МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА

Факультет прикладної математики та інформатики

ЗВІТ

до індивідуального завдання №5

з дисципліни «Моделі статистичного навчання»

Виконав

студент групи ПМіМ-12:

Зелінський Олександр

Перевірив:

Проф. Заболоцький Т. М.

Львів – 2021

**Хід виконання**

**1. Вибір найкращої підмножини**

**1.1**

**Зображення, що містить текст

Автоматично згенерований опис**

**1.2**

Зображення, що містить текст

Автоматично згенерований опис

**1.3**

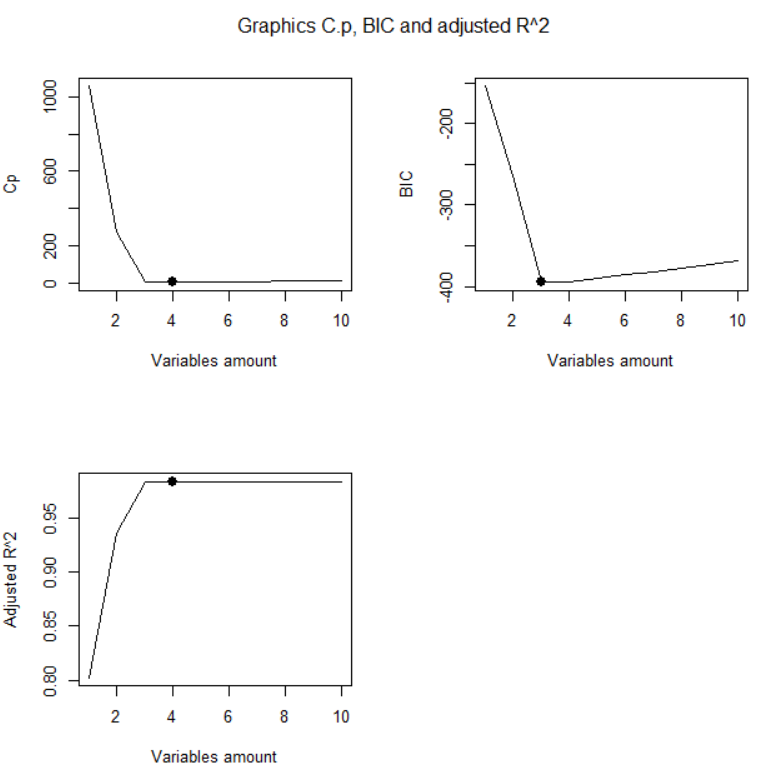
Зображення, що містить текст

Автоматично згенерований опис

Зображення, що містить стіл

Автоматично згенерований опис

В результаті можемо побачити найкращий набір змінних для кожної розмірності моделі, також видно, що \* позначають те, що змінна включена у відповідну модель.

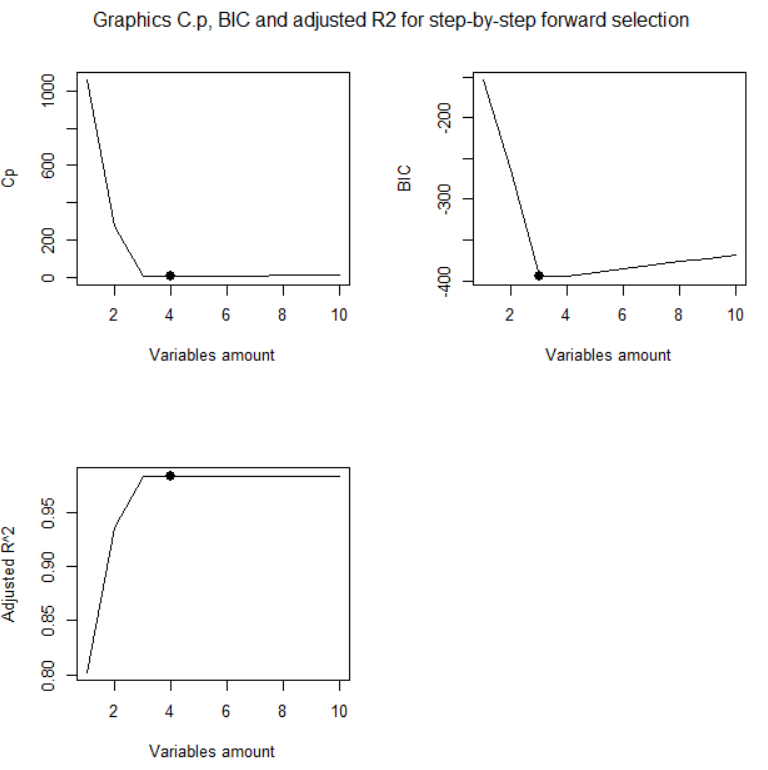


З огляду на результати видно, що найкраща модель за показниками та скорегованим – це модель з 4 змінними , а для показника BIC бачимо, що вже найкращою буде модель зі змінними .

**1.4**

Зображення, що містить текст

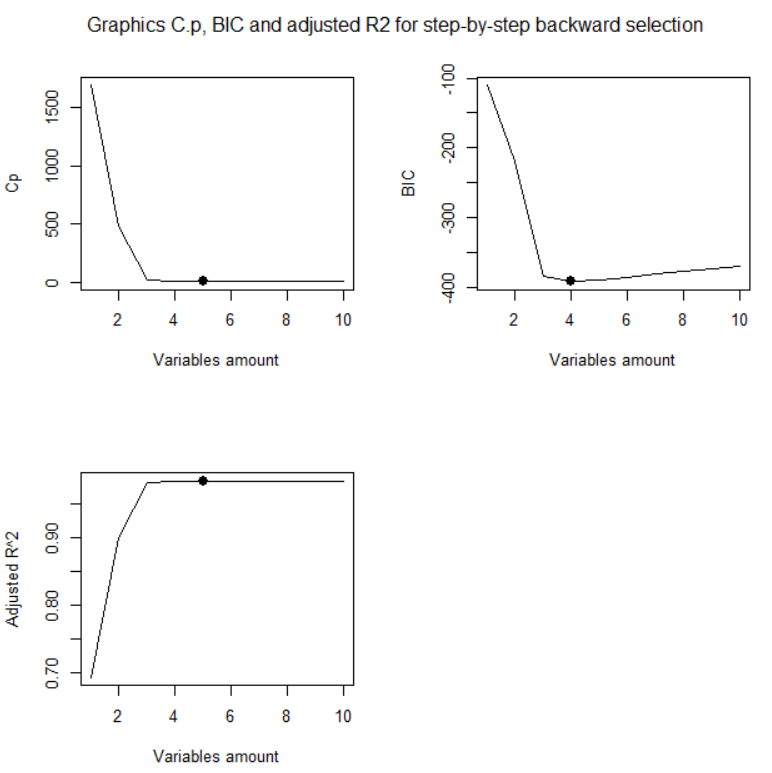
Автоматично згенерований опис



З огляду на результати видно, що для покрокового вибору вперед найкраща модель за показниками та скорегованим – це модель з 4 змінними , а для показника BIC бачимо, що вже найкращою буде модель зі змінними .

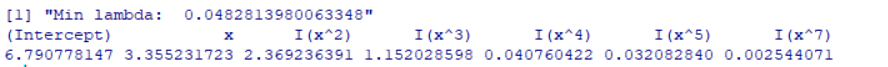
Зображення, що містить текст

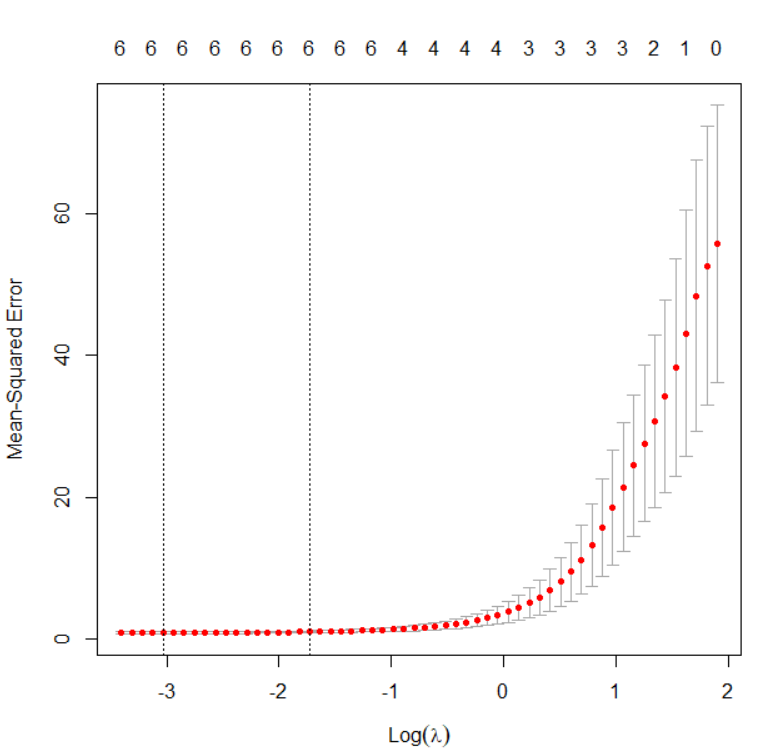
Автоматично згенерований опис



З огляду на результати видно, що для покрокового вибору назад найкраща модель за показниками та скорегованим – це модель з 5 змінними , а для показника BIC бачимо, що вже найкращою буде модель з 4 змінними .

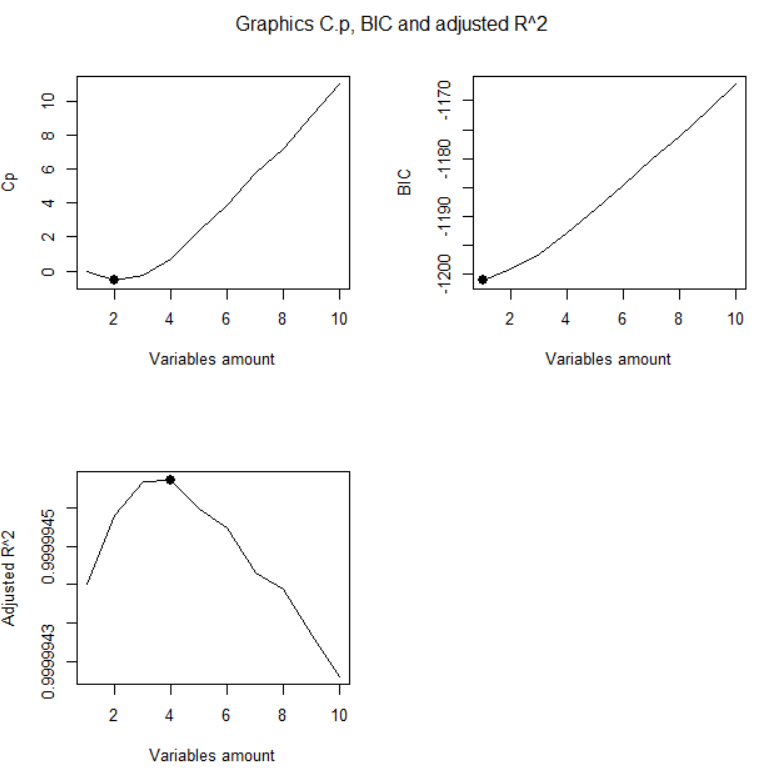
**1.5**





Згідно з графіком, при значеннях логарифма від лямбди більших за нуль помилка зростає. Лямба для якої помилка буде найменшою дорівнює 0.048, а модель включає в себе змінні .

**1.6**



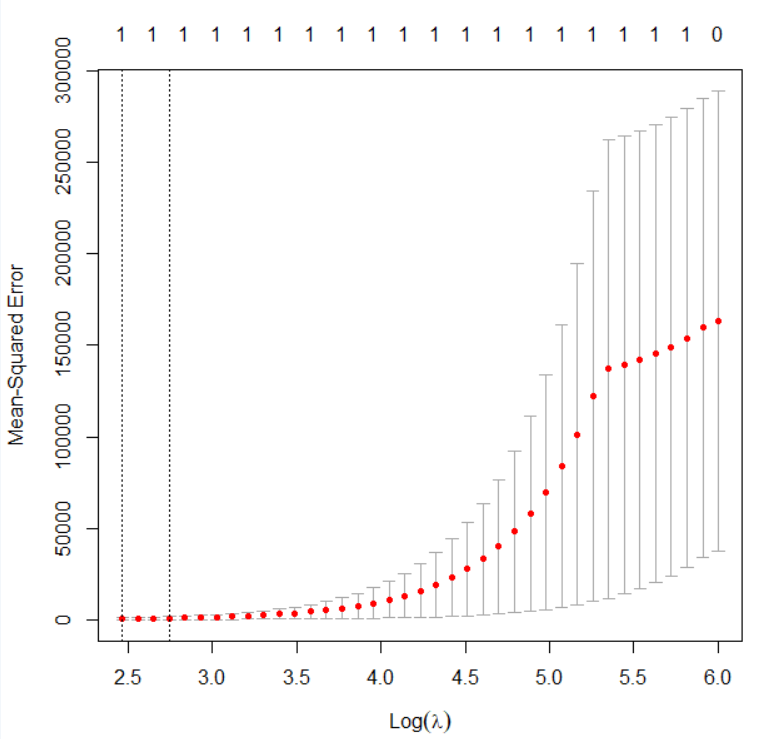
З огляду на результати видно, що найкраща модель за показниками – , за скорегованим – це модель , а для показника BIC бачимо, що вже найкращою буде модель зі змінними .

Оскільки , то можемо з впевненістю сказати, що метод найкращого вибору підмножини з використанням BIC визначає найбільш точну модель з однією змінною.

Зображення, що містить текст

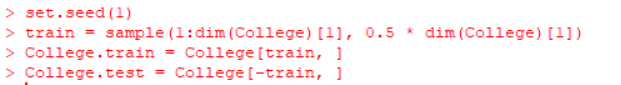
Автоматично згенерований опис

В методі лассо, ми знайшли лямбду для якої помилка мінімальна, яка дорівнює 7.41 і з допомогою неї знайшли найкращу модель, яка включає одну змінну .



**2. Collage**

**2.1**



**2.2**

Зображення, що містить текст

Автоматично згенерований опис

**2.3**

Зображення, що містить текст

Автоматично згенерований опис

**2.4**

Зображення, що містить текст

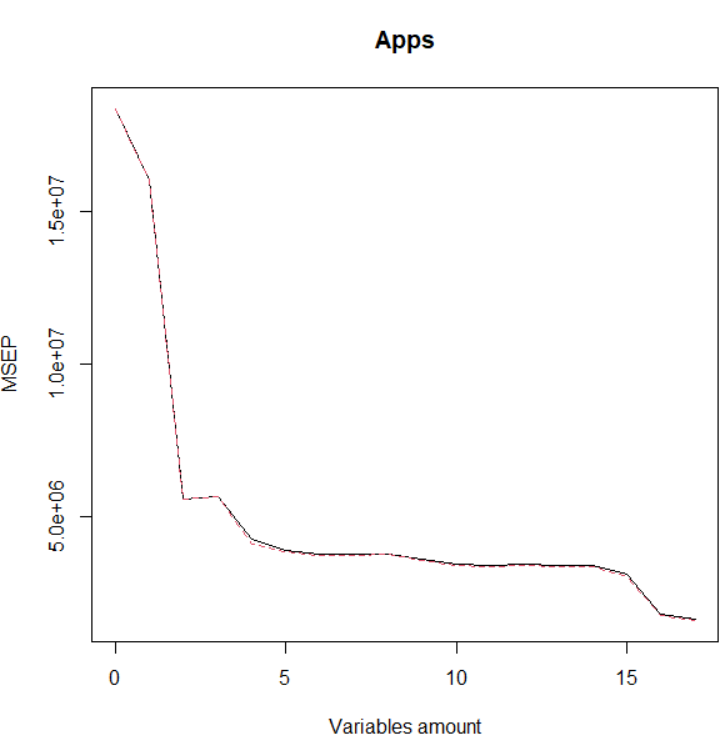
Автоматично згенерований опис

Можемо бачити, що жоден з коефіцієнтів не рівний нулю

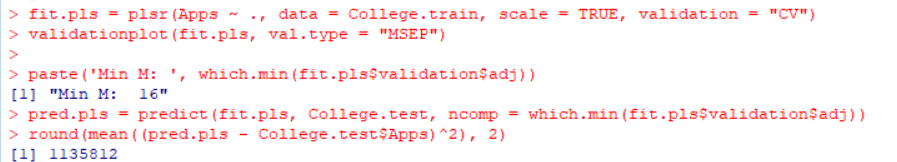
**2.5**

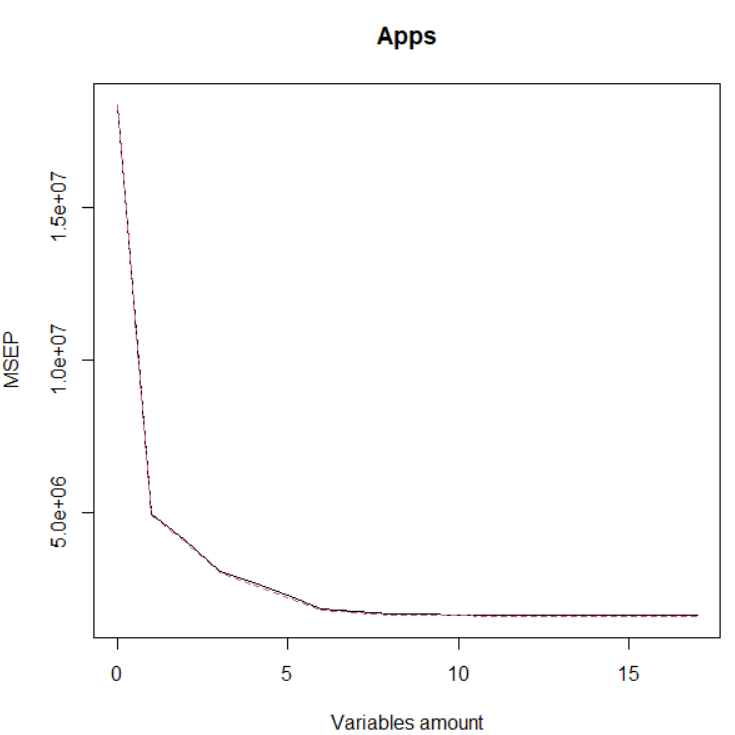
Зображення, що містить текст

Автоматично згенерований опис



**2.6**





**2.7**

Зображення, що містить текст

Автоматично згенерований опис

З обчислених коефіцієнтів детермінації видно, що є найближчийм (90.1744%) для моделі лассо і найнижчим для моделі PLS (90,1536%). Різниця між тестовими помилками не значна (~2000).

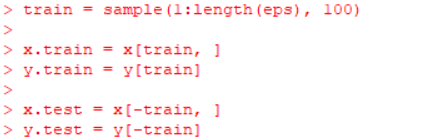
**3. Collage**

**3.1**

Зображення, що містить текст

Автоматично згенерований опис

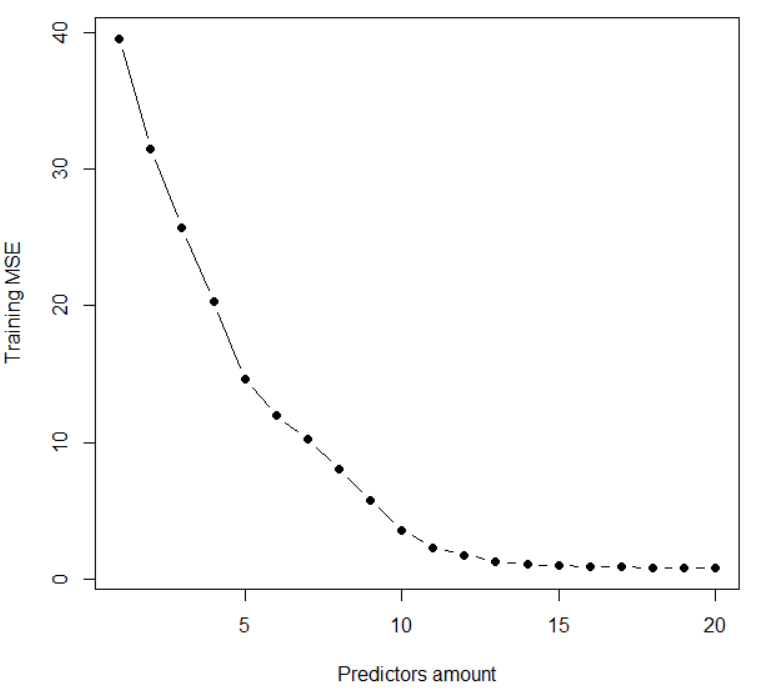
**3.2**



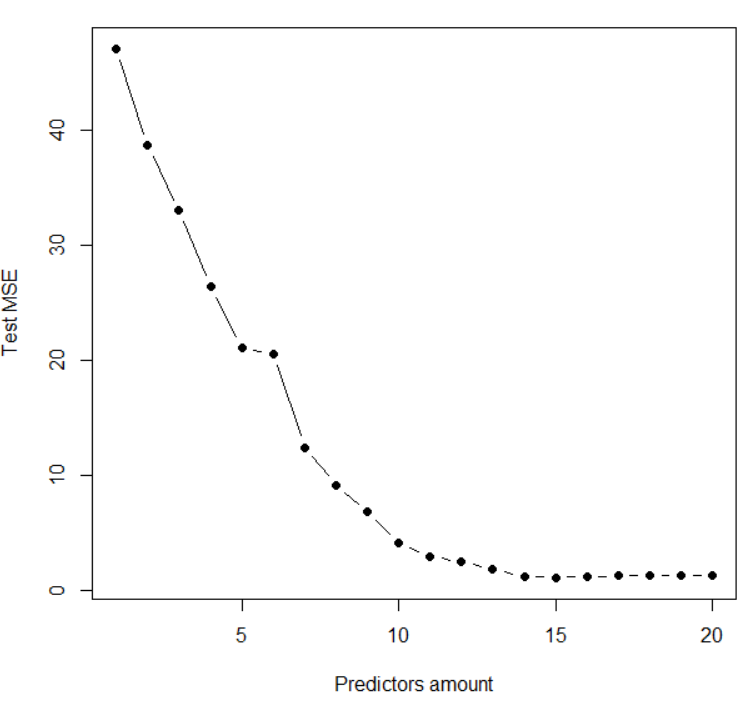
**3.3**

Зображення, що містить текст

Автоматично згенерований опис



**3.4**

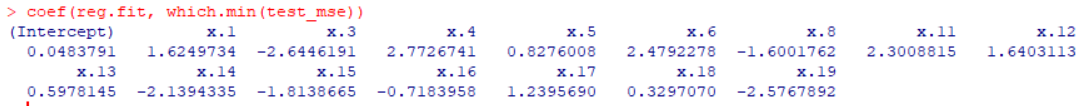


**3.5**



Видно, що модель має 15 змінних, а MSE дорівнює 1.17.

**3.6**

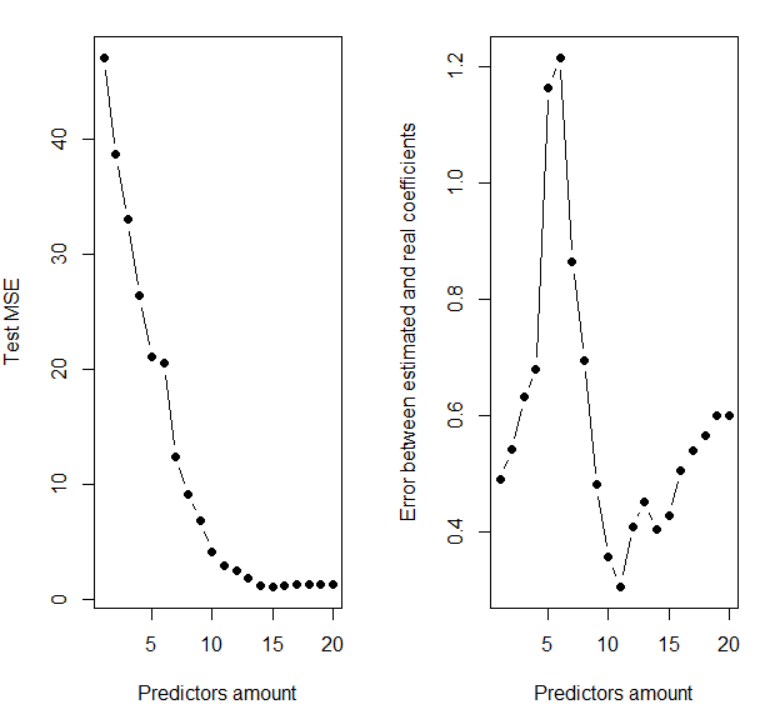


З результату видно, що модель правильно визначила нульові коефіцієнти і виключила їх з моделі, а решта оцінок є точною.

**3.7**

Зображення, що містить текст

Автоматично згенерований опис

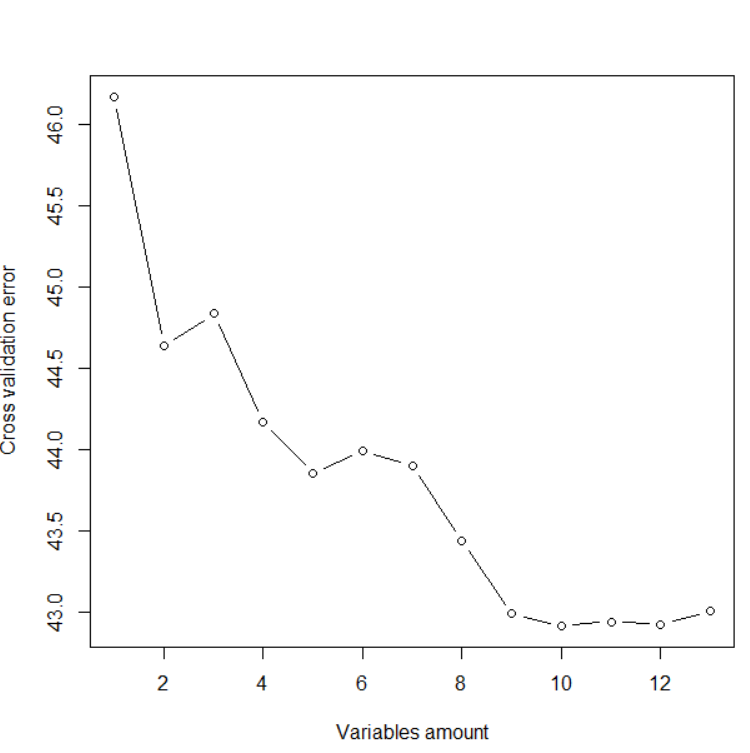


В результаті видно, що модель з одинадцятьма предикторами мінімізує помилку між оціночними та справжніми значеннями коефіцієнтів.

**4. Boston**

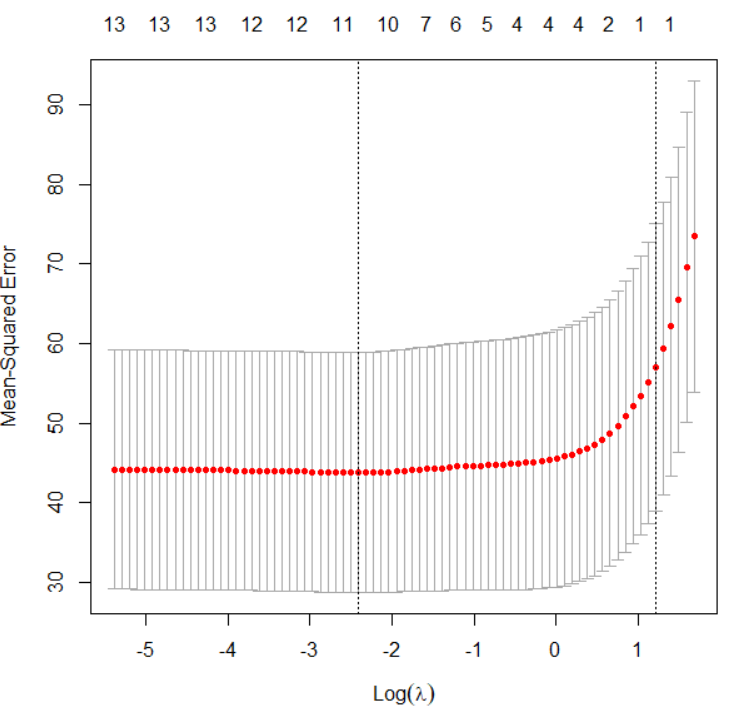
Зображення, що містить текст

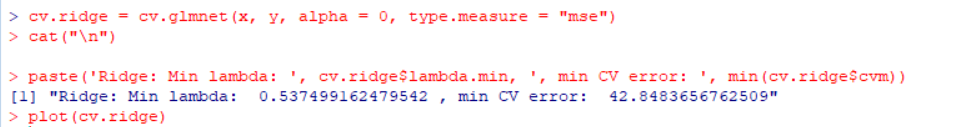
Автоматично згенерований опис

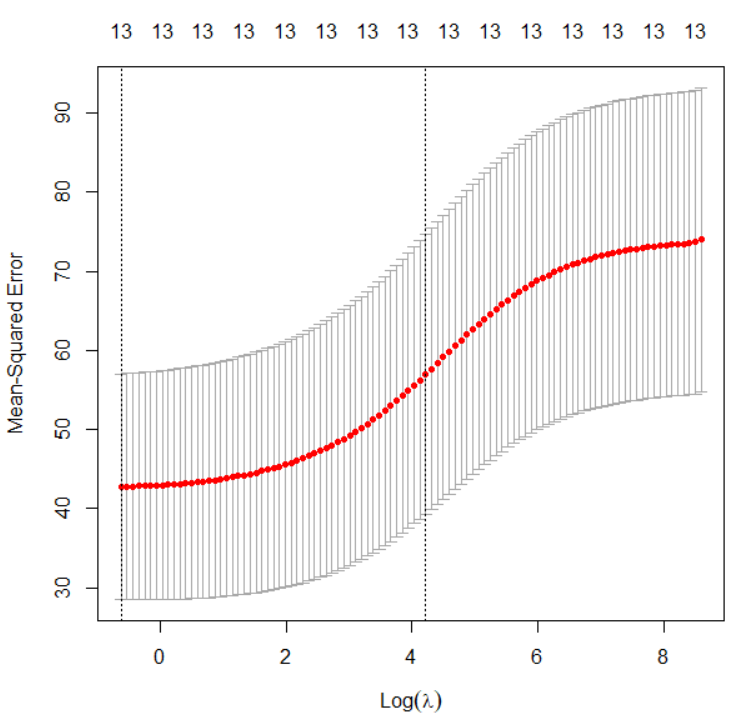


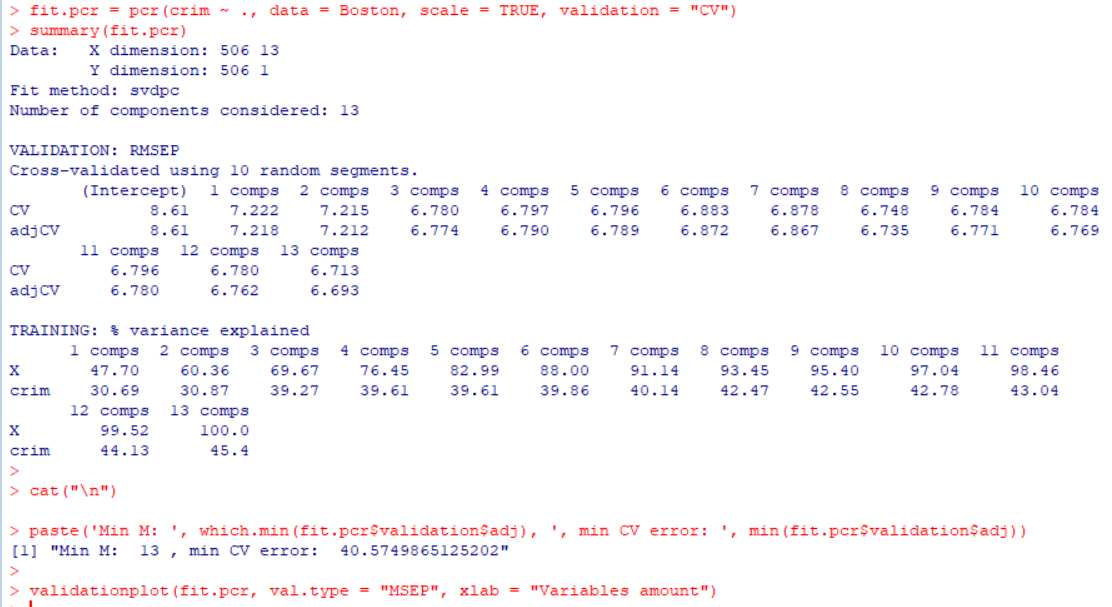
Зображення, що містить текст

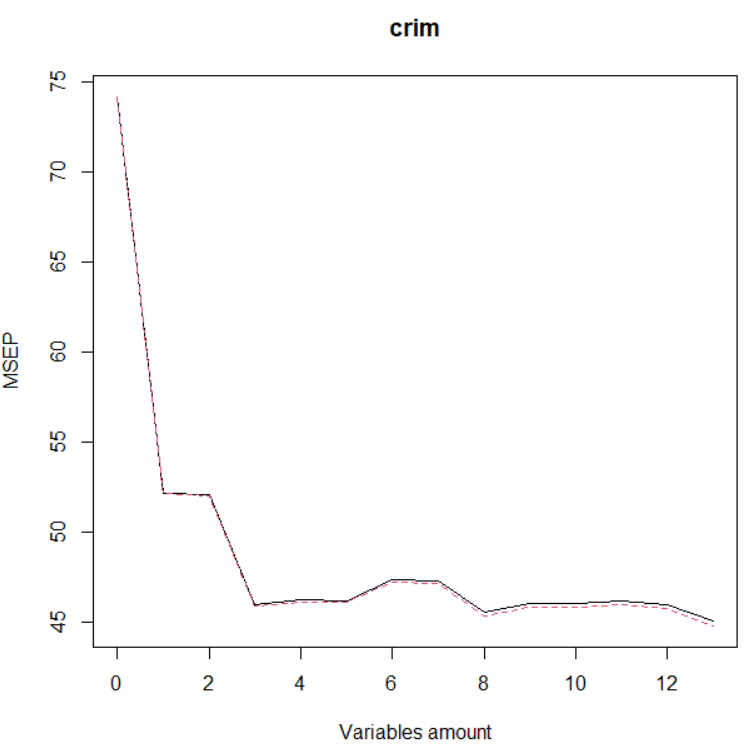
Автоматично згенерований опис











Зважаючи на результати, можна сказати, що найнижчу помилку має метод PCR, а саме 40.575, найвищу помилку має метод лаcо – 43.796. При чому, модель з найнижчою помилкою для методу найкращого вибору підмножини має 13 предикторів.