FALL 2011

STAT 8003: Statistical Method I

Lecture 2

Sep. 15, 2011

1. Definition

A matrix is a rectangle array of numbers. e.g.

$$M = \begin{pmatrix} 2 & 7 \\ 10 & -5 \\ 8 & 6 \end{pmatrix}$$

We refer M as a 3x2 matrix. An nxp matrix is of the form

$$M = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix}$$

A vector is a matrix that has only one row (called a row vector) or one column (called a column vector). e.g.

$$a = \begin{pmatrix} 7 \\ 2 \\ 3 \end{pmatrix}, \quad b = (7,2,3)$$

A scalor is a matrix with only one row and one column

A square matrix has as many rows as it has cohemn. e.g

$$M = \begin{pmatrix} 1 & 2 & 5 \\ 4 & 2 & 6 \\ 7 & 3 & 2 \end{pmatrix}$$

A symmetric matrix is a square matrix with $a_{ij} = a_{ji}$ for all i and j e.g.

$$M = \begin{pmatrix} 4 & 3 & 2 \\ 3 & 1 & 8 \\ 2 & 8 & 3 \end{pmatrix}$$

A Diagonal matrix is a symmetric matrix where all the off-diagonal elements are 0. e.g.

$$M = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

An identity matrix is a diagonal matrix with 1s and only 1s on the diagonal. e.g.

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

2. Matix Operation

2.1 Matrix addition

If we have two matrices of the same size, then we can add them together by adding their corresponding elements. e.g.

$$A = \begin{pmatrix} 10 & 2 \\ -3 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 2 & -5 \\ 2 & 6 \end{pmatrix}$$

then $A + B = \begin{pmatrix} 12 & -3 \\ -1 & 7 \end{pmatrix}$

2.2 Matrix Multiplication

When we multiply a matrix by a scalor we multiply each element of the matrix by that number. e.g.

$$2\begin{pmatrix} 3 & -4 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 6 & -8 \\ 4 & 2 \end{pmatrix}$$

To multiply two matrices together, calculate the dot product of each row of the first matrix with each column of the second matrix. The dot product is calculated by multiplying each element of the row with the corresponding element of the column and adding up the result.

$$A_{n\times p} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix} \qquad B_{p\times k} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pk} \end{pmatrix}$$

then $C_{n\times k} = AB = \begin{pmatrix} c_{11} & \cdots & c_{1k} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nk} \end{pmatrix}$, where $c_{ij} = \sum_{l=1}^{k} a_{il} b_{lk}$

eg. $$(10 \quad -1 \quad 3)\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} = 10\times 2 + (-1)\times 0 + 3\times 1 = 23$$

$$\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}(10 \quad -1 \quad 3) = \begin{pmatrix} 20 & -2 & 6 \\ 0 & 0 & 0 \\ 10 & -1 & 3 \end{pmatrix}$$

Note: ① The mutiplication of AB exists only when the column number of A is equal to the row number of B so that A and B are comformable.

② If A and B are comformable (AB exists) as well as B and A (BA exists), it is NOT always true that $AB = BA$.

When multiplying a scalor and a matrix, do the multiplication elementwisely.

$$2\begin{pmatrix}1 & 0\\ 2 & 4\end{pmatrix} = \begin{pmatrix}2 & 0\\ 4 & 8\end{pmatrix}$$

For identity matrix I, if the multipliaation is comformable, then

$AI = A$, $IB = B$.

△ From now on, unless otherwise specified, we assume all the multiplications are comformable.

2.3 Transposition

To transpose a matrix, interchange its rows and columns. e.g.

$$\begin{pmatrix}3 & 0 & 6\\ 4 & 9 & 5\end{pmatrix}^T = \begin{pmatrix}3 & 4\\ 0 & 9\\ 6 & 5\end{pmatrix}$$

$(AB)^T = B^T A^T$

2.4 Trace of a square matrix

The trace of a square matrix is the sum of its diagonal elements

Let $A = (a_{ij})_{n\times n}$ $\quad tr(A) = \sum_{i=1}^{n} a_{ii}$

① $tr(A+B) = tr(A) + tr(B)$ ② $tr(AB) = tr(BA)$ ③ $tr(ABC) = tr(CAB)$

2.5 The determinant of a square matrix

The determinant is a scalor-valued function of a matrix.

$A = (a_{ij})_{n\times n}$ then $\det(A) = |A| = \sum_{i_1,\cdots,i_n} \varepsilon_{i_1,\cdots,i_n} a_{1,i_1} \cdots a_{n,i_n}$

where $\varepsilon_{i_1,\cdots i_n} = \begin{cases} 1 & \text{if } (i_1,\cdots,i_n) \text{ is an even permutation of } \{1,\cdots,n\} \\ -1 & \text{if } (i_1,\cdots,i_n) \text{ is an odd permutation of } \{1,\cdots,n\} \end{cases}$

so that $\varepsilon_{i_1,\cdots,i_n} = \prod_{j=1}^{n-1} \left(\frac{1}{j!} \prod_{k=j+1}^{n} (i_j - i_k) \right)$

e.g. $A_{3\times3} = \begin{pmatrix} 1 & 0 & 4 \\ 2 & 3 & 1 \\ 4 & 6 & 0 \end{pmatrix} = (a_{ij})$

Permutions of $\{1,2,3\}$ total # of permutations: $3\times2\times1=6$

even permutition $\{1,2,3\}$ $\{2,3,1\}$ $\{3,1,2\}$

odd permutation $\{2,1,3\}$ $\{3,2,1\}$ $\{1,3,2\}$

$$\det(A) = |A| = \underset{\{1,2,3\}}{1\times3\times0} + \underset{\{2,3,1\}}{0\times1\times4} + \underset{\{3,1,2\}}{4\times2\times6}$$

$$- \underset{\{2,1,3\}}{0\times2\times0} - \underset{\{3,2,1\}}{4\times3\times4} - \underset{\{1,3,2\}}{1\times1\times6}$$

$$= 0+0+48-0-48-6$$

$$= -6$$

Tricks: even permutions $\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}_{3\times3}$

odd permutations $\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}_{3\times3}$

Laplace's formula and the adjacent matrix

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} A_{ij} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} A_{ij}$$

A_{ij} is the determinant of the matrix deleting ith row and jth column in A.

e.g $A = \begin{pmatrix} 1 & 0 & 4 \\ 2 & 3 & 1 \\ 4 & 6 & 0 \end{pmatrix}$

Take $i=1$, $\det(A) = 1\cdot\begin{vmatrix} 3 & 1 \\ 6 & 0 \end{vmatrix} - 0\cdot\begin{vmatrix} 2 & 1 \\ 4 & 0 \end{vmatrix} + 4\begin{vmatrix} 2 & 3 \\ 4 & 6 \end{vmatrix}$

$$= -6 + 4\times0 = -6$$

26 Rank

Linear independency: A family of vectors is linearly independent if none of them can be written as a linear combination of finitely many other vectors in the collection. A family of vectors is not linearly independent is called linearly dependent. e.g.

$$\underbrace{\overbrace{\begin{matrix}x_1\\ \parallel \\ \begin{pmatrix}0\\0\\1\end{pmatrix}\end{matrix}\quad \begin{matrix}x_2\\ \parallel \\ \begin{pmatrix}0\\2\\-2\end{pmatrix}\end{matrix}\quad \begin{matrix}x_3\\ \parallel \\ \begin{pmatrix}1\\-2\\1\end{pmatrix}\end{matrix}}^{\text{independent}}\quad \begin{matrix}x_4\\ \parallel \\ \begin{pmatrix}4\\2\\3\end{pmatrix}\end{matrix}}_{\text{dependent}}$$

$$\boxed{\begin{aligned} &x_4 = 4x_3 + 5x_2 + 9x_1 \\ &\Updownarrow \\ &x_1 = \left(-\tfrac{5}{9}\right)x_2 + \left(-\tfrac{4}{9}\right)x_3 + \tfrac{1}{9}x_4 \\ &\Updownarrow \\ &x_1 + \tfrac{5}{9}x_2 + \tfrac{4}{9}x_3 - \tfrac{1}{9}x_4 \end{aligned}}$$

$x_1, \cdots, x_n$ are linearly dependent if and only if there exists a set of n scalors, $a_1, \cdots, a_n$, not all zero, such that

$$a_1 x_1 + \cdots + a_n x_n = \underset{\uparrow}{0}$$

zero vector, not a scalor

The column rank of A is the maximum number of linearly independant column vectors of A. The row rank of a matrix A is the maximum number of a matrix A is the maximum number of linearly independent row vectors of A. The column rank and row rank are always equal. e.g

$$A = \begin{pmatrix} 0 & 0 & 1 & 4 \\ 0 & 2 & -2 & 2 \\ 1 & -2 & 1 & 3 \end{pmatrix} \qquad \left.\begin{aligned} \text{column rank} &= 3 \\ \text{row rank} &= 3 \end{aligned}\right\} \Rightarrow \text{rank}(A) = 3$$

Some properties of the rank of a matrix:

1. For an $m \times n$ matrix A

$$\text{rank}(A) \leq \min(m,n)$$

If $\text{rank}(A) = \min(m,n)$, then A is called a full rank matrix.

2. $\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$
3. $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$
4. For $n \times n$ matrix A, $|A| = 0$ if and only if $\text{rank}(A) < n$.
5. For nonsingular matrix A, B and any matrix C, then

$\text{rank}(A_{n \times n}) = n$

$$\text{rank}(C) = \text{rank}(AC) = \text{rank}(CB) = \text{rank}(ACB)$$

6. $\text{rank}(A) = \text{rank}(A^T) = \text{rank}(A^TA) = \text{rank}(AA^T)$
7. For an $m \times n$ matrix A and an $m \times 1$ vector b.

$$\text{rank}([A, b]) \geq \text{rank}(A)$$

(adding a column to A can't reduce its rank)

2.7 Inverse of a square matrix

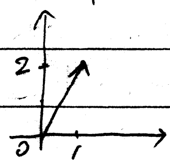
If $\text{rank}(A_{n \times n}) = n$, then $A_{n \times n}$ can be inversed. Use A^{-1} to denote A's inverse. $AA^{-1} = A^{-1}A = I$

To solve linear equations $A_{n \times n} x_{n \times 1} = b_{n \times 1}$

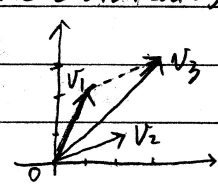
if $\text{rank}(A) = n$, then $x = A^{-1}b$

2.8 Geometric concepts:

vector: directed line segment extending from the origin (the point 0) to the point indicated by the coordinates (elements) of the vector



$v = (1,2)$



$v_1 = (1,2)$ $v_2 = (2,1)$

$v_3 = v_1 + v_2 = (3,3)$

3. Linear spaces:

All the vector of length n generate the R^n space.

$R^n = \{v : v = (v_1, \dots, v_n),\ v_i \in R\}$

R^n is a special case of the more general concept of a vector space:

Let V denote a set of n-dimensional vectors. $\forall\ v_i, v_j \in V$

if $v_i + v_j \in V$ and $c v_i \in V$ for any $c \in \mathbb{R}$.

then V is said to be a vector space of order n.

3.1 Spanning set, Linear Independence, and Basis.

$V = \mathcal{L}(v_1, \dots, v_k)$ if V contains and only contains the linear combinations of $v_1, \dots, v_k$: $c_1 v_1 + c_2 v_2 + \dots + c_k v_k$ $\quad c_i \in R \quad i = 1, \dots, k$.

If the spanning set of vectors $(v_1, \dots, v_k)$ also has the property of linear independence, then $\{v_1, \dots, v_k\}$ is called a basis of V.

Recall: Linearly independent (LIN)

$$\Updownarrow$$

no redundancy among $\{v_1, \dots, v_k\}$

Every basis of a given vector space V has the same number of elements. The number is called dimension or rank of V.

e.g. $x_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ $\quad x_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ $\quad x_3 = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}$ are all in $\mathbb{R}^3$

$$V = \mathcal{L}(x_1, x_2, x_3) = \{a x_1 + b x_2 + c x_3 \mid a, b, c \in \mathbb{R}\}$$

$$= \left\{ \begin{pmatrix} a+b+3c \\ 2a-b \\ 0 \end{pmatrix} \middle| a, b, c \in \mathbb{R} \right\} = \left\{ \begin{matrix} d \\ e \\ 0 \end{matrix} \middle| d, e \in \mathbb{R} \right\}$$

Note that x_1, x_2, x_3 are not LIN, since $x_3 = x_1 + 2x_2$

The LIN can be removed by eliminating any one of the three vectors from the set.

What are the basis of V? $\{x_1, x_2\}$ or $\{x_1, x_3\}$ or $\{x_2, x_3\}$

V is of order 3 but $V \neq \mathbb{R}^3$.

3.2 Subspaces.

Let W be a vector space and V be a set with $V \subset W$. Then V is a subspace of W iff V is also a vector space

e.g. In the last example $V \subset \mathbb{R}^3$, $V = \mathcal{L}(x_1, x_2, x_3)$. V is a subspace of $\mathbb{R}^3$.

3.3 Column Space, Rank of a Matrix

The column space of a matrix A is denoted $C(A)$, and defined as the space spanned by the columns of A. i.e.

$$A_{n \times m} = \left(a_1, \cdots, a_m \right) \qquad a_i \in \mathbb{R}^n, \ i=1,\dots,m$$

then $C(A) = \mathcal{L}(a_1, \cdots, a_m)$

$\dim(C(A)) = \text{rank}(A)$

4. Inner Products, Norm, Orthogonality and Projections.

4.1 Inner product

$\forall x, y \in \mathbb{R}^n$. We define $\langle x,y \rangle$ to be the inner product operation given by $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i = x^T y = y^T x$.

Inner product is also called dot product.

Properties of the inner product:

① $\langle x, y \rangle = \langle y, x \rangle$

② $\langle ax, y \rangle = a \langle x, y \rangle \qquad a \in \mathbb{R}$

③ $\langle x_1 + x_2, y \rangle = \langle x_1, y \rangle + \langle x_2, y \rangle$

4.2 Norm

The ℓ_p norm of a vector $x \in \mathbb{R}^n$ is defined to be

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Specially, the ℓ_2 norm of x is $\|x\|_2 = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} = \left(\langle x, x \rangle \right)^{1/2} = \left(x^T x \right)^{1/2}$

4.3 Angle

The inner product between two vectors $x, y \in \mathbb{R}^n$ quantifies the angle between them. In particular, if θ is the angle formed between x and y then $\cos\theta = \frac{\langle x, y \rangle}{\|x\|\|y\|}$.

4.4 Orthogonality

$x, y \in \mathbb{R}^n$, they are said to be orthogonal (i.e. perpendicular) if $\langle x, y \rangle = 0$. The orthogonality of x and y is denoted with the notation $x \perp y$.

e.g

$$X = (x_1, x_2, x_3, x_4) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

Then $x_2 \perp x_3$, $x_2 \perp x_4$, $x_3 \perp x_4$ and the other pairs are not orthogonal, i.e. x_2, x_3, x_4 are mutually orthogonal.

The length of the vectors are

$\|x_1\| = \sqrt{1^2+1^2+1^2+1^2+1^2+1^2} = \sqrt{6}$

$\|x_2\| = \|x_3\| = \|x_4\| = \sqrt{2}$

Pythagorean Theorem: Let $v_1, v_2, \ldots, v_k$ be mutually orthogonal vectors in a vector space V; Then

$$\left\| \sum_{i=1}^{k} v_i \right\|^2 = \sum_{i=1}^{k} \|v_i\|^2$$

Proof: $\left\| \sum_{i=1}^{k} v_i \right\|^2 = \left\langle \sum_{i=1}^{k} v_i, \sum_{j=1}^{k} v_j \right\rangle$

$$= \sum_{i=1}^{k} \left\langle v_i, \sum_{j=1}^{k} v_j \right\rangle$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{k} \langle v_i, v_j \rangle = \sum_{i=1}^{k} \langle v_i, v_i \rangle + \sum_{i=1}^{k} \sum_{j \neq i} \langle v_i, v_j \rangle$$

Since $v_1, \ldots, v_k$ are mutually orthogonal, $\forall i \neq j$ $\langle v_i, v_j \rangle = 0$

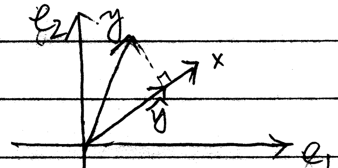
And therefore $\sum_{i=1}^{k} \sum_{j=1}^{k} \langle v_i, v_j \rangle = \sum_{i=1}^{k} \langle v_i, v_i \rangle = \sum_{i=1}^{k} \|v_i\|^2$.

4.5 Projections.

The (orthogonal) projection of a vector y on a vector x is the vector $\hat{y}$ such that

① $\hat{y} = bx$ for some constant b; and

② $(y-\hat{y}) \perp x$ (or equivalently, $\langle \hat{y}, x \rangle = \langle y, x \rangle$)



$\hat{y} = bx$.

The notation $p(y|x)$ will denote the projection of y on x.

How to find $\hat{y} = p(y|x)$?

$$\hat{y}^T x = y^T x = (b^T x)^T x = b\|x\|^2$$

$$b\|x\|^2 = y^T x \Rightarrow b = \frac{y^T x}{\|x\|^2} \quad \text{unless } x = 0$$

$$\text{So } \hat{y} = \begin{cases} 0 & \text{when } x = 0 \\ \left(\frac{y^T x}{\|x\|^2}\right) x & \text{otherwise} \end{cases}$$

Cauchy-Schwartz Inequality:

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$$

Proof: Let $\hat{y} = \mathrm{Proj}(y|x) = bx$

$$\text{LHS} = |\langle x, y \rangle|^2 = |\langle x, bx + (y-\hat{y}) \rangle|^2 = b^2 (\langle x, x \rangle)^2$$

$$\begin{aligned} \text{RHS} &= \langle x, x \rangle \langle y, y \rangle \\ &= \langle x, x \rangle \langle bx + (y-\hat{y}), bx + (y-\hat{y}) \rangle \\ &= \langle x, x \rangle \left(b^2 \langle x, x \rangle + \langle y-\hat{y}, y-\hat{y} \rangle \right) \\ &= b^2 (\langle x, x \rangle)^2 + \langle x, x \rangle \langle y-\hat{y}, y-\hat{y} \rangle \end{aligned}$$

$\therefore$ LHS $\leq$ RHS and the equality holds iff $x = 0$ or $y = \hat{y} \in C(x)$.

4.6 Projection onto a Subspace

The projection of a vector y on a subspace V of $\mathbb{R}^n$ is the vector $\hat{y} \in V$ such that $(y-\hat{y}) \perp V$. The vector $e = y - \hat{y}$ will be called the residual vector for y relative to V.

$y = \text{proj}(y|V) \Rightarrow y - \hat{y} \perp V \quad \hat{y} \in V$

Let $V = \mathcal{L}(X) = \mathcal{L}(x_1, \dots, x_p)$

$\hat{y} = P(y|V)$ iff $\langle \hat{y}, x_i \rangle = \langle y, x_i \rangle$ for $i = 1, \dots, p$

If $\{x_1, \dots, x_n\}$ is an orthogonal basis of V, then

$$\hat{y} = \sum_{i=1}^{n} \frac{\langle x_i, y \rangle}{\|x_i\|} x_i$$

4.7 Projection Matrices

Definition: P is a (orthogonal) projection matrix onto V iff

(i) $v \in V$ implies $Pv = v$ (projection); and

(ii) $w \perp V$ implies $Pw = 0$ (orthogonality)

For X an $n \times p$ matrix of full rank, we can show $P = X(X^TX)^{-1}X^T_{(n\times n)}$ is a projection matrix onto $V = C(X)$.

Check: (i) $\forall v \in V = C(x_1, \dots, x_p) = a_1x_1 + \dots + a_px_p = Xa$

$$Pv = P(a^TX) = X(X^TX)^{-1}X^TXa = Xa = v$$

(ii) $\forall w \perp V = C(X)$ then $X^Tw = 0$

$$Pw = X(X^TX)^{-1}X^Tw = 0$$

Theorem: P is a projection matrix onto its column space $C(P) \in \mathbb{R}^n$

iff i) $PP = P$ (idempotent), and

(ii) $P = P^T$ (symmetric)

Now, for $X_{n\times p}$ full rank, $P = X(X^TX)^{-1}X^T$ is a projection matrix onto $C(P)$

because (i) P is symmetric

(ii) P is idempodent.

It leads to the conclusion: $C(X) = C(X(X^TX)^{-1}X^T)$

Theorem: Projection matrices are uniqeee.

5. Eigenvalues and Eigenvectors

An $n \times n$ square matrix A is said to have an eigenvalue λ, with a corresponding eigenvector $v \neq 0$ if $Av = \lambda v \iff (A - \lambda I)v = 0$

How to get eigenvalues

Set $|A - \lambda I| = 0$. $\because$ $|A-\lambda I|$ is a polynomial in λ of order n, so there will be n (not necessarily, and not necessarily distinct) eigenvalues (solutions).

Definition: A is called orthogonal matrix if $A^{-1} = A^T$

Spectral Decomposition: If $A_{n\times n}$ is symmetric then there exists an orthogonal matrix U, such that

$$A = U_{n\times n} \Lambda_{n\times n} U_{n\times n}^T \quad \text{where}$$

$$\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \qquad \lambda_i \in \mathbb{R}$$

$$A = U\Lambda U^T = (\lambda_1 u_1 \cdots \lambda_n u_n) \begin{pmatrix} u_1^T \\ \vdots \\ u_n^T \end{pmatrix} = \sum_{i=1}^n \lambda_i u_i u_i^T$$

Note that $u_i u_i^T = P_i$ $i=1,\dots,n$ are projections onto the one-dimensional subspace $\mathcal{L}(u_i)$

Results:

① $tr(A) = \sum_{i=1}^n \lambda_i$ and $|A| = \prod_{i=1}^n \lambda_i$

② $A^{-1} = (U\Lambda U^T)^{-1} = U\Lambda^{-1}U^T$

③ If A is positive semidefinite (p.s.d) i.e. $\lambda_i \geq 0$ $i=1,\dots,n$ then $A^{1/2} = U\Lambda^{1/2}U^T$ with $\Lambda^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{pmatrix}$

6 Quadratic Forms:

For a symmetric matrix $A_{n\times n}$ a quadratic form in $x_{n\times 1}$ is defined by

$$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

Quadratic forms are going to arise frequently in linear models as squared lengths of projections, or sum of squares.

The spectral decomposition can be used to "diagonalize" the matrix in a quadratic form $\forall x \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$

$$x^T A x = x^T U \Lambda U^T x = (Ux)^T \Lambda \underbrace{(Ux)}_{y} = y^T \Lambda y$$

For projection matrix P

(i) $\forall v \in V$, $Pv = v$ $\forall v \in V$ are eigenvectors of P with eigenvalues equal to 1.

(ii) $\forall w \in V^\perp$ $Pw = 0$ $\forall w \in V^\perp$ are eigenvectors of P with eigenvalues equal to 0.

7 Positive Definite Matrix

For an $n\times n$ symmetric matrix A.

(i) $A_{n\times n}$ is positive definite iff all A's eigenvalues > 0

(ii) $A_{n\times n}$ is semi positive definite iff all A's eigenvalues ≥ 0.

Properties

(1) Let $A_{n\times n}$ to be p.d and $B_{k\times n}$ with $\text{rank}(B) = k \leq n$ then BAB^T is p.d.

(2) Let $A_{n\times n}$ to be p.d and $B_{k\times p}$ with $\text{rank}(B) < \min(k,n)$, then BAB^T is p.s.d

(3) A symmetric matrix A is p.d iff there exists a non-singular matrix M such as $A = M^T M$

(4) A p.d. matrix is non-singular.

(5) Let $B \in \mathbb{R}^{n\times p}$

if rank $(B) = p$. then $B^T B$ is p.d.

If rank $(B) < p$ then $B^T B$ is p.s.d

(6) If A is p.d. then A^{-1} is p.d.

(7) If A is p.d and is partitioned in the form $\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, where A_{11} and A_{22} are square, then A_{11} and A_{22} are both p.d.

8. Generalized Inverse

$\forall$ $n\times k$ matrix A, it's generalized inverse A^- is any $k\times n$ matrix satisfying $AA^-A = A$. (1)

① A^- always exists

② A^- is not unique

③ If A is non-singular, then $A^- = A^{-1}$ (unique)

e.g. $A = \begin{pmatrix} 2 & 2 & 3 \\ 1 & 0 & 1 \\ 3 & 2 & 4 \end{pmatrix} = (x_1, x_2, x_3)$

x_1, x_2 LIN, $x_3 = x_1 + x_2 \Rightarrow$ rank$(A) = 2$

$A_1^- = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $A_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -3/2 & 1/2 \\ 0 & 0 & 0 \end{pmatrix}$

Some Properties of generalized inverse

① $(X^-)^T$ is a generalized inverse of X^T. Furthermore, if X is symmetric, then $(X^-)^T$ is a generalized inverse of X

proof: $(X^T)(X^-)^T(X^T) = (XX^-X)^T = X^T$

For symmetric matrix $X(X^-)^TX = X$

(2) If $\text{rank}(X_{n \times k}) = r$ then $\text{rank}(X^{-}X) = \text{rank}(XX^{-}) = \text{rank}(X) = r$.

(3) If G and H are two generalized inverse of X^TX. then

$$X = XGX^TX = XHX^TX$$

Proof: For $v \in \mathbb{R}^n$ Let $v = v_1 + v_2$ where $v_1 \in C(X)$ and $v_2 \perp C(X)$

Suppose $v_1 = Xb$ for some $b \in \mathbb{R}^n$

Then $v^TXGX^TX = v_1^TXGX^TX = b^T(X^TX)G(X^TX) = b^T(X^TX) = v^TX$

Since this is true for all v. then $XGX^TX = X$.

(4) G is a generalized inverse of X^TX, then GX^T is a generalized inverse of X.

(5) G, H are two generalized inverse of X^TX. then XGX^T is idempodent, symmetric and is invariant to the choice of G. i.e: $XGX^T = XHX^T$

Result (4) says that $X(X^TX)^-X$ is symmetric and idempodent for any generalized inverse $(X^TX)^-$. Therefore $X(X^TX)^-X^T$ is the unique projection matrix onto $C(X(X^TX)^-X^T)$

Theorem: $X(X^TX)^-X^T$ is the projection matrix onto $C(X)$

A generalized inverse X^- of X which satisfies (1) and has the additional properties

(2) $X^-XX^- = X^-$

(3) X^-X is symmetric and

(4) XX^- is symmetric

is unique, and is known as Moore-Penrose Inverse.

For linear equations $Ax=c$. it is not necessary that there exists a solution of x. When there is a solution of x, we call the system to be consistent.
A necessary and sufficient condition for $Ax=c$ to be consistent is
$rank(A) = rank([A\ c])$
If the system of equations $Ax=c$ is consistent and if A^- is any generalized inverse of A, then $x = A^-c$ is a solution
Proof: $AA^-Ax = Ax \Rightarrow AA^-c = c \Rightarrow A(A^-c) = c$
For consistent systems of equations with >1 solutions, different choices of A^- will yield different solutions of $Ax=c$.
Using this result, we can prove another necessary and sufficient condition for: $Ax=c$ to be consistent.
Theorem: The system of equations $Ax=c$ has a solution if and only if for any genalized inverse A^- of A, it is true that $AA^-c=c$.