Homework 6

Wenda Zhou

June 26, 2017

1. (2 points) Model selection

   We fit a linear on a fictitious dataset to obtain the following result:

```
Call:
lm(formula = y ~ x1 + x2, data = data.homework)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.3335 -0.7237 -0.1705  0.5762  1.9115 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)   0.2990     0.2546   1.174  0.25641   
x1            0.6538     0.2458   2.660  0.01650 * 
x2            0.7397     0.2136   3.464  0.00297 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.03 on 17 degrees of freedom
Multiple R-squared:  0.4633,Adjusted R-squared:  0.4002 
F-statistic: 7.337 on 2 and 17 DF,  p-value: 0.005044
```

   (a) Suppose we wish to select a submodel using best subset selection. Write down the list of models we will consider.

   (b) Suppose we fit a model with only the x_1 variable.

```
Call:
lm(formula = y ~ x1, data = data.homework)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0831 -0.8359  0.2331  0.8785  1.7681 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.001078   0.304097   0.004    0.997
x1          0.381063   0.295527   1.289    0.214

Residual standard error: 1.307 on 18 degrees of freedom
Multiple R-squared:  0.08456,Adjusted R-squared:  0.0337 
F-statistic: 1.663 on 1 and 18 DF,  p-value: 0.2136
```

   Suppose we wish to use the BIC to select the model. Compute the BIC for each model. Which would you select? The residual sum of squares is 18.0399 for the full model (with two variables). The RSS is 30.77 for the reduced model (with one variable).

2. (3 points) Bayes Estimator

   Define the Gamma distribution with parameters α and β to have p.d.f. (for $\lambda > 0$).

$$f(\lambda) = C(\alpha, \beta)\lambda^{\alpha-1}e^{-\beta\lambda}, \tag{1}$$

   where $C(\alpha, \beta)$ is a normalizing constant. The Gamma distribution has mean α/β.

   Suppose that we observe $X \sim \text{Poisson}(\lambda)$, and we have a prior on λ given by $\lambda \sim \text{Gamma}(\alpha, \beta)$.

   Compute the posterior distribution of λ given x. Compute the Bayes estimator (posterior mean) $\hat{\lambda}^B$.

3. (5 points) Contingency tables

   Suppose that we wish to model the probability of being cured as a function of whether the patient is treated or not.

   We have a trial with 20 patients total, 10 who have been assigned to the treatment, and 10 who have been assigned to the control group.

	Treated	
Cured	Yes	No
Yes	6	4
No	4	6

   (a) Compute the chi-square statistic for testing association between the treatment and whether the patient was cured.

   (b) Suppose that instead we wish to use a logistic regression to model the probability of being cured. Write down the equation for the logistic regression for the treatment group and the control group.

   (c) Compute the observed log-odds ratio from the contingency table. Solve the system to obtain the estimates of the coefficient.