Optimization

A. Lotsi, E. Wit

University of Groningen
JBI of Mathematics and Computer Science

June 3, 2013

rijksuniversiteit
groningen

Outline: APTS

Lectures

- Introduction
- Eigen-decomposition(spectral-decomposition) 2.4.3
- Singular value decomposition
- The QR decomposition

Introduction

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$$

works using "lm" function.

However

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

did not work?

Something has gone badly wrong here !!!!!

Why is that "lm" function works but !!!!

Some terminology

Now consider some more interesting things to do with matrices. First recall the following basic terminology.

$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}$ square	$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{pmatrix}$ symmetric	$\mathbf{y}^\mathsf{T}\mathbf{X}\mathbf{y} > 0 \; \forall \, \mathbf{y} \neq \mathbf{0}$ $(\mathbf{y}^\mathsf{T}\mathbf{X}\mathbf{y} \geq 0 \; \forall \, \mathbf{y} \neq \mathbf{0})$ positive (semi−)definite
$\begin{pmatrix} x_{11} & 0 & 0 \\ x_{21} & x_{22} & 0 \\ x_{31} & x_{32} & x_{33} \end{pmatrix}$ lower triangular	$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ 0 & x_{22} & x_{23} \\ 0 & 0 & x_{33} \end{pmatrix}$ upper triangular	$\begin{pmatrix} x_{11} & 0 & 0 \\ 0 & x_{22} & 0 \\ 0 & 0 & x_{33} \end{pmatrix}$ diagonal

$\mathbf{Q}$ is an **orthogonal matrix** iff $\mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^\mathsf{T} = \mathbf{I}$.

- Choleski decomposition: A matrix square root
- Eigen-decomposition:

Matrix inversion, rank and condition

Consider

$$A = \mathbf{U}\Lambda\mathbf{U}^T$$

With conditions on $\mathbf{U}$ and Λ we have

$$A^{-1} = \mathbf{U}\Lambda^{-1}\mathbf{U}^T$$

Challenges

- if $\lambda_i = 0$ then Λ^{-1} is impossible to find $\Rightarrow$ A is rank deficiency
- A matrix with $\lambda_i \neq 0$ is called FULL RANK MATRIX (FRM).
- Number of non zero eigen values is called ROM
- Generalized Inverse/Pseudo Inverse

Consequence of Near Rank deficiency

Consider the system
$Ax = y \Rightarrow$
$x = A^{-1}y = \mathbf{U}\Lambda^{-1}\mathbf{U}^T y$
Solution not defined if $\lambda_i = 0$.

Rank deficiency matrix can not be inverted

Need to calculate

$$\kappa = max|\lambda_i|/min|\lambda_i|$$

if $\kappa \rightarrow 1/m$ then trouble !!!!!

A system with large κ is referred to as **ill-conditioned**

Orthogonal matrices: $\kappa = 1$

$>$?kappa

EXAMPLE

Want to understand why $\hat{\beta} = (X^T X)^{-1} X^T y$
fail?
higher κ? :
SEE code 2.4.3
Questions

- Could we have noticed this problem directly from X
- How does the “lm” function avoid this problem?

Answers SOON!!!!!!!!!!!!!

preconditioning

Question: How to reduce κ?
System : Given that **D** is diagonal, then

$$\mathbf{D}y = x$$

is solvable however large $\kappa(D)$ is.

Note: $(X^T X)^{-1}$ was impossible
Diagonal preconditioning: $D_{ii} = 1/\sqrt{(X^T X)_{ii}}$.
Cleraly

$$(X^T X)^{-1} = D(DX^T XD)^{-1}D$$

Note that $\kappa(DX^T XD) < \kappa(X^T X)$ SEE code 2.4.4

Asymmetric eigen-decomposition

- Asymmetric matrix ⇒ complex numbers ⇒ complex eigen values and vectors
- Eigen-decomp. of Asym. mat. $= O(n^3)$. BUT more expensive than Sym. case.
- Better to have a decomposition that provides some of the useful properties of the eigen decomposition without the inconvenience of complex numbers.

SVD

SVD

Definition

Suppose $A \sim (r \times c) : r > c$
$SV(A) = \sqrt{(}EV(A^T A))$

If A is psd, then $SV(A) = EV(A)$
Related to SV is SVD

The (SVD) is a *factorization* of a *real* or *complex matrix*.

SVD theorem

Given A whose entries come from field K, either the field of *real numbers* or field of *complex numbers*, there exists a factorization of the form

$$A_{r \times c} = U_{r \times c} D_{c \times c} V^T_{c \times c}$$

-col(U) are left SV (gene coefficient vectors)
-V^T has rows that are right SV (expression level vectors)

SVD

The SVD represents an expansion of the original data in a coordinate system where the covariance matrix is diagonal.

Number of non-zero SV = Rank

SV=most reliable method for numerical rank determination.

$\kappa = d_1/d_c$

Example: Go to R SEE code. 2.5

- We now know the cause of the original problem
- SVD not only provides a diagnosis of the pb. but one possible solution

$$
\begin{aligned}
(X^T X)^{-1} X^T y &= (VDU^T UDV^T)^{-1} VDU^T y \\
&= (VD^2 V^T)^{-1} VDU^T y \\
&= VD^{-2} V^T VDU^T y \\
&= VD^{-1} U^T y
\end{aligned}
$$

Two things:

- κ is that of X
- RHS is a sort of pseudoinverse

SVD has many uses:
Rank approximation: *rank* $k \leq rank(X)$ with

$$\tilde{X} = U\tilde{D}V^T$$

Used to find LRA to observed Σ in high dim. reduction technique in MS

The QR decomposition

SVD: Stable solution for LMF but at higher computational cost.
QR: provides an alternative solution.

$X \sim (r * c)$ rectangular matrix $(r > c)$

$$X = QR$$

QR has cost of $O(rc^2) \approx 1/3 cost(SVD)$
Given the LM example:

$$\begin{aligned}
(X^T X)^{-1} X^T y &= (R^T Q^T Q R)^{-1} R^T Q^T y \\
&= (R^T R)^{-1} R^T Q^T y \\
&= R^{-1} R^{-T} R^T Q^T y \\
&= R^{-1} Q^T y
\end{aligned}$$

SEE code. 2.5

The QR decomposition

Note:

SV(R)=SV(X) system has same κ as$X$.

R routine "lm" uses QR approach

Another application of QR decomposition: determinant calculation.
Given A is square with $A = QR$ then
$|A| = |Q||R| = |R| = \prod R_{ii}$
$\ln |A| = \sum \ln |R_{ii}|$

Conclusion

- Clear idea of the importance of *stability* in matrix computation
- Understanding of key *decompositions* and *applications*
- Dealing with matrices think about *flops* and *condition numbers*

The End