6. Orthogonality and Least Squares

1) Inner Product, Length, orthogonality
2) Orthogonal Sets
3) Orthogonal Projection
4) The Gram-Schmidt Process
5) Least Square problem

1. Inner Product, Length, Orthogonality

1) Inner Product

(1) Definition : Inner product

If $\boldsymbol{u}$, $\boldsymbol{v}$ is in R^{n}

$\boldsymbol{u} \bullet \boldsymbol{v} = \boldsymbol{v}^T \boldsymbol{u}$

: The inner product of $\boldsymbol{u}$ and $\boldsymbol{v}$

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_n \end{bmatrix},\ \boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \cdots \\ v_n \end{bmatrix}$$

$$\boldsymbol{u} \bullet \boldsymbol{v} = \boldsymbol{v}^T \boldsymbol{u} = \begin{bmatrix} v_1\, v_2 \cdots v_n \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

ex) $\boldsymbol{u} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix},\ \boldsymbol{v} = \begin{bmatrix} 4 \\ 2 \\ -1 \end{bmatrix}$

(2) Property of the inner product

If $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ is in R^{n} and k is a scalar

① $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$
② $(\boldsymbol{u}+\boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$
③ $(c\boldsymbol{u}) \cdot \boldsymbol{v} = c(\boldsymbol{u} \cdot \boldsymbol{v})$
④ $\boldsymbol{u} \cdot \boldsymbol{u} \geqq 0$, $\boldsymbol{u} \cdot \boldsymbol{u} = 0$ if and only if $\boldsymbol{u} = \mathbf{0}$

2) The length of a Vector

(1) Definition : Length

The length(or norm) of $\boldsymbol{v}$ is the non negative scalar $\|\boldsymbol{v}\|$

A vector whose length is 1 is called a unit vector

(2) properties of length

① $\|c\boldsymbol{v}\| = |c| \|\boldsymbol{v}\|$

② normalizing : divide a nonzero vector $\boldsymbol{v}$ by its length – unit vector $\dfrac{1}{\|\boldsymbol{v}\|}\boldsymbol{v}$

ex) $\boldsymbol{v} = (1, -2, 2, 0)$

3) Distance

(1) Definition : distance

If $\boldsymbol{u}$, $\boldsymbol{v}$ is in R^{n}

distance between $\boldsymbol{u}$ and $\boldsymbol{v}$: length of the vector $\boldsymbol{u}-\boldsymbol{v}$

$dist(\boldsymbol{u},\boldsymbol{v}) = \|\boldsymbol{u}-\boldsymbol{v}\|$

ex) $\boldsymbol{u}=(u_1, u_2, u_3)$, $\boldsymbol{v}=(v_1, v_2, v_3)$

$dist(\boldsymbol{u},\boldsymbol{v}) = \|\boldsymbol{u}-\boldsymbol{v}\|$

4) Orthogonal Vectors

(1) Definition : Orthogonal Vectors

Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in R^{n} are orthogonal if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$

the zero vector is orthogonal to every vector in R^{n}

(2) Theorem

Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal if and only if $\|\boldsymbol{u}+\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$

5) Orthogonal Complements

(1) Definition : Orthogonal Complements

If a vector $\boldsymbol{z}$ is orthogonal to every vector in a subspace W of R^{n}, then $\boldsymbol{z}$ is orthogonal to W

The set of all vectors $\boldsymbol{z}$ that are orthogonal to W is orthogonal complement of W

$$W^{\perp} = \{\boldsymbol{z} | \boldsymbol{z} \cdot \boldsymbol{w} = 0 \text{ for } all \ \boldsymbol{w} \in W\}$$

(2) Theorem

① A vector $\boldsymbol{x}$ is in $W^{\perp}$ if and only if $\boldsymbol{x}$ is orthogonal to every vector in a set that spans W

② $W^{\perp}$ is a subspace of R^{n}

③ Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A

The orthogonal complement of the column space of A is the null space of A^T

$$(RowA)^{\perp} = NulA,\ (ColA)^{\perp} = NulA^T$$

④ Let W be a subspace of R^{n}. then

$$\dim W + \dim W^{\perp} = n$$

2. Orthogonal Sets

1) Orthogonal Set

(1) Definition : Orthogonal set

A set of vectors $\{u_1, u_2, \cdots, u_p\}$ in R^n is an orthogonal set if each pair of distinct vectors from the set is orthogonal,

$u_i \cdot u_j = 0$ for all $i \neq j$

ex) $u_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$, $u_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$, $u_3 = \begin{bmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{bmatrix}$

(2) Theorem

If $S = \{u_1, u_2, \cdots, u_p\}$ is an orthogonal set of nonzero vectors in R^n, then S is linearly independent and hence is a basis for the subspace spanned by S

(3) Definition : Orthogonal basis

An orthogonal basis for a subspace W of R^{n} is a basis for W that is also an orthogonal set

(4) Theorem

Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_p\}$ be an orthogonal basis for a subspace W of R^{n}. For each $\boldsymbol{y}$ in W, the weights in the linear combination

$$\boldsymbol{y} = c_1 \boldsymbol{u}_1 + \cdots + c_p \boldsymbol{u}_p$$

and $c_j = \dfrac{\boldsymbol{y} \bullet \boldsymbol{u}_j}{\boldsymbol{u}_j \bullet \boldsymbol{u}_j}$ $(j = 1, \cdots, p)$

ex) $\boldsymbol{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$, $\boldsymbol{u}_3 = \begin{bmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{bmatrix}$, $\boldsymbol{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$

2) Orthonormal Set

(1) Definition : Orthonormal Sets

A set $\{u_1, u_2, \cdots, u_p\}$ is an orthonormal set if it is an orthogonal set of unit vectors

ex) Standard basis $\{e_1, e_2, \cdots, e_n\}$ for R^{n}

ex) $u_1 = \begin{bmatrix} \frac{3}{\sqrt{11}} \\ \frac{1}{\sqrt{11}} \\ \frac{1}{\sqrt{11}} \end{bmatrix}$, $u_2 = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}$, $u_3 = \begin{bmatrix} -\frac{1}{\sqrt{66}} \\ -\frac{2}{\sqrt{66}} \\ \frac{7}{\sqrt{66}} \end{bmatrix}$

(2) Theorem

① An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$

② Let U be an $m \times n$ matrix with orthonormal columns, and let $\boldsymbol{x}$ and $\boldsymbol{y}$ be in R^{n}

a. $\| U\boldsymbol{x} \| = \| \boldsymbol{x} \|$
b. $(U\boldsymbol{x}) \cdot (U\boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{y}$
c. $(U\boldsymbol{x}) \cdot (U\boldsymbol{y}) = 0$ if and only if $\boldsymbol{x} \cdot \boldsymbol{y} = 0$

(3) Definition : Orthogonal matrix

An orthogonal matrix is a square matrix U such that $U^{-1} = U^T$

Such a matrix has orthonormal columns

(4) Theorem

An orthogonal matrix have orthonormal rows

ex) ex) $\boldsymbol{u}_1 = \begin{bmatrix} \frac{3}{\sqrt{11}} \\ \frac{1}{\sqrt{11}} \\ \frac{1}{\sqrt{11}} \end{bmatrix}$, $\boldsymbol{u}_2 = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}$, $\boldsymbol{u}_3 = \begin{bmatrix} -\frac{1}{\sqrt{66}} \\ -\frac{2}{\sqrt{66}} \\ \frac{7}{\sqrt{66}} \end{bmatrix}$

3. Orthogonal Projection

1) Orthogonal projection

(1) Purpose : Orthogonal Projection

Given a vector $\boldsymbol{y}$ and a subspace W in $\mathrm{R^n}$, there is a vector $\hat{\boldsymbol{y}}$ in W such that

① $\hat{\boldsymbol{y}}$ is the unique vector in W for which $\boldsymbol{y} - \hat{\boldsymbol{y}}$ is orthogonal to W
② $\hat{\boldsymbol{y}}$ is the unique vector in W closest to $\boldsymbol{y}$

(2) How to find $\hat{\boldsymbol{y}}$

When $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_p\}$ is an orthogonal basis for $\mathrm{R^n}$, then the vector $\boldsymbol{y}$ in $\mathrm{R^n}$ can be written as

$\boldsymbol{y} = \boldsymbol{z}_1 + \boldsymbol{z}_2$

$\boldsymbol{z}_1$ is a linear combination of some of the $\boldsymbol{u}_i$ and $\boldsymbol{z}_2$ is a linear combination of the rest of the $\boldsymbol{u}_i$

Theorem

Lew W be a subspace of R^{n}. Then each $\boldsymbol{y}$ in R^{n} can be written uniquely in the form

$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{z}$

where $\hat{\boldsymbol{y}}$ is in W and $\boldsymbol{z}$ is in $W^{\perp}$. In fact, if $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_p\}$ is any orthogonal basis for W, then

$$\hat{\boldsymbol{y}} = \frac{\boldsymbol{u}_1 \bullet \boldsymbol{y}}{\boldsymbol{u}_1 \bullet \boldsymbol{u}_1}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{u}_p \bullet \boldsymbol{y}}{\boldsymbol{u}_p \bullet \boldsymbol{u}_p}\boldsymbol{u}_p$$

and $\boldsymbol{z} = \boldsymbol{y} - \hat{\boldsymbol{y}}$

$\hat{\boldsymbol{y}}$: the orthogonal projection of $\boldsymbol{y}$ onto W : $proj_W \boldsymbol{y}$

ex) Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_5\}$ be an orthogonal basis for R^5 and let

$\boldsymbol{y} = c_1\boldsymbol{u}_1 + \cdots + c_5\boldsymbol{u}_5$

Consider the subspace $W = Span\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$

ex) $\boldsymbol{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\boldsymbol{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $W = Span\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$

2) Properties of orthogonal projection

(1) Theorem

Let $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_p}\}$ be an orthogonal basis for a subspace W of R^n. For each $\boldsymbol{y}$ in W, the weights in the linear combination

$$\boldsymbol{y} = c_1\boldsymbol{u_1} + \cdots + c_p\boldsymbol{u_p}$$

where $c_i = \dfrac{\boldsymbol{u_i} \cdot \boldsymbol{y}}{\boldsymbol{u_i} \cdot \boldsymbol{u_i}}\boldsymbol{u_i}$

In this case, $proj_W\boldsymbol{y} = \boldsymbol{y}$

(2) Theorem : The best approximation theorem

Let W be a subspace of R^n, let $\boldsymbol{y}$ be any vector in R^n, and let $\hat{\boldsymbol{y}}$ be the orthogonal projection of $\boldsymbol{y}$ onto W. Then $\hat{\boldsymbol{y}}$ is the closest point in W to $\boldsymbol{y}$, in the sense that

$$\| \boldsymbol{y} - \hat{\boldsymbol{y}} \| < \| \boldsymbol{y} - \boldsymbol{v} \|$$

for all $\boldsymbol{v}$ in W distinct from $\hat{\boldsymbol{y}}$

(3) notation

$\hat{y}$: the best approximation to y by elements of W

$\| y - v \|$: error of using v in place of y

error is minimized when $v = \hat{y}$

$\hat{y}$ does not depend on the particular orthogonal basis used to compute it

(4) Theorem

If $\{u_1, u_2, \cdots, u_p\}$ is an orthonormal basis for a subspace W of R^n, then

$\hat{y} = (u_1 \bullet y)u_1 + \cdots + (u_p \bullet y)u_p$

If $U = [u_1 \ u_2 \ \cdots u_p]$, then

$proj_W y = UU^T y$ for all $y \in \mathrm{R}^n$

4. The Gram-Schmidt Process

1) The Gram-Schmidt Process

(1) The Gram-Schmidt Process

The Gram-Schmidt process is a simple algorithm for producing an orthogonal or orthonormal basis

ex) Let $W = Span\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$, where $\boldsymbol{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$, $\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Construct an orthogonal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ for W

ex) $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$, then $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ is linearly independent and is a basis for a subspace W for R^4. Construct an orthogonal basis for W

(2) Theorem : The Gram–Schmidt Process

Given a basis $\{x_1, \cdots, x_p\}$ for a nonzero subspace W of R^{n}, define

$$v_1 = x_1$$

$$v_2 = x_2 - \frac{x_2 \bullet v_1}{v_1 \bullet v_1} v_1$$

$$v_3 = x_3 - \frac{x_3 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_3 \bullet v_2}{v_2 \bullet v_2} v_2$$

…

$$v_p = x_p - \frac{x_p \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_p \bullet v_2}{v_2 \bullet v_2} v_2 - \cdots - \frac{x_p \bullet v_{p-1}}{v_{p-1} \bullet v_{p-1}} v_{p-1}$$

Then $\{v_1, v_2, \cdots, v_p\}$ is an orthogonal basis for W. In addition

$Span\{v_1, \cdots, x_k\} = Span\{x_1, \cdots, x_k\}$ for $1 \leqq k \leqq p$

(3) Orthonormal Bases

An orthonormal basis is constructed easily from an orthogonal basis $\{v_1, v_2, \cdots, v_p\}$: simply normalize all the v_k

5. Least-Squares Problems

1) Least-Squares Problems

(1) Least-Squares problem

Consider the system $\boldsymbol{Ax} = \boldsymbol{b}$ that has no solution

When a solution is demanded and none exists, the best one can do is to find an $\boldsymbol{x}$ that makes $A\boldsymbol{x}$ as close as possible to $\boldsymbol{b}$

Think of $A\boldsymbol{x}$ as an approximation to $\boldsymbol{b}$. The smaller the distance between $\boldsymbol{b}$ and $A\boldsymbol{x}$, given by $\| \boldsymbol{b} - A\boldsymbol{x} \|$, the better the approximation

The general least-squares problem is to find an $\boldsymbol{x}$ that makes $\| \boldsymbol{b} - A\boldsymbol{x} \|$ as small as possible

(2) Definition : Least-squares solution

If A is $m \times n$ matrix and $\boldsymbol{b}$ is in R^{m}, a least-squares solution of $A\boldsymbol{x} = \boldsymbol{b}$ is an $\hat{\boldsymbol{x}}$ in R^{n} such that
$\| \boldsymbol{b} - A\hat{\boldsymbol{x}} \| \leqq \| \boldsymbol{b} - A\boldsymbol{x} \|$
for all $\boldsymbol{x}$ in R^{n}
point : for all $\boldsymbol{x}$ in R^{n}, $A\boldsymbol{x}$ is in the $Col\,A$ - We seek an $\boldsymbol{x}$ that makes $A\boldsymbol{x}$ the closest point in $Col\,A$ to $\boldsymbol{b}$

(3) How to find the solution

least-squares solution of $A\boldsymbol{x}=\boldsymbol{b}$ is an $\hat{\boldsymbol{x}}$ such that

$\| \boldsymbol{b}-A\hat{\boldsymbol{x}} \| \leqq \| \boldsymbol{b}-A\boldsymbol{x} \|$ for all $\boldsymbol{x}$ in $\mathrm{R^n}$

Let $\hat{\boldsymbol{b}}=proj_{ColA}\boldsymbol{b}$

then $\hat{\boldsymbol{b}}$ is in the column space A, the equation $A\boldsymbol{x}=\hat{\boldsymbol{b}}$ is consistent, and there is an $\hat{\boldsymbol{x}}$ in $\mathrm{R^n}$ such that

$A\hat{\boldsymbol{x}}=\hat{\boldsymbol{b}}$

Since $\hat{\boldsymbol{b}}$ is the closest point in $Col\,A$ to $\boldsymbol{b}$, a vector $\hat{\boldsymbol{x}}$ is a least-squares solution of $A\boldsymbol{x}=\boldsymbol{b}$
Suppose $\hat{\boldsymbol{x}}$ satisfies $A\hat{\boldsymbol{x}}=\hat{\boldsymbol{b}}$

By the Orthogonal Decomposition Theorem, the projection $\hat{\boldsymbol{b}}$ has the property that $\boldsymbol{b}-\hat{\boldsymbol{b}}$ is orthogonal to $ColA$, so $\boldsymbol{b}-A\hat{\boldsymbol{x}}$ is orthogonal to each column of A
If $\boldsymbol{a}_j$ is any column of A, then $\boldsymbol{a}_j \cdot (\boldsymbol{b}-A\hat{\boldsymbol{x}})=0$, and $\boldsymbol{a}_j{}^T(\boldsymbol{b}-A\hat{\boldsymbol{x}})=0$
since $\boldsymbol{a}_j{}^T$ is a row of A^T,

$A^T(\boldsymbol{b}-A\hat{\boldsymbol{x}})=0$

$A^TA\hat{\boldsymbol{x}}=A^T\boldsymbol{b}$

Finding the least-squares solution : solve the equation $A^TA\hat{\boldsymbol{x}}=A^T\boldsymbol{b}$

(4) Theorem

① The set of least-squares solutions of $A\boldsymbol{x}=\boldsymbol{b}$ coincides with the nonempty set of solutions of the normal equation $A^{T}A\boldsymbol{x}=A^{T}\boldsymbol{b}$

② Let A be an $m\times n$ matrix. The following statements are logically equivalent

a. The equation $A\boldsymbol{x}=\boldsymbol{b}$ has a unique least-squares solution for each $\boldsymbol{b}$ in R^{m}
b. The columns of A are linearly independent.
c. The matrix $A^{T}A$ is invertible

When these staements are true, the least-squares solution $\hat{\boldsymbol{x}}$ is given by

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

ex) $A=\begin{bmatrix}4&0\\0&2\\1&1\end{bmatrix}$, $\boldsymbol{b}=\begin{bmatrix}2\\0\\11\end{bmatrix}$

2) Application : linear regression

(1) linear regression

given the data set (x_1, y_1), (x_2, y_2), …, (x_n, y_n), find the line $y = \beta_0 + \beta_1 x$ that is close to the points

Finding β_0, β_1 : linear regression

(2) Finding β_0, β_1

let $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix}$, $Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$

Find the least squares solution of the equation

$Y = X\beta$