Homework 6

Due: November 28th, 11:59 pm.

Instructions: Please scan or typeset your solutions and upload them as a single pdf file to Canvas.

Readings: Notes 7, Notes 8, Chapter 15, and Section 4.3 in *An Introduction to Statistical Learning with Applications in R, 2nd Edition* (`https://www.statlearning.com`).

Question I: Multivariate Linear Regression

We consider a regression model for the `carprice` data set (`carprice.csv` can be found in Canvas).

1. Let y_i be the car price (in hundreds), x_{i1} be the age, and x_{i2} be the mileage (in thousands) of the ith car. Consider the following model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \quad i = 1, \ldots, n; \quad \epsilon_1, \ldots, \epsilon_n \overset{iid}{\sim} N(0, \sigma^2).$$

Write down the pdf $f(y_i)$.

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

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

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

5. The maximum likelihood estimators (MLEs) of $\beta_0, \beta_1, \beta_2$ and σ are

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}) = \arg\max L(\beta_0, \beta_1, \beta_2, \sigma).$$

Explain that it is equivalent to the following

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}) = \arg\max l(\beta_0, \beta_1, \beta_2, \sigma) = \arg\min -l(\beta_0, \beta_1, \beta_2, \sigma).$$

6. Explain that the MLE and least squared estimator of $\beta_0, \beta_1, \beta_2$ are the same.

7. In class, we show that $\hat{\beta} = (X^TX)^{-1}X^Ty$. Now it is your turn to use it. Carry out this computation in `R`. What is your $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$?

8. How do you interpret $\hat{\beta}_1$ and $\hat{\beta}_2$?

9. Given your estimates, what is your price prediction of a used car with 4 years old, and 50,000 miles?

10. Use `lm` command, what is your $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$? Are they the same as your answers in Question 7?

Question II: Logistic Regression

We consider the logistic regression model for the `Default` data set (see Notes 8).

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

$$p_i = P(Y_i = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i})}}, \quad i = 1, \ldots, n$$

Write down the pdf $f(y_i)$.

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

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

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 $-l(\beta_0, \beta_1, \beta_2, \beta_3)$.

5. The maximum likelihood estimators (MLEs) of $\beta_0, \beta_1, \beta_2$ and β_3 are

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = \arg\max L(\beta_0, \beta_1, \beta_2, \beta_3).$$

Explain that it is equivalent to the following

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = \arg\max l(\beta_0, \beta_1, \beta_2, \beta_3) = \arg\min -l(\beta_0, \beta_1, \beta_2, \beta_3).$$

6. In class we used the Newton-Raphson iteration to obtain the estimates

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

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

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

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

9. Use `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?

Problems by learning objectives (plus rubric):

Question	Points	Goal: to reinforce estimation in regression models
I	20 pt	Linear Regression (MLE and Least Squares)
II	20 pt	Logistic Regression (Numerical approximation)
	40 pts.	