ST 661 Note

Frances Lin

1/11/2022

Outline

pd matrices

Determinant

Orthogonal matrices

Thm (*)

Suppose A is symmetric, A is pd iff ∃ a nonsingular matrix P s.t. $A = P^T P$.

Proof. <=

Suppose $A = P^T P$ for some nonsingular P, then

$$y^T A y = y^T P^T P y = (Py)^T P y \geq 0$$

The last step sort of works like a sum of square. In addition,

$$(Py)^T P y = 0$$

only if $Py = 0$.

=> $y = 0$ because P is nonsingular (or full rank)

=> A is pd since whenever $y \neq 0$, $y^T A y > 0$ □

Proof. => is included in the HW and requires eigenvalues.

If we relax the assumption such that P is nonsingular, then we get psd.

Comment. P is not unique, unless… see the following

Cholesky Decomposition

Suppose that A is pd, then it is possible to find a nonsingular upper triangular matrix

$$T = \begin{bmatrix} . & * & * \\ 0 & . & * \\ 0 & 0 & . \end{bmatrix}$$

s.t. $A = T^T T$.

Thm

Let B be a nxp matrix,

i) If $rank(B) = p$, then $B^T B$ is pd.

ii) If $rank(B) < p$, then $B^T B$ is psd, not pd. (psd can be pd but not in this case.)

Proof. Exercise. Use quadric form from previous proof and rank.

Thm

If A is pd, A^{-1} is also pd.

Proof. Use Thm $(*)$,

A is pd => $A = P^T P$ for nonsingular P, then

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

=> A^{-1} is pd.

Determinant

Check textbook for full definition. As a demonstration, suppose

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

then $det(A) = a_{11}a_{22} - a_{12}a_{21}$ is a degree 2 polynomial function, for example. In general, let A be nxn, then $det(A)$ (or $|A|$) is a scalar degree n polynomial function of A.

Thm

i) If

$$D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & d_n \end{bmatrix},$$

is diagonal, then $det(D) = d_1 d_2 ... d_n$.

If D_2 is upper (or lower) triangular matrix such as

$$D_2 = \begin{bmatrix} d_1 & * & * \\ 0 & \cdots & * \\ 0 & 0 & d_n \end{bmatrix},$$

then $det(D) = d_1 d_2 ... d_n$.

ii) A is square matrix and A is singular <=> $det(A) = 0$.

iii) If A is pd, then $det(A) > 0$. (Not the inverse.)

If A is psd, then $det(A) \leq 0$. (Again, not the inverse.)

iv) $det(A) = det(A^T)$

v) If A is singular ($det(A) \neq 0$), then $det(A^{-1}) = \frac{1}{det(A)}$.

vi) If cA where c is scalar and A is nxn, then $det(cA) = c^n det(A)$.

e.g.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

and

$$2A = \begin{bmatrix} 2a_{11} & 2a_{12} \\ 2a_{21} & 2a_{22} \end{bmatrix}.$$

Thm

If A and B are both square nxn matrices, then

$$det(AB) = det(A)det(B).$$

Corollary

i)

$$det(BA) = det(BA)$$

ii) (special case of thm)

$$det(A^k) = (det(A))^k$$

e.g.

$$det(A^2) = (det(A))^2$$

Thm

If A is square and partitioned as

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

and A_{11} and A_{22} are both square and nonsingular (can be of different size), then

$$det(A) = det(A_{11})det(A_{22} - A_{21}(A_{11})^{-1}A_{12}) = det(A_{22})det(A_{11} - A_{12}(A_{22})^{-1}A_{21}).$$

Intuition.

$$det(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}) = a_{11}a_{22} - a_{12}a_{21} = a_{11}(a_{22} - \frac{a_{12}a_{21}}{a_{11}}) = a_{22}(a_{11} - \frac{a_{12}a_{21}}{a_{22}}).$$

Corollary

Suppose

$$A = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}$$

or

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

and assume A_{11} and A_{22} are square, then

$$det(A) = det(A_{11})det(A_{22}).$$

This is the extension of the previous thm where

i) If

$$D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & d_n \end{bmatrix},$$

is diagonal, then $det(D) = d_1 d_2 ... d_n$.

Thm

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

and assume

1) A_{11} & A_{12} are square
2) A_{11} is nonsingular
3) $B = A_{22} - A_{21}(A_{11})^{-1}A_{12}$ is nonsingular, then

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} - B^{-1}A_{21}A_{11}^{-1} + A_{11}^{-1}A_{12}B^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}B^{-1} \\ -B^{-1}A_{21}A_{11}^{-1} & B^{-1} \end{bmatrix}.$$

? $dim(A_{11}) = dim(3\ terms)$

Note. We only need B^{-1} and A_{11}^{-1} and the rest are matrix multiplication.

In addition, this is helpful when we need to compare $A_{11}^{-1} - B^{-1}A_{21}A_{11}^{-1} + A_{11}^{-1}A_{12}B^{-1}A_{21}A_{11}^{-1}$ to A_{11}^{-1}.

e.g.

$$X_{new} = \begin{bmatrix} X & Z \end{bmatrix}$$

$${X_{new}}^T X_{new} = \begin{bmatrix} X^T \\ Z^T \end{bmatrix} \begin{bmatrix} X & Z \end{bmatrix} = \begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z \end{bmatrix}$$

Orthogonal Matrices

Suppose

$$a = \begin{bmatrix} a_1 & a_2 \end{bmatrix}^T$$

length of a (or $|a|$) $= \sqrt(a_1^2 + a_2^2) = \sqrt(a^T a)$.

a is an unit vector iff length of length of a (or $|a|$) $= 1$.

Suppose

$$b = \begin{bmatrix} b_1 & b_2 \end{bmatrix}^T,$$

if a and b are both unit vectors, then

$$cos(\theta) = a_1 b_1 + a_2 b_2 = a^T b = b^T a = < a, b > .$$

The last term is called an inner product.

In general, if a and b are not unit vectors, then

$$cos(\theta) = \frac{a^T b}{\sqrt(a^T a)\sqrt(b^T b)},$$

where $\sqrt(a^T a)$ is length of a and $\sqrt(b^T b)$ is length of b.

a and b are perpendicular (i.e. $\theta = 90^0$) iff

$$cos(\theta) = 0 = a^T b \quad (or \quad b^T a = 0)$$

Definition

In general, if a is an $n-$dim vector, then length of a

$$= \sqrt(a^T a) = \sqrt(a_1^2 + a_2^2 + ... + a_n^2).$$

Definition

Suppose $a, b \in \mathbb{R}^n$, a and b are said to be orthogonal ("perpendicular" in 2D) iff

$$a^T b \quad (or \quad b^T a) = a_1 b_1 + a_2 b_2 + ... + a_n b_n = 0.$$

Definition

If $a^T a = 1 = |a|$, then a is said to be normalized. (Any vector b can be normalized by

$$c = \frac{b}{\sqrt(b^T b)}.$$

Proof. $c^T c = ... = 1$.

Definition

A set of $p-$dim vectors $c_1, c_2, ...c_p$ that are normalized and mutually orthogonal are said to be an orthonormal set of vectors.

i.e.

i) $c_i^T c_i = 1$ for any i

ii) $c_i^T c_j = 0$ for any $i \neq j$

Definition

If a pxp matrix $C = (c_1, c_2, ...c_p)$ has orthonormal columns, then C is called an orthogonal matrix.

Corollary

If C is orthogonal, then $C^TC = I$.

$$C^TC = \begin{bmatrix} c_1^T \\ c_2^T \\ \vdots \\ c_p^T \end{bmatrix} \begin{bmatrix} c_1^T & c_2^T & \cdots & c_p^T \end{bmatrix} =$$

(i, j) entry of $C^TC = c_i^T c_j = 0$ when $i \neq j$. $C^TC = c_i^T c_j = 1$ when $i = j$, so

$$C^TC = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I.$$

Remark.

$$C^TC = I <=> CC^T = I$$

$$C \textit{ is orthogonal} <=> CC^T = I$$

$CC^T = I$ means that the rows of C are orthonormal.

End. Verify A^{-1} and review the final definition.