STAT5030 Midterm Linear Models (13-14)

1. Consider K $(k = 1, ..., K)$ regression models

$$Y_{ki} = \alpha_k + \beta_k x_{ki} + \epsilon_{ki}, \quad (i = 1, 2, ..., n_k)$$

where the ϵ_{ki} are independently and identically distributed as $N(0, \sigma^2)$.

   (a) Find the least squares estimates of α_k and β_k.

   (b) To conduct the test of equal y-intercept (all K regression lines meet at the same point when $x = 0$), what are the null and alternative hypotheses? What is the reduced model under the null? Derive the SSE of the full model and the SSE of the reduced model, and the details of the testing procedure.

2. Let

$$Y_i = \theta_i + \epsilon_i$$

where $i = 1, 2, 3, 4$ and ϵ_i are independent $N(0, \sigma^2)$. Let $\theta_1 + \theta_2 + \theta_3 + \theta_4 = 0$.

   (a) Derive the least squares estimates of the parameters.

   (b) Find the SSE when $Y_1 = 1, Y_2 = 2, Y_3 = 3$, and $Y_4 = 4$.

3. Suppose that the regression curve

$$E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2$$

have a local maximum at $x = x_m$ where x_m is near the origin. If Y is observed at n points $x_i, (i = 1, 2, ..., n)$ in $[-a, a]$, $\bar{x} = 0$, and Y_i are independent normal random variables with variance equal to σ^2, Using the random variable $U = \hat{\beta}_1 + 2x_m\hat{\beta}_2$ where $\hat{\beta}_1$ and $\hat{\beta}_2$ are LSE of β_1 and β_2 respectively, outline a method for finding a confidence interval for x_m.

4. The hat matrix is $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \{h_{ij}\}$ (Let $\mathbf{X}$ be a matrix with full column rank and with $\mathbf{1}$ as its first column). From class notes, we have $h_{ii} = 1/n + (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)'(\mathbf{X}_c'\mathbf{X}_c)^{-1}(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)$, where $\mathbf{x}_{1i}' = (x_{i1}, x_{i2}, ..., x_{ik})$, $\bar{\mathbf{x}}_1' = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_k)$, and $(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)'$ is the ith row of the centered matrix $\mathbf{X}_c$. Prove that we can also express h_{ii} as the following:

$$h_{ii} = 1/n + (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)'(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)\sum_{r=1}^{k}\frac{1}{\lambda_r}cos^2\theta_{ir},$$

where θ_{ir} is the angle between $\mathbf{x}_{1i} - \bar{\mathbf{x}}_1$ and $\mathbf{a}_r$, the rth normalized eigenvector (λ_r is the corresponding eigenvalue) of $\mathbf{X}_c'\mathbf{X}_c$.

Problem from STAT5030

1. Consider the linear model

$$Y_{n\times 1} = X\beta + \epsilon, \qquad E(\epsilon) = 0, \quad Cov(\epsilon) = \sigma^2 V$$

with $\beta = (\beta_1, ..., \beta_p)'$ and sample data $(X, Y_{n\times 1})$. Let V be a known positive definite matrix.

$(X'V^{-1}X)^{-1}X'V^{-1}Y$

(a) What is the Generalized Least Squares estimator of β?

(b) (Prediction) Let

$$X_0 = \begin{bmatrix} X_{n+1,1} & X_{n+1,2} & \cdots & X_{n+1,p} \\ X_{n+2,1} & X_{n+2,2} & \cdots & X_{n+2,p} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ X_{n+m,1} & X_{n+m,2} & \cdots & X_{n+m,p} \end{bmatrix}$$

$Xa=$ $\quad X'P=X_0'$

∃可逆 $XA=X_0$

$XP=X_0$

and Y_0 be the values that we are interested to predict given X_0. The column space of X_0' is a subset of the column space of X'. Assume that ∃A s.t. $X'A=X_0'$

$$Y_0 = X_0\beta + \epsilon_0, \qquad E(\epsilon_0) = 0, \quad Cov(\epsilon_0) = \sigma^2 V_0$$

$A'VA$

with V_0 be a known positive definite matrix. In addition, assume that $Cov(\epsilon, \epsilon_0) = 0$ and $X_0\beta$ is estimable. $=\sigma^2[X_0(X'V^{-1}X)^{-1}X_0'+V_0]$

$X_0(X'V^{-1}X)^{-1}X'V^{-1}Y$

i. What is the prediction of Y_0 (denoted by $\hat{Y}_0$)? $\Rightarrow \triangleq \hat{Y}_0$ $\quad Cov(\hat Y_0 - Y_0) = X_0(X'V^{-1}X)^{-1}X'V^{-1}\sigma^2 V V^{-1}X(X'V^{-1}X)^{-1}X_0' + \sigma^2 V_0$

ii. What is $Cov(\hat{Y}_0 - Y_0)$? $\hat Y_0 - Y_0 = X_0(X'V^{-1}X)^{-1}X'V^{-1}(X\beta+\epsilon) - X_0\beta - \epsilon_0 = X_0(X'V^{-1}X)^{-1}X'V^{-1}\epsilon - \epsilon_0$

(c) Let X_0 be the same as given in Part (b), and the model is also the same EXCEPT that $Cov(\epsilon, \epsilon_0) = \sigma^2 W$.

Now, let $Y_0^* = CY$ be a linear unbiased predictor of Y_0. Define the prediction mean squared error (PMSE) of Y_0^* as $\quad Y_0^* - Y_0 = CY - Y_0 = CX\beta + C\epsilon - X_0\beta - \epsilon_0 \sim ((CX-X_0)\beta,$

$$E(Y_0^* - Y_0)'A(Y_0^* - Y_0)$$

where A is a positive definite matrix.

$\beta'X'C'AX_0\beta$ 而 $\sigma^2 tr(ACW)$

i. Prove that the PMSE of Y_0^* is $\quad Cov(C\epsilon - \epsilon_0) =$

$$\beta'(CX - X_0)'A(CX - X_0)\beta + \sigma^2 tr[A(CVC' + V_0 - 2CW)]$$

$\sigma^2[CVC'+V_0-2CW]$

ii. The best (minimum PMSE) linear unbiased estimator of Y_0 is

$$Y_0^* = \hat{Y}_0 + D.$$

Find D. $\quad Y_0^* = X_0(X'V^{-1}X)^{-1}X'V^{-1}Y + D \triangleq CY + D$; $\hat Y_0$; $Y_0^* - Y_0 \sim (D, \sigma^2[CVC'+V_0-2CW])$

PMSE of $Y_0^* = D'AD + \sigma^2 tr[A(CVC'+V_0-2CW)]$

attains minimum at $D=0$

2. Let

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \qquad Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

and assume that X and X_1 have full column rank. Consider the linear model

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

$- X_2'X_2(X_1'X_1+X_2'X_2)^{-1}$
$(X_1'Y_1+X_2'Y_2$

where $\epsilon \sim N(0, \sigma^2 I)$. Let $\hat\beta$ be the least squares estimator of β and $\hat Y = X\hat\beta = (\hat Y_1, \hat Y_2)'$. Further, for the linear model

$$Y_1 = X_1\beta^* + \epsilon^*$$

$X_1'X_1(X_1'X_1+X_2'X_2)^{-1}(X_1'Y_1+X_2'Y_2$

where $\epsilon^* \sim N(0, \sigma^2 I)$, the least squares estimator of β^* is $\hat\beta^*$. Let

$= X_1'X_1' + X_2'Y_2$

$$\hat Y^* = X\hat\beta^* = \begin{bmatrix} \hat Y_1^* \\ \hat Y_2^* \end{bmatrix}$$

$\hat\beta = (X'X)^{-1}(X'Y)$

$= (X_1' \ X_2')\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}(X_1'Y_1+X_2'Y_2)$

$= (X_1'X_1+X_2'X_2)^{-1}$

Define

$$Y - \hat Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} - \begin{bmatrix} \hat Y_1 \\ \hat Y_2 \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$$Y - \hat Y^* = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} - \begin{bmatrix} \hat Y_1^* \\ \hat Y_2^* \end{bmatrix} = \begin{bmatrix} e_1^* \\ e_2^* \end{bmatrix}$$

$\hat Y = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} X_1\hat\beta \\ X_2\hat\beta \end{pmatrix}$

(a) Prove that

$$\hat\beta - \hat\beta^* = M_1^{-1}X_2'e_2$$ √ How

$e_1 = Y_1 - X_1\hat\beta$

where $M_1 = X_1'X_1$.

$(X_1 \ )$

(b) Express e_2 in terms of e_2^* and rewrite the expression of $\hat\beta - \hat\beta^*$ in Part (a)

X_1' X_2' $(\quad)$

(c) The following is a data set with sample size = 7

M_1 $\quad \hat\beta^* = (X_1'X_1)^{-1}(X_1'Y_1)$

-3

x	-3	-2	-1	0	1	2	3
y	14	7	.	.	.	.	-2

$X_1'Y_1$

For the above data and with a simple linear regression model, the parameter estimate $\hat\beta^* = (6, -2)'$.

$\hat\beta^*$

$X_2'Y_2 - X_2'X_2(X_1'X_1+X_2'X_2)^{-1}$
$(X_1'Y_1+X_2'Y_2)$

Suppose an additional observation $(x, y) = (4, 4)$ is obtained (You now have 8 pairs of (x, y) in your updated dataset), compute the new parameter estimate $\hat\beta$. (Hint: use Parts (a) and (b))

$X_1 = x \qquad Y_1 = y$
$X_2 = 4 \qquad Y_2 = 4$

√

$e_2 = Y_2 - X_2\hat\beta$

$e_2^* = Y_2 - X_2\hat\beta^*$
$= Y_2 - X_2(X_1'X_1)^{-1}(X_1'Y_1)$
$\hat\beta = (X_1'X_1)^{-1}(X_1')(Y - Y_2)$.

$K = X(X'X)^- X'$ $KX = X$

Problem from STAT5030

1. Consider the model

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \qquad i = 1,2,3,4, \quad j = 1,2,3,4,$$

where ϵ_{ij} are independently distributed as $N(0, \sigma^2)$.

(a) Let $\beta = (\mu, \tau_1, \tau_2, \tau_3, \tau_4)'$. Find a set of 4 linearly independent estimable functions of β.

(b) Derive a test to test the null hypothesis $H_0 : \tau_1 - \tau_2 = \tau_3 - \tau_4$.

(c) Is $\tau_1 + 2\tau_2$ estimable? Why?

2. (a) Let $A_{m\times m}$ and $B_{n\times n}$ be two nonsingular matrices. Further, assume that the matrices U and V are $m \times n$ and $n \times m$ respectively. Prove that

左右边直接 def

$$(A + UBV)^{-1} = (A^{-1} - A^{-1}UB(B + BVA^{-1}UB)^{-1}BVA^{-1}.)$$

$(A + UBV)$

(b) Consider a regression model,

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

H.

$X(X'X)^- X'$

where X, $n \times p$, is full column rank. $Y = (Y_1, ..., Y_n)'$. Further, assume that the $Var(\epsilon) = \sigma^2 I$. Let e_i be the ith residual and h_i be the ith diagonal element of the hat matrix. Let $\hat{\beta}$ and $\hat{\beta}_{(i)}$ be the least squares estimate of β with and without the ith case included in the data respectively.

i. Show that

$$(X'_{(i)}X_{(i)})^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1}x_i x_i'(X'X)^{-1}}{1 - h_i},$$

where $X_{(i)}$ denotes the regression matrix with the ith row (x_i') deleted. (Hint: $X'X = X'_{(i)}X_{(i)} + x_i x_i'$)

$X' = X_{(i)} + (0, \dots, 0, X_i, 0 \dots 0)$

ii. Prove that

$$\hat{\beta} - \hat{\beta}_{(i)} = \frac{(X'X)^{-1}x_i e_i}{1 - h_i}.$$

— END —

Qualifying Exam. (Linear Models) Dec. 2013

1. Consider a regression model,

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

where X, $n \times p$, is full column rank. Let $Y = (Y_1, ..., Y_n)'$, $X' = (x_1, x_2, ..., x_n)$ and let $Y_{(i)}$ be the corresponding Y vector and $X_{(i)}$ be the corresponding X matrix after deleting the i-th case. Further, assume that $Var(\epsilon) = \sigma^2 I$. Let e_i be the ith residual and h_i be the ith diagonal element of the hat matrix. Let $\hat{\beta}$ and $\hat{\beta}_{(i)}$ be the least squares estimate of β with and without the ith case included in the data respectively. Also, we let SSE and $SSE_{(i)}$ be the error Sum of squares with and without the ith case included in the data respectively.

You are given the result that

$$\hat{\beta} - \hat{\beta}_{(i)} = \frac{(X'X)^{-1} x_i e_i}{1 - h_i}.$$

(a) Show that

$$Y'_{(i)} Y_{(i)} = Y'Y - Y_i^2$$

(b) Show that

$$Y'_{(i)} X_{(i)} \hat{\beta}_{(i)} = Y'X\hat{\beta} - y_i^2 + \frac{e_i^2}{1 - h_i}$$

(c) Find the value of $SSE_{(i)}$ if $SSE = 20$, $e_i = 0.8$ and $h_i = 0.2$.

2. Consider a linear regression model,

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_{p-1} x_i^{p-1} + \epsilon_i,$$

$i = 1, ..., n$. Also, $\epsilon_1, \epsilon_2, ..., \epsilon_n$ are i.i.d. $N(0, \sigma^2)$. Let $P_k(x)$ be a polynomial of order k. Then the above model could be rewritten as

$$y_i = \alpha_0 P_0(X_i) + \alpha_1 P_1(X_i) + \cdots + \alpha_{p-1} P_{p-1}(X_i) + \epsilon_i,$$

Assume that

$$\sum_{i=1}^{n} P_l(x_i) P_m(x_i) = 0, \quad l \neq m, \quad \text{for all } l \text{ and } m,$$

(a) Derive the least squares estimator of α_j, $j = 0, 1, ..., p-1$.

(b) Derive the test for testing the null hypothesis H_0: $\alpha_j = 0$.

To prove:

$$\frac{S_n}{B_n} \xrightarrow{d} N(0,1)$$

5. S_n is submartingale, τ is stopping time, to prove

   (1)
$$E(S_{\tau\wedge n}) \leq ES_n$$

   (2)
$$P(\max_k S_k > x) \leq E(|S_n|1(\max_k S_k > x))$$

STAT 5030

1. Prove XGX^T is invariant of generalized inverse G of X^TX

2. Consider the model

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, i = 1,2,3,4, j = 1,2,3,4,$$

   where ε_{ij} are independently distributed as $N(0,\sigma^2)$.

   (1) Let $\beta = (\mu, \tau_1, \tau_2, \tau_3, \tau_4)'$. Find a set of 4 linearly independent estimable functions of β.

   (2) Derive a test to test the null hypothesis $H_0 : \tau_1 - \tau_2 = \tau_3 - \tau_4$.

   (3) Is $\tau_1 + 2\tau_2$ estimable? Why?

3. There are two groups of data (Y_1, X_1), (Y_2, X_2) with

$$Y_1 = X_1\beta + \varepsilon_1,$$

$$Y_2 = X_2\beta + \varepsilon_2,$$

   X_1 and X_2 is not necessarily full-rank. And suppose that $\lambda^T\beta$ is estimable.

   (1) T_1 and T_2 are BLUE of $\lambda^T\beta$ for data (Y_1, X_1) and (Y_2, X_2), respectively. Give T_1 and T_2 and calculate $Var(T_1)$ and $Var(T_2)$

   (2) Let $T(\alpha) = \alpha T_1 + (1-\alpha)T_2$. Find α to minimize $Var(T(\alpha))$

   (3) Let $Y = (Y_1^T, Y_2^T)^T, X = (X_1^T, X_2^T)^T$, give the BLUE T_3 of $\lambda^T\beta$ for data (Y, X). And calculate $Var(T_3)$.

   (4) Explain $Var(T_3) \leq Var(T(\alpha))$ with equality when $r(X_1) = 1$ or $r(X_2) = 1$

(5) A and B are symmetric and nonnegtive matrix and a is in the vector space of A and B, that's to say, there exist x and y s.t. $a = Ax = By = (A+B)z$, then

$$a^T A^- a a^T B^- a \geq a^T (A+B)^- a (a^T A^- a + a^T B^- a)$$

with equality if $r(A) = 1$ or $r(B) = 1$.

Hint:

$$\begin{aligned}
a^T A^- a &= x^T A x = x^T P P^T x \\
a^T B^- a &= y^T B y = y^T Q Q^T y \\
a^T (A+B)^- a &= z^T (A+B) z = z^T A z + z^T B z \\
(z^T A x)^2 = (z^T P P^T x)^2 &\leq (z^T P P^T z)(x^T P P^T x) = (z^T A z)(x^T A x) \\
(z^T B y)^2 = (z^T Q Q^T y)^2 &\leq (z^T Q Q^T y)(x^T Q Q^T y) = (z^T B z)(y^T B y) \\
z^T A z + z^T B z &\geq (z^T (A+B) z)^2 (\frac{1}{x^T A x} + \frac{1}{y^T B y}) \\
z^T A z + z^T B z &\leq \frac{x^T A x y^T B y}{x^T A x + y^T B y}
\end{aligned}$$

If $r(A) = 1$, then $A = uu^T$, where u is a vector. Then

$$(z^T u u^T x)^2 = (z^T u u^T z)(x^T u u^T x)$$

$$\begin{aligned}
By &= Ax = u u^T x = ku \\
Bz &= Ax - Az = (k-l)u
\end{aligned}$$

where $k = u^T x, l = u^T z$. Thus, we have $Bz = QQ^T z = \frac{k-l}{k} QQ^T y$, then $Q^T z = \frac{k-l}{k} Q^T y$ s.t.

$$(z^T Q Q^T y)^2 = (z^T Q Q^T y)(x^T Q Q^T y).$$

4. About ridge regression and LASSO.