

CS381: Numerical Computation & Softwares

Ordinary least square problems

Jakramate Bootkrajang

Department of Computer Science

Chiang Mai University

Implementing simple closed-form solution and Cholesky approach

Recap

- กำหนด data matrix X และ ตัวแปรตาม Y ต้องการหา \vec{a} ที่อธิบายความสัมพันธ์เชิงเส้นระหว่าง X และ Y ได้แม่นยำที่สุด
- วัดความสามารถในการอธิบายความสัมพันธ์เชิงเส้นโดย Square error: $\|X\vec{a} - Y\|_2^2$
- เวกเตอร์ที่ดีที่สุดหาได้จาก $\vec{a} = (X^T X)^{-1} X^T Y$

Naive method

ให้สร้างฟังก์ชันที่รับเมทริกซ์ X เวกเตอร์ Y แล้วคืนค่าเป็น เวกเตอร์ \vec{a} ที่เป็นคำตอบที่ดีที่สุดของปัญหา linear least square.

```
function naive_lle(X, Y)
    ...
    ...
    ...
    return a
end
```

ทดสอบฟังก์ชัน

ลองทำนายราคาบ้านดู (Boston house price prediction) download ข้อมูลจาก

`www.cs.science.cmu.ac.th/person/jakramate/housing.csv`

ข้อมูลเก็บในรูปแบบ csv ที่ 1-13 column แรกคือค่า X และ column สุดท้ายคือค่า Y (ราคาบ้าน)

การอ่าน CSV ใน Julia

ต้องติดตั้ง package CSV ก่อน โดย

```
import Pkg; Pkg.add("CSV")
```

จากนั้นทำการอ่านไฟล์ แล้วแปลงให้อยู่ในรูป matrix และ vector

```
using CSV  
  
data = CSV.read("housing.csv")  
  
X = convert(Matrix, data[:,1:13])  
  
Y = convert(Vector, data[:,end])
```

จากนั้นเรียกใช้ฟังก์ชันที่สร้างไปตอนต้น

```
a = naive_lls(X, Y)
# then evaluate the model
error = (X*a - Y).^2
```

ได้ error เท่าไร ? เราสามารถแสดง ค่าประมาณ และค่าจริงเพื่อเปรียบเทียบโดย

$[X*a \ Y]$

Cholesky least square algorithm

- 1 Compute $X^T X$ and $X^T Y$
- 2 Factorise $X^T X \rightarrow R^T R$
- 3 Solve lower triangular system $R^T w = X^T Y$ for w
- 4 Solve upper triangular system $R\vec{\alpha} = w$ for $\vec{\alpha}$

Choleskey decomposition

ก่อนอื่น import library สำหรับ linear algebra

```
using LinearAlgebra
# factor X'X using Cholesky decomposition
C = cholesky(X'X)
# upper triangular is C.U, lower is C.L
# C.L * C.U is approximately X'X
```

Backward substitution operator

การคำนวณ $Ax = b$ หาก matrix A เป็น lower หรือ upper triangular matrix

$$x = A \setminus b$$

จะคือการทำ backward substitution โดยอัตโนมัติ

Cholesky least square algorithm

1 Compute $X^T X$ and $X^T Y$

2 Factorise $X^T X \rightarrow R^T R$

$$C = \text{cholesky}(X'X)$$

3 Solve lower triangular system $R^T w = X^T Y$ for w

$$w = C.L \setminus X'Y$$

4 Solve upper triangular system $R\vec{\alpha} = w$ for $\vec{\alpha}$

$$b = C.U \setminus w$$

Cholesky VS Naive

- a (คำตอบขอวิธี naive) และ b คำตอบที่ได้จาก Cholesky เท่ากันหรือไม่ ?
- วิธีไหนเร็วกว่า (hint ใช้ tic() toc() จับเวลาดู)