Згадаємо постановку задачі у класичному регресійному аналізі. Вважаємо, що для регресійної моделі

$$y = X\alpha + e$$

справедливі усі припущення класичного регресійного аналізу:

I. $e \sim \mathcal{N}\left(\theta_N, \sigma^2 E_N\right), \sigma^2 \in \mathbb{R}_+$,

II. $rank(X) = p$,

III. немає ніяких обмежень на α, тобто $\alpha \in \mathbb{R}^p$.

Введемо позначення $Q(\alpha) = \|y - X\alpha\|^2$. Тоді відомо, що при справедливості припущень I, II, III оцінка МНК є розв'язком такої оптимізаційної задачі

$$\hat{\alpha} = \arg\min_{\alpha} Q(\alpha) = \left(X^T X\right)^{-1} X^T y.$$

Послабимо тепер припущення II, а саме будемо вважати, що матриця X задовольняє умові:

$$\text{II}'.\ rank(X) = p - r, \quad r \geq 1. \tag{0.39}$$

Припущення II' еквівалентно умові, що стовпчики матриці X лінійно залежні, тобто що

$$\exists a_i \neq \vec{\theta}_p : \quad Xa_i = \vec{\theta}_N, i = \overline{1,r}. \tag{0.40}$$

Якщо справедливі припущення I, II', III, тобто умова (0.40) виконується строго, то кажуть, що знаходяться в умовах ***строгої мультиколінеарності***, якщо ж умова (0.40) виконується не строго, тобто добутки Xa_i попадають тільки в окіл початку координат, то кажуть, що знаходяться в умовах ***мультиколінеарності***.

Проаналізуємо обидва ці випадки.

Випадок строгої мультиколінеарності. У цьому випадку оцінка МНК існує і вона буде не єдина. А множина цих усіх оцінок МНК задається як множина усіх розв'язків *системи нормальних рівнянь для оцінки МНК*, а саме:

$$\left(X^T X\right)\hat{\alpha} = X^T y.$$

Зауважимо, що в умовах строгої мультиколінеарності оцінка МНК у вигляді

$$\hat{\alpha} = \left(X^T X\right)^{-1} X^T y \tag{0.41}$$

не існує.

Випадок мультиколінеарності. В умовах мультиколінеарності оцінка МНК у вигляді (0.41) теоретично існує, бо матриця $\left(X^T X\right)$ є не виродженою, але її практичне використання буде проблематичним. Дійсно справедливість умов мультиколінеарності приводить до того, що існують власні значення матриці $\left(X^T X\right)$, які попадають в окіл нуля, а це означає, що $\det\left(X^T X\right)$ буде близьким до нуля, тобто матриця $\left(X^T X\right)$ буде погано обумовленою. А це у свою чергу приводить до таких негативних наслідків для оцінки МНК у вигляді (0.41):

- Оцінка МНК у вигляді (0.41) буде вже нестійкою. Тобто незначні зміни в елементах матриці X можуть привести до значної зміни значень оцінок параметрів $\hat{\alpha} = \left(X^T X\right)^{-1} X^T y$. Свій внесок до цього додадуть також похибки машинного округлення. А це у свою чергу приводить до втрати точності прогнозу згідно нашої моделі.

- Оцінка МНК у вигляді (0.41) буде вже мало ефективною, бо її характеристика розсіювання $\sigma^2\left(X^T X\right)^{-1}$ різко зростає. Згадайте властивість І.а. $\hat{\alpha} \sim \mathcal{N}\left(\alpha, \sigma^2\left(X^T X\right)^{-1}\right)$.
- Параметри регресійної моделі будуть сильно корельовані між собою згідно властивості І.а. $\hat{\alpha} \sim \mathcal{N}\left(\alpha, \sigma^2\left(X^T X\right)^{-1}\right)$. А це вже позбавляє змісту їх інтерпретацію.

Для подолання цих проблем при оцінюванні параметрів регресійної моделі в умовах мультиколінеарності використовують:

- Попереднє центрування незалежних змінних.
- При поліноміальній апроксимації функції регресії переходять від регресорів $\left\{x^i\right\}_{i=0}^p$ до регресорів у вигляді набору ортогональних поліномів $\left\{\varphi_i(x)\right\}_{i=0}^p$.
- Гребеневі оцінки.

Найбільш вживаним підходом є використання гребеневих оцінок.

Гребенева оцінка (ridge estimate).

Гребенева оцінка була запропонована Хоерлом (A. Hoerl) у 1962 році і визначається таким чином:

$$\hat{\alpha}(\varepsilon) = \left(X^T X + \varepsilon E_p\right)^{-1} X^T y, \varepsilon > 0, \qquad (0.42)$$

де ε – мале позитивне дійсне число.

Ця оцінка дозволяє подолати погано обумовленість матриці $\left(X^T X\right)$. Але з'ясувалося, що гребенева оцінка є зміщеною і це призвело до відтермінування її широкого використання на практиці. Але згодом з'ясувалося, що гребенева оцінка при коректному виборі параметра ε дозволяє досягти навіть більшої точності у середньо квадратичному розумінні у порівнянні з оцінкою МНК $\hat{\alpha} = \left(X^T X\right)^{-1} X^T y$. Саме це відновило інтерес до використання на практиці гребеневої оцінки.

Введемо позначення: $B(\varepsilon) = \left(X^T X + \varepsilon E_p\right)^{-1}$. Це дозволяє гребеневу оцінку записати у такому вигляді

$$\hat{\alpha}(\varepsilon) = B(\varepsilon) X^T y, \varepsilon > 0. \tag{0.43}$$

Теорема. Для гребеневої оцінки $\hat{\alpha}(\varepsilon)$ справедливо:

1) $\hat{\alpha}(\varepsilon) = \left[E_p - \varepsilon B(\varepsilon)\right]\hat{\alpha}$, де $\hat{\alpha} = \left(X^T X\right)^{-1} X^T y$,

2) $M\hat{\alpha}(\varepsilon) = \alpha + \Delta\left(\hat{\alpha}(\varepsilon)\right)$, де $\Delta\left(\hat{\alpha}(\varepsilon)\right) = -\varepsilon B(\varepsilon)\alpha$,

3) $m(\varepsilon) = M\left\|\hat{\alpha}(\varepsilon) - \alpha\right\|^2 = M\left\|\hat{\alpha}(\varepsilon) - M\hat{\alpha}(\varepsilon)\right\|^2 + \left\|\Delta\left(\hat{\alpha}(\varepsilon)\right)\right\|^2$.

гребеневої оцінки

$$m(\varepsilon) = M\left\|\hat{\alpha}(\varepsilon) - \alpha\right\|^2 = M\left\|\hat{\alpha}(\varepsilon) - M\hat{\alpha}(\varepsilon) + M\hat{\alpha}(\varepsilon) - \alpha\right\|^2 =$$

$$= M\left\|\left(\hat{\alpha}(\varepsilon) - M\hat{\alpha}(\varepsilon)\right) + \underbrace{\left(M\hat{\alpha}(\varepsilon) - \alpha\right)}_{\Delta(\hat{\alpha}(\varepsilon))}\right\|^2 =$$

$$= M\left\|\hat{\alpha}(\varepsilon) - M\hat{\alpha}(\varepsilon)\right\|^2 + 2M\left\{\left(\hat{\alpha}(\varepsilon) - M\hat{\alpha}(\varepsilon)\right)^T \Delta\left(\hat{\alpha}(\varepsilon)\right)\right\} +$$

$$+\left\|\Delta\left(\hat{\alpha}(\varepsilon)\right)\right\|^2 = M\left\|\hat{\alpha}(\varepsilon) - M\hat{\alpha}(\varepsilon)\right\|^2 +$$

$$+2M\left\{\left(\hat{\alpha}(\varepsilon) - M\hat{\alpha}(\varepsilon)\right)^T\right\}\Delta\left(\hat{\alpha}(\varepsilon)\right) + \left\|\Delta\left(\hat{\alpha}(\varepsilon)\right)\right\|^2 =$$

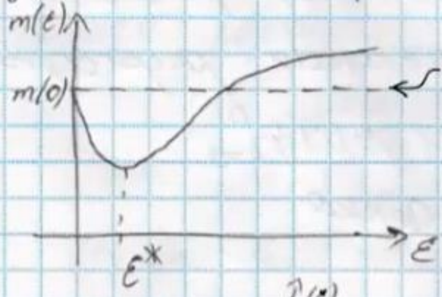
$$= M\left\|\hat{\alpha}(\varepsilon) - M\hat{\alpha}(\varepsilon)\right\|^2 + 2\left\{\underbrace{\left(M\hat{\alpha}(\varepsilon) - M\hat{\alpha}(\varepsilon)\right)^T}_{\theta_p}\right\}\Delta\left(\hat{\alpha}(\varepsilon)\right) +$$

$$+\left\|\Delta\left(\hat{\alpha}(\varepsilon)\right)\right\|^2 = M\left\|\hat{\alpha}(\varepsilon) - M\hat{\alpha}(\varepsilon)\right\|^2 + \left\|\Delta\left(\hat{\alpha}(\varepsilon)\right)\right\|^2. \quad \blacksquare$$

Зауваження 1. Згідно п. 3 теореми

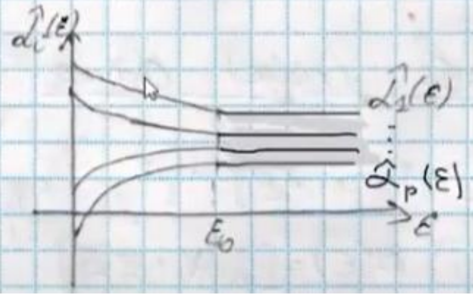
$$m(\varepsilon) = M\left\|\hat{\alpha}(\varepsilon) - \alpha\right\|^2 = \underbrace{M\left\|\hat{\alpha}(\varepsilon) - M\hat{\alpha}(\varepsilon)\right\|^2}_{m_1(\varepsilon)} + \underbrace{\left\|\Delta\left(\hat{\alpha}(\varepsilon)\right)\right\|^2}_{m_2(\varepsilon)}.$$

Зауваження 1 (Херл, Кенард; 1970).



середньоквадратичне відхилення оцінки МНК

Зауваження 2.



Графіки гребіневих слідів

де ε_0 – найменший ε при якому відбувається стабілізація усіх графіків $\hat{\alpha}_i(\varepsilon)$

Методи вибору структури регресійної моделі

η: $y(1), \ldots, y(N)$

Ми маємо деякий набір регресорів

$\varphi_i(\vec{\xi})$: $x_i(1), \ldots, x_i(N)$, $i = \overline{1,p}$

I. Метод включення

II. Метод виключення

III. Пряма крокова регресія

IV. Обернена крокова регресія

Зауваження. Залежна змінна η матиме індекс 0. $\eta \equiv \vec{\xi}_0$

Позн.: $I = \{0, 1, 2, \ldots, p\}$

I. Метод включення

Почат. модель: $y(k) = \alpha_0 + e_+(k)$, $k = \overline{1,N}$

Ідея: ...

1. $I_- := \{1, 2, \ldots, p\}$, $\tilde{p} := 1$, $\gamma_+ > 0$;
2. [illegible]

$I_i := I \setminus \{0, i\}$

$\hat{r}_{0i}(I_i)$, $i \in I_-$

3. $i_+ = \arg\max\limits_{i \in I_-} |\hat{r}_{0i}(I_i)|$

4. $F_{i_+} < F_{\gamma_+}(1, N-(\tilde{p}+1))$

⊕ — $\varphi_{i_+}(\vec{\xi})$ - не включ;
stop;

⊖ — $\varphi_{i_+}(\vec{\xi})$ - включаємо;
$\tilde{p} := \tilde{p} + 1$;
$I_- : I_- \setminus \{i_+\}$;
if $\tilde{p} = p+1$ then stop;
goto п. 2;

II. Метод виключення

Почат. модель: $y(k) = a_0 + \sum_{i=1}^{p} a_i x_i(k) + e_-(k)$, $k = \overline{1, N}$
Ідея: ...

1. $I_+ := \{1, 2, ..., p\}$; $\tilde{p} := p+1$; $\gamma_- > 0$;
2. F_i, $i \in I_+$
3. $i_- = \arg\min_{i \in I_+} F_i$;
4. $F_{i_-} < F_{\gamma_-}(1, N-\tilde{p})$

⊕ — $\varphi_{i_-}(\vec{\xi})$ - виключити
$\tilde{p} := \tilde{p} - 1$;
$I_+ := I_+ \setminus \{i_-\}$;
if $\tilde{p} = 1$ then stop;
goto п. 2;

⊖ — $\varphi_{i_-}(\vec{\xi})$ - не викл;
stop;

Порівняння $\overline{\underline{I}}$, $\overline{\underline{II}}$.

III. Пряма крокова регресія

Почат. модель: $y(k) = \alpha_0 + e_+(k), \; k = \overline{1, N}.$

Ідея: ...

1. $I_- := \{1, 2, \dots, p\}; \; I_+ := \{\};$

   $M := \varnothing;$ // кількість попер. послідовних невдалих спроб

   $\tilde{p} := 1; \quad \gamma_- \geq \gamma_+ > 0;$

2. $\hat{r}_{0i(I_i)}, \quad i \in \left[\begin{array}{l} I_- \text{ // на першій ітерації} \\ I_- \setminus \{i_-\} \end{array}\right.,$

   де $I_i := I \setminus \{0, i\}.$

3. $i_+ = \arg\max\limits_{i \in \left[\begin{array}{l} I_- \text{ // на першій ітерації} \\ I_- \setminus \{i_-\} \end{array}\right.} |r_{0i(I_i)}|$

4. $F_{i_+} < F_{\gamma_+}(1, N - (\tilde{p} + 1))$

+

−

$\varphi_{i_+}(\vec{\xi})$ - не включ.;

if $M = 1$ then stop;

$M := 1;$

if $\tilde{p} = 1$ then stop;

$\varphi_{i_+}(\vec{\xi})$ - включити;

$\tilde{p} := \tilde{p} + 1; \quad M := \varnothing;$

$I_- := I_- \setminus \{i_+\};$

$I_+ := I_+ \cup \{i_+\};$

if $\tilde{p} = 2$ then goto п. 2;

5. F_i, $i \in I_+ \setminus \{i_+\}$;

6. $i_- := \arg\min\limits_{i \in I_+ \setminus \{i_+\}} F_i$;

7. $F_{i_-} < F_{\gamma_-}(1, \nu - \tilde{p})$

(+)

$\varphi_{i_-}(\vec{\xi})$ – виключити;

$\tilde{p} := \tilde{p} - 1$; $M := \emptyset$;

$I_- := I_- \cup \{i_-\}$;

$I_+ := I_+ \setminus \{i_-\}$;

if $\tilde{p} = p$ then goto п. 5;

else goto п. 2;

(−)

$\varphi_{i_-}(\vec{\xi})$ – не виключ.;

if $M = 1$ then stop;

$M := 1$;

if $\tilde{p} = p + 1$ then stop;

goto п. 2;

IV Обернена крокова регресія

Почат. модель: $y(k) = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_i(k) + e(k)$, $k = \overline{1, N}$

Ідея: ...

1. $I_- := \{\}$; $I_+ = \{1, 2, \dots, p\}$;

$M := \emptyset$; $\tilde{p} := p + 1$; $\gamma_- \geq \gamma_+ > 0$;

2. F_i, $i \in \left[\begin{array}{l} I_+ \quad \text{// перша ітерація} \\ I_+ \setminus \{i_+\}; \end{array}\right.$

3. $i_- = \arg\min F_i$, $i \in \left[\begin{array}{l} I_+ \quad \text{// перша ітерація} \\ I_+ \setminus \{i_+\}; \end{array}\right.$

4. $F_{i_-} < F_{\gamma-}(1, N-\tilde{p})$

⊕

$\varphi_{i_-}(\vec{\xi})$ – виключити;

$\tilde{p} := \tilde{p} - 1$; $M := \varnothing$;

$I_- := I_- \cup \{i_-\}$;

$I_+ := I_+ \setminus \{i_-\}$;

if $\tilde{p} = p$ then goto п.2;

⊖

$\varphi_{i_-}(\vec{\xi})$ – не виключ;

if $M = 1$ then stop;

$M := 1$;

if $\tilde{p} = p+1$ then stop;

5. $\hat{r}_{0i}(I_i)$, $i \in I_- \setminus \{i_-\}$,

де $I_i := I \setminus \{0, i\}$;

6. $i_+ = \arg\max\limits_{i \in I_- \setminus \{i_-\}} |\hat{r}_{0i}(I_i)|$

7. $F_{i_+} < F_{\gamma+}(1, N-(\tilde{p}+1))$

⊕

$\varphi_{i_+}(\vec{\xi})$ – не включ;

if $M = 1$ then stop;

$M := 1$;

if $\tilde{p} = 1$ then stop;

goto п.2;

⊖

$\varphi_{i_+}(\vec{\xi})$ – включити;

$\tilde{p} := \tilde{p} + 1$; $M := \varnothing$;

$I_- := I_- \setminus \{i_+\}$;

$I_+ := I_+ \cup \{i_+\}$;

if $\tilde{p} = 2$ then goto п.5.

goto п.2;

Нелінійний регресійний аналіз.

$$y(k) = g(X(k), \vec{\alpha}) + e(k), \quad k=\overline{1,N} \qquad (15)$$

Нехай (15) є нелінійною по $\vec{\alpha}$.

Нелінійні регресійні моделі класифікуються на два класи:

1) внутрішньолінійні;
2) внутрішньонелінійні.

Озн. Модель називається внутрішньолінійною, якщо шляхом перетворень її можна звести до лінійної по параметрам, інакше модель є внутрішньонелінійною.

1)... Приклад. $y(k) = \ln(\alpha_1)\, X(k)^{2^{\alpha_2}} + e(k), \; k=\overline{1,N}.$

$$\ln y(k) = \ln\ln(\alpha_1) + 2^{\alpha_2}\ln(X(k)) + \tilde{e}(k), \; k=\overline{1,N}.$$

Нехай $\tilde{y}(k) = \ln(y(k))$, $\beta_1 = \ln\ln(\alpha_1)$, $\beta_2 = 2^{\alpha_2}$.

$\tilde{x}(k) = \ln(X(k))$, $\tilde{e}(k)$ - ..., тоді

$$\tilde{y}(k) = \beta_1 + \beta_2\tilde{x}(k) + \tilde{e}(k), \; k=\overline{1,N}.$$

За допомогою МНК знаходимо $\hat{\beta}_1, \hat{\beta}_2$, тоді

$$\begin{cases} \hat{\beta}_1 = \ln\ln(\hat{\alpha}_1) \\ \hat{\beta}_2 = 2^{\hat{\alpha}_2} \end{cases} \Rightarrow \begin{cases} \hat{\alpha}_1 = e^{e^{\hat{\beta}_1}} \\ \hat{\alpha}_2 = \log_2(\hat{\beta}_2) \end{cases}$$

Зауваження. Так як ми працювали з $\tilde{e}(k)$, а не $e(k)$, то такі оцінки варто сприймати як деяке наближення до істинних оцінок МНК фізичної моделі.

Маю модель (15) внутрішньолінійна, тоді існують такі функції $g_1(\cdot)$, $\{g_{2j}(\cdot)\}$, $\{g_{3j}(\cdot)\}$, що справедливо

$$g_1(y(k)) = \sum_j \beta_j\, g_{2j}(x(k)) + \tilde{e}(k), \quad k=\overline{1,N}, \qquad (16)$$

де $\{\beta_j = g_{3j}(\alpha), \forall j$ (17)

З системи (16) за допомогою МНК знаходимо $\{\hat{\beta}_j\}$

А з (17), вже відомими $\{\hat{\beta}_j\}$ знаходять $\hat{\alpha}$.

2) Випадок внутрішньонелінійної системи (15)

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{k=1}^{N} e^2(k) = \arg\min_{\alpha} I(\alpha)$$

де $I(\alpha) = \sum_{k=1}^{N} e^2(k)$

За допомогою методів безумовної нелінійної оптимізації мінімізуємо функціонал $I(\alpha)$.

Наприклад. Метод Марквардта (градієнтний + лінеаризації)