error. Comment on the results.

8. We will now perform cross-validation on a simulated data set.

(a) Generate a simulated data set as follows:

```
> set.seed(1)
> y=rnorm(100)
> x=rnorm(100)
> y=x-2*x^2+rnorm(100)
```

In this data set, what is n and what is p ? Write out the model used to generate the data in equation form.

- (b) Create a scatterplot of X against Y . Comment on what you find.
- (c) Set a random seed, and then compute the LOOCV errors that result from fitting the following four models using least squares:

i. $Y = \beta_0 + \beta_1 X + \epsilon$

ii. $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$

iii. $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$

iv. $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 X^4 + \epsilon$.

Note you may find it helpful to use the `data.frame()` function to create a single data set containing both X and Y .

- (d) Repeat (c) using another random seed, and report your results. Are your results the same as what you got in (c)? Why?
- (e) Which of the models in (c) had the smallest LOOCV error? Is this what you expected? Explain your answer.
- (f) Comment on the statistical significance of the coefficient estimates that results from fitting each of the models in (c) using least squares. Do these results agree with the conclusions drawn based on the cross-validation results?

5.4 Exercises 201

hw2

Edit good question | 0

Updated 4 years ago by Charles Hwang

S the students' answer, where students collectively construct a single answer

Actions ▾

You can use the `plot()` function to plot x and y as defined in (a)!

Edit undo thanks | 1

Updated 4 years ago by Brian Ho

i the instructors' answer, where instructors collectively construct a single answer

When you post a question on Piazza, can you post like a screen shot of the question from the book? (I know I could just look it up, but it would be easier for me if I just see it in the question).

Cheers,
Greg

[undo thanks](#) | 1

Updated 4 years ago by Gregory J. Matthews