Лабораторна робота №3
з комп'ютерної статистики

Горбунов Даніел Денисович
1 курс магістратури
група "Прикладна та теоретична статистика"

15 листопада 2021 р.

1 Вступ.

У цій роботі ми продовжимо викладення результатів застосування різних підходів до побудови хорошої лінійної регресійної моделі залежності цін однієї компанії від цін інших компаній (тобто тим, чим займалися у перших двох роботах). У звіті наводяться результати застосування гребеневої регресії на повних та свіжих даних, результати прогнозування були порівняні з прогнозами найкращих моделей з перших двох робіт.

2 Хід роботи.

2.1 Зауваження.

Ми будемо розглядати лінійну регресійну модель з адитивною константою:

$$Y_j \approx \sum_{i=1}^{d} \beta_i X_j^i + \beta_0$$

І застосуємо рідж-регресію як з штрафуванням цієї константи так і без нього:

- Де адитивна константа штрафується, це класична рідж регресія. Її використаємо на нормованих даних:

  $$Y_j \approx \sum_{i=1}^{d} \beta_i' \frac{X_j^i}{\mathcal{S}(X^i)} + \beta_0',\ \mathcal{S}(X^i) = \sqrt{\overline{(X^i)^2}}$$

  У такому разі оцінки рідж-оцінки для початкових змінних зовсім прості:

  $$\beta_0 = \beta_0',\ \beta_i = \frac{\beta_i'}{\mathcal{S}(X^i)},\ i = \overline{1, d}$$

- Якщо адитивна константа не штрафується, то розглянемо перетворену модель вигляду:

  $$\tilde{Y}_j \approx \sum_{i=1}^{d} \beta_i' \frac{\tilde{X}_j^i}{\mathcal{S}(\tilde{X}^i)} + \beta_0',\ \tilde{Y}_j = Y_j - \overline{Y},\ \tilde{X}_j = X_j - \overline{X} \Rightarrow \mathcal{S}(\tilde{X}^i) = S(X^i)$$

  У такому разі цільовий функціонал матиме іншу форму. Розпишемо його:

У записах $\vec{1} = (1, \dots, 1)^T \in \mathbb{R}^N$, $\beta = (\beta_1, \dots, \beta_d)^T$, $I = \text{diag}\left((1, \dots, 1)\right) \in \mathbb{R}^{d \times d}$, $\lambda \geq 0$.

$$
\begin{aligned}
J(\beta, \beta_0) &= ||Y - (X\beta + \vec{1}\beta_0)||^2 + \lambda||\beta||^2 = \\
&= (Y^T - (\beta^T X^T + \vec{1}^T\beta_0))(Y - (X\beta + \vec{1}\beta_0)) + \lambda\beta^T\beta = \\
&= ||Y||^2 + \beta^T(A + \lambda I)\beta + \beta_0 N \beta_0 - 2Y^T X\beta - 2Y^T\vec{1}\beta_0 + 2\beta^T X^T\vec{1}\beta_0 = \\
&= ||Y||^2 + \left\langle Q\hat{\beta}, \hat{\beta} \right\rangle - \left\langle \vec{c}, \hat{\beta} \right\rangle, \text{ де} \\
\hat{\beta} &= \begin{pmatrix} \beta_0 \\ \beta \end{pmatrix}, Q = \begin{pmatrix} N & \sum_j X_j \\ (\sum_j X_j)^T & A + \lambda I \end{pmatrix}, \sum_j X_j = (\sum_{j=1}^N X_j^1, \dots, \sum_{j=1}^N X_j^d),\ A = X^T X \\
\vec{c} &= 2 \cdot \begin{pmatrix} \sum_{j=1}^N Y_j \\ X^T Y \end{pmatrix}
\end{aligned}
$$

J - опуклий функціонал, бо Q - симетрична невід'ємна матриця, тому якщо існує мінімум, то він глобальний. Функція гладка на $\mathbb{R}^{d+1}$, тому екстремуми знаходимо з умови оптимальності:

$$J'(\hat{\beta}) = 2Q\beta - \vec{c} = 0 \Leftarrow Q\beta = \frac{1}{2}\vec{c}$$

Розв'язок єдиний, коли існує Q^{-1}, і матиме вигляд:

$$\hat{\beta}_* = \frac{1}{2}Q^{-1}\vec{c} = \begin{pmatrix} N & \sum_j X_j \\ (\sum_j X_j)^T & A + \lambda I \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^N Y_j \\ X^T Y \end{pmatrix} \tag{1}$$

Якщо X, Y – центровані, то (1) спрощується:

$$\hat{\beta}_* = \begin{pmatrix} N & \vec{0} \\ \vec{0}^T & A + \lambda I \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ X^T Y \end{pmatrix},\ \vec{0} = (0, \dots, 0) \in \mathbb{R}^d$$

А тому зсув у перетвореній моделі $\hat{\beta}_{*,0}$ рівний нулю, звідси маємо, що у початковій моделі зсув визначається виключно через середнє Y, вибіркові характеристики X^i та коефіцієнти $\hat{\beta}_{*,j}$ при перетворених змінних. Остаточно, коефіцієнти у початковій моделі будуть такими:

$$\beta_0 = \overline{Y} - \sum_{i=1}^d \beta_i \overline{X^i},\ \beta_i = \frac{\beta_i'}{S(X^i)},\ i = \overline{1, d}$$

Маючи певне розуміння того, як реалізована підгонка рідж-моделі в пакеті MASS, можна переходити до практичної частини.

2.2 Підгонка регресійної моделі.

Вважаємо, що значення штрафувального множника $\lambda \in \Lambda \subset \mathbb{R}^+$ є оптимальним, якщо функціонал крос-валідації $\text{CV}(\lambda)$ набуває найменшого значення на Λ. Ми будемо розглядати $\Lambda = [0.01, 50]$.

2.2.1 Підгонка за всіма сесіями.

Штраф на зсув	Наявний	Відсутній
λ	0.027	0.739
Зсув	3.639276267	3.607062048
clf	0.005068145	0.004956264
clx	0.454634888	0.456197540
cma	-0.328103570	-0.327703610
cmcsa	-0.014915118	-0.013732471
cme	0.055481190	0.055391471
cmg	-0.011029711	-0.010904072
cmi	0.032482657	0.032539101
cms	0.853266287	0.842294877
cnp	0.111596075	0.115621303

Табл. 1: Оптимальні множники та значення оцінок коефіцієнтів. Повні дані.

Як видно з таблиці, значення оцінок за двома підходами досить схожі між собою. Крім того, ми бачимо, що ці оцінки є близькими до оцінок класичного МНК. А як ми пам'ятаємо з результатів у першій роботі, за повними даними підгонка виходила невдалою.

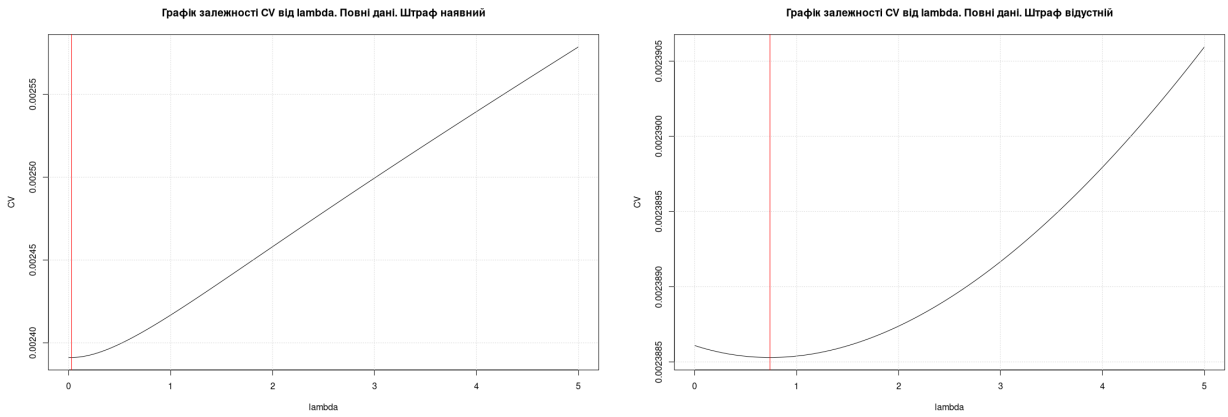

Рис. 1: Графік функціонала крос-валідації для повних даних.

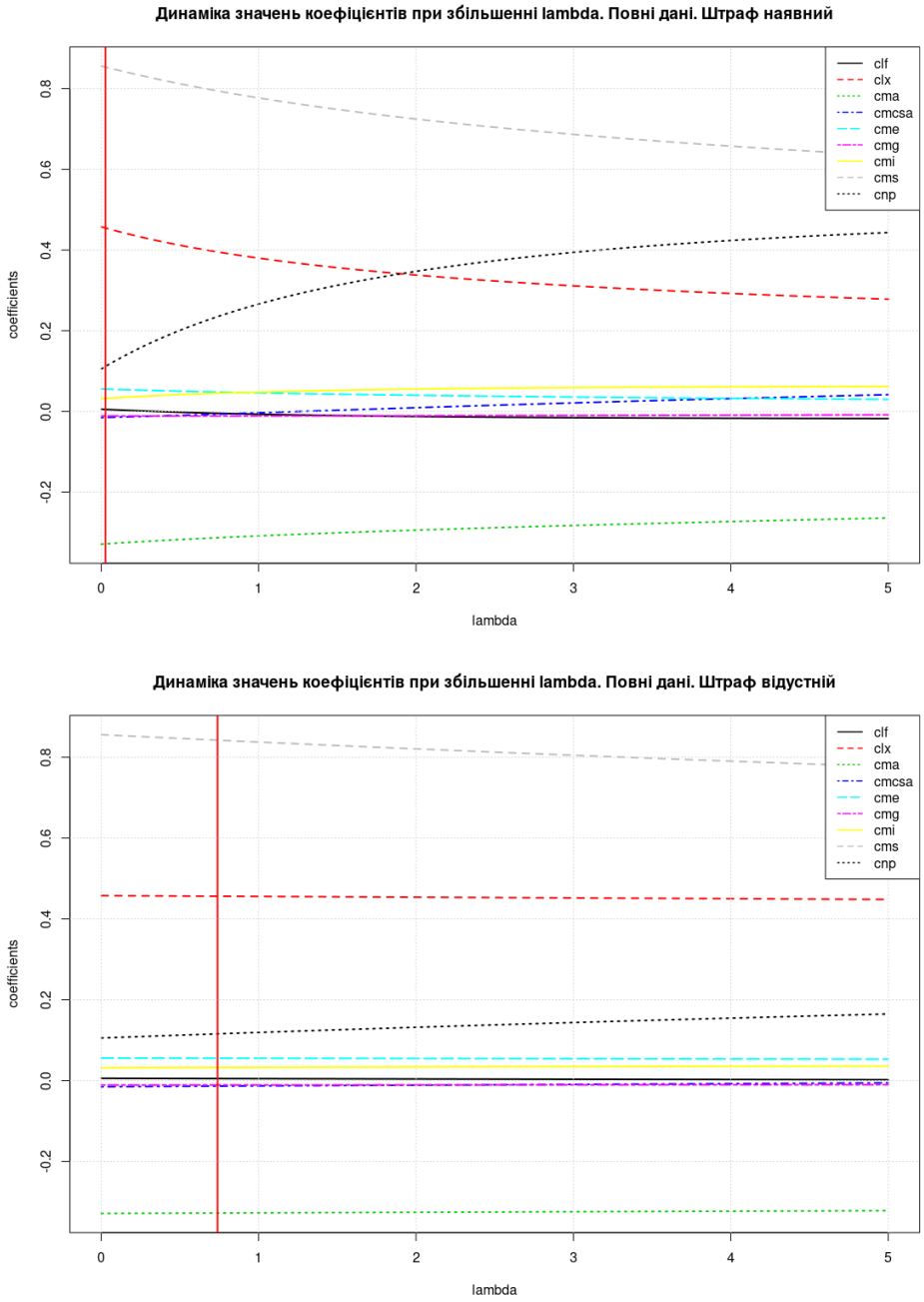

Рис. 2: Графік зміни значень коефіцієнтів при змінних для повних даних.

2.2.2 Підгонка за "свіжими" даними.

Штраф на зсув	Наявний	Відсутній
λ	0.001	0.236
Зсув	9.1977737632	9.210109838
clf	-0.0156799669	-0.023422840
clx	0.4283173516	0.524676843
cma	0.1505901689	0.162487257
cmcsa	-0.1701064968	-0.275094900
cme	-0.1607642659	-0.187447135
cmg	0.0135449677	0.004674947
cmi	0.0009470528	-0.032642131
cms	-0.3304507067	-0.835167543
cnp	1.2802751700	2.071487163

Табл. 2: Оптимальні множники та значення оцінок коефіцієнтів. Свіжі дані.

Якщо штрафувати зсув, то оптимальність досягається при зовсім малих значеннях штрафувального множника, що свідчить про те, що отримана оцінка близька до класичної. Якщо не штрафувати зсув, то результати виходять цікавіші.

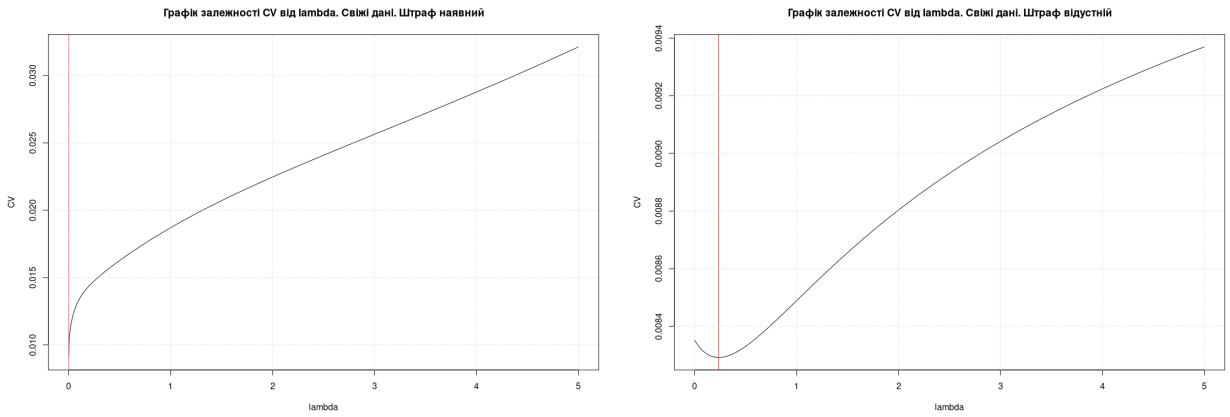

Рис. 3: Графік функціонала крос-валідації для свіжих даних.

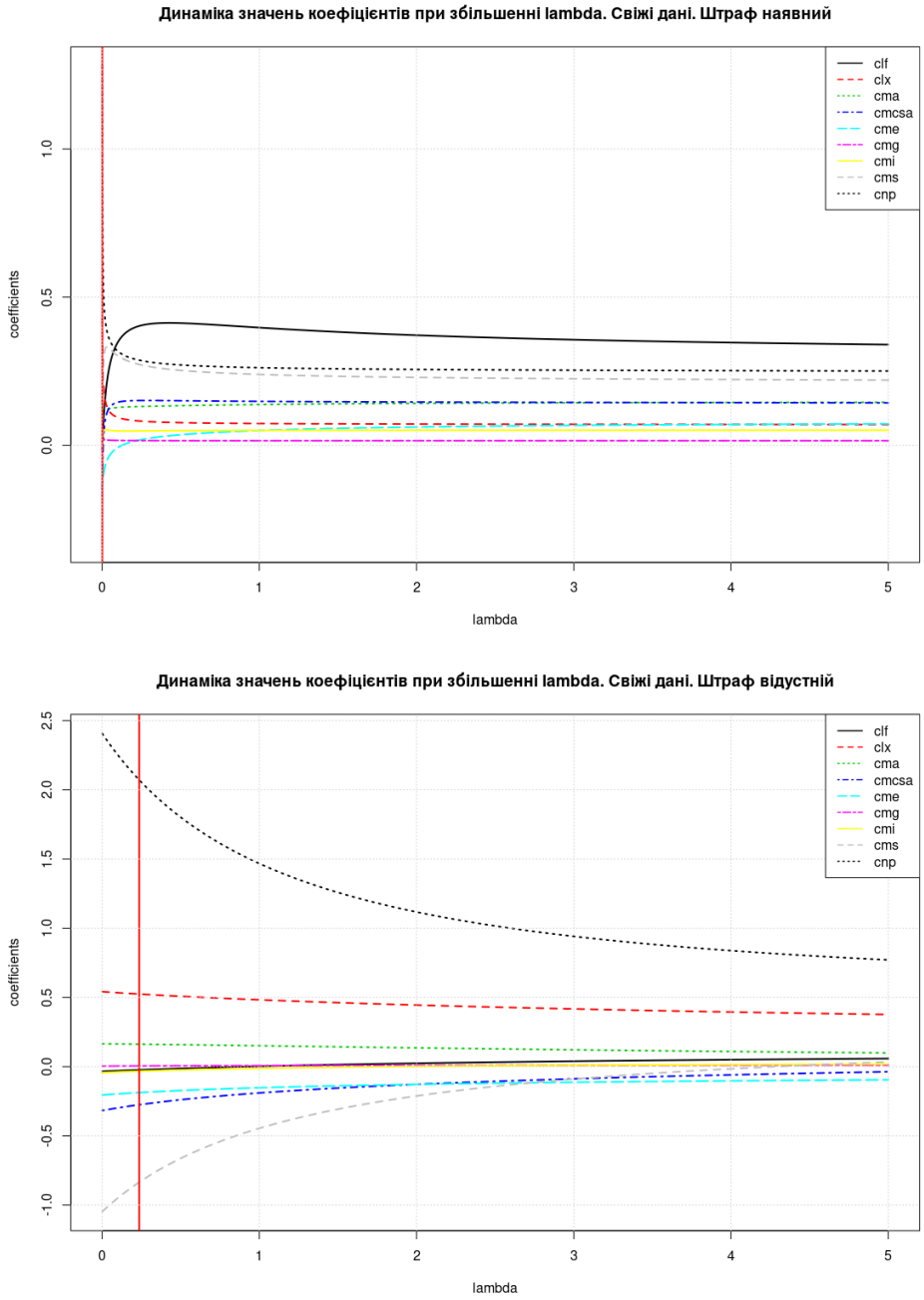

Рис. 4: Графік зміни значень коефіцієнтів при змінних для свіжих даних.

Цікава картина коли штраф накладається на зсув. При зменшенні значення штрафувального множника, значення при змінній cms ”летить” вгору.

2.3 Прогнозування.

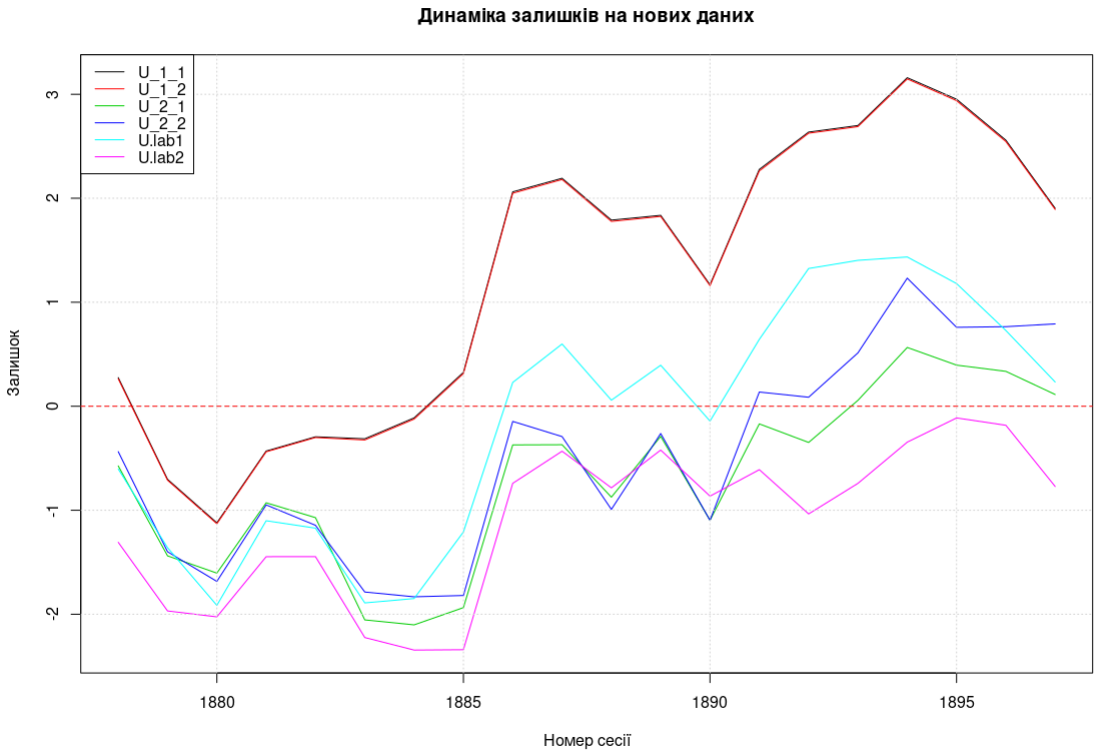

Рис. 5: Порівняння залишків прогнозу на нових даних для різних моделей.

Поведінка залишків для рідж-моделей на повних даних досить схожа з тим, що спостерігали для класичної моделі на повних даних з першої роботи. Кращим вийшов розкид для рідж-моделей на свіжих даних. Навіть можна побачити, що коливання залишків трохи менше за коливання залишків у найкращій моделі з першої роботи. Але, зауважимо, що для рідж-моделей оцінка коефіцієнтів є зсунутою, тому оцінка похибок теж буде зсунутою. Хоча, як бачимо, це дещо компенсується меншим розкидом.

3 Висновки.

У рідж-моделі ми жертвуємо незміщеністю оцінки, аби отримати менший розкид під час прогнозування, як нам вдалося, у першому наближенні, переконатися у даній роботі.