LINEAR REGRESSION MODELS W4315

HOMEWORK 5 ANSWERS

October 28, 2009

Due: 10/22/29
Instructor: Frank Wood (10:35-11:50)

1. (15 points) In order to get a maximum likelihood estimate of the parameters of a Box-Cox transformed simple linear regression model ($Y_i^\lambda = \beta_0 + \beta_1 X_i + \epsilon_i$), we need to find the gradient of the likelihood with respect to its parameters (the gradient consists of the partial derivatives of the likelihood function w.r.t. all of the parameters). Derive the partial derivatives of the likelihood w.r.t all parameters assuming that $\epsilon_i \sim N(0, \sigma^2)$. (N.B. the parameters here are $\lambda, \beta_0, \beta_1, \sigma$)
(Extra Credit: Given this collection of partial derivatives (the gradient), how would you then proceed to arrive at final estimates of all the parameters? Hint: consider how to increase the likelihood function by making small changes in the parameter settings.)

Answer:

The gradient of a multi-variate function is defined to be a vector consisting of all the partial derivatives w.r.t every single variable. So we need to write down the full likelihood first:
$L = \prod \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\sum(y_i^\lambda - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}}$
Then the log-likelihood function is:
$l = -\frac{n}{2} log\sigma^2 - \frac{\sum(y_i^\lambda - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}$
Take derivatives w.r.t to all the four parameters, we have the followings:

$$\frac{\partial l}{\partial \lambda} = -\frac{1}{\sigma^2} \sum (y_i^\lambda - \beta_0 - \beta_1 x_i) y_i^\lambda ln y_i \tag{1}$$
$$\frac{\partial l}{\partial \beta_0} = \frac{1}{\sigma^2} \sum (y_i^\lambda - \beta_0 - \beta_1 x_i) \tag{2}$$
$$\frac{\partial l}{\partial \beta_1} = \frac{1}{\sigma^2} \sum (y_i^\lambda - \beta_0 - \beta_1 x_i) x_i \tag{3}$$
$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum (y_i^\lambda - \beta_0 - \beta_1 x_i)^2}{2\sigma^4} \tag{4}$$

From the above equations array, we can have the gradient.
Extra credit: It's easily seen that equation (1) has no analytical form of the solution. So in

order to find the λ we are seeking for, we should try a sequence of λ's and figure out the MLE of all other parameters when λ taking corresponding value. Then using the MLE of all the parameters we evaluate the corresponding log-likelihood and find the largest one. Then λ matching this max log-likelihood generates the appropriate Box-Cox transformation.

2. (10 points) Problem 4.22 in the book.

Answer:

Following the thread on Page 155 in the textbook, we have:
$Suppose P(A_1) = P(A_2) = P(A_3) = \alpha$, then
$P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) = P(A_1 \cup \bar{A}_2 \cup A_3) = 1 - P(A_1 \cup A_2 \cup A_3) = 1 - P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_1$
$1 - 3\alpha + P(A_1A_2) + P(A_1A_3) + P(A_2A_3) - P(A_1A_2A_3)$
So we have $P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) \geq 1 - P(A_1) - P(A_2) - P(A_3)$

3. (10 points) Do problem 5.3 in the book.

Answer:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}.$$

Define $\mathbf{X}$ and $\mathbf{e}$ in the similar way. Then we have
(1)$\mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{e}$
(2)$\mathbf{X}'\mathbf{e} = \mathbf{0}$

4. (10 points) Do problem 5.15 in the book.

Answer:

(1) The matrix form of the equations is:

$$\begin{pmatrix} 5 & 2 \\ 23 & 7 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 8 \\ 28 \end{pmatrix}$$

(2) The solution is given as:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 5 & 2 \\ 23 & 7 \end{pmatrix}^{-1} \begin{pmatrix} 8 \\ 28 \end{pmatrix} = \begin{pmatrix} -0.6364 & 0.1818 \\ 2.0909 & -0.4545 \end{pmatrix} \begin{pmatrix} 8 \\ 28 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \end{pmatrix}$$

5. (10 points) Do problem 5.17 in the book.

Answer:

(1) The matrix form is:

$$\begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

where we denote the 3*3 matrix as $\mathbf{A}$

(2) $E\mathbf{W} = \mathbf{AEY}$

(3) $cov(\mathbf{W}) = \mathbf{Acov}(\mathbf{Y})\mathbf{A}'$

6. (15 points) Do problem 5.24 in the book.

Answer:

(a)The Matlab code is as follows:

$X = NaN(6, 2);$

$X(:, 1) = ones(6, 1);$

$X(:, 2) = [412334];$

$Y = [16510151322]';$

$beta = inv(X' * X) * X' * Y;$

$Y\_hat = X * beta;$

$res = Y - Y\_hat;$

$SSE = res' * res;$

$SSR = beta' * X' * Y - 1/6 * Y' * ones(6, 6) * Y;$

$v\_hat = SSE/4;$

$var\_matrix = v\_hat. * inv(X' * X);$

$[1 - 6] * var\_matrix * [1; -6]$

$$(1) \quad \begin{pmatrix} 0.4390 \\ 4.6098 \end{pmatrix} (2) \begin{pmatrix} -2.8780 \\ -0.0488 \\ 0.3415 \\ 0.7317 \\ -1.2683 \\ 3.1220 \end{pmatrix}$$

(3) 145.2073 (4) 20.29

$$(5)\quad \begin{pmatrix} 6.8055 & -2.1035 \\ -2.1035 & 0.7424 \end{pmatrix}$$

(6) -27.2195 (7) 58.7745

(b) (1) It's just the element at (1,2) of the variance-covariance matrix.

(2) It's just the very first element of the variance-covariance matrix.

(3) It's just the square root of the very last element of the variance-covariance matrix, so 0.8616. (c) $H = X(X'X)^{-1}X'$, by Matlab, we have the hat matrix is

$$\begin{pmatrix} 0.3659 & -0.1463 & 0.0244 & 0.1951 & 0.1951 & 0.3659 \\ -0.1463 & 0.6585 & 0.3902 & 0.1220 & 0.1220 & -0.1463 \\ 0.0244 & 0.3902 & 0.2683 & 0.1463 & 0.1463 & 0.0244 \\ 0.1951 & 0.1220 & 0.1463 & 0.1707 & 0.1707 & 0.1951 \\ 0.1951 & 0.1220 & 0.1463 & 0.1707 & 0.1707 & 0.1951 \\ 0.3659 & -0.1463 & 0.0244 & 0.1951 & 0.1951 & 0.3659 \end{pmatrix}$$

(d) The value is from (5.80) from the book, $MSE(I-H)$, so from Matlab we get:

$$\begin{pmatrix} 3.2171 & 0.7424 & -0.1237 & -0.9899 & -0.9899 & -1.8560 \\ 0.7424 & 1.7323 & -1.9798 & -0.6187 & -0.6187 & 0.7424 \\ -0.1237 & -1.9798 & 3.7121 & -0.7424 & -0.7424 & -0.1237 \\ -0.9899 & -0.6187 & -0.7424 & 4.2070 & -0.8662 & -0.9899 \\ -0.9899 & -0.6187 & -0.7424 & -0.8662 & 4.2070 & -0.9899 \\ -1.8560 & 0.7424 & -0.1237 & -0.9899 & -0.9899 & 3.2171 \end{pmatrix}$$

7. (15 points) Do problem 5.29 in the book.

Answer:

It's actually pretty straight-forward:

Since $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y}$ and $\mathbf{y} \sim \mathbf{N}(\mathbf{X}\beta, \sigma^2\mathbf{I})$, then we have:

$$\begin{aligned} E\mathbf{b} &= (\mathbf{X'X})^{-1}\mathbf{X'Ey} \\ &= (\mathbf{X'X})^{-1}\mathbf{X'X}\beta \\ &= \beta \end{aligned}$$

And from here, the unbiasedness just flows.

8. (15 points) Consider the following linear regression model:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

where $\mathbf{X}$ is the design matrix whose first column consists of all ones, β is a vector including all the regression parameters, ϵ is the error term which follows the standard Gaussian assumption (independent, equal-variance normal distribution, i.e. $\epsilon \sim \mathbf{N}(\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{I})$ where $\mathbf{I}$ is identity matrix). Derive the maximum likelihood estimator for β and σ using matrix calculation.

Answer:

In order to find MLE of the parameters of interest, we always need to write down the full likelihood and take derivatives w.r.t to all the parameters. So in the problem's setting, we have $\mathbf{y} \sim \mathbf{N}(\mathbf{X}\beta, \sigma^{\mathbf{2}}\mathbf{I})$
then it gives us the normal likelihood function which goes like:
$L(\beta, \sigma^{\mathbf{2}}) = (\frac{\mathbf{1}}{\sqrt{\mathbf{2}\pi\sigma}})^{\mathbf{n}} \mathbf{exp}\{-\frac{\mathbf{1}}{\mathbf{2}\sigma^{\mathbf{2}}}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)\}$
Then we consider the log-likelihood function which is :

$$\begin{aligned} l(\beta, \sigma^{\mathbf{2}}) &= C - \frac{n}{2} log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) \\ &= C - \frac{n}{2} log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y}' - \beta'\mathbf{X}')(\mathbf{y} - \mathbf{X}\beta) \\ &= C - \frac{n}{2} log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y}'\mathbf{y} - \mathbf{2}\mathbf{y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta) \end{aligned}$$

where C is a constant which is independent of the parameters.
Then notice that in matrix calculation, we have the following rules of taking derivatives:

$$\begin{aligned} \frac{d(\mathbf{A}\mathbf{x})}{d\mathbf{x}} &= A' \\ \frac{d(\mathbf{x}'\mathbf{A}\mathbf{x})}{d\mathbf{x}} &= (\mathbf{A} + \mathbf{A}')\mathbf{x} \end{aligned}$$

Applying the above rules, we have:
$\frac{dl}{d\beta} = -\frac{1}{2\sigma^2}(-2\mathbf{X}'\mathbf{y} + \mathbf{2}\mathbf{X}'\mathbf{X}\beta) \quad \Rightarrow \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{y}$
$\frac{dl}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) \quad \Rightarrow \hat{\sigma^{\mathbf{2}}} = \frac{\mathbf{1}}{\mathbf{n}}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$