LINEAR REGRESSION MODELS W4315

HOMEWORK 5 ANSWERS

March 9, 2010

Due: 03/04/10
Instructor: Frank Wood

1. (20 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.

2. (15 points) [1] Derive an extension of Bonferroni inequality (4.2a) which is given as

$$P(\bar{A}_1 \bigcap \bar{A}_2) \geq 1 - \alpha - \alpha = 1 - 2\alpha$$

for the case of three statements, each with statement confidence coefficient $1 - \alpha$.

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. (25 points) [2] Refer to **Consumer finance** Problems 5.5 and 5.13.
a. Using matrix methods, obtain the following: (1) vector of estimated regression coefficients, (2) vector of residuals, (3) SSR, (4) SSE, (5) estimated variance-covariance matrix of **b**, (6) point estimate of $E\{Y_h\}$ when $X_h = 4$, (7) $s^2\{\text{pred}\}$ when $X_h = 4$
b. From your estimated variance-covariance matrix in part (a5), obtain the following: (1) $s\{b_0, b_1\}$; (2) $s^2\{b_0\}$; (3) $s\{b_1\}$
c. Find the hat matrix $\mathbf{H}$
d. Find $s^2\{\mathbf{e}\}$

Answer:

(a) $\mathbf{X} = \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$; $\mathbf{Y} = \begin{bmatrix} 16 \\ 5 \\ 10 \\ 15 \\ 13 \\ 22 \end{bmatrix}$; $\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 2 & 3 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 6 & 17 \\ 17 & 55 \end{bmatrix}$
$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{41} \begin{bmatrix} 55 & 17 \\ -17 & 6 \end{bmatrix}$;
$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \frac{1}{41} \begin{bmatrix} 55 & 17 \\ -17 & 6 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 2 & 3 & 3 & 4 \end{bmatrix} = \frac{1}{41} \begin{bmatrix} -13 & 38 & 21 & 4 & 4 & -13 \\ 7 & -11 & -5 & 1 & 1 & 7 \end{bmatrix}$

[1] This is problem 4.22 in 'Applied Linear Regression Models(4th edition)' by Kutner etc.
[2] This is problem 5.24 in 'Applied Linear Regression Models(4th edition)' by Kutner etc.

$$\mathbf{H}=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'=\tfrac{1}{41}\begin{bmatrix}1&4\\1&1\\1&2\\1&3\\1&3\\1&4\end{bmatrix}\begin{bmatrix}-13&38&21&4&4&-13\\7&-11&-5&1&1&7\end{bmatrix}=\tfrac{1}{41}\begin{bmatrix}15&-6&1&8&8&15\\-6&27&16&5&5&-6\\1&16&11&6&6&1\\8&5&6&7&7&8\\8&5&6&7&7&8\\15&-6&1&8&8&15\end{bmatrix}$$

(1): $\hat{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}=\tfrac{1}{41}\begin{bmatrix}-13&38&21&4&4&-13\\7&-11&-5&1&1&7\end{bmatrix}\begin{bmatrix}16\\5\\10\\15\\13\\22\end{bmatrix}=\tfrac{1}{41}\begin{bmatrix}18\\189\end{bmatrix}=\begin{bmatrix}0.4390\\4.6098\end{bmatrix}$

(2): Residual$=\mathbf{Y}-\mathbf{X}\hat{\beta}=\begin{bmatrix}16\\5\\10\\15\\13\\22\end{bmatrix}-\begin{bmatrix}1&4\\1&1\\1&2\\1&3\\1&3\\1&4\end{bmatrix}\begin{bmatrix}0.4390\\4.6098\end{bmatrix}=\begin{bmatrix}-2.8780\\-0.0488\\0.3415\\0.7317\\-1.2683\\3.1220\end{bmatrix}$

(3): $SSR=\mathbf{Y}'[\mathbf{H}-\frac{1}{n}\mathbf{J}]\mathbf{Y}=145.2073$

(4): $SSE=\mathbf{Y}'(\mathbf{I}-\mathbf{H})\mathbf{Y}=20.2927$

(5): The estimated variance-covariance matrix of $\mathbf{b}=\mathbf{s}^2\{\mathbf{b}\}=MSE(\mathbf{X}'\mathbf{X})^{-1}=\begin{bmatrix}6.8055&-2.1035\\-2.1035&0.7424\end{bmatrix}$

(6): The point estimate of $E\{Y_h\}=\mathbf{X}_h'\mathbf{b}=\begin{bmatrix}1&4\end{bmatrix}\begin{bmatrix}0.4390\\4.6098\end{bmatrix}=18.8780$

(7): At $X_h=4, s^2\{pred\}=MSE(1+\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h)=6.9292$

(b) $s\{b_0,b_1\}=-2.1035; s^2\{b_0\}=6.8055; s\{b_1\}=\sqrt{0.7424}=0.8616$

(c) As calculated in part(a), the hat matrix $\mathbf{H}=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'=\frac{1}{41}\begin{bmatrix}1&4\\1&1\\1&2\\1&3\\1&3\\1&4\end{bmatrix}\begin{bmatrix}-13&38&21&4&4&-13\\7&-11&-5&1&1&7\end{bmatrix}$

$$= \frac{1}{41}\begin{bmatrix} 15 & -6 & 1 & 8 & 8 & 15 \\ -6 & 27 & 16 & 5 & 5 & -6 \\ 1 & 16 & 11 & 6 & 6 & 1 \\ 8 & 5 & 6 & 7 & 7 & 8 \\ 8 & 5 & 6 & 7 & 7 & 8 \\ 15 & -6 & 1 & 8 & 8 & 15 \end{bmatrix} = \begin{bmatrix} 0.3659 & -0.1463 & 0.0244 & 0.1951 & 0.1951 & 0.3659 \\ -0.1463 & 0.6585 & 0.3902 & 0.1220 & 0.1220 & -0.1463 \\ 0.0244 & 0.3902 & 0.2683 & 0.1463 & 0.1463 & 0.0244 \\ 0.1951 & 0.1220 & 0.1463 & 0.1707 & 0.1707 & 0.1951 \\ 0.1951 & 0.1220 & 0.1463 & 0.1707 & 0.1707 & 0.1951 \\ 0.3659 & -0.1463 & 0.0244 & 0.1951 & 0.1951 & 0.3659 \end{bmatrix}$$

(d) $s^2\{\mathbf{e}\} = MSE(\mathbf{I} - \mathbf{H}) = \begin{bmatrix} 3.2171 & 0.7424 & -0.1237 & -0.9899 & -0.9899 & -1.8560 \\ 0.7424 & 1.7323 & -1.9798 & -0.6187 & -0.6187 & 0.7424 \\ -0.1237 & -1.9798 & 3.7121 & -0.7424 & -0.7424 & -0.1237 \\ -0.9899 & -0.6187 & -0.7424 & 4.2070 & -0.8662 & -0.9899 \\ -0.9899 & -0.6187 & -0.7424 & -0.8662 & 4.2070 & -0.9899 \\ -1.8560 & 0.7424 & -0.1237 & -0.9899 & -0.9899 & 3.2171 \end{bmatrix}$

Matlab Code:

X=[1 4;1 1;1 2;1 3;1 3;1 4]

Y=[16;5;10;15;13;22]

J=ones(6,6)

I=eye(6,6)

$[n, m] = size(Y)$

$Z = inv(X' * X)$

H=X*Z*X'

beta=Z*X'*Y

residual=Y-H*Y

SSR=Y'*(H-(1/n)*J)*Y

SSE=Y'*(I-H)*Y

MSE=SSE/(n-2)

cov=MSE*Z

$s2\_e = MSE * (I - H)$

Xh=[1;4]

Yhhat=Xh'*beta

$s2\_pred = MSE * (1 + Xh' * Z * Xh)$

4. (25 points) [3] In a small-scale regression study, the following data were obtained: Assume

[3]This is problem 6.27 in 'Applied Linear Regression Models(4th edition)' by Kutner etc.

i:	**1**	**2**	**3**	**4**	**5**	**6**
X_{i1}	7	4	16	3	21	8
X_{i2}	33	41	7	49	5	31
Y_i	42	33	75	28	91	55

that regression model (1) which is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta X_{i2} + \epsilon_i \tag{5}$$

with independent normal error terms is appropriate. Using matrix methods, obtain (a) $\mathbf{b}$; (b) $\mathbf{e}$; (c) $\mathbf{H}$; (d) SSR; (e) $s^2\{\mathbf{b}\}$; (f) $\hat{Y}_h$ when $X_{h1} = 10$, $X_{h2} = 30$; (g) $s^2\{\hat{Y}_h\}$ when $X_{h1} = 10$, $X_{h2} = 30$

Answer:

(a) $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 33.9321 \\ 2.7848 \\ -0.2644 \end{bmatrix}$

(b) $\mathbf{e} = \mathbf{Y} - \mathbf{X}\mathbf{b} = \begin{bmatrix} -2.6996 \\ -1.2300 \\ -1.6374 \\ -1.3299 \\ -0.0900 \\ 6.9868 \end{bmatrix}$

(c) $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \begin{bmatrix} 0.2314 & 0.2517 & 0.2118 & 0.1489 & -0.0548 & 0.2110 \\ 0.2517 & 0.3124 & 0.0944 & 0.2663 & -0.1479 & 0.2231 \\ 0.2118 & 0.0944 & 0.7044 & -0.3192 & 0.1045 & 0.2041 \\ 0.1489 & 0.2663 & -0.3192 & 0.6143 & 0.1414 & 0.1483 \\ -0.0548 & -0.1479 & 0.1045 & 0.1414 & 0.9404 & 0.0163 \\ 0.2110 & 0.2231 & 0.2041 & 0.1483 & 0.0163 & 0.1971 \end{bmatrix}$

(d) $SSR = \mathbf{Y}'[\mathbf{H} - \frac{1}{n}\mathbf{J}]\mathbf{Y} = 3009.926$

(e) $s^2\{\mathbf{b}\} = MSE(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 715.4711 & -34.1589 & -13.5949 \\ -34.1589 & 1.6617 & 0.6441 \\ -13.5949 & 0.6441 & 0.2625 \end{bmatrix}$

(f) $\hat{Y}_h = \mathbf{X}'_h\mathbf{b} = \begin{bmatrix}1 & 10 & 30\end{bmatrix}\begin{bmatrix}33.9321\\2.7848\\-0.2644\end{bmatrix} = 53.8471$

(g) At X_{h1}=10 and $X_{h2} = 30$, $s^2\{\hat{Y}_h\} = \mathbf{X}'_h s^2\{\mathbf{b}\}\mathbf{X}_h = 5.4246$

Matlab Code:

```
X=[1 7 33;1 4 41;1 16 7;1 3 49;1 21 5; 1 8 31]
Y=[42;33;75;28;91;55]
J=ones(6,6)
I=eye(6,6)
[n,m]=size(Y)
Z=inv(X'*X)
H=X*Z*X'
beta=Z*X'*Y
residual=Y-H*Y
SSR=Y'*(H-(1/n)*J)*Y
SSE=Y'*(I-H)*Y
MSE=SSE/(n-3)
cov=MSE*Z
s2_e=MSE*(I-H)
Xh=[1;10;30]
Yhhat=Xh'*beta
s2_yhat=Xh'*cov*Xh
```

5. (15 points) Consider the classic regression model using matrix, i.e.

$$Y = X\beta + \epsilon$$

where X is a $n * p$ design matrix whose first column is an all 1 vector, $\epsilon \sim N(\mathbf{0}, \mathbf{I})$ and I is an identity matrix. Prove the followings:

a. The residual sum of squares $RSS = \hat{\mathbf{e}}'\hat{\mathbf{e}}$ can be written in a matrix form:

$$RSS = \mathbf{y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \tag{6}$$

b. We call the RHS of (2) a 'sandwich'. Prove the matrix in the middle layer of the 'sandwich' $\mathbf{N} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is an idempotent matrix.

c. Prove that the rank of $\mathbf{N}$ defined in part (b) is $n-p$.
N.B. p columns in design matrix means there are $p-1$ predictors plus 1 intercept term. Before handling the problem, make clear of the dimensions of all the matrices here.

Answer:

(a) $SSE = \mathbf{e}'\mathbf{e} = (\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb}) = (\mathbf{y}'-\mathbf{b}'\mathbf{X}')(\mathbf{y}-\mathbf{Xb}) = \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{I}\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} = \mathbf{y}'(\mathbf{I}-\mathbf{b}'\mathbf{X}')\mathbf{y} = \mathbf{y}'(\mathbf{I}-((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')'\mathbf{X}')\mathbf{y} = \mathbf{y}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}$

(b)$\mathbf{A}^2 = \mathbf{AA} = (\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \mathbf{I}-2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{I}-2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}\mathbf{I}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = A$
Therefore, A is an idempotent matrix.

(c) Since A is a symmetric and idempotent matrix, rank(A)=trace(A)
Let $H = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$
$trace(A) = trace(\mathbf{I}_{n\times n} - \mathbf{H}_{n\times n}) = trace(\mathbf{I}) - trace(\mathbf{H}) = n - trace(\mathbf{H}) = n - trace(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = n - trace((\mathbf{X}'\mathbf{X})^{-1}_{p\times p}\mathbf{X}'_{p\times n}\mathbf{X}_{n\times p}) = n - trace(\mathbf{I}_{p\times p}) = n-p$
So rank(A)=n-p