

Моделі статистичного навчання: вибір лінійної моделі та
регуляризація

Точність прогнозування:

Точність прогнозування: за умови, що справжня залежність між залежною змінною і предикторами є приблизно лінійною, оцінки отримані методом найменших квадратів матимуть невелике зміщення.

Точність прогнозування: за умови, що справжня залежність між залежною змінною і предикторами є приблизно лінійною, оцінки отримані методом найменших квадратів матимуть невелике зміщення. Якщо $n \gg p$ тоді ці оцінки матимуть низьку дисперсію.

Точність прогнозування: за умови, що справжня залежність між залежною змінною і предикторами є приблизно лінійною, оцінки отримані методом найменших квадратів матимуть невелике зміщення. Якщо $n \gg p$ тоді ці оцінки матимуть низьку дисперсію. Однак якщо n не набагато більше за p , то оцінкам може бути притаманна велика дисперсія, що призводить в свою чергу до переоцінювання і, як наслідок, поганих прогнозів для майбутніх спостережень, що не використовувалися при навчанні моделі.

Точність прогнозування: за умови, що справжня залежність між залежною змінною і предикторами є приблизно лінійною, оцінки отримані методом найменших квадратів матимуть невелике зміщення. Якщо $n \gg p$ тоді ці оцінки матимуть низьку дисперсію. Однак якщо n не набагато більше за p , то оцінкам може бути притаманна велика дисперсія, що призводить в свою чергу до переоцінювання і, як наслідок, поганих прогнозів для майбутніх спостережень, що не використовувалися при навчанні моделі. Якщо ж $p > n$, то не існує єдиної оцінки коефіцієнтів методом найменших квадратів. Шляхом накладання обмежень або стиснення коефіцієнтів, ми часто можемо істотно зменшити дисперсію за рахунок незначного збільшення зміщення. Це часто призводить до значного покращення точності прогнозування.

Інтерпретація моделі:

Інтерпретація моделі: Часто буває, що деякі або багато змінних насправді не пов'язані з залежною змінною. Включення таких нерелевантних змінних призводить до непотрібної складності моделі. Видаливши ці змінні - тобто встановивши значення оцінок відповідних коефіцієнтів, що дорівнює нулю - можемо отримати модель, яка легше інтерпретується.

Ми обговоримо три важливі класи методів.

Ми обговоримо три важливі класи методів.

- Вибір підмножини. Цей підхід передбачає виявлення підмножини з p предикторів, які ми вважаємо пов'язаними з залежною змінною. Тоді ми оцінюємо модель методом найменших квадратів на скороченому наборі змінних.

Ми обговоримо три важливі класи методів.

- Вибір підмножини. Цей підхід передбачає виявлення підмножини з p предикторів, які ми вважаємо пов'язаними з залежною змінною. Тоді ми оцінюємо модель методом найменших квадратів на скороченому наборі змінних.
- Стиснення. Цей підхід передбачає оцінювання моделі, що включає всі p предикторів. Однак оцінені коефіцієнти стискаються до нуля відносно оцінок найменших квадратів. Ця процедура (також відома як регуляризація) має наслідком зменшення дисперсії. Залежно від типу стискання, деякі коефіцієнти можуть бути оцінені нулем.

Ми обговоримо три важливі класи методів.

- Вибір підмножини. Цей підхід передбачає виявлення підмножини з p предикторів, які ми вважаємо пов'язаними з залежною змінною. Тоді ми оцінюємо модель методом найменших квадратів на скороченому наборі змінних.
- Стиснення. Цей підхід передбачає оцінювання моделі, що включає всі p предикторів. Однак оцінені коефіцієнти стискаються до нуля відносно оцінок найменших квадратів. Ця процедура (також відома як регуляризація) має наслідком зменшення дисперсії. Залежно від типу стискання, деякі коефіцієнти можуть бути оцінені нулем.
- Зменшення вимірності. Цей підхід передбачає проектування p -предикторів в M -мірний підпростір, де $M < p$. Це досягається шляхом обчислення M різних лінійних комбінацій або проекцій змінних. Тоді ці M -проекції використовуються як предиктори при оцінці моделі методом найменших квадратів.

Алгоритм. Вибір найкращої підмножини

Алгоритм. Вибір найкращої підмножини

1. Нехай M_0 позначає нульову модель, яка не містить предикторів. Це модель, яка просто передбачає середнє значення вибірки для кожного спостереження.

Алгоритм. Вибір найкращої підмножини

1. Нехай M_0 позначає нульову модель, яка не містить предикторів. Це модель, яка просто передбачає середнє значення вибірки для кожного спостереження.

2. Для $k = 1, 2, \dots, p$:

(а) Оцінюють всі C^p_k моделі, які містять точно k предикторів.

Алгоритм. Вибір найкращої підмножини

1. Нехай M_0 позначає нульову модель, яка не містить предикторів. Це модель, яка просто передбачає середнє значення вибірки для кожного спостереження.

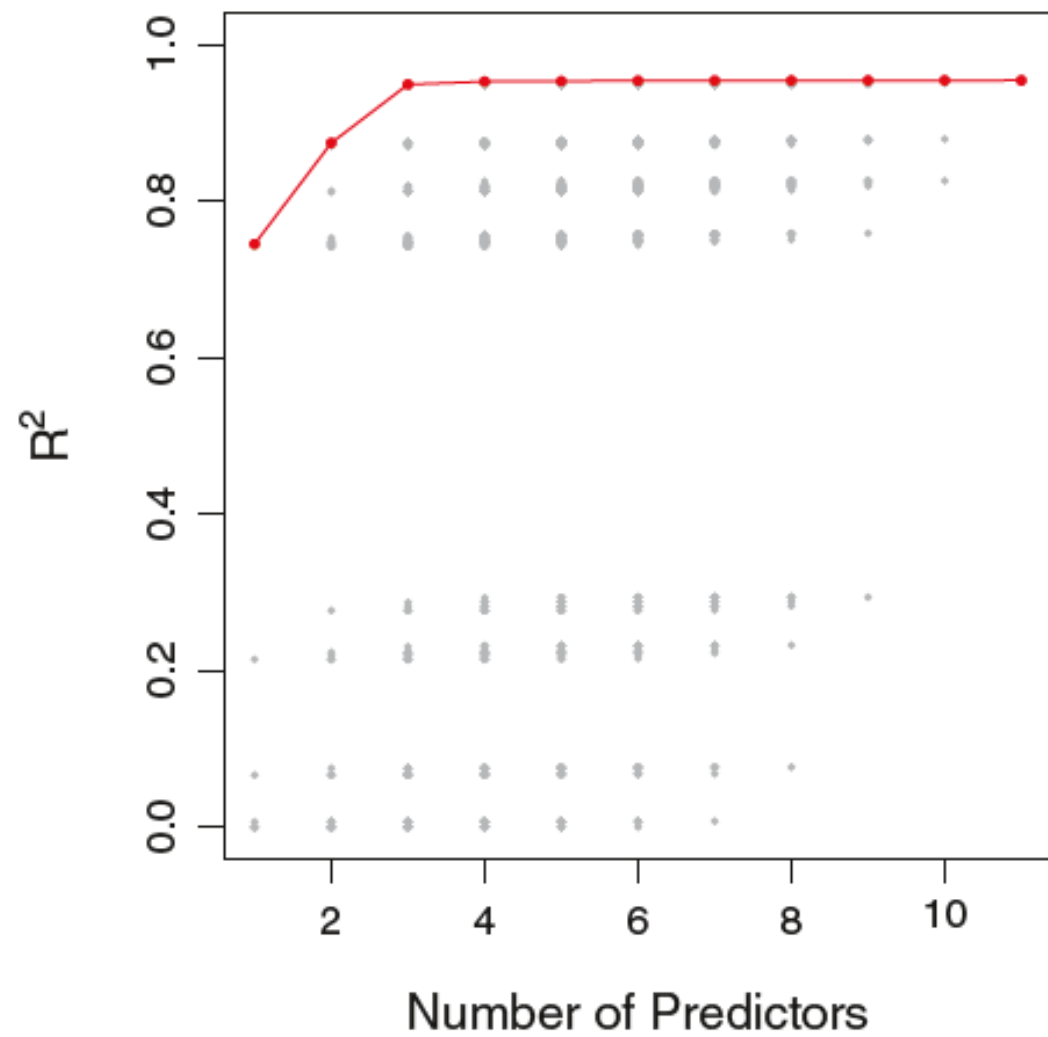
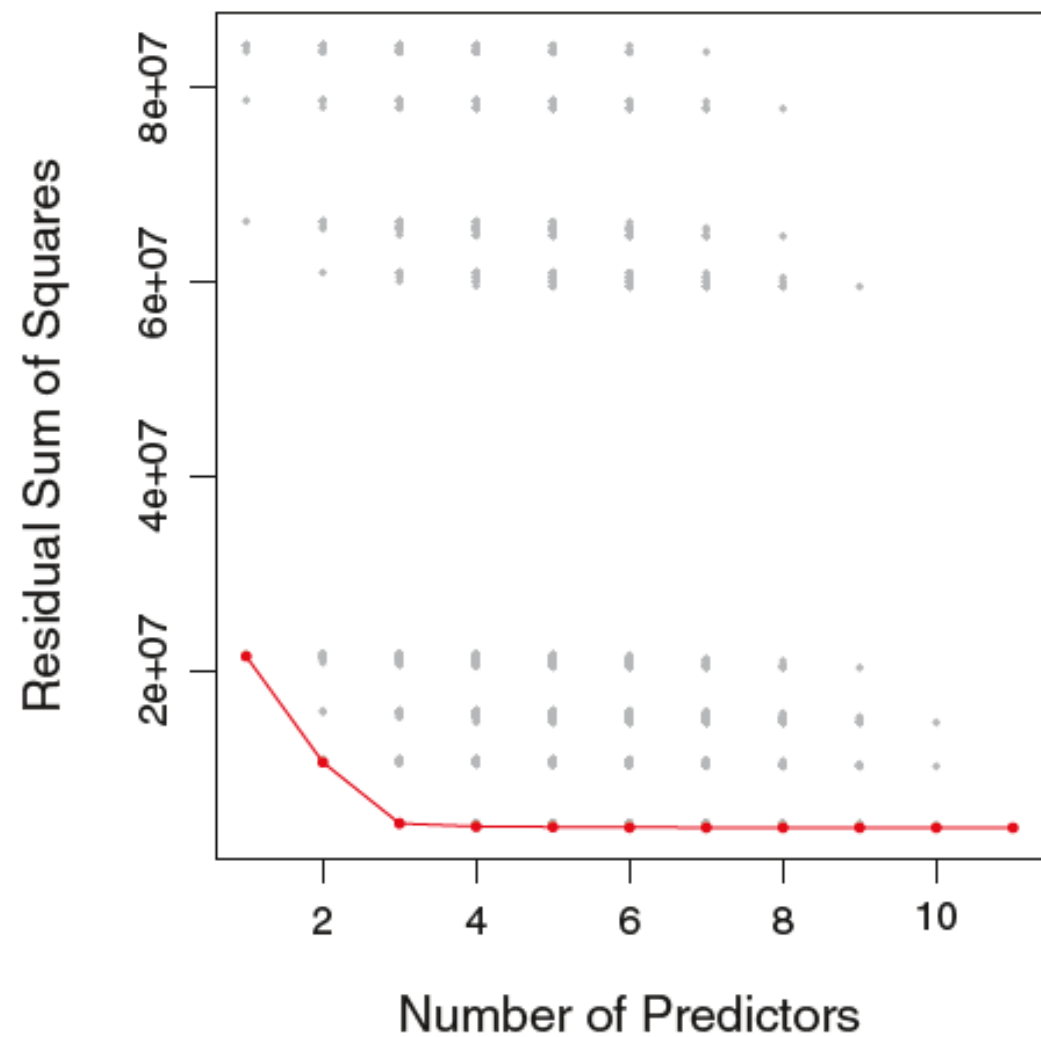
2. Для $k = 1, 2, \dots, p$:

(a) Оцінюють всі C^p_k моделі, які містять точно k предикторів.

(b) Вибирають найкращу модель серед розглянутих C^p_k моделей, і називають це M_k .

Алгоритм. Вибір найкращої підмножини

1. Нехай M_0 позначає нульову модель, яка не містить предикторів. Це модель, яка просто передбачає середнє значення вибірки для кожного спостереження.
2. Для $k = 1, 2, \dots, p$:
 - (a) Оцінюють всі C^p_k моделі, які містять точно k предикторів.
 - (b) Вибирають найкращу модель серед розглянутих C^p_k моделей, і називають це M_k .
3. Вибирають одну найкращу модель із M_0, \dots, M_p .



Алгоритм. Покроковий вибір вперед

Алгоритм. Покроковий вибір вперед

1. Нехай M_0 позначає нульову модель, яка не містить предикторів. Це модель, яка просто передбачає середнє значення вибірки для кожного спостереження.

Алгоритм. Покроковий вибір вперед

1. Нехай M_0 позначає нульову модель, яка не містить предикторів. Це модель, яка просто передбачає середнє значення вибірки для кожного спостереження.

2. Для $k = 0, 2, \dots, p - 1$:

(а) Оцінюють всі $p - k$ моделі, які додають до M_k один предиктор.

Алгоритм. Покроковий вибір вперед

1. Нехай M_0 позначає нульову модель, яка не містить предикторів. Це модель, яка просто передбачає середнє значення вибірки для кожного спостереження.

2. Для $k = 0, 2, \dots, p - 1$:

(a) Оцінюють всі $p - k$ моделі, які додають до M_k один предиктор.

(b) Вибирають найкращу модель серед розглянутих $p - k$ моделей, і називають її M_{k+1} .

Алгоритм. Покроковий вибір вперед

1. Нехай M_0 позначає нульову модель, яка не містить предикторів. Це модель, яка просто передбачає середнє значення вибірки для кожного спостереження.
2. Для $k = 0, 2, \dots, p - 1$:
 - (a) Оцінюють всі $p - k$ моделі, які додають до M_k один предиктор.
 - (b) Вибирають найкращу модель серед розглянутих $p - k$ моделей, і називають її M_{k+1} .
3. Вибирають одну найкращу модель із M_0, \dots, M_p .

# Variables	Best subset	Forward stepwise
One	rating	rating
Two	rating, income	rating, income
Three	rating, income, student	rating, income, student
Four	cards, income student, limit	rating, income, student, limit

Алгоритм. Покроковий вибір назад

Алгоритм. Покроковий вибір назад

1. Нехай M_p позначає модель, яка містить всі p предикторів.

Алгоритм. Покроковий вибір назад

1. Нехай M_p позначає модель, яка містить всі p предикторів.
2. Для $k = p, p - 1, \dots, 1$:
 - (а) Оцінюють всі k моделі, які містять на одного предиктора менше ніж M_k , тобто $k - 1$ предиктор.

Алгоритм. Покроковий вибір назад

1. Нехай M_p позначає модель, яка містить всі p предикторів.
2. Для $k = p, p - 1, \dots, 1$:
 - (а) Оцінюють всі k моделі, які містять на одного предиктора менше ніж M_k , тобто $k - 1$ предиктор.
 - (б) Вибирають найкращу модель серед розглянутих k моделей, і називають її M_{k-1} .

Алгоритм. Покроковий вибір назад

1. Нехай M_p позначає модель, яка містить всі p предикторів.
2. Для $k = p, p - 1, \dots, 1$:
 - (a) Оцінюють всі k моделі, які містять на одного предиктора менше ніж M_k , тобто $k - 1$ предиктор.
 - (b) Вибирають найкращу модель серед розглянутих k моделей, і називають її M_{k-1} .
3. Вибирають одну найкращу модель із M_0, \dots, M_p .

Два підходи

Два підходи

1. Непрямо оцінити тестову помилку пристосувавши навчальну помилку до врахування ефекту зміщення (C_p , AIC , BIC , скорегований R^2).

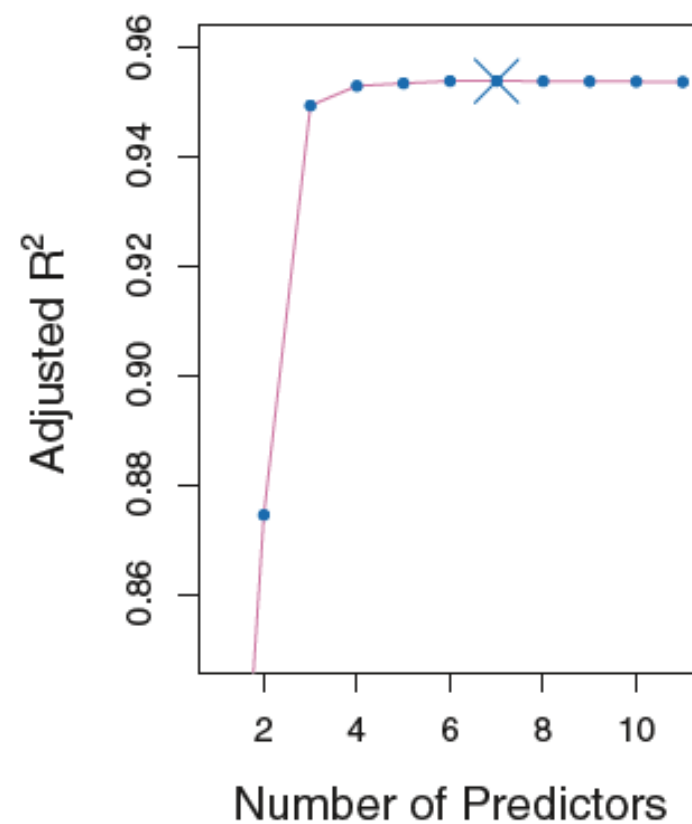
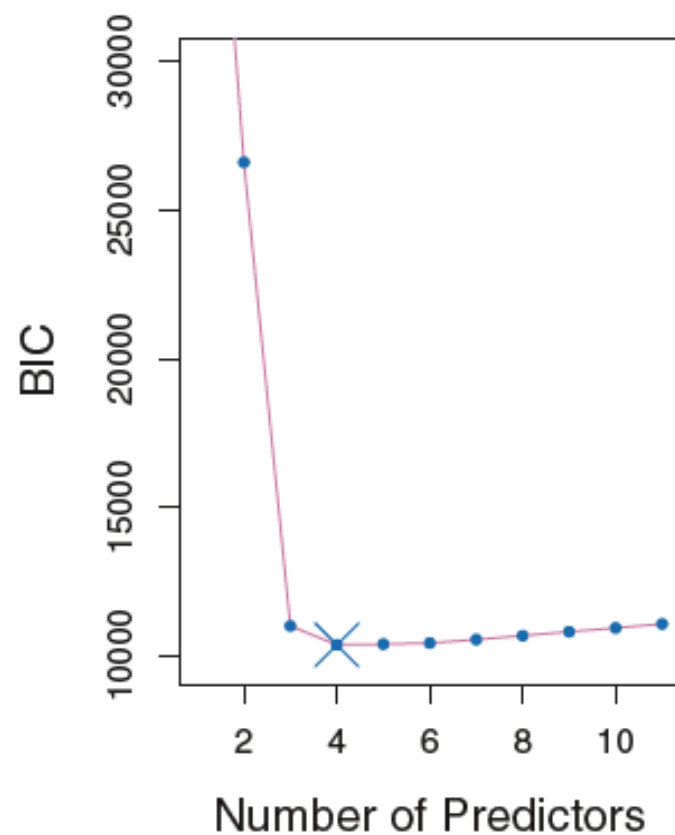
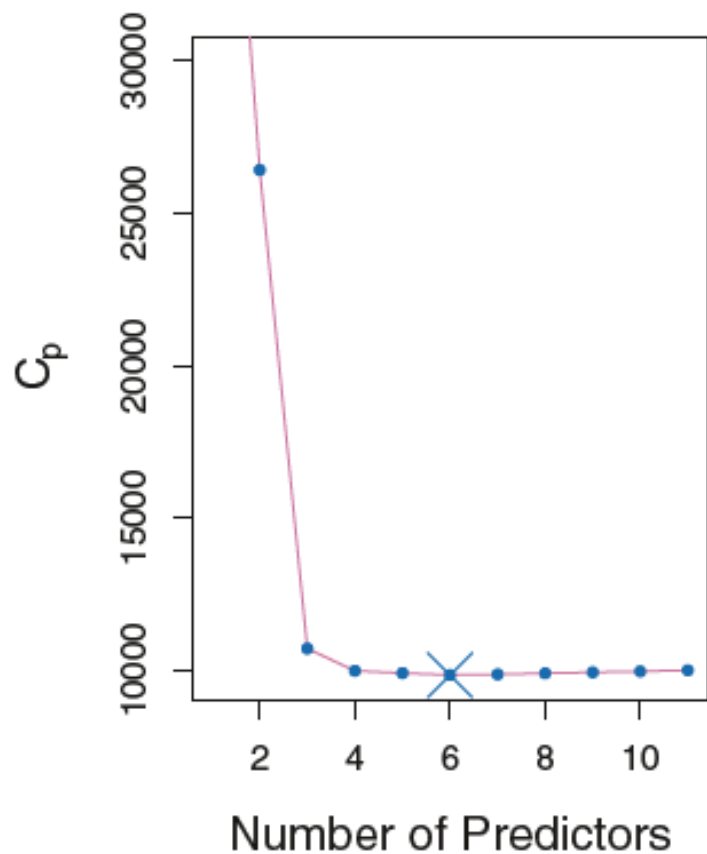
Два підходи

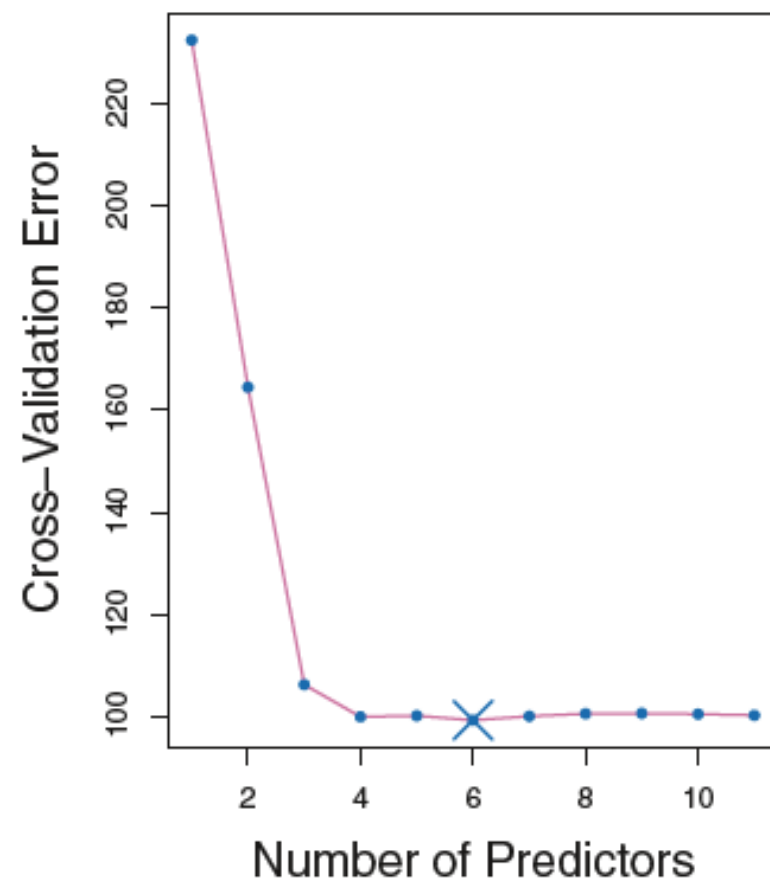
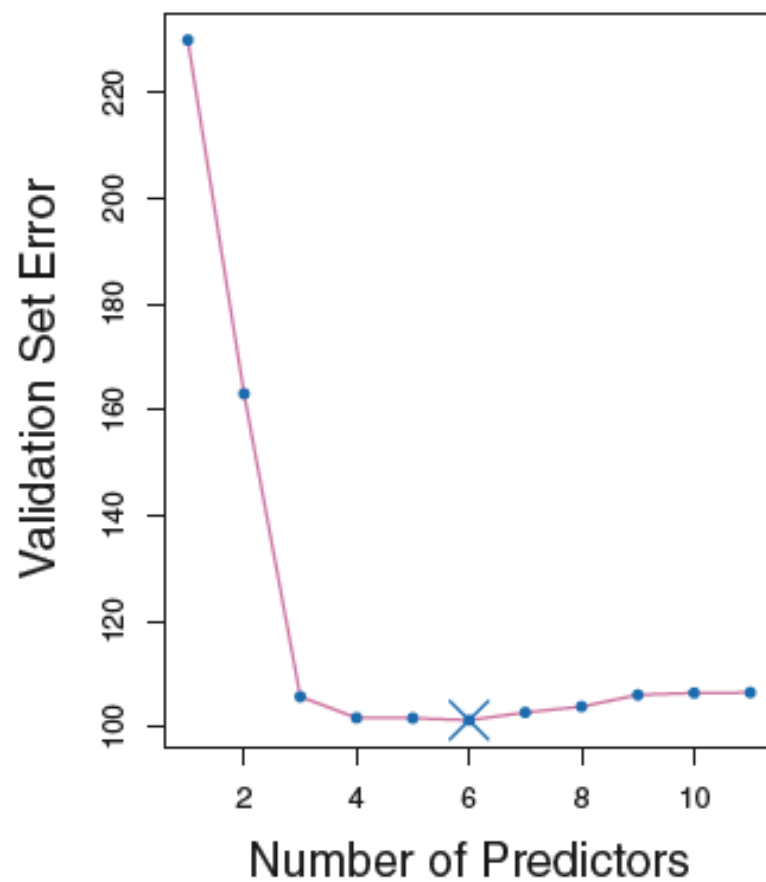
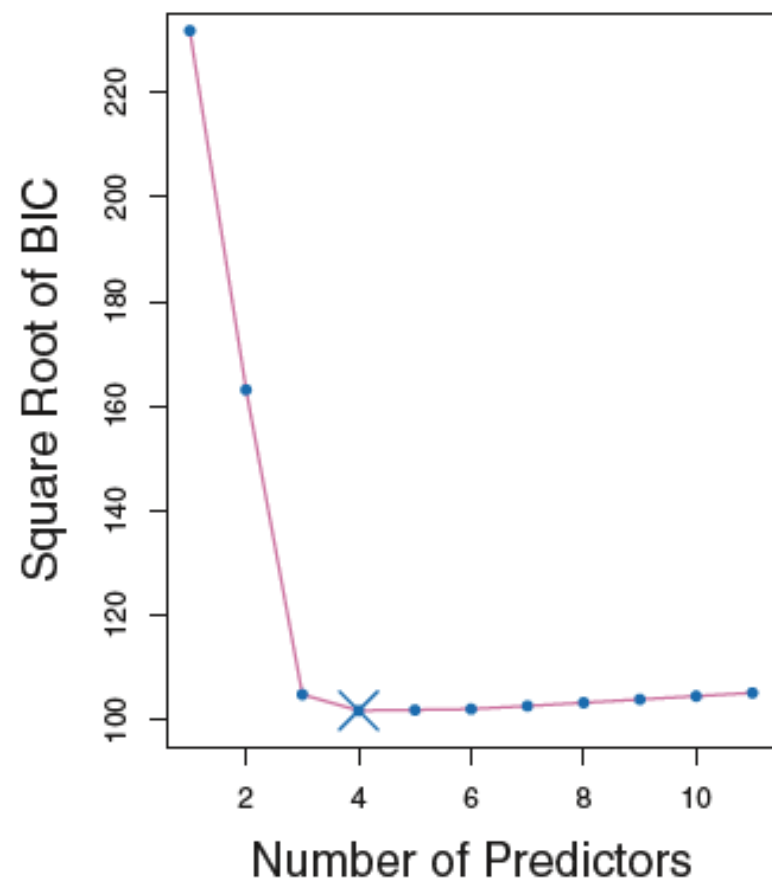
1. Непрямо оцінити тестову помилку пристосувавши навчальну помилку до врахування ефекту зміщення (C_p , AIC , BIC , скорегований R^2).
2. Прямо оцінити тестову помилку використовуючи метод валідаційної множини чи перехресної перевірки.

$$C_p = \frac{1}{n} (\text{RSS} + 2d\hat{\sigma}^2)$$

$$\text{BIC} = \frac{1}{n} (\text{RSS} + \log(n)d\hat{\sigma}^2)$$

$$\text{Adjusted } R^2 = 1 - \frac{\text{RSS}/(n - d - 1)}{\text{TSS}/(n - 1)}$$





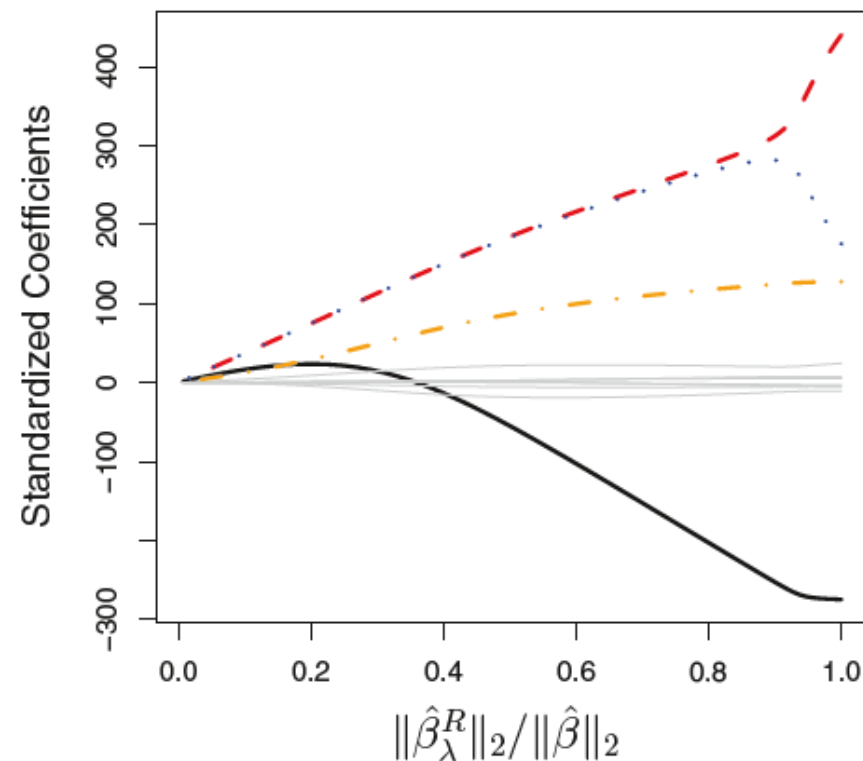
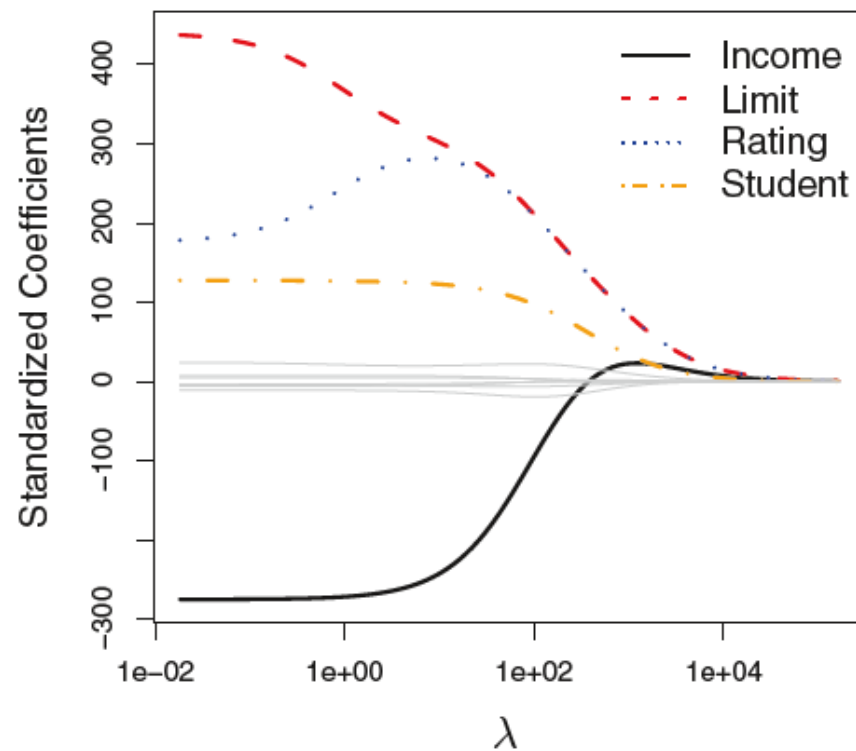
Методи стиснення. Гребенева регресія

Методи стиснення. Гребенева регресія

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 = \text{RSS} + \lambda \sum_{j=1}^p \beta_j^2$$

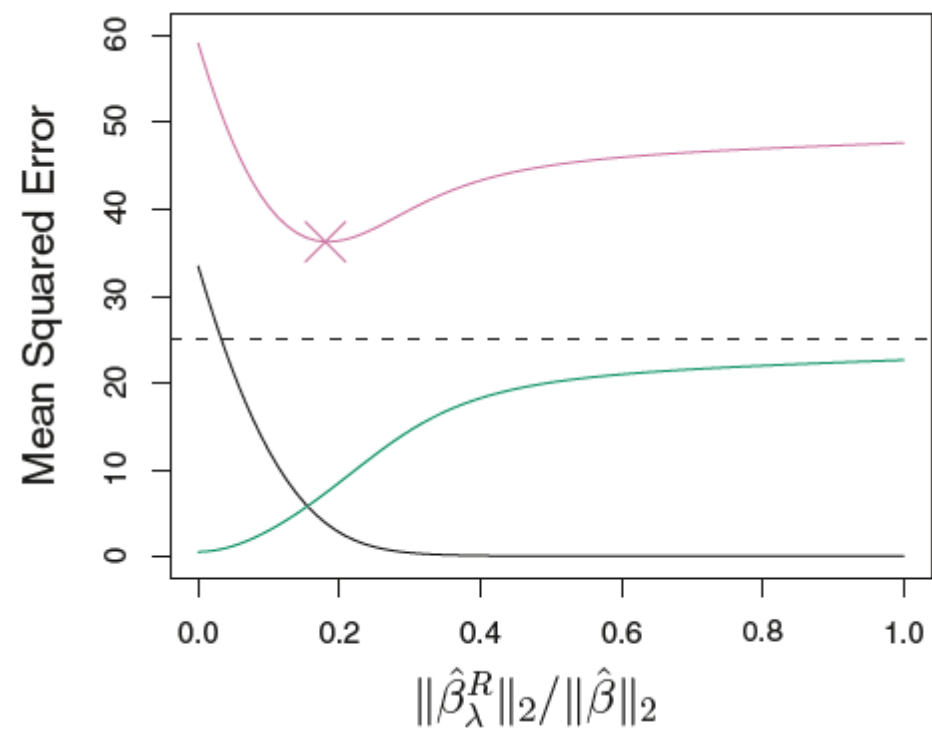
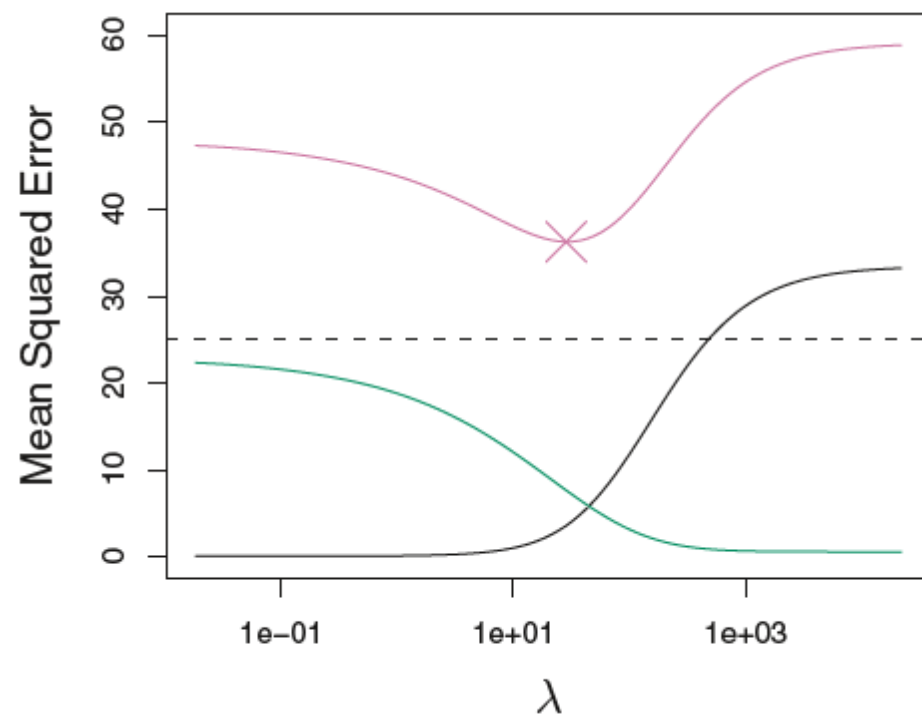
Методи стиснення. Гребенева регресія

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 = \text{RSS} + \lambda \sum_{j=1}^p \beta_j^2$$



$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}}$$

$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}}$$



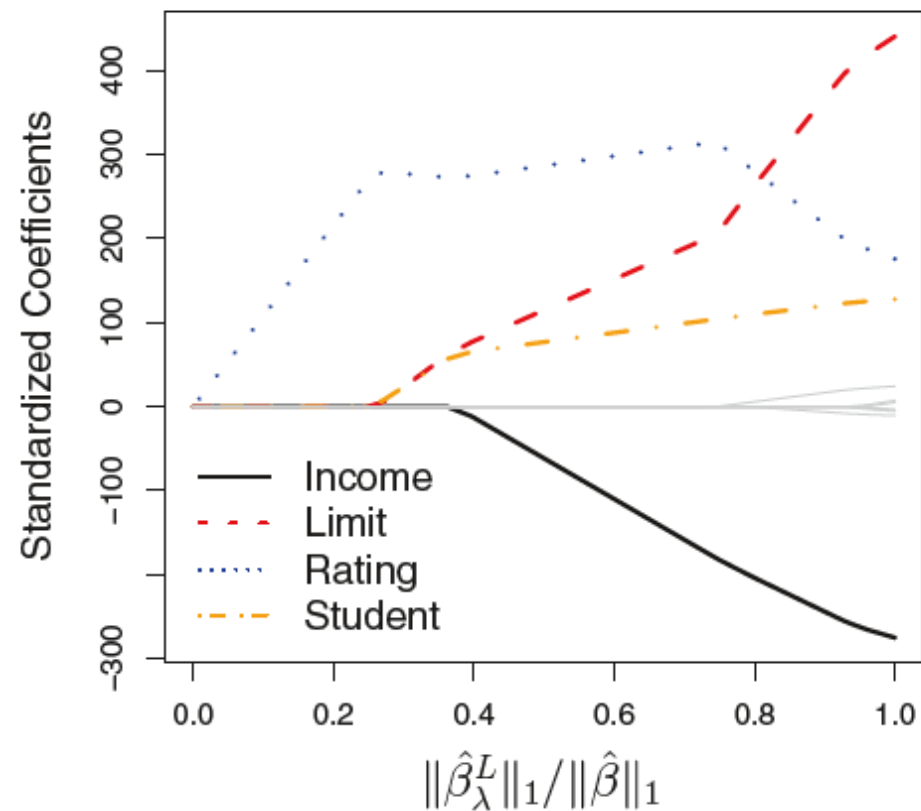
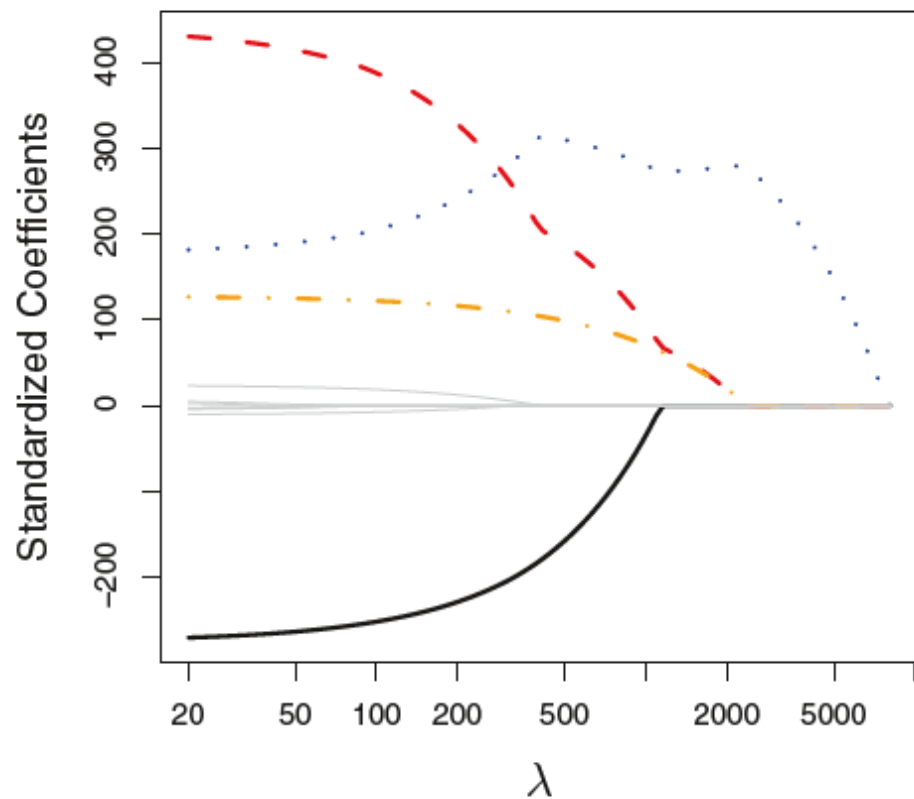
Методи стиснення. Лассо

Методи стиснення. Лассо

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| = \text{RSS} + \lambda \sum_{j=1}^p |\beta_j|$$

Методи стиснення. Лассо

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| = \text{RSS} + \lambda \sum_{j=1}^p |\beta_j|$$



$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^p |\beta_j| \leq s$$

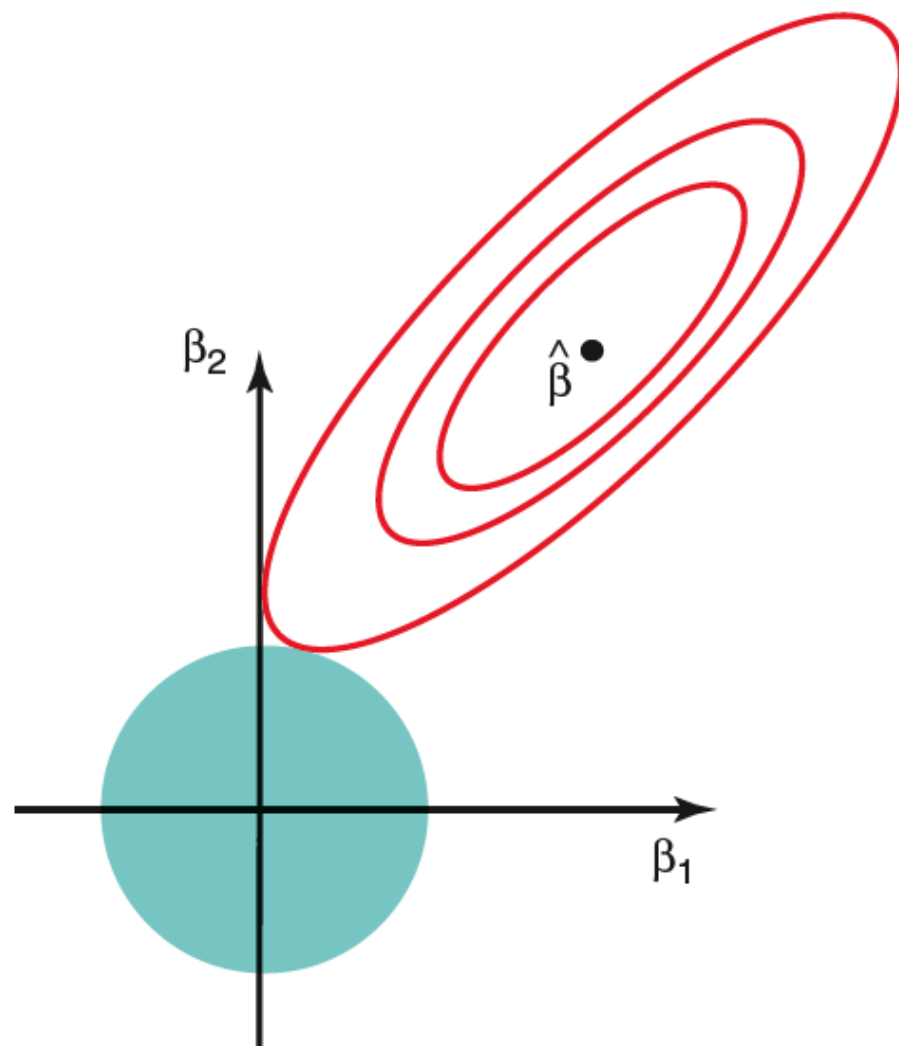
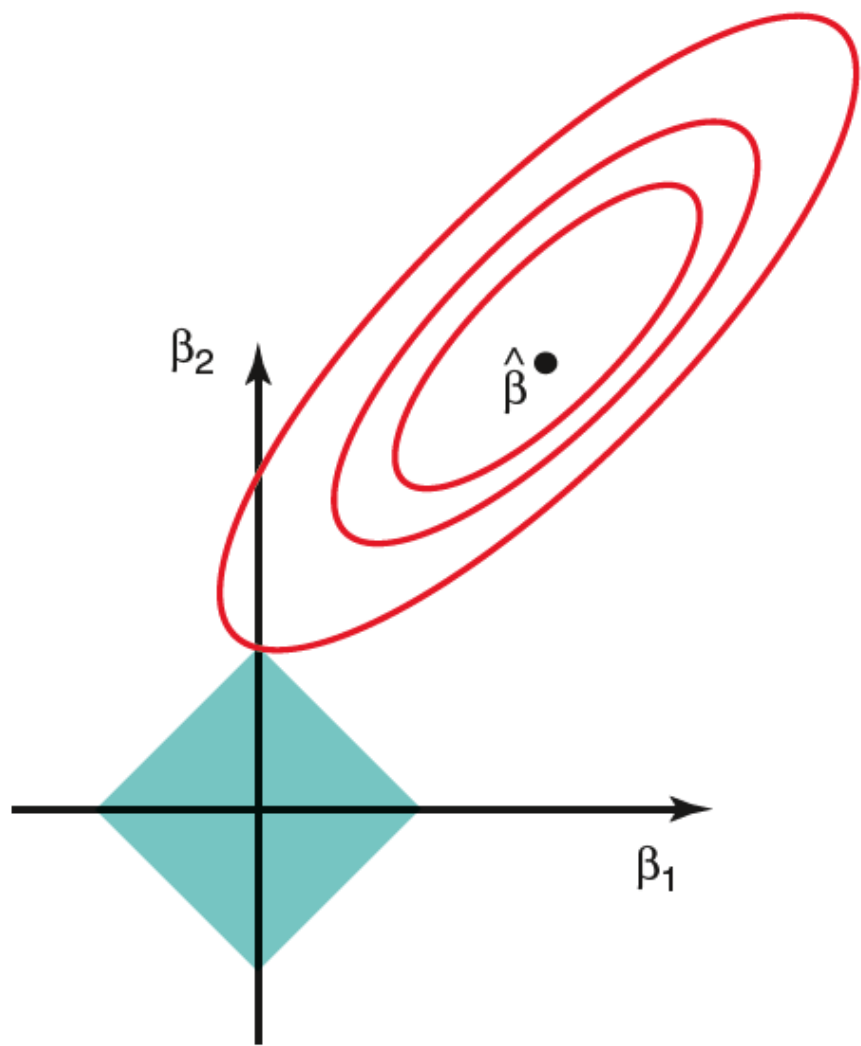
$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^p |\beta_j| \leq s$$

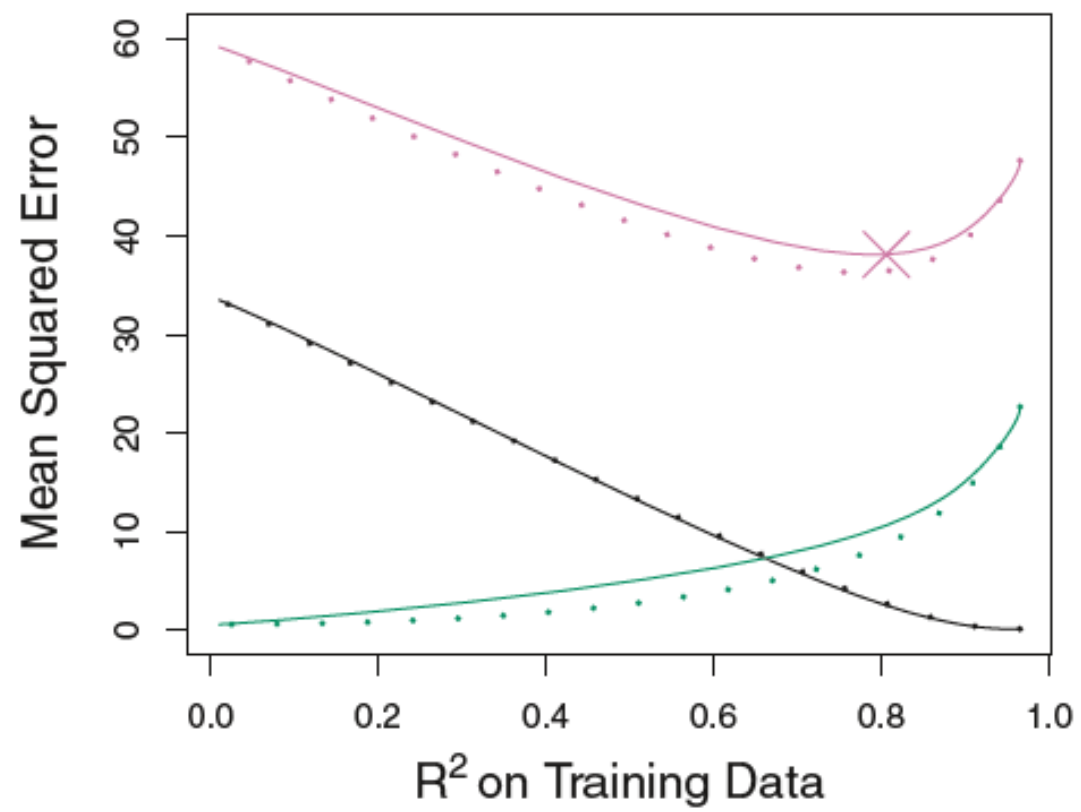
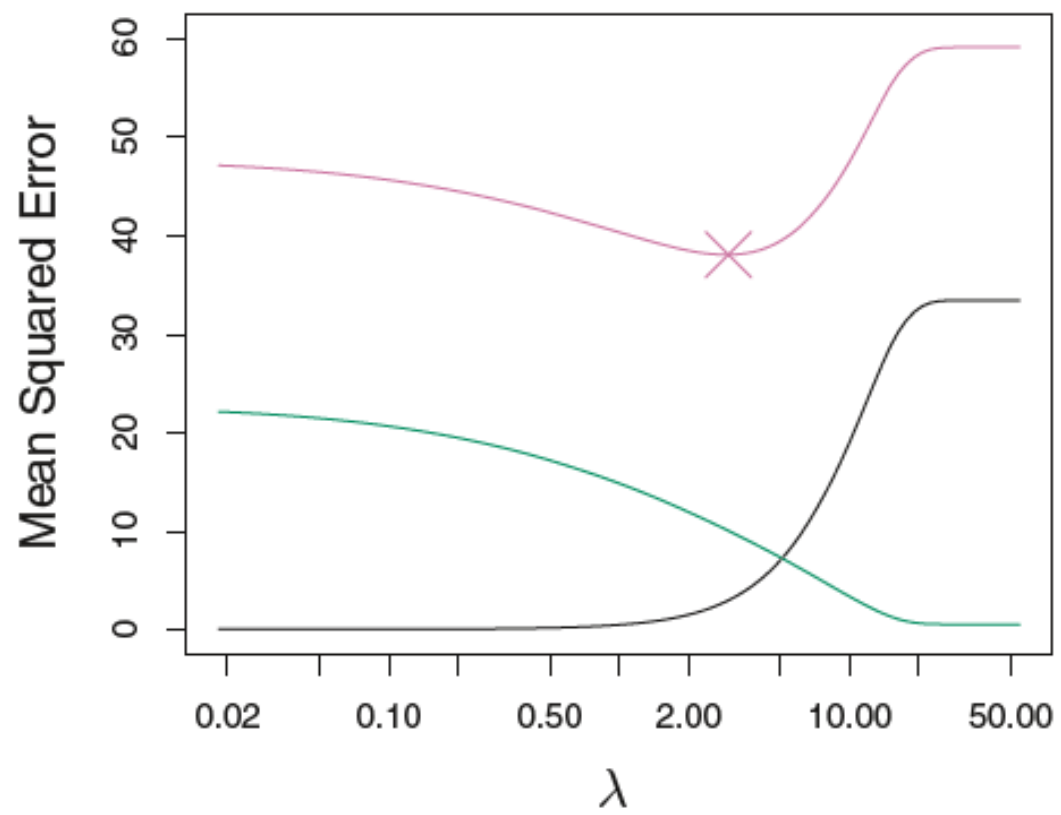
$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^p \beta_j^2 \leq s,$$

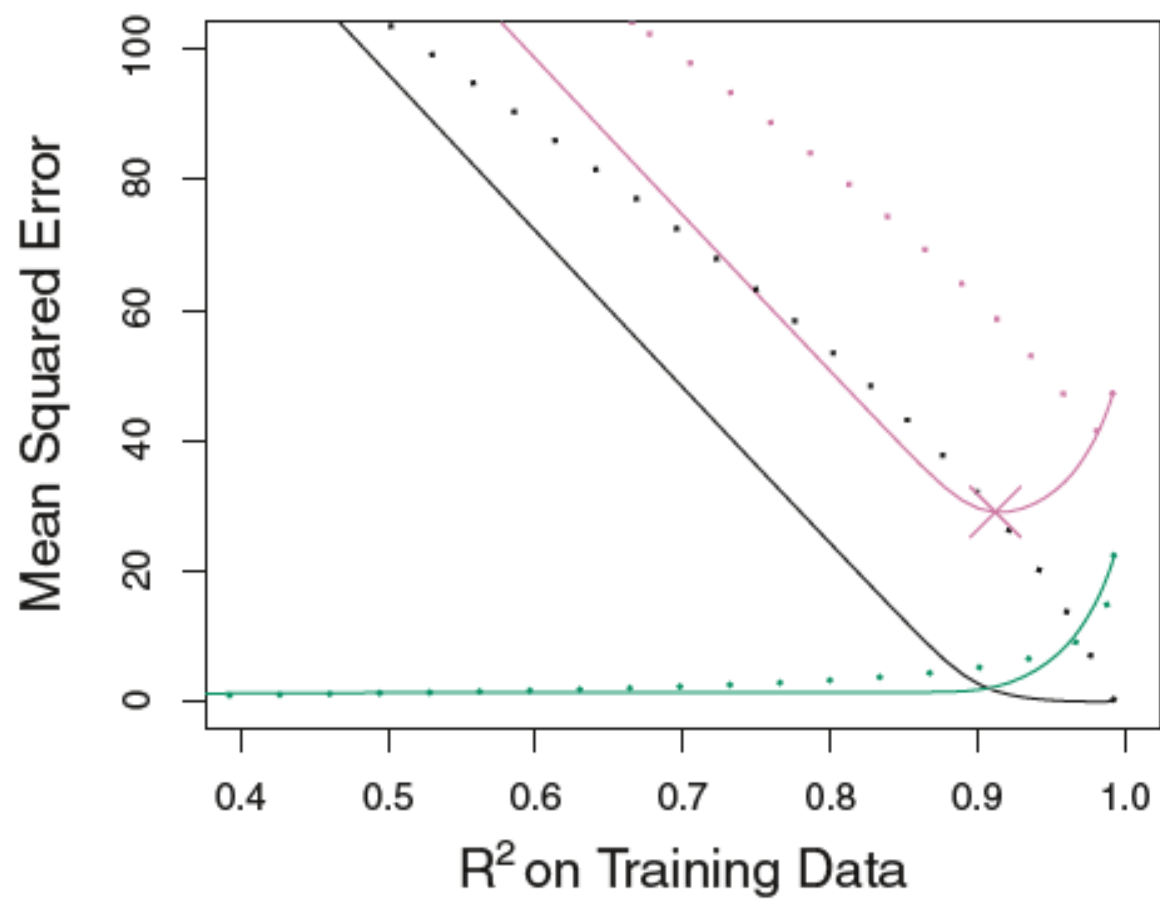
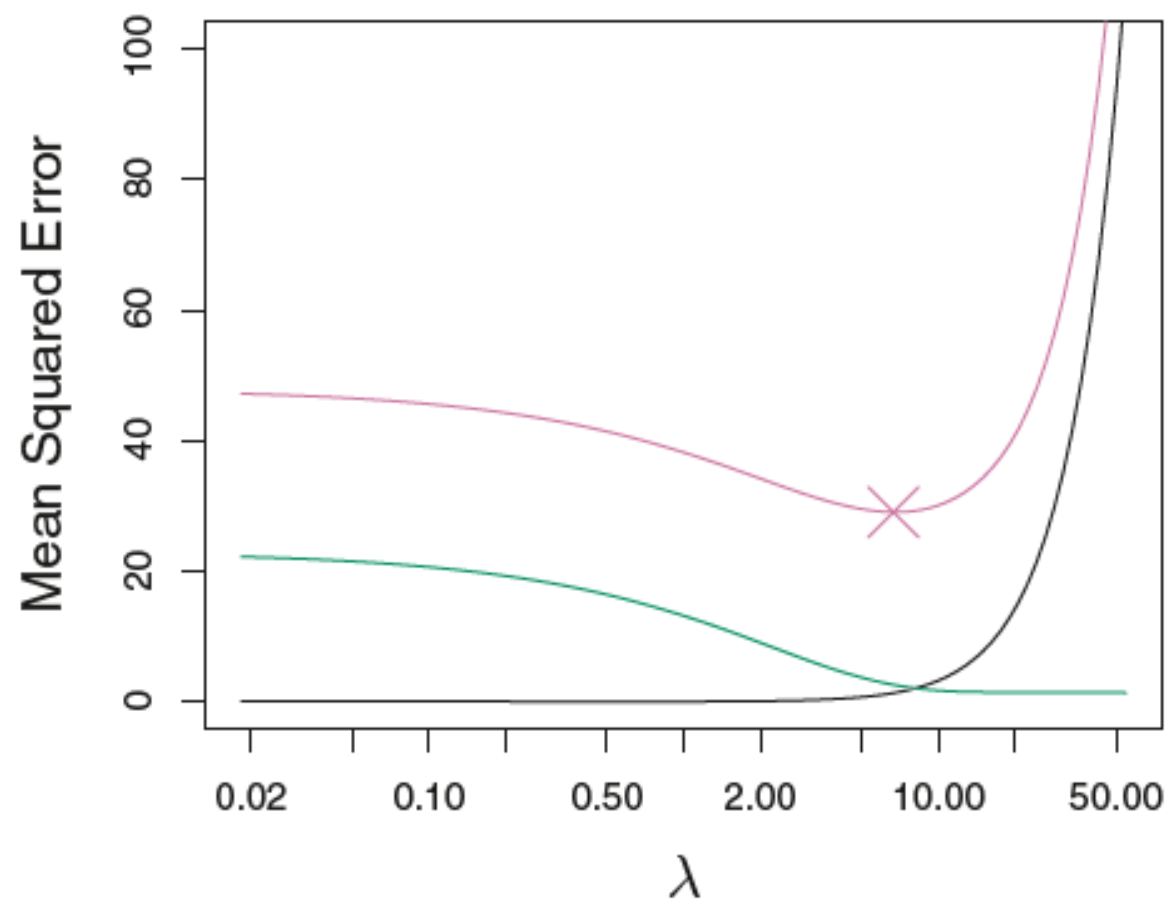
$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^p |\beta_j| \leq s$$

$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^p \beta_j^2 \leq s,$$

$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^p I(\beta_j \neq 0) \leq s$$







Приклад.

Приклад. Нехай $n = p$, матриця X – діагональна, $\beta_0 = 0$.

Приклад. Нехай $n = p$, матриця X – діагональна, $\beta_0 = 0$.

Маємо:

оцінка методом найменших квадратів $\hat{\beta}_j = y_j$

Приклад. Нехай $n = p$, матриця X – діагональна, $\beta_0 = 0$.

Маємо:

оцінка методом найменших квадратів $\hat{\beta}_j = y_j$

гребенева регресія $\hat{\beta}_j^R = y_j / (1 + \lambda)$

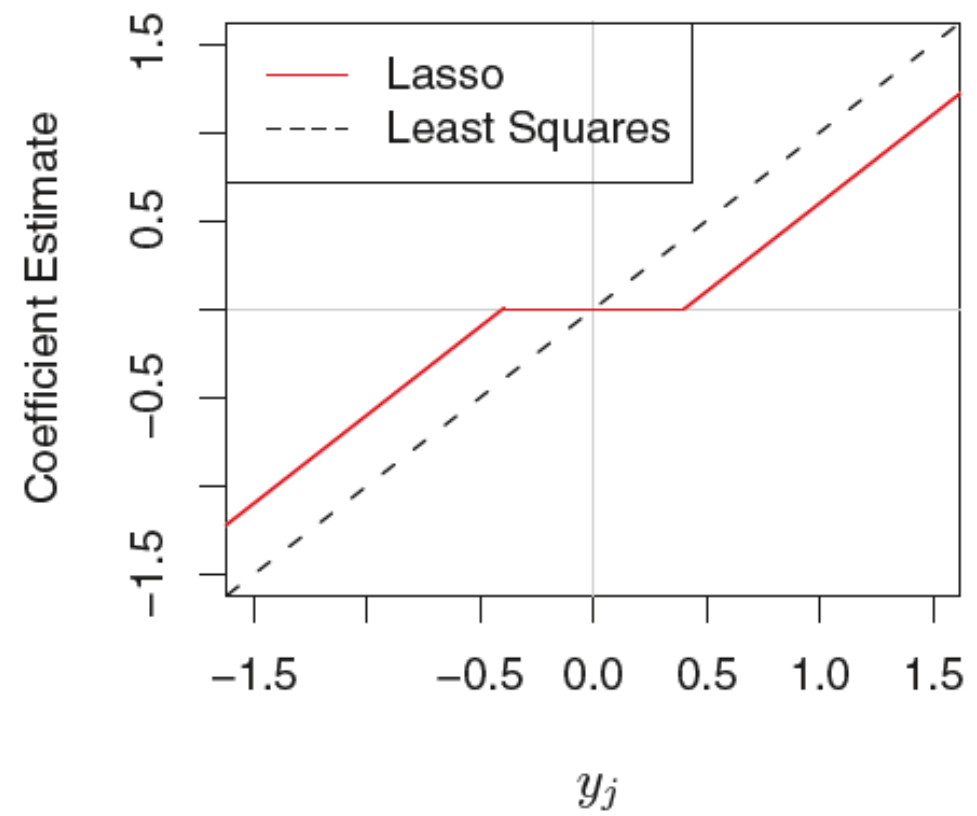
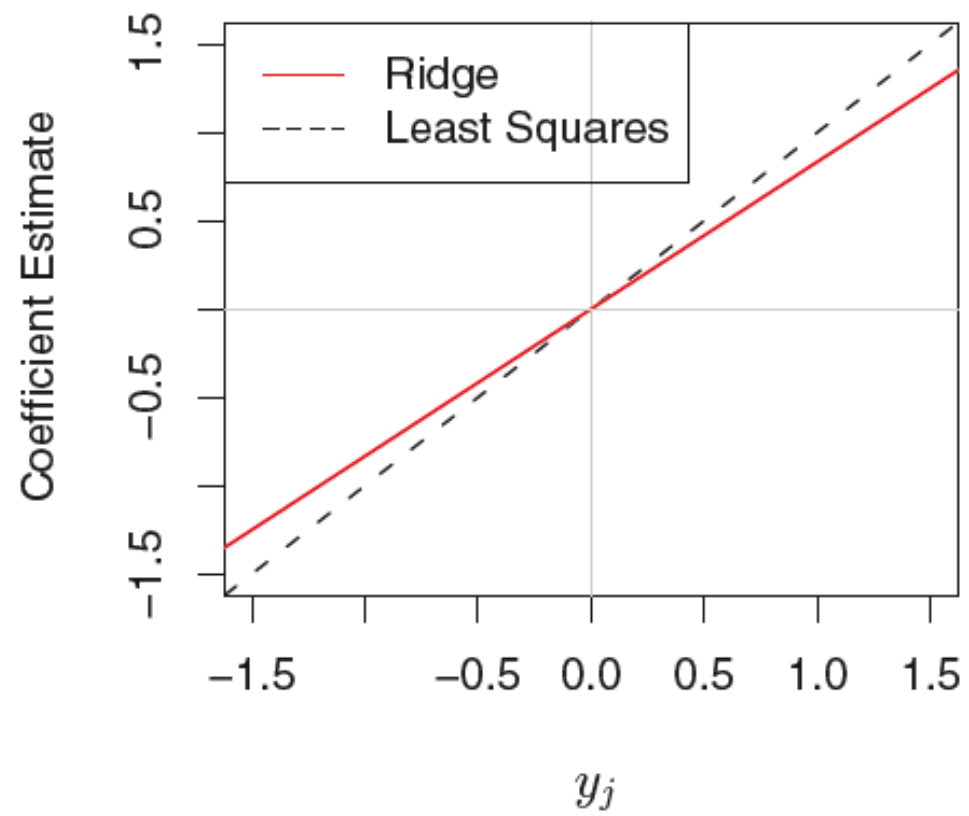
Приклад. Нехай $n = p$, матриця X – діагональна, $\beta_0 = 0$.

Маємо:

оцінка методом найменших квадратів $\hat{\beta}_j = y_j$

гребенева регресія $\hat{\beta}_j^R = y_j / (1 + \lambda)$

лассо
$$\hat{\beta}_j^L = \begin{cases} y_j - \lambda/2 & \text{if } y_j > \lambda/2; \\ y_j + \lambda/2 & \text{if } y_j < -\lambda/2; \\ 0 & \text{if } |y_j| \leq \lambda/2. \end{cases}$$



Байєсівська інтерпретація.

Байєсівська інтерпретація.

Позначимо $p(\beta)$ – апріорний розподіл вектора коефіцієнтів, а $f(Y \mid X, \beta)$ – правдоподібність.

Байєсівська інтерпретація.

Позначимо $p(\beta)$ – апріорний розподіл вектора коефіцієнтів, а $f(Y \mid X, \beta)$ – правдоподібність. Тоді апостеріорний розподіл вектора коефіцієнтів матиме вигляд

$$p(\beta|X, Y) \propto f(Y|X, \beta)p(\beta|X) = f(Y|X, \beta)p(\beta).$$

Байєсівська інтерпретація.

Позначимо $p(\beta)$ – апріорний розподіл вектора коефіцієнтів, а $f(Y | X, \beta)$ – правдоподібність. Тоді апостеріорний розподіл вектора коефіцієнтів матиме вигляд

$$p(\beta|X, Y) \propto f(Y|X, \beta)p(\beta|X) = f(Y|X, \beta)p(\beta).$$

Припустивши, що $p(\beta) = \prod_{j=1}^p g(\beta_j)$, отримаємо

1. Якщо g – розподіл Гауса із середнім нуль та стандартним відхиленням, що є функцією від λ , то звідси випливає, що найбільш вірогідне значення для β , враховуючи дані – задається гребеневою регресією.

Байєсівська інтерпретація.

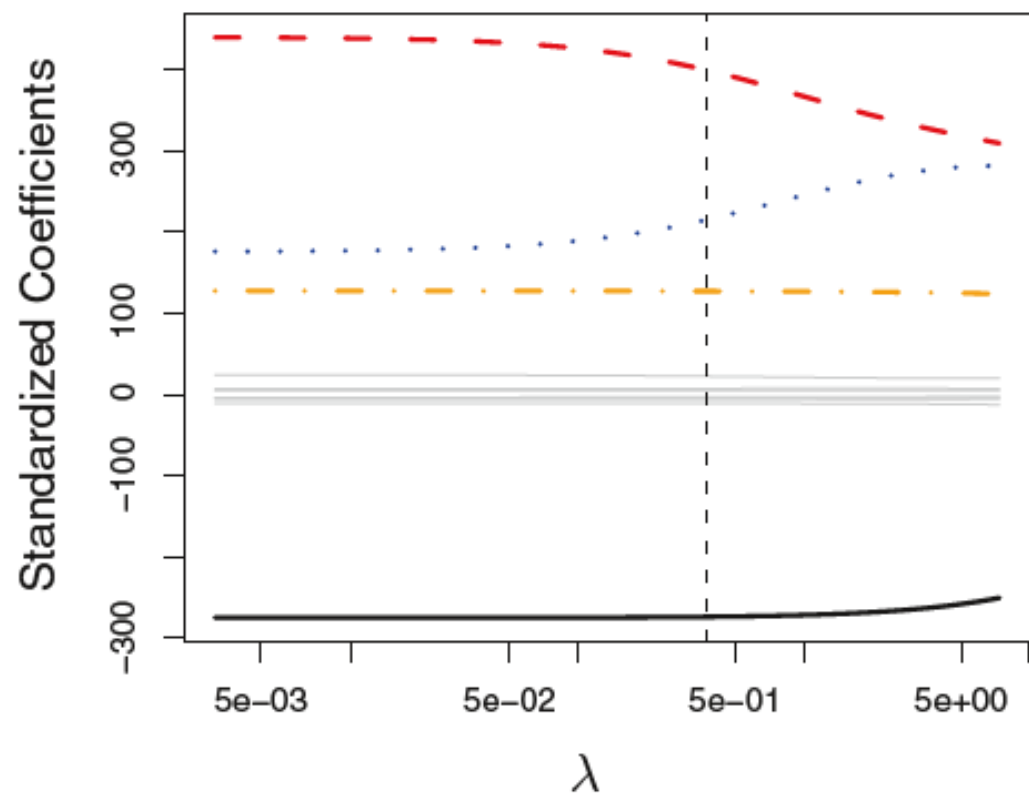
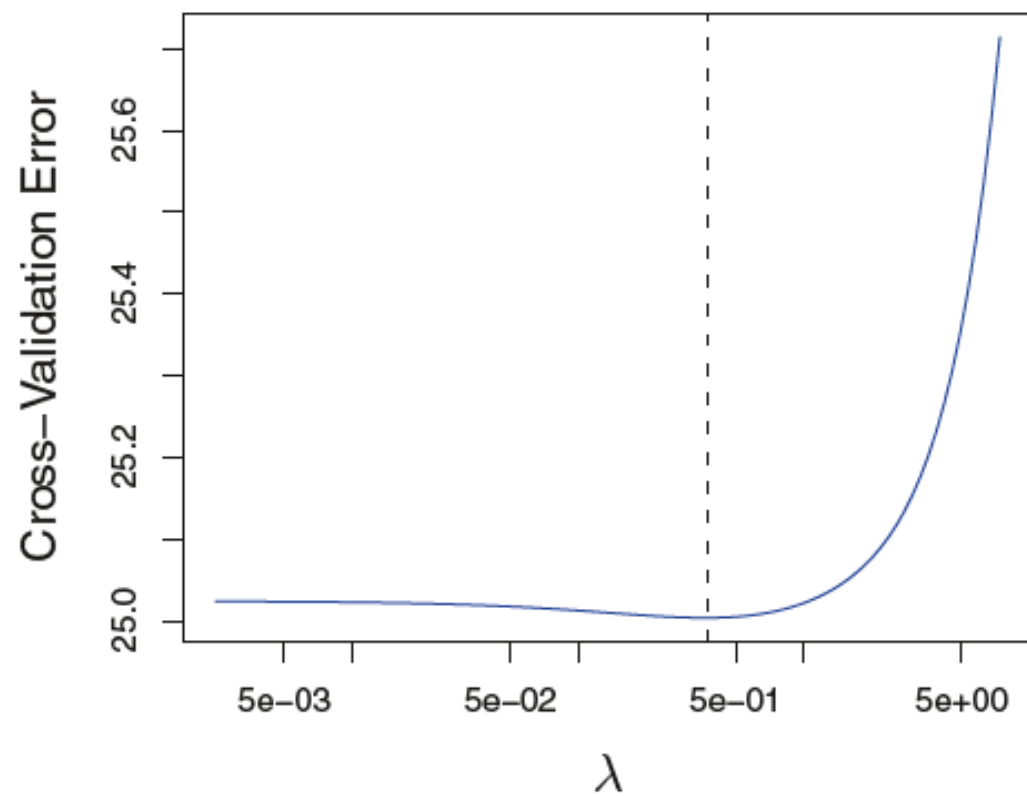
Позначимо $p(\beta)$ – апріорний розподіл вектора коефіцієнтів, а $f(Y | X, \beta)$ – правдоподібність. Тоді апостеріорний розподіл вектора коефіцієнтів матиме вигляд

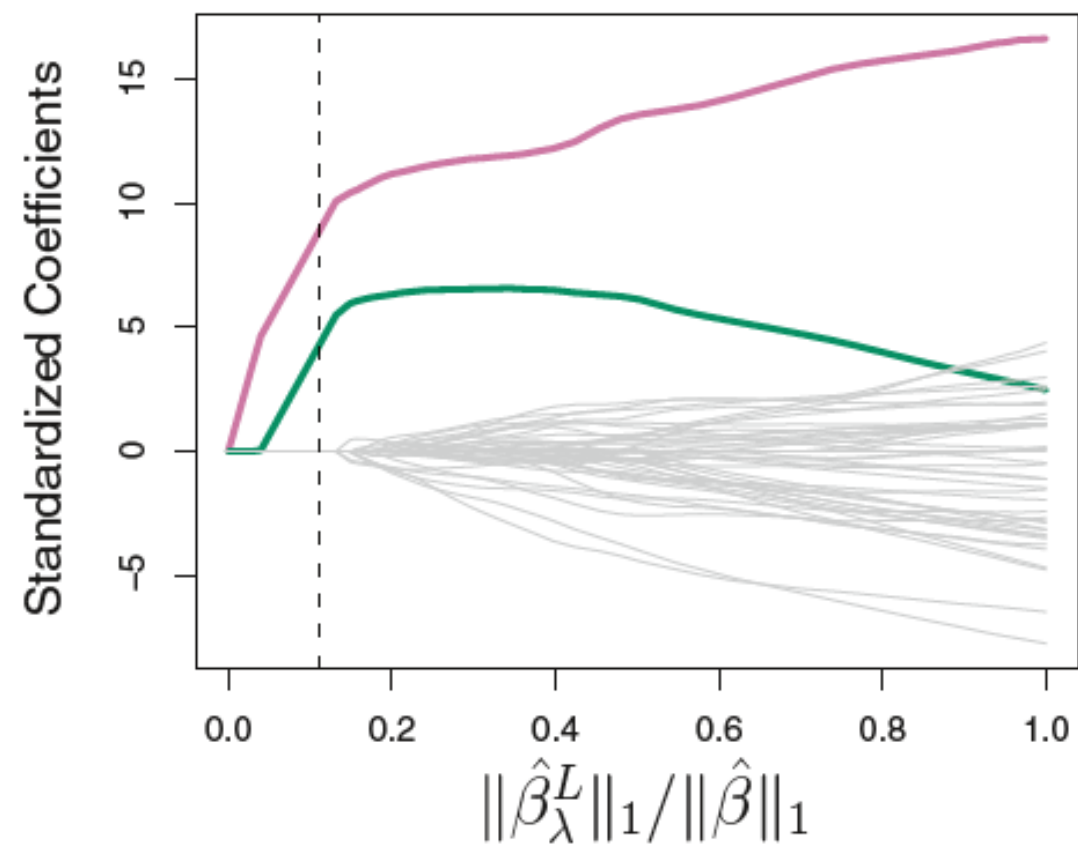
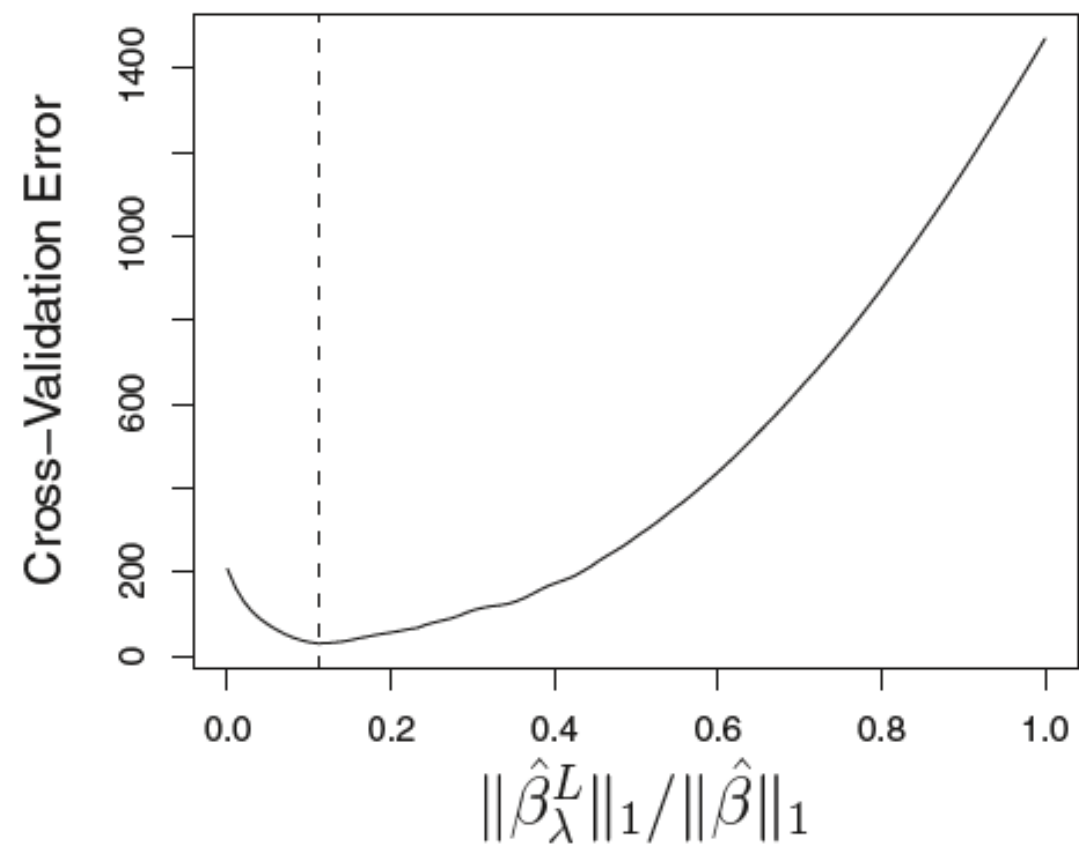
$$p(\beta|X, Y) \propto f(Y|X, \beta)p(\beta|X) = f(Y|X, \beta)p(\beta).$$

Припустивши, що $p(\beta) = \prod_{j=1}^p g(\beta_j)$, отримаємо

1. Якщо g – розподіл Гауса із середнім нуль та стандартним відхиленням, що є функцією від λ , то звідси випливає, що найбільш вірогідне значення для β , враховуючи дані – задається гребеневою регресією.
2. Якщо g – розподіл Лапласа із середнім нуль та параметром масштабу, що є функцією від λ , то звідси випливає, що найбільш вірогідне значення для β , враховуючи дані – задається методом лассо.

Вибір значення для λ .





Методи зниження розмірності.

Методи зниження розмірності.

Нехай Z_1, Z_2, \dots, Z_M — позначають $M < p$ лінійних комбінацій початкових предикторів X_1, X_2, \dots, X_p

$$Z_m = \sum_{j=1}^p \phi_{jm} X_j$$

для деяких констант $\phi_{1m}, \phi_{2m}, \dots, \phi_{pm}$.

Методи зниження розмірності.

Нехай Z_1, Z_2, \dots, Z_M – позначають $M < p$ лінійних комбінацій початкових предикторів X_1, X_2, \dots, X_p

$$Z_m = \sum_{j=1}^p \phi_{jm} X_j$$

для деяких констант $\phi_{1m}, \phi_{2m}, \dots, \phi_{pm}$.

Розглянемо регресію

$$y_i = \theta_0 + \sum_{m=1}^M \theta_m z_{im} + \epsilon_i, \quad i = 1, \dots, n,$$

Зауважимо, що

$$\sum_{m=1}^M \theta_m z_{im} = \sum_{m=1}^M \theta_m \sum_{j=1}^p \phi_{jm} x_{ij} = \sum_{j=1}^p \sum_{m=1}^M \theta_m \phi_{jm} x_{ij} = \sum_{j=1}^p \beta_j x_{ij}$$

Зауважимо, що

$$\sum_{m=1}^M \theta_m z_{im} = \sum_{m=1}^M \theta_m \sum_{j=1}^p \phi_{jm} x_{ij} = \sum_{j=1}^p \sum_{m=1}^M \theta_m \phi_{jm} x_{ij} = \sum_{j=1}^p \beta_j x_{ij}$$

Тобто

$$\beta_j = \sum_{m=1}^M \theta_m \phi_{jm}$$

Зауважимо, що

$$\sum_{m=1}^M \theta_m z_{im} = \sum_{m=1}^M \theta_m \sum_{j=1}^p \phi_{jm} x_{ij} = \sum_{j=1}^p \sum_{m=1}^M \theta_m \phi_{jm} x_{ij} = \sum_{j=1}^p \beta_j x_{ij}$$

Тобто

$$\beta_j = \sum_{m=1}^M \theta_m \phi_{jm}$$

Загалом, зниження розмірності відбувається в два етапи: обчислюються трансформовані предиктори Z_1, Z_2, \dots, Z_M ; оцінюється модель з M предикторами.

Зауважимо, що

$$\sum_{m=1}^M \theta_m z_{im} = \sum_{m=1}^M \theta_m \sum_{j=1}^p \phi_{jm} x_{ij} = \sum_{j=1}^p \sum_{m=1}^M \theta_m \phi_{jm} x_{ij} = \sum_{j=1}^p \beta_j x_{ij}$$

Тобто

$$\beta_j = \sum_{m=1}^M \theta_m \phi_{jm}$$

Загалом, зниження розмірності відбувається в два етапи: обчислюються трансформовані предиктори Z_1, Z_2, \dots, Z_M ; оцінюється модель з M предикторами.

Розглянемо два способи вибору Z_1, Z_2, \dots, Z_M : метод головних компонент та метод часткових найменших квадратів

Метод головних компонент

Метод головних компонент

Перший головний компонент – це напрямок, уздовж якого спостереження різняться найбільше.

Метод головних компонент

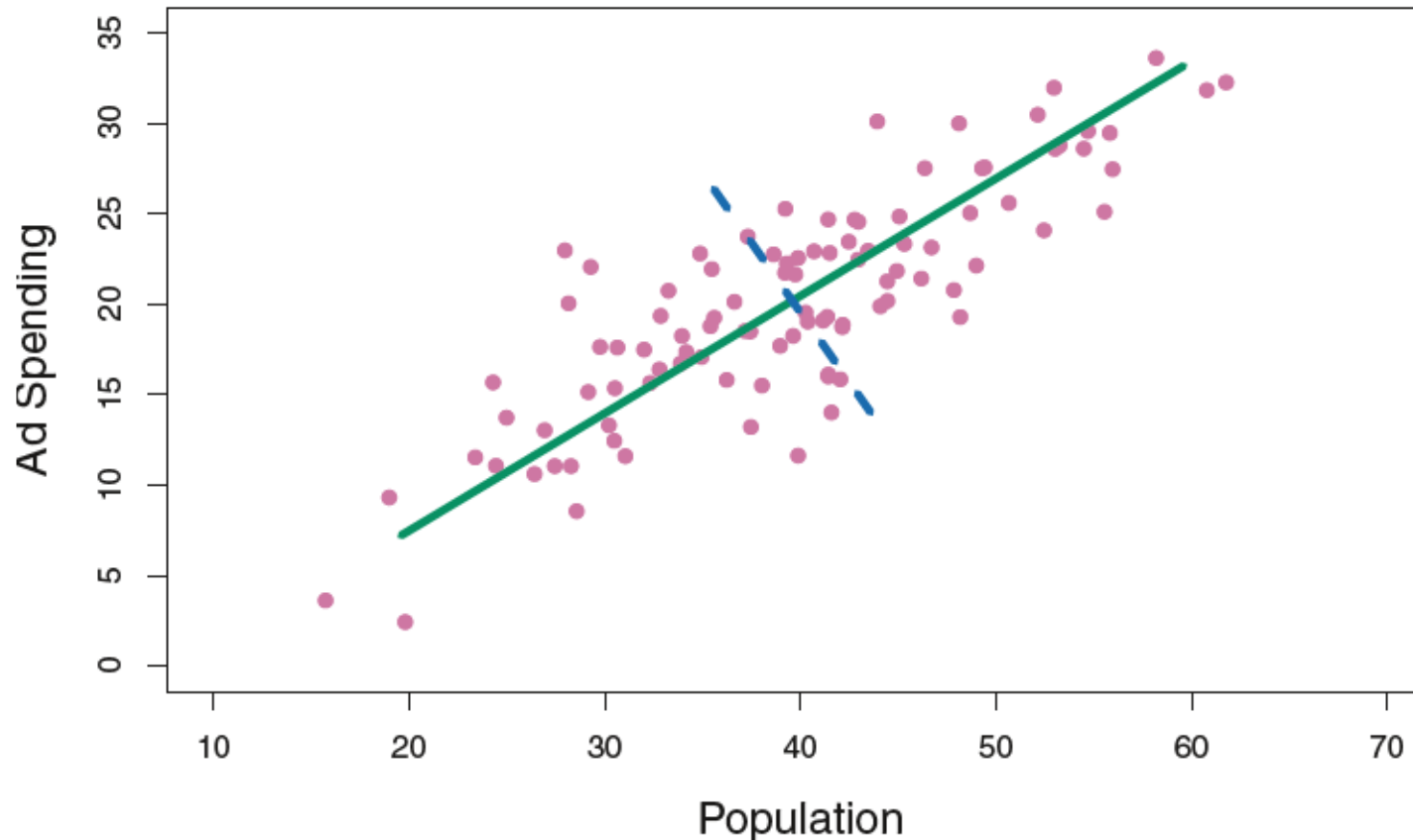
Перший головний компонент – це напрямок, уздовж якого спостереження різняться найбільше. Або еквівалентно – це лінійна комбінація змінних з максимальною дисперсією.

Метод головних компонент

Перший головний компонент – це напрямок, уздовж якого спостереження різняться найбільше. Або еквівалентно – це лінійна комбінація змінних з максимальною дисперсією. Або еквівалентно – перший головний компонент визначає пряму, що знаходиться найближче до даних.

Метод головних компонент

Перший головний компонент – це напрямок, уздовж якого спостереження різняться найбільше. Або еквівалентно – це лінійна комбінація змінних з максимальною дисперсією. Або еквівалентно – перший головний компонент визначає пряму, що знаходиться найближче до даних.

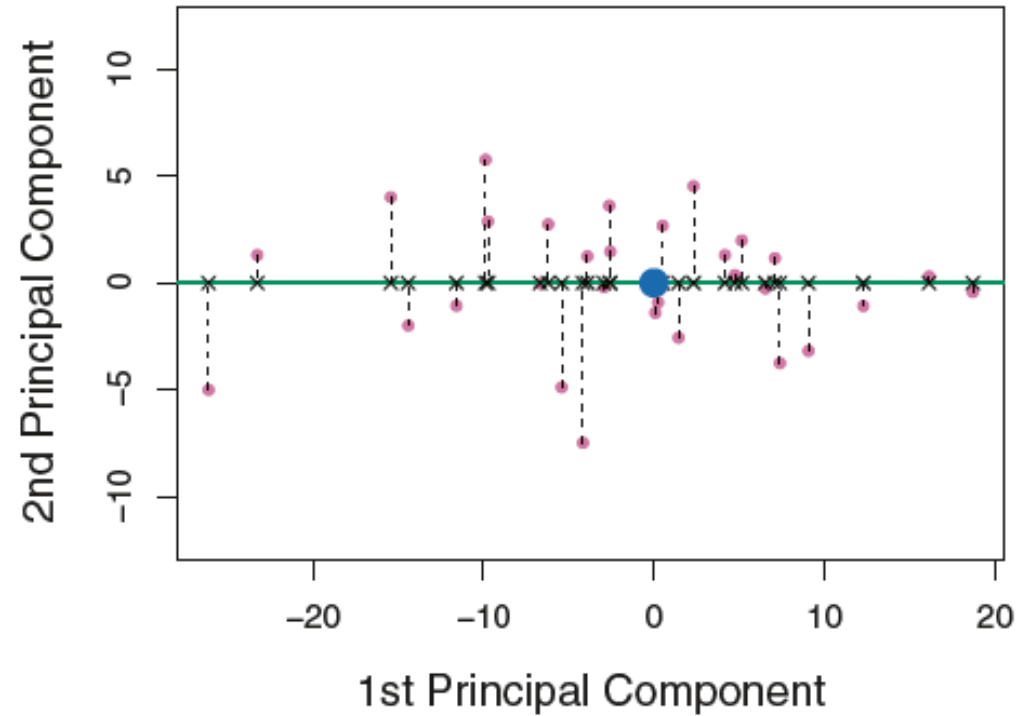
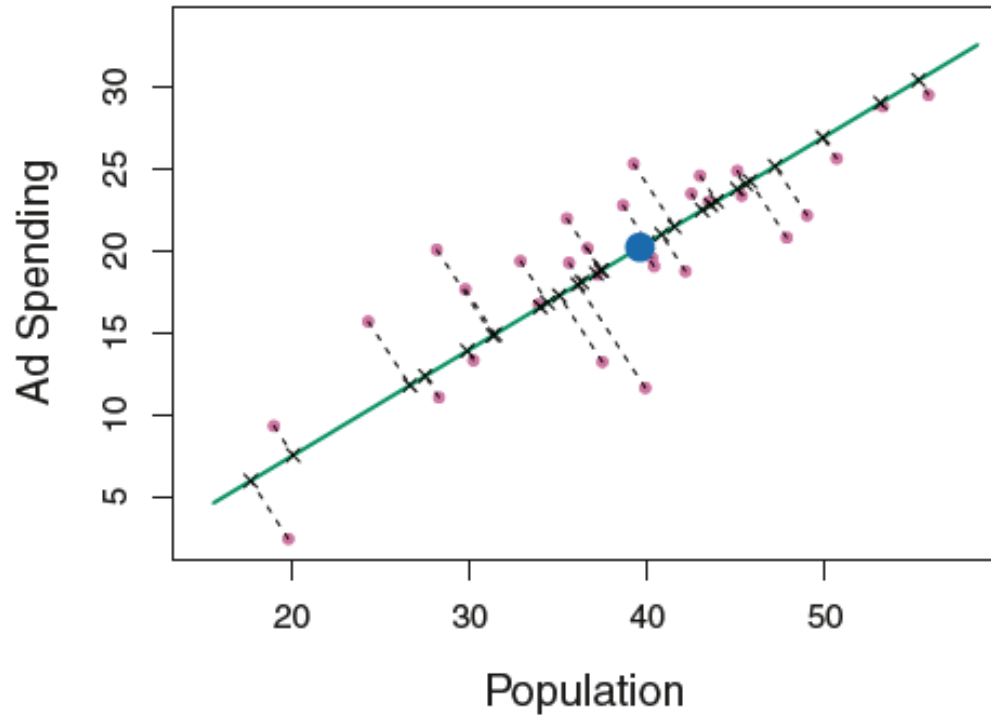


$$Z_1 = 0.839 \times (\text{pop} - \overline{\text{pop}}) + 0.544 \times (\text{ad} - \overline{\text{ad}})$$

$$z_{i1} = 0.839 \times (\text{pop}_i - \overline{\text{pop}}) + 0.544 \times (\text{ad}_i - \overline{\text{ad}})$$

$$Z_1 = 0.839 \times (\text{pop} - \overline{\text{pop}}) + 0.544 \times (\text{ad} - \overline{\text{ad}})$$

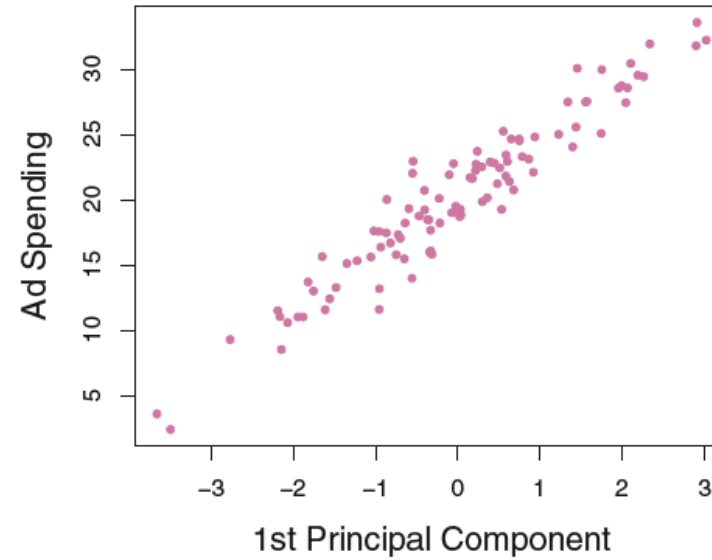
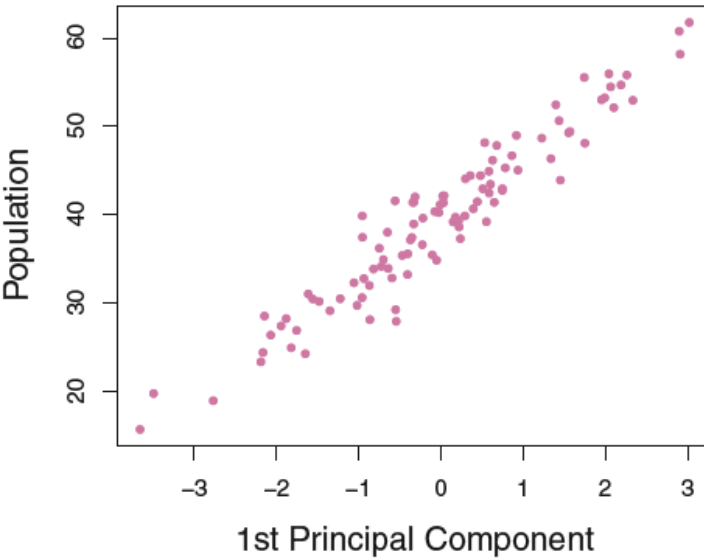
$$z_{i1} = 0.839 \times (\text{pop}_i - \overline{\text{pop}}) + 0.544 \times (\text{ad}_i - \overline{\text{ad}})$$



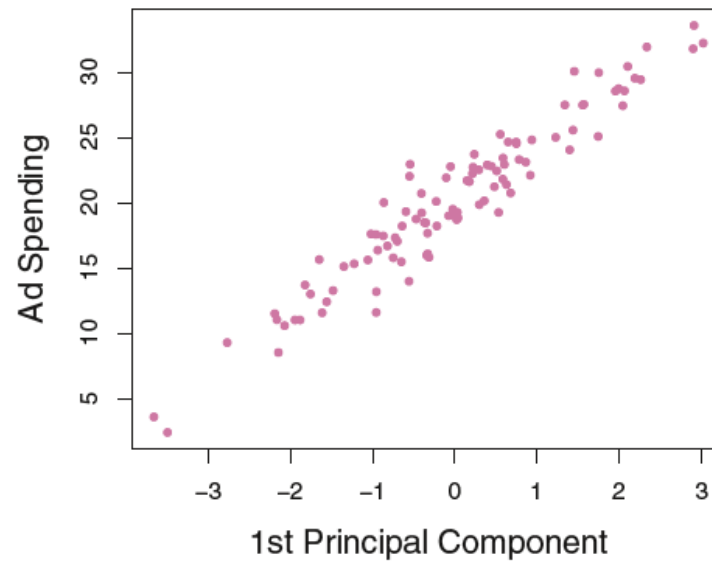
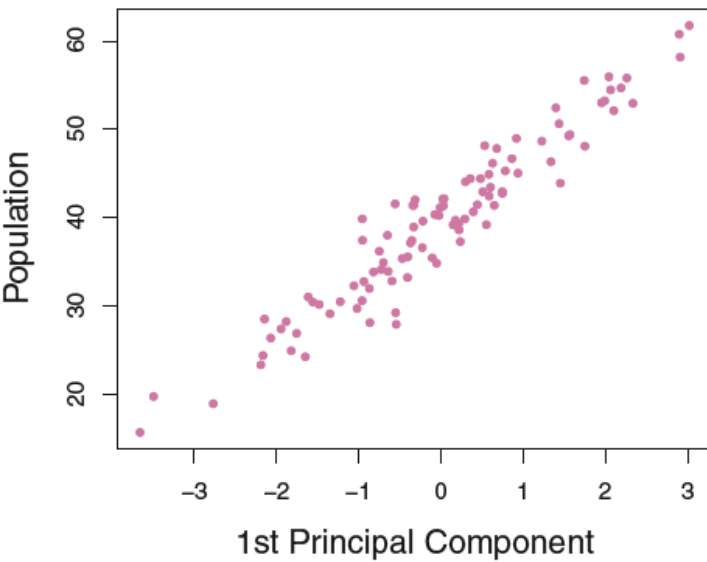
Загалом можна побудувати до p головних компонент.

Загалом можна побудувати до p головних компонент. Наприклад, другий головний компонент будується так, що це лінійна комбінація змінних з максимальною дисперсією і є некорельований з першим головним компонентом.

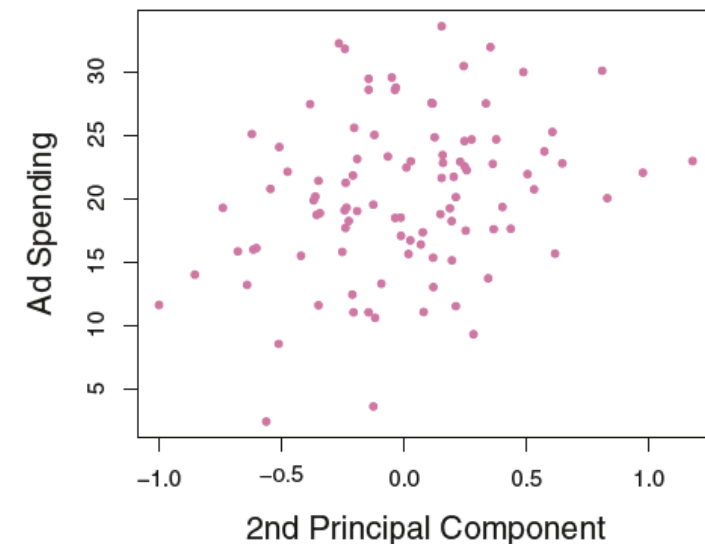
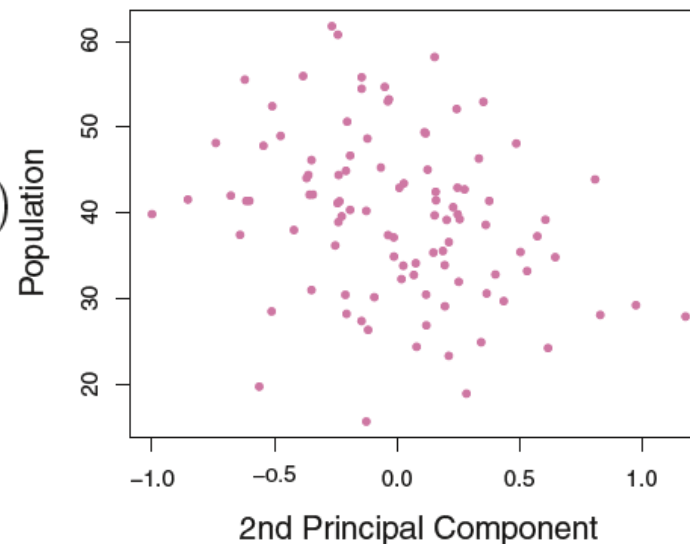
Загалом можна побудувати до p головних компонент. Наприклад, другий головний компонент будується так, що це лінійна комбінація змінних з максимальною дисперсією і є некорельований з першим головним компонентом.

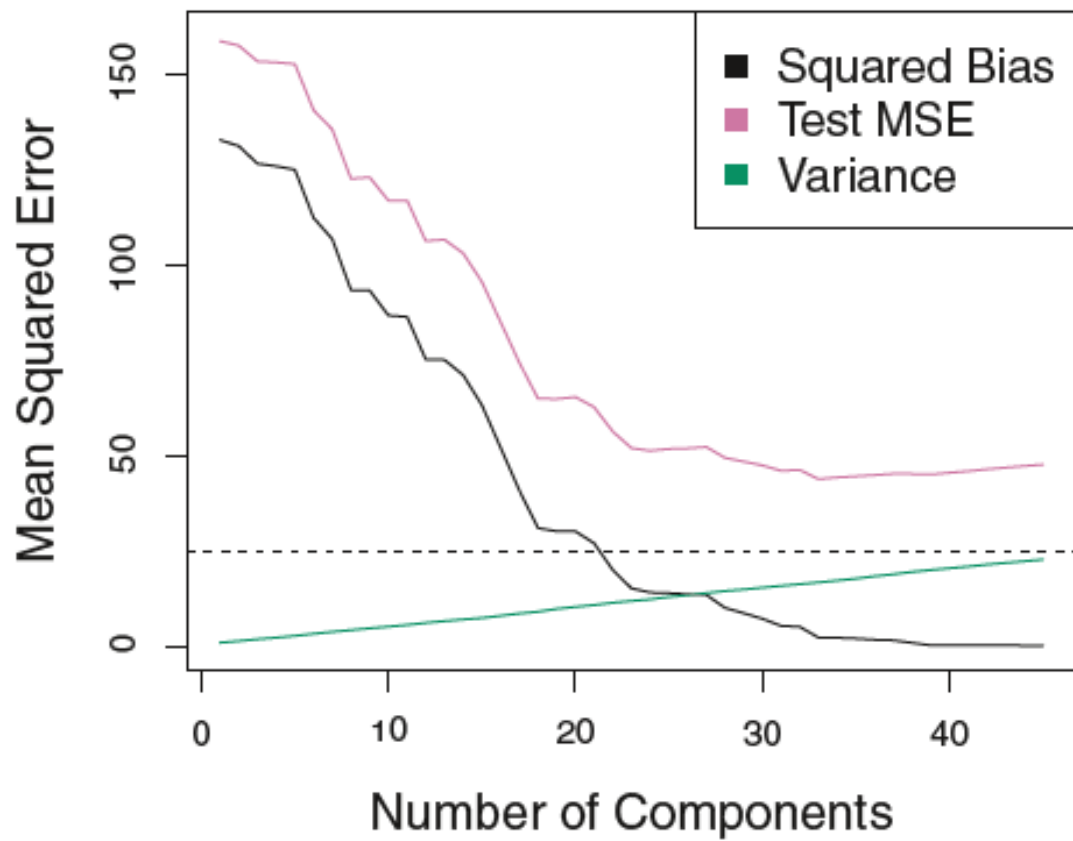
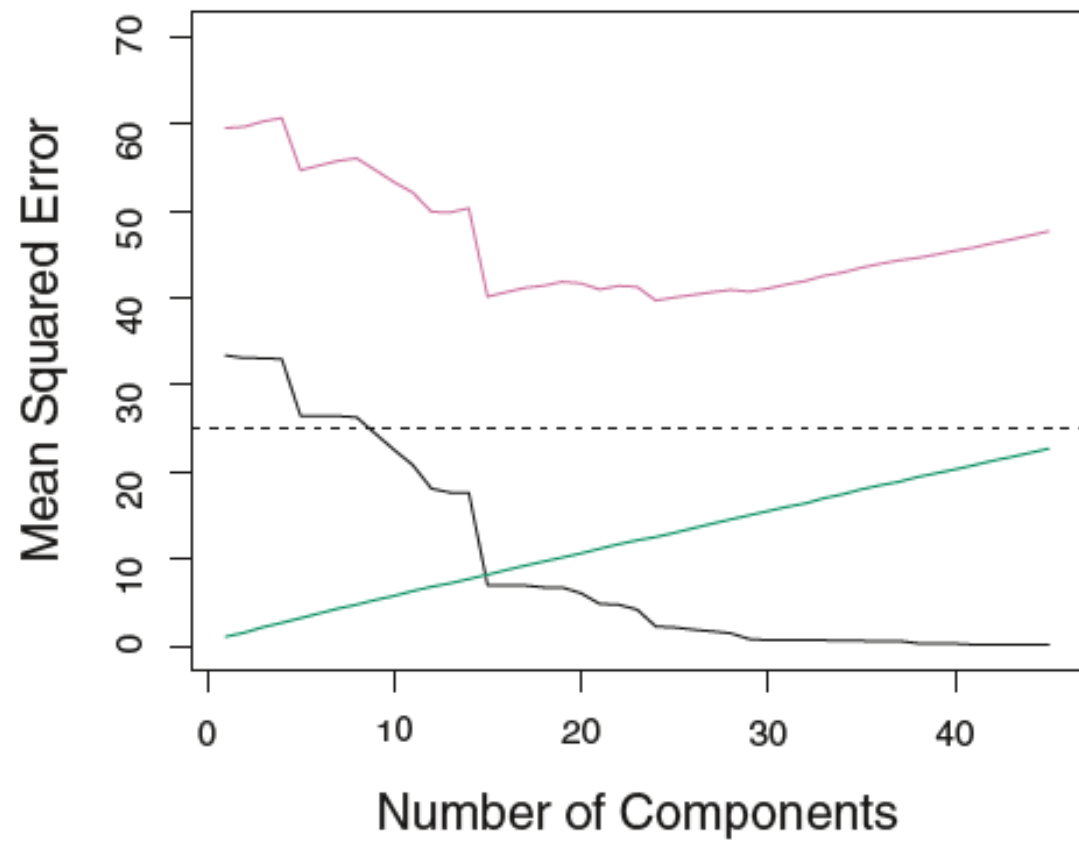


Загалом можна побудувати до p головних компонент. Наприклад, другий головний компонент будується так, що це лінійна комбінація змінних з максимальною дисперсією і є некорельований з першим головним компонентом.

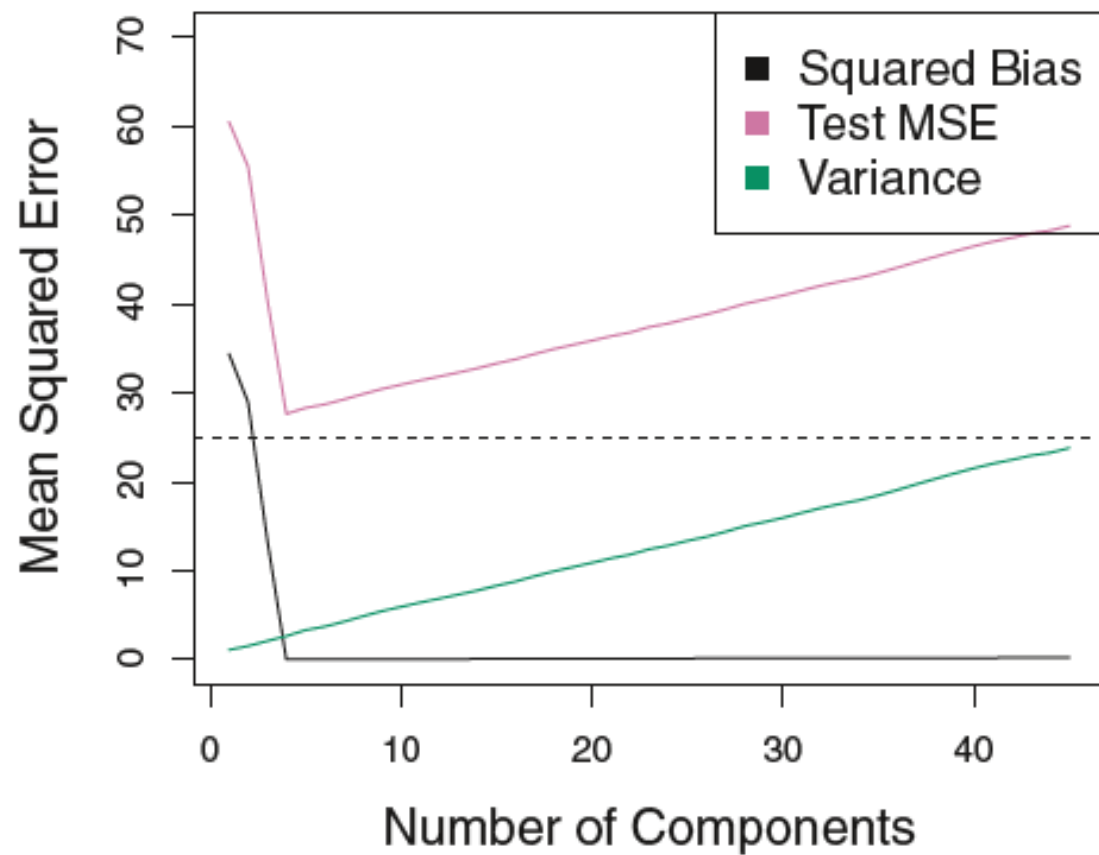


$$Z_2 = 0.544 \times (\text{pop} - \overline{\text{pop}}) - 0.839 \times (\text{ad} - \overline{\text{ad}})$$

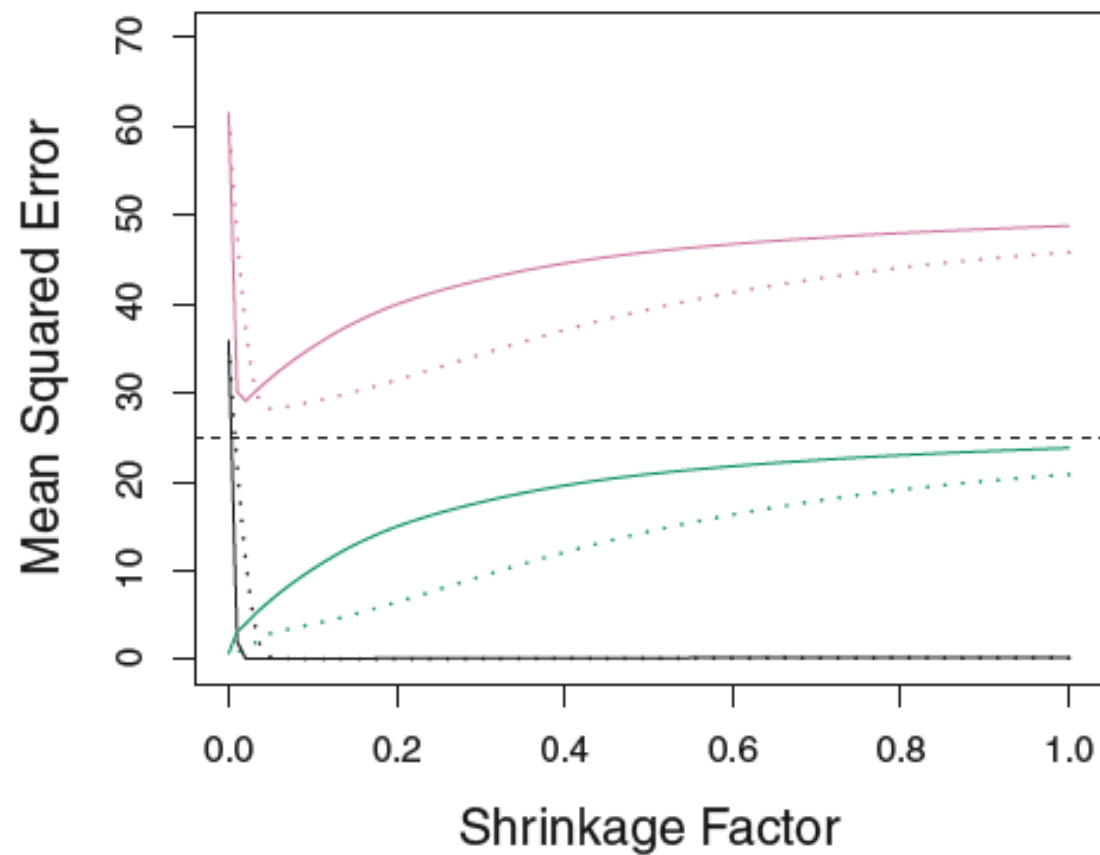


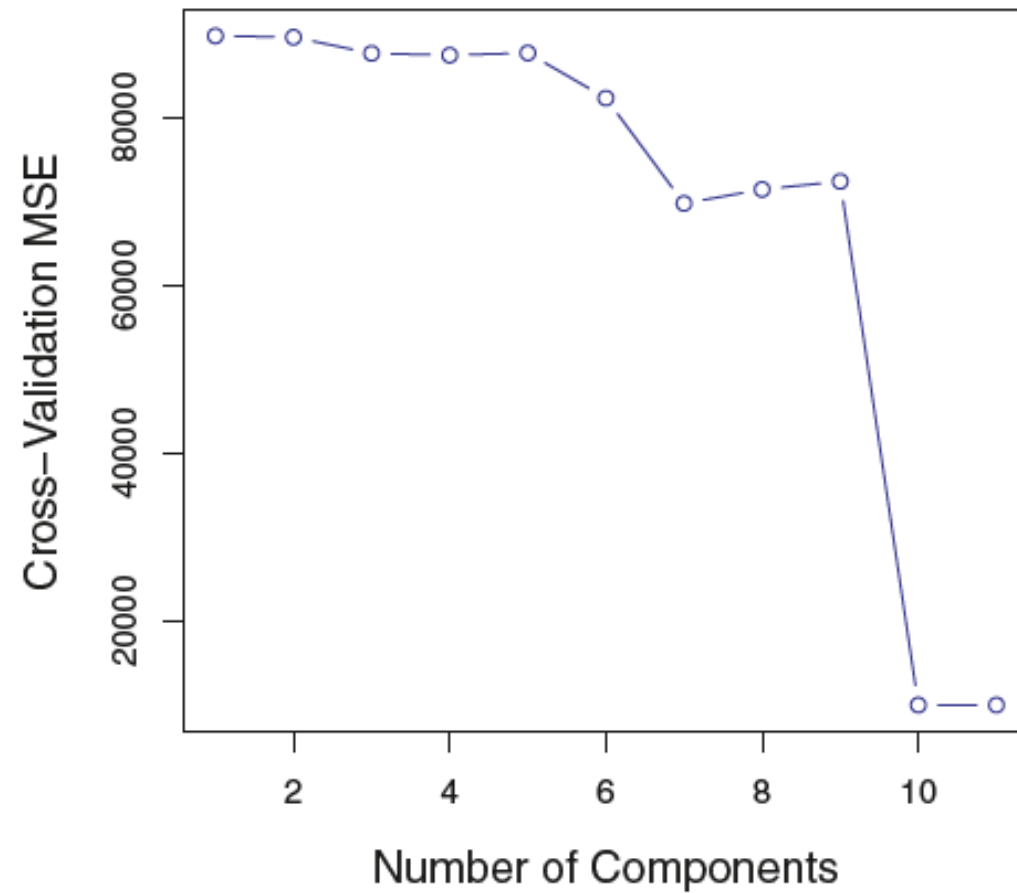
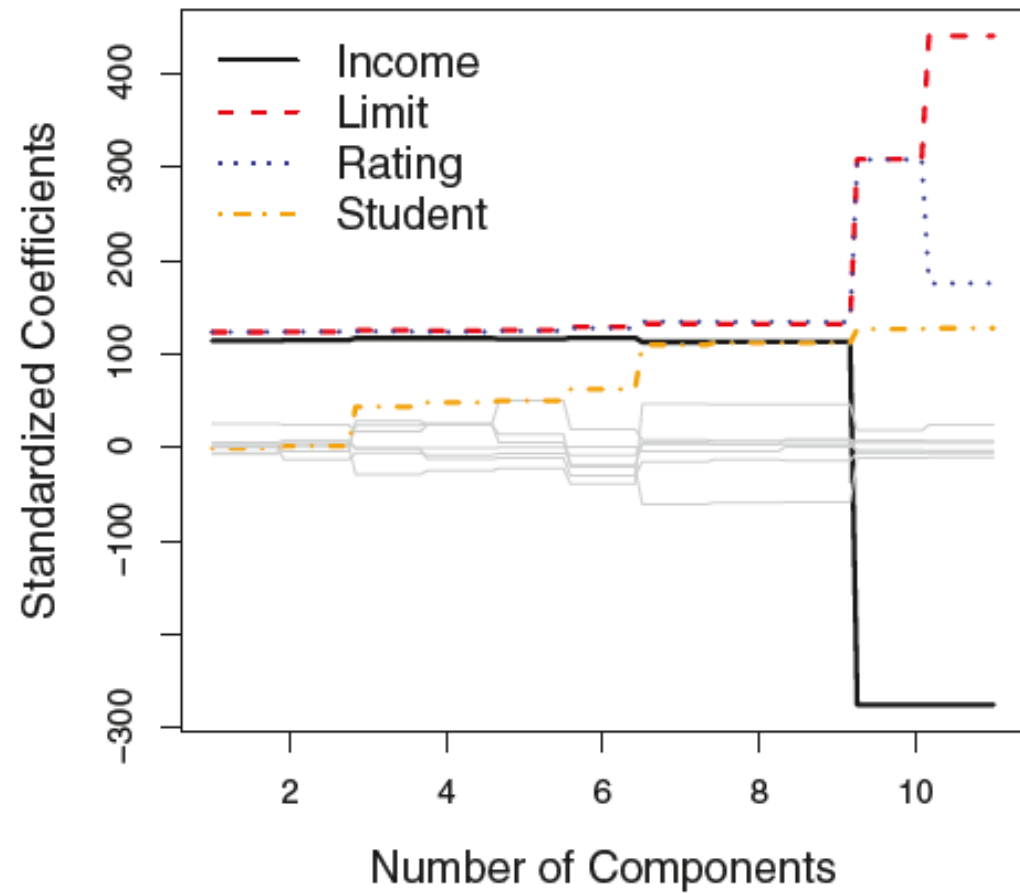


PCR



Ridge Regression and Lasso





Часткових найменших квадратів (PLS).

Часткових найменших квадратів (PLS).

Даний метод шукає напрямки, що дозволяє описати як предиктори так і залежну змінну.

Часткових найменших квадратів (PLS).

Даний метод шукає напрямок, що дозволяє описати як предиктори так і залежну змінну.

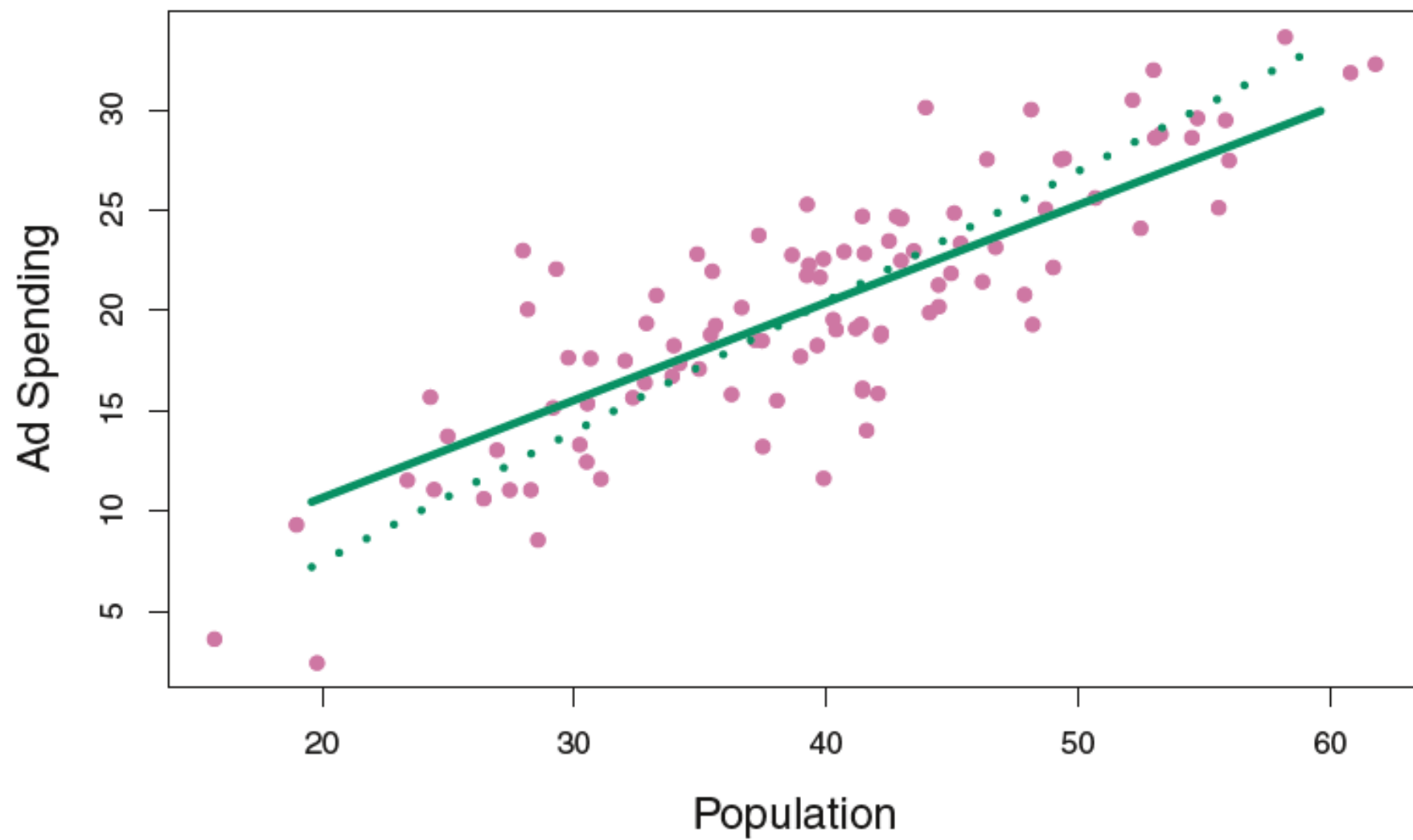
Перший PLS напрямок отримують обчисленням відповідного коефіцієнта φ_{j1} як оцінку коефіцієнта лінійної регресії Y на X_j отриману методом найменших квадратів.

Часткових найменших квадратів (PLS).

Даний метод шукає напрямок, що дозволяє описати як предиктори так і залежну змінну.

Перший PLS напрямок отримують обчисленням відповідного коефіцієнта φ_{j1} як оцінку коефіцієнта лінійної регресії Y на X_j отриману методом найменших квадратів.

Для побудови другого PLS напрямку спочатку корегують предиктори з врахуванням першого напрямку, а саме, будують регресію кожного предиктора на першу PLS і обчислюють залишки. Далі використовують алгоритм аналогічний до алгоритму побудови першого PLS напрямку тільки замість X_j використовують залишки відповідної моделі.



Дані високої вимірності.

Дані високої вимірності.

Що відбувається з моделями коли $p > n$?

1. Передбачення тиску на основі віку, статі та ІМТ + 500000 однонуклеотидних поліморфізмів.

Дані високої вимірності.

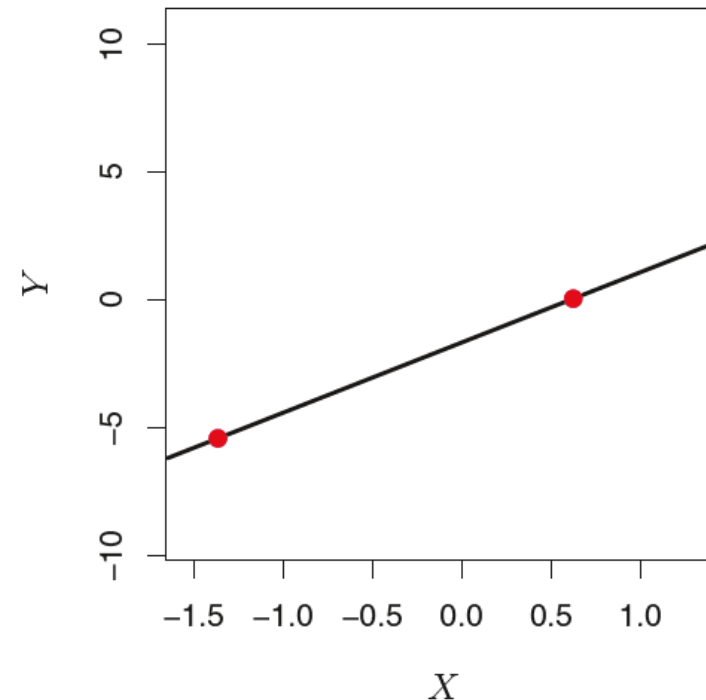
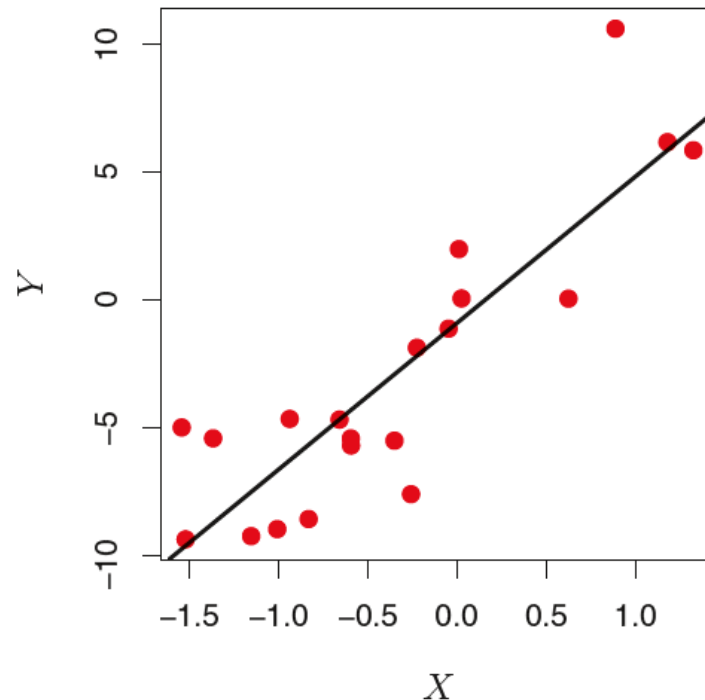
Що відбувається з моделями коли $p > n$?

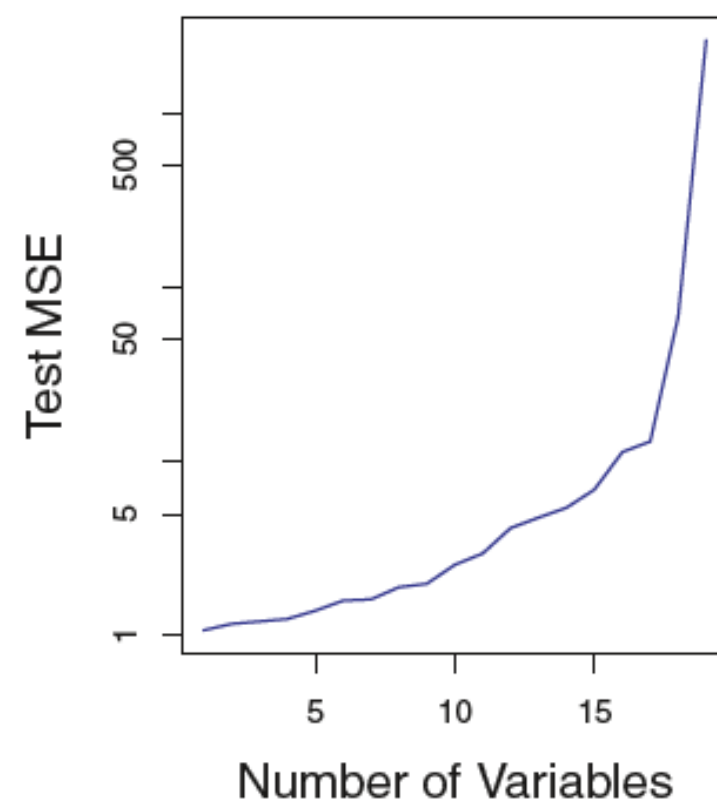
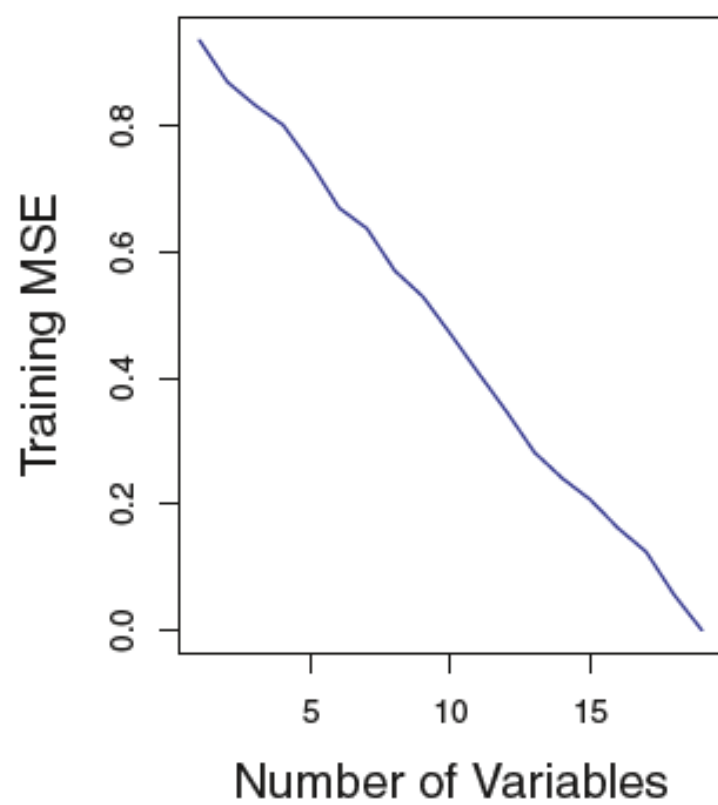
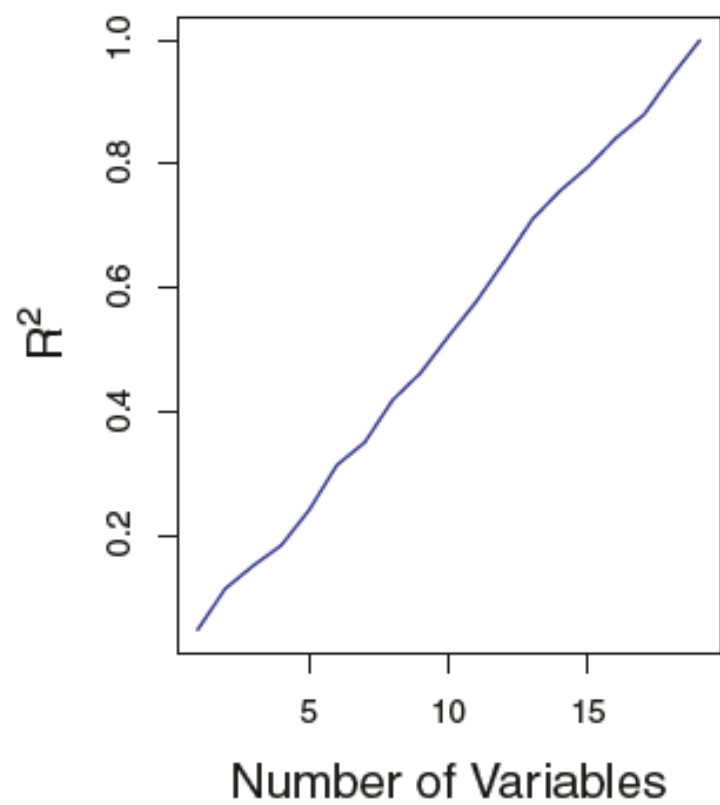
1. Передбачення тиску на основі віку, статі та ІМТ + 500000 однонуклеотидних поліморфізмів.
2. Передбачення онлайн покупок на основі слів введених для пошуку.

Дані високої вимірності.

Що відбувається з моделями коли $p > n$?

1. Передбачення тиску на основі віку, статі та ІМТ + 500000 однонуклеотидних поліморфізмів.
2. Передбачення онлайн покупок на основі слів введених для пошуку.





Приклад:

Приклад: Застосуємо метод лассо до даних з $p = 20, 50, 2000$ з яких лише 20 мають вплив на залежну змінну і $n = 100$.

