Maths/LA/Tut7
Least Squares

16 Nov 2020

CES

Last updated: 19 Oct 2021

Tutorial 7 Help links

Youtube link: playlist

https://www.youtube.com/playlist?list=PLki3aFwg-9exa_oECiSJtTtaei7zTKwbI

PDF

Q1-6: **https://www.dropbox.com/s/pc33morjzp3fmxm/Tut7_Q1_6_ces.pdf?dl=0**

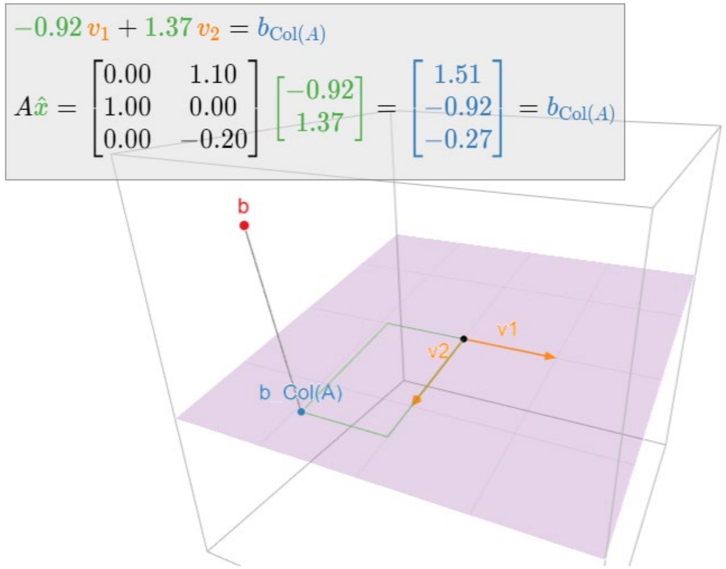
Overview of Least Squares

1) Cornell's CS3220 class:

https://www.cs.cornell.edu/~bindel/class/cs3220-s12/notes/lec10.pdf

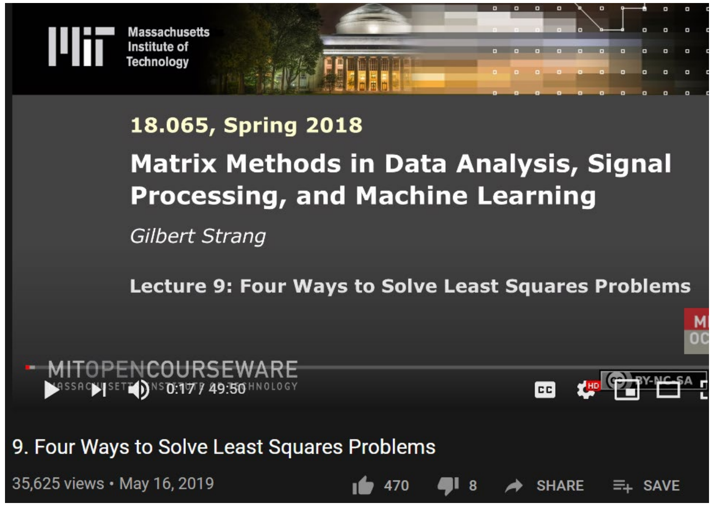
2) GaTech's online book with nice visualization applet

https://textbooks.math.gatech.edu/ila/least-squares.html



How many ways to solve the least squares

- There are several ways to solve the least squares solution.
- See:
  - https://stats.stackexchange.com/questions/160179/do-we-need-gradient-descent-to-find-the-coefficients-of-a-linear-regression-mode/164164#164164
  - Strang's 4 ways to solve the least squares (Advance):
    - https://www.youtube.com/watch?v=ZUU57Q3CFOU



Why is $A^T A$ invertible when A has full col rank (related to Q5-17e NS 5-18d)

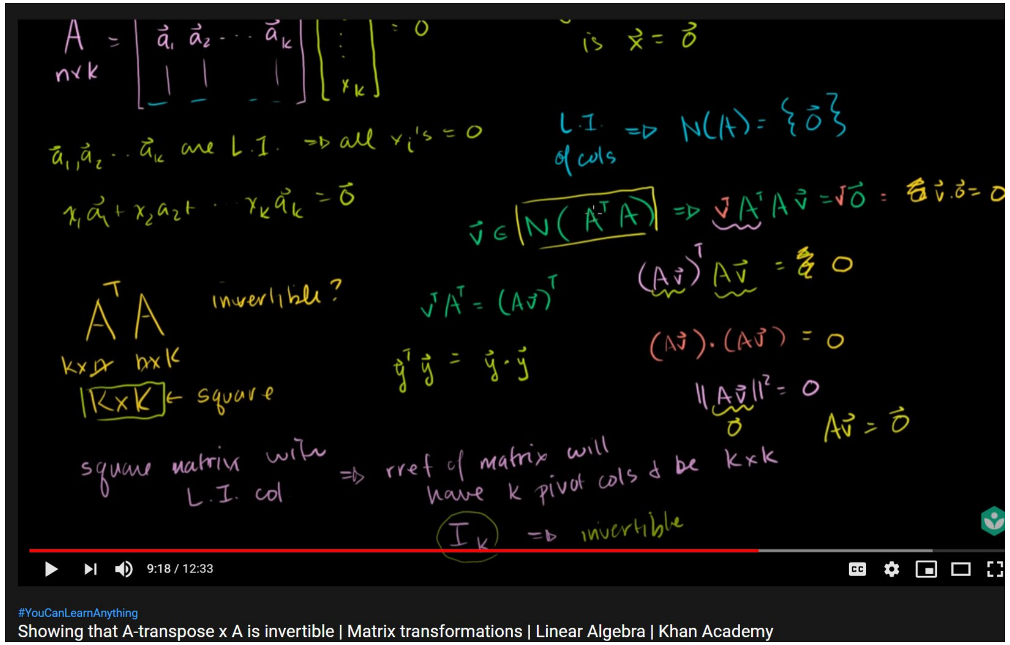
1) Khan Academy:

https://www.youtube.com/watch?v=ESSMQH6Y5OA

2) See rank of $A^T A$ vs A

https://math.stackexchange.com/questions/349738/prove-operatornamerankata-operatornameranka-for-any-a-in-m-m-times-n

3) To proof that rank(A) == rank(A^TA)
So that inverse (A^TA) exist, so that least squares is unique when the columns of A Is independent
https://yutsumura.com/rank-and-nullity-of-a-matrix-nullity-of-transpose/



Proof of Tut 7/Q6

The reader may have noticed that we have been careful to say "the least-squares solutions" in the plural, and "a least-squares solution" using the indefinite article. This is because a least-squares solution need not be unique: indeed, if the columns of A are linearly dependent, then $Ax = b_{\text{Col}(A)}$ has infinitely many solutions. The following theorem, which gives equivalent criteria for uniqueness, is an analogue of this corollary in Section 6.3.

Theorem. *Let A be an $m \times n$ matrix and let b be a vector in $\mathbf{R}^m$. The following are equivalent:*

1. *$Ax = b$ has a unique least-squares solution.*
2. *The columns of A are linearly independent.*
3. *$A^T A$ is invertible.*

In this case, the least-squares solution is

$$\widehat{x} = (A^T A)^{-1} A^T b.$$

Proof.

The set of least-squares solutions of $Ax = b$ is the solution set of the consistent equation $A^T Ax = A^T b$, which is a translate of the solution set of the homogeneous equation $A^T Ax = 0$. Since $A^T A$ is a square matrix, the equivalence of 1 and 3 follows from the invertible matrix theorem in Section 5.1. The set of least squares-solutions is also the solution set of the consistent equation $Ax = b_{\text{Col}(A)}$, which has a unique solution if and only if the columns of A are linearly independent by this important note in Section 2.5.

Important Note 2.5.9 (Recipe: Checking linear independence).

A set of vectors $\{v_1, v_2, \ldots, v_k\}$ is linearly independent if and only if the vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_k v_k = 0$$

has only the trivial solution, if and only if the matrix equation $Ax = 0$ has only the trivial solution, where A is the matrix with columns $v_1, v_2, \ldots, v_k$:

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & & | \end{pmatrix}.$$

This is true if and only if A has a pivot position in every column. Solving the matrix equatiion $Ax = 0$ will either verify that the columns $v_1, v_2, \ldots, v_k$ are linearly independent, or will produce a linear dependence relation by substituting any nonzero values for the free variables.

Ref:
https://textbooks.math.gatech.edu/ila/1553/least-squares.html

Uniqueness of least squares solution ONLY when the columns of A are independent

- https://courses.math.tufts.edu/math70/Section%20Summaries/Chapter6/sect%206.5.pdf

Theorem 14 Let A be an $m \times n$ matrix. The following statements are logically equivalent.
(a) The equation $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution for each $\mathbf{b}$ in $\mathbb{R}^n$.
(b) The columns of A are linearly independent.
(c) The matrix $A^T A$ is invertible.
When these statements are true, the least squares solution $\hat{\mathbf{x}}$ is given by:

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}.$$

(This provides a fast solution method when $(A^T A)^{-1}$ is easy to find.)

Proof of Theorem 14:

$a \to b$ Recall that we proved (or will prove) that $\mathrm{Nul}(A) = \mathrm{Nul}(A^T A)$. Suppose the columns of A are linearly independent. Then $\mathrm{Nul}(A) = \{\mathbf{0}\}$. Since $\mathrm{Nul}(A) = \mathrm{Nul}(A^T A)$, $\mathrm{Nul}(A^T A) = \{\mathbf{0}\}$ also. This means that $(A^T A)\mathbf{x} = \mathbf{0} \to \mathbf{x} = \mathbf{0}$. Thus the columns of $A^T A$ are linearly independent, and since $A^T A$ is a square matrix, $A^T A \sim I$ so $A^T A$ is invertible.

$b \to c$ Suppose the $n \times n$ matrix $A^T A$ is invertible. We can always find least squares solutions using the normal equations:

$$A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$$

But since $A^T A$ is invertible we can apply the inverse to both sides of the previous equation to get:

$$(A^T A)^{-1}(A^T A)\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$$

$$I\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$$

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$$

This is a unique solution.

$c \to a$ Suppose the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b} \in \mathbb{R}^m$. Then since $\mathbf{0}_{\mathbb{R}^m} \in \mathbb{R}^m$, $A\mathbf{x} = \mathbf{0}_{\mathbb{R}^m}$ has a unique solution. Since $\mathbf{x} = \mathbf{0}_{\mathbb{R}^n}$ *is* a solution, it must be the only one. Thus $A\mathbf{x} = \mathbf{0} \to \mathbf{x} = \mathbf{0}$ and so the columns of A are linearly independent.

Infinitely many solutions for least squares (When col of A are dependent)

Example (Infinitely many least-squares solutions).

Find the least-squares solutions of $Ax = b$ where:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 1 & 2 & -3 \end{pmatrix} \qquad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

Solution

We have

$$A^T A = \begin{pmatrix} 3 & 3 & -3 \\ 3 & 5 & -7 \\ -3 & -7 & 11 \end{pmatrix} \qquad A^T b = \begin{pmatrix} 6 \\ 0 \\ 6 \end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\left(\begin{array}{ccc|c} 3 & 3 & -3 & 6 \\ 3 & 5 & -7 & 0 \\ -3 & -7 & 11 & 6 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 5 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

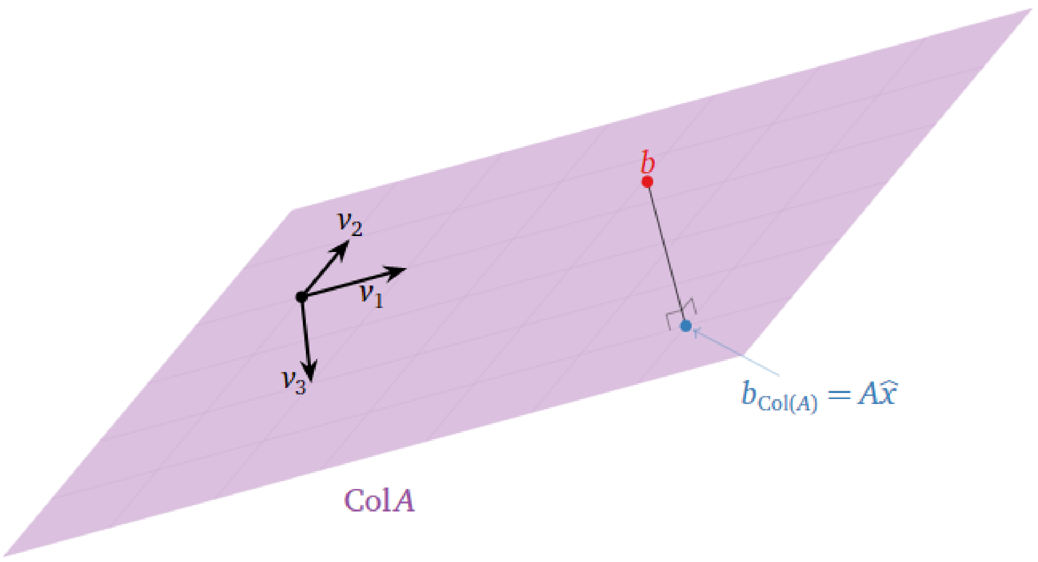
The free variable is x_3, so the solution set is

$$\begin{cases} x_1 = -x_3 + 5 \\ x_2 = 2x_3 - 3 \\ x_3 = \quad x_3 \end{cases} \xrightarrow[\text{vector form}]{\text{parametric}} \hat{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 5 \\ -3 \\ 0 \end{pmatrix}.$$

For example, taking $x_3 = 0$ and $x_3 = 1$ gives the least-squares solutions

$$\hat{x} = \begin{pmatrix} 5 \\ -3 \\ 0 \end{pmatrix} \quad \text{and} \quad \hat{x} = \begin{pmatrix} 4 \\ -1 \\ 1 \end{pmatrix}.$$

Geometrically, we see that the columns v_1, v_2, v_3 of A are coplanar:



Therefore, there are many ways of writing $b_{\text{Col}(A)}$ as a linear combination of v_1, v_2, v_3.

https://textbooks.math.gatech.edu/ila/1553/least-squares.html

Is $A^T b$ in the column space of A?

Corollary:

1) Why does $A^T Ax = A^T b$
(the normal equation)

always have a solution?

Ref:

http://staff.imsa.edu/~fogel/LinAlg/PDF/33%20Least%20Squares.pdf

Let's ask how close we can come to solving the equation. $\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} r \\ h \end{bmatrix}$ is guaranteed to be in the column space of the matrix. So instead of using the real **b**, let's find the thing in the column space that is as close to **b** as possible, and solve for that instead! Let **p** be the projection of **b** into the column space. Then the error vector $\mathbf{e} = \mathbf{b} - \mathbf{p}$ is as small as possible. Let's call the solution to this new problem $\hat{\mathbf{x}}$ so we are solving $A\hat{\mathbf{x}} = \mathbf{p}$. The one thing we know about **e** is that it is orthogonal to the column space, so it is in the left nullspace. That is, $A^T\mathbf{e} = \mathbf{0}$. This means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$, or $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

So instead of $A\mathbf{x} = \mathbf{b}$, we solve the *normal equations* $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$. (We will show later that this *always* has a solution). In this case, we multiply both sides by $A^T = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$ to obtain the system $\begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix}\begin{bmatrix} \hat{r} \\ \hat{h} \end{bmatrix} = \begin{bmatrix} 31 \\ 14 \end{bmatrix}$. A little elimination shows that $\hat{r} = 3/2$ and $\hat{h} = 5/3$.

So we guess our little plant started our 5/3 cm tall and grew at a rate of 3/2 cm/day. This is clearly wrong, since it would predict heights of 19/6, 14/3, and 37/6 instead of 3, 5,and 6, so we're off by –1/6, 1/3, and –1/6 respectively. The is, $\mathbf{e} = (-1/6, 1/3, -1/6)$.

So why is $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$ always solvable? Well, we use our Fundamental Theorem of Linear Algebra. The column space $\boldsymbol{C}(A^T A)$ is the orthogonal complement of the left nullspace of $A^T A$. Well, this is easier in symbols: $\boldsymbol{C}(A^T A) = \left(\boldsymbol{N}(A^T A)^T\right)^{\perp} = \left(\boldsymbol{N}(A^T A)\right)^{\perp} = \left(\boldsymbol{N}(A)\right)^{\perp} = \boldsymbol{C}(A^T)$ (we've seen that A and $A^T A$ have the same nullspace because if $A\mathbf{x} = \mathbf{0}$ certainly $A^T A\mathbf{x} = \mathbf{0}$, but if $A^T A\mathbf{x} = \mathbf{0}$, we multiply on both sides by $\mathbf{x}^T$ and find the $\|A\mathbf{x}\| = 0$, so $A\mathbf{x} = \mathbf{0}$). But since the column spaces of $A^T A$ and A^T are the same, and $A^T\mathbf{b}$ is in the column space of A^T we can certainly always solve $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

Another explanation: why $A^TAc = A^Tx$ is consistent

Ref: https://textbooks.math.gatech.edu/ila/projections.html#projections-ATA-formula

Theorem. *Let A be an $m \times n$ matrix, let $W = \mathrm{Col}(A)$, and let x be a vector in $\mathbf{R}^m$. Then the matrix equation*

$$A^TAc = A^Tx$$

in the unknown vector c is consistent, and x_W is equal to Ac for any solution c.

Proof. Let $x = x_W + x_{W^\perp}$ be the orthogonal decomposition with respect to W. By definition x_W lies in $W = \mathrm{Col}(A)$ and so there is a vector c in $\mathbf{R}^n$ with $Ac = x_W$. Choose any such vector c. We know that $x - x_W = x - Ac$ lies in $W^\perp$, which is equal to $\mathrm{Nul}(A^T)$ by this important note in Section 6.2. We thus have

$$0 = A^T(x - Ac) = A^Tx - A^TAc$$

and so

$$A^TAc = A^Tx.$$

This exactly means that $A^TAc = A^Tx$ is consistent. If c is any solution to $A^TAc = A^Tx$ then by reversing the above logic, we conclude that $x_W = Ac$.

Related: space of $A^T A$ vs A

- https://math.stackexchange.com/questions/1272572/row-space-and-column-space-of-at-a-and-a-at

2

If $Ax = 0$, then $A^T Ax = 0$, which means $N(A) \subset N(A^T A)$, $N(A)$ is the null space of A.

On the other hand, if $A^T Ax = 0$, then

$$x^T A^T Ax = 0, \quad \text{or} \quad \|Ax\| = 0$$

which means $Ax = 0$, and thus

$$N(A^T A) \subset N(A) \quad \text{and} \quad N(A^T A) = N(A)$$

Since $\text{rank}(A) = n - N(A)$, there is

$$\text{rank}(A) = \text{rank}(A^T A)$$

Suppose $A = [\alpha_1, \cdots, \alpha_n]$ (α_i is the column vector of A), then

$$A^T A = A^T[\alpha_1, \cdots, \alpha_n] = [A^T\alpha_1, \cdots, A^T\alpha_n]$$

For each column of $A^T A$

$$\begin{aligned} A^T\alpha_i &= [\beta_1 \cdots \beta_n]\alpha_i \\ &\qquad (\beta_i \text{ is the column of } A^T \text{ and row of } A) \\ &= [\beta_1 \cdots \beta_n]\begin{bmatrix} a_{i1} \\ \vdots \\ a_{in} \end{bmatrix} \\ &= \sum_{j=1}^{n} a_{ij}\beta_j \end{aligned}$$

So column of $A^T A$ is the linear combination of rows of A, or

$$\text{col}(A^T A) = \text{row}(A)$$

Obviously $\text{rank}(A^T) = \text{rank}(A)$, so

$$\text{row}(A^T A) = \text{col}(A^T A) = \text{row}(A)$$

Similarly we have

$$\text{row}(AA^T) = \text{col}(AA^T) = \text{row}(A^T) = \text{col}(A)$$

Nullity of A and $A^T A$

- https://yutsumura.com/rank-and-nullity-of-a-matrix-nullity-of-transpose/

Problem 140

Let A be an $m \times n$ matrix. The nullspace of A is denoted by $\mathcal{N}(A)$. The dimension of the nullspace of A is called the nullity of A.

Prove the followings.

(a) $\mathcal{N}(A) = \mathcal{N}(A^{\mathrm{T}} A)$.

(b) $\mathrm{rank}(A) = \mathrm{rank}(A^{\mathrm{T}} A)$.

Proof.

(a)$\mathcal{N}(A) = \mathcal{N}(A^{\mathrm{T}} A)$.

Show $\mathcal{N}(A) \subset \mathcal{N}(A^{\mathrm{T}} A)$

Consider any $\mathbf{x} \in \mathcal{N}(A)$. Then we have $A\mathbf{x} = \mathbf{0}$. Multiplying it by A^{T} from the left, we obtain

$$A^{\mathrm{T}} A\mathbf{x} = A^{\mathrm{T}}\mathbf{0} = \mathbf{0}.$$

Thus $\mathbf{x} \in \mathcal{N}(A^{\mathrm{T}} A)$, and hence $\mathcal{N}(A) \subset \mathcal{N}(A^{\mathrm{T}} A)$.

Show $\mathcal{N}(A) \supset \mathcal{N}(A^{\mathrm{T}} A)$

On the other hand, let $\mathbf{x} \in \mathcal{N}(A^{\mathrm{T}} A)$. Thus we have

$$A^{\mathrm{T}} A\mathbf{x} = \mathbf{0}.$$

Multiplying it by $\mathbf{x}^{\mathrm{T}}$ from the left, we obtain

$$\mathbf{x}^{\mathrm{T}} A^{\mathrm{T}} A\mathbf{x} = \mathbf{x}^{\mathrm{T}}\mathbf{0} = \mathbf{0}.$$

This implies that we have

$$\mathbf{0} = (A\mathbf{x})^{\mathrm{T}}(A\mathbf{x}) = ||A\mathbf{x}||^2$$

and the length of the vector $A\mathbf{x}$ is zero, thus the vector $A\mathbf{x} = \mathbf{0}$. Hence $\mathbf{x} \in \mathcal{N}(A)$, and we obtain $\mathcal{N}(A) \supset \mathcal{N}(A^{\mathrm{T}} A)$.

(b) $\mathrm{rank}(A) = \mathrm{rank}(A^{\mathrm{T}} A)$

We use the rank-nullity theorem and obtain

$$\mathrm{rank}(A) = n - \dim(\mathcal{N}(A)) = n - \dim(\mathcal{N}(A^{\mathrm{T}} A)) = \mathrm{rank}(A^{\mathrm{T}} A).$$

(Note that the size of the matrix $A^{\mathrm{T}} A$ is $n \times n$.)

Rank of A and $A^T A$ are same

- https://math.stackexchange.com/questions/349738/prove-operatornamerankata-operatornameranka-for-any-a-in-m-m-times-n

- And therefore if A is tall and full rank, then $A^T A$ is invertible

Method1: Using dimension and rank

127

Let $\mathbf{x} \in N(A)$ where $N(A)$ is the null space of A.

So,

$$\begin{aligned} A\mathbf{x} &= \mathbf{0} \\ \implies A^T A\mathbf{x} &= \mathbf{0} \\ \implies \mathbf{x} &\in N(A^T A) \end{aligned}$$

Hence $N(A) \subseteq N(A^T A)$.

Again let $\mathbf{x} \in N(A^T A)$

So,

$$\begin{aligned} A^T A\mathbf{x} &= \mathbf{0} \\ \implies \mathbf{x}^T A^T A\mathbf{x} &= \mathbf{0} \\ \implies (A\mathbf{x})^T (A\mathbf{x}) &= \mathbf{0} \\ \implies A\mathbf{x} &= \mathbf{0} \\ \implies \mathbf{x} &\in N(A) \end{aligned}$$

Hence $N(A^T A) \subseteq N(A)$.

Therefore

$$\begin{aligned} N(A^T A) &= N(A) \\ \implies \dim(N(A^T A)) &= \dim(N(A)) \\ \implies \operatorname{rank}(A^T A) &= \operatorname{rank}(A) \end{aligned}$$

Method2: Using SVD

Let r be the rank of $A \in \mathbb{R}^{m \times n}$. We then have the SVD of A as

$$A_{m\times n} = U_{m\times r}\Sigma_{r\times r}V_{r\times n}^T$$

This gives $A^T A$ as

$$A^T A = V_{n\times r}\Sigma_{r\times r}^2 V_{r\times n}^T$$

which is nothing but the SVD of $A^T A$. From this it is clear that $A^T A$ also has rank r. In fact the singular values of $A^T A$ are nothing but the square of the singular values of A.

share cite improve this answer follow

answered Apr 3 '13 at 3:25

user17762

Why QR is more stable? (related to Q5-18f)

See explanation in slide 8.16- 8.17 in

- http://www.seas.ucla.edu/~vandenbe/133A/lectures/ls.pdf

Projection Matrix Proof

$P = A(A^T A)^{-1} A^T$

Assume that A has full column rank, (all columns independent)
Show that above P has two properties

a) $P = P^T$
b) $PP = P$

Pre-amples:

1) We need to proof $(A^{-1})^T = (A^T)^{-1}$

- see: https://math.stackexchange.com/questions/340233/transpose-of-inverse-vs-inverse-of-transpose

I would derive the formula step by step this way.

Lets have invertible matrix A, so you can write following equation (definition of inverse matrix):

$$AA^{-1} = I$$

Lets transpose both sides of equation. (using $I^T = I$, $(XY)^T = Y^T X^T$)

$$(AA^{-1})^T = I^T$$

$$(A^{-1})^T A^T = I$$

From the last equation we can say (based on the definition of inverse matrix) that A^T is inverse of $(A^{-1})^T$. So we can write following.

$$((A^{-1})^T)^{-1} = A^T$$

By inverting both sides of equation we obtain the desired formula.

$$(A^{-1})^T = (A^T)^{-1}$$