Applications of Inner Product Spaces

M. Rajesh[1]

Department Of Mathematics, S.R. R Govt. Arts & Science College (A), Karimnagar– 505001

Abstract

The aim of the present paper is to know the applications of Inner product spaces. We can approximate a line or a polynomial for a set of points in the plane. The method of approximating a line or a polynomial for a set of points in the plane s the method of least squares. An inner product of two vectors in a vector space is a scalar which is the product of transpose of first vector and second vector. The cosine angle between two vectors can be finding by using the inner product and the length of vectors. We defined orthogonal vectors, orthonormal vectors, orthogonal basis and orthonormal basis and Gram-Schmidt orthogonalization process.

By using the least squares method, we can find a line or polynomial for a given set of points in the plane. By using the method of least squares we can approximate exponential, logarithmic and trigonometric functions to a line or a polynomial.

Keywords: Inner product, Norm of a vector, angle between two vectors, orthogonal vectors, orthonormal vectors, Schwarz's inequality, Triangle inequality,Pythagoras theorem, Gram-Schmidt orthogonalization process, the method of least squares.

INTRODUCTION

(1) Inner product of two vectors: Inner product of two vectors **u** and**v**of a vector space

is denoted by **u. v** anddefined as$\mathbf{u}^{\mathbf{T}}\mathbf{v}$.

If **u**=$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ - \\ - \\ - \\ u_n \end{bmatrix}$ **v**= $\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ - \\ - \\ - \\ v_n \end{bmatrix}$ then **u. v**= $\mathbf{u}^{\mathbf{T}}\mathbf{v}$=$[\, u_1\ u_2\ u_3 - - - -\ \ u_n] \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ - \\ - \\ - \\ v_n \end{bmatrix}$

$= u_1 v_1 + u_2 v_{2+} u_3 v_3$+---------+ $u_n v_n$

Example: If **u**=$\begin{bmatrix} 1 \\ 4 \\ -6 \end{bmatrix}$,**v**=$\begin{bmatrix} 2 \\ 6 \\ 9 \end{bmatrix}$ are vectors in R^3 then **u. v**= $\mathbf{u}^{\mathbf{T}}\mathbf{v}$=$[1\ \ 4\ \ -6] \begin{bmatrix} 2 \\ 6 \\ 9 \end{bmatrix}$ =1.2+4.6+(-6).9 =2+24-54=-28

Hence the inner product of **u, v**=**u. v**=-28

(2) Inner Product Space

An inner product on a vector space V is a function f that, to each pair of vectors **u** and **v** in V, associates a real number f(**u.v)** and satisfies the following axiomsfor all u, v, and w in V and all scalars c

1. f (**u**.**v**) =f (**v**.**u**) 2. f(**u**+**v**, **w**)=f(**u**,**w**)+f(**v**,**w**) 3. f (c**u**,**v**) =c f(**u**,**v**) 4.f (**u**.**u**) ≥0 and **u**.**u**=o if and only if u=0. A vector space with an inner product f is called an inner product space.

(3) Length of a Vector: The length of a vector **u** is denoted by **u** and defined as || **u** || =$\sqrt{\mathbf{u}.\mathbf{u}}$

Example: If **u**= $\begin{bmatrix}1\\-2\\3\end{bmatrix}$ then || **u** || = $\sqrt{[1-\ 2\ \ 3]\begin{bmatrix}1\\-2\\3\end{bmatrix}}$=$\sqrt{(1)^2+(-2)^2+(3)^2}$=$\sqrt{1+4+9}$ =$\sqrt{14}$. Hence || **u** ||= $\sqrt{14}$

(4) Unit Vector: A vector **u** in R^n is said to be a unit vector if length of **u** is 1 i.e. || **u** || =1

As $\|\frac{\boldsymbol{u}}{\|\mathbf{u}\|}\|$ =1,$\frac{\boldsymbol{u}}{\|\mathbf{u}\|}$ is unit vector in direction of **u**.

Ex:$If\ \boldsymbol{u}=\begin{bmatrix}\frac{1}{\sqrt{3}}\\-\frac{1}{\sqrt{3}}\\-\frac{1}{\sqrt{3}}\end{bmatrix}$ then **u** is a unit vector because || **u** || =$\sqrt{(\frac{1}{\sqrt{3}})^2+(\frac{-1}{\sqrt{3}})^2+(\frac{-1}{\sqrt{3}})^2}=\sqrt{\frac{1}{3}+\frac{1}{3}+\frac{1}{3}}=\sqrt{\frac{3}{3}}=1$

(5) Angle between two vectors

Definition: If **u**,**v** are two vectors and θis angle between **u, v** thencos θ=$\frac{\boldsymbol{u}.\boldsymbol{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$

Example: If **u**=$\begin{bmatrix}1\\-3\\2\end{bmatrix}$,**v**=$\begin{bmatrix}1\\6\\9\end{bmatrix}$ then cos θ=$\frac{[1-3\ \ 2]\begin{bmatrix}1\\6\\9\end{bmatrix}}{\|\begin{bmatrix}1\\-3\\2\end{bmatrix}\|\|\begin{bmatrix}1\\6\\9\end{bmatrix}\|}=\frac{1(1)+(-3)(6)+2(9)}{(\sqrt{(1)^2+(-3)^2+(2)^2})(\sqrt{(1)^2+(6)^2+(9)^2})}$ =$\frac{1-18+18}{\sqrt{14}\sqrt{118}}$ =$\frac{1}{2\sqrt{413}}\Rightarrow\cos\theta=\frac{1}{2\sqrt{413}}$

Hence the angle between two vectors **u**= $\begin{bmatrix}1\\-3\\2\end{bmatrix}$,**v**=$\begin{bmatrix}1\\6\\9\end{bmatrix}$ $is\ \ \theta=\cos^{-1}(\frac{1}{2\sqrt{413}})$

(6) Orthogonal vectors

Definition: Orthogonal vectors two vectors **u,v** are said to be orthogonal vectors if angle between **u, v** is90^0 i.e.cos θ=cos90^0=0.Hence two vectors **u, v** are said to be orthogonal vectors if $\cos\theta=0$, $where$ θ is the angle between the vectors **u**, **v**.

Example: If **u**= $\begin{bmatrix}1\\4\\-6\end{bmatrix}$,v=$\begin{bmatrix}2\\-5\\-3\end{bmatrix}$ then **u. v**=$\mathbf{u}^{\mathbf{T}}$**v** =$1(2)+4(-5)+(-6)(-3)$=2-20+18=0

Hence **u. v**=0. ∴**u, v** are orthogonal vectors as **u. v**=0.

(7) Orthogonal set

Definition: A set {$\mathbf{u}_1, \mathbf{u}_2$ ....$\mathbf{u}_n$} of vectors in R^n is said to be orthogonal set if every two different vectors of {$\mathbf{u}_1, \mathbf{u}_2$ ....$\mathbf{u}_n$} are orthogonal.

Example: If S= $\{\begin{bmatrix}1\\-1\\2\end{bmatrix}, \begin{bmatrix}1\\-1\\-1\end{bmatrix}, \begin{bmatrix}3\\3\\0\end{bmatrix}\}$ R^3then S is an orthogonal set.

Because $\begin{bmatrix}1\\-1\\2\end{bmatrix}, \begin{bmatrix}1\\-1\\-1\end{bmatrix}; \begin{bmatrix}1\\-1\\-1\end{bmatrix}, \begin{bmatrix}3\\3\\0\end{bmatrix}; \begin{bmatrix}1\\-1\\2\end{bmatrix}, \begin{bmatrix}3\\3\\0\end{bmatrix}$are orthogonal vectors

as the inner product of 1)$\begin{bmatrix}1\\-1\\2\end{bmatrix}, \begin{bmatrix}1\\-1\\-1\end{bmatrix}$= [1 -1 2]$\begin{bmatrix}1\\-1\\-1\end{bmatrix}$=1(1) +(-1) (-1) +2(-1) =1+1-2=0

2) $\begin{bmatrix}1\\-1\\-1\end{bmatrix}\begin{bmatrix}3\\3\\0\end{bmatrix}$ =[1 -1 -1]$\begin{bmatrix}3\\3\\0\end{bmatrix}$=1(3) +(-1)3+(-1)0=3-3+0=0

3) $\begin{bmatrix}1\\-1\\2\end{bmatrix}, \begin{bmatrix}3\\3\\0\end{bmatrix}$ =[1 -1 2] $\begin{bmatrix}3\\3\\0\end{bmatrix}$ =1(3) +(-1)3+2(0) =3-3+0=0

(8) Orthonormal set

Definition: A set {$\mathbf{u_1}, \mathbf{u_2}$ .... $\mathbf{u_n}$} of is said vectors in R^n orthonormal set if it is orthogonal set and every vector of it is a unit vector.

Example:S={$\begin{bmatrix}\frac{1}{\sqrt{3}}\\-\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{bmatrix}, \begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\\0\end{bmatrix}$}is an orthonormal set in R^3as it is orthogonal set and every vector in it is unit vector. Because the vectors$\begin{bmatrix}\frac{1}{\sqrt{3}}\\-\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{bmatrix}, \begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\\0\end{bmatrix}$are orthogonal and unit vectors.

Because Inner product of =$\begin{bmatrix}\frac{1}{\sqrt{3}}\\-\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{bmatrix}, \begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\\0\end{bmatrix}$=$[\frac{1}{\sqrt{3}} \; -\frac{1}{\sqrt{3}} \frac{1}{\sqrt{3}}]\begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\\0\end{bmatrix}$=$\frac{1}{\sqrt{3}}\left(\frac{1}{\sqrt{2}}\right) + (-\frac{1}{\sqrt{3}})(\frac{1}{\sqrt{2}})+(\frac{1}{\sqrt{3}})(0)=\frac{1}{\sqrt{6}} - \frac{1}{\sqrt{6}}+0=0$.

Also $\left\|\begin{bmatrix}\frac{1}{\sqrt{3}}\\-\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{bmatrix}\right\| = \sqrt{(\frac{1}{\sqrt{3}})^2 + (-\frac{1}{\sqrt{3}})^2 + (\frac{1}{\sqrt{3}})^2} = \sqrt{\frac{1}{3}+\frac{1}{3}+\frac{1}{3}} = \sqrt{\frac{3}{3}} = \sqrt{1} = 1$

$\left\|\begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\\0\end{bmatrix}\right\| = \sqrt{(\frac{1}{\sqrt{2}})^2 + (\frac{1}{\sqrt{2}})^2 + (0)^2}=\sqrt{\frac{1}{2}+\frac{1}{2}+0} = \sqrt{1+0} = \sqrt{1}$

(9) Orthogonal basis

Definition: A set $B = \{\propto_1, \propto_2, \propto_3, ..., \propto_n\}$ of vectors in R^n is said to be orthogonal basis if

1) $B = \{\propto_1, \propto_2, \propto_3, \ldots\ldots\ldots\ldots, \propto_n\}$ is orthogonal set in R^n 2) $B = \{\propto_1, \propto_2, \propto_3, \ldots\ldots\ldots\ldots, \propto_n\}$ is a basis of R^n. **Example:** The set S={$\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$ $\begin{bmatrix} 4 \\ -7 \\ 1 \end{bmatrix}$} is an orthogonal basis of R^3.

Because $\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$; $\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$, $\begin{bmatrix} 4 \\ -7 \\ 1 \end{bmatrix}$; $\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 4 \\ -7 \\ 1 \end{bmatrix}$ are orthogonal vectors

and the vectors $\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$ $\begin{bmatrix} 4 \\ -7 \\ 1 \end{bmatrix}$ forms a basis for R^3.($\because$ $\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$ $\begin{bmatrix} 4 \\ -7 \\ 1 \end{bmatrix}$ are Linearly Independent vectors, every vector of R^3 is a linear combination of vectors

$\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$ $\begin{bmatrix} 4 \\ -7 \\ 1 \end{bmatrix}$)

Example: S={$\begin{bmatrix} 1 \\ 1 \\ 8 \end{bmatrix}$, $\begin{bmatrix} 3 \\ 5 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$}is an orthogonal set but not an orthogonal basis. Because S is not a linearly independent set so that S is not a basis.

(10) Parallelogram Law

Theorem: If **u, v** are two vectors then $||\mathbf{u} + \mathbf{v}||^2 + ||\mathbf{u} - \mathbf{v}||^2 = 2(||\mathbf{u}||^2 + (||\mathbf{v}||^2)$

Example: Consider the two vectors $\boldsymbol{u} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$, **v**=$\begin{bmatrix} 1 \\ 4 \\ -5 \end{bmatrix}$. Here **u + v**=$\begin{bmatrix} 3 \\ 3 \\ -2 \end{bmatrix}$, **u-v**=$\begin{bmatrix} 1 \\ -5 \\ 8 \end{bmatrix}$

$||\mathbf{u}+\mathbf{v}|| = \sqrt{(3)^2 + (3)^2 + (-2)^2} = \sqrt{9 + 9 + 4} = \sqrt{22}$.

$||\mathbf{u}-\mathbf{v}|| = \sqrt{(1)^2 + (-5)^2 + (8)^2} = \sqrt{1 + 25 + 64} = \sqrt{90}$. Hence $||\mathbf{u} + \mathbf{v}||^2 + ||\mathbf{u} - \mathbf{v}||^2 = 112$.

$||\mathbf{u} + \mathbf{v}||^2 + ||\mathbf{u} - \mathbf{v}||^2 = (\sqrt{22})^2 + (\sqrt{90})^2 = 22 + 90 = 112$ …………………... (1)

$||\mathbf{u}|| = \sqrt{(2)^2 + (-1)^2 + (3)^2} = \sqrt{4 + 1 + 9} = \sqrt{14} \Rightarrow ||\mathbf{u}||^2 = 14$.

$||\mathbf{v}|| = \sqrt{(1)^2 + (4)^2 + (-5)^2} = \sqrt{1 + 16 + 25} = \sqrt{42} \Rightarrow ||\mathbf{v}||^2 = 42$.

Hence $2(||\mathbf{u}||^2 + (||\mathbf{v}||^2) = 2(14+42) = 112$…………… (2)

∴ From equations (1), (2) we get $||\mathbf{u} + \mathbf{v}||^2 + ||\mathbf{u} - \mathbf{v}||^2 = 2(||\mathbf{u}||^2 + (||\mathbf{v}||^2)$

(11) Schwarz's Inequality

Theorem: If **u, v** are two vectors in R^n then $|\mathbf{u}.\ \mathbf{v}| \le ||\mathbf{u}||\,||\mathbf{v}||$

Example: Consider the two vectors $\boldsymbol{u} = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}$, **v**=$\begin{bmatrix} 2 \\ 6 \\ 5 \end{bmatrix}$. Here $\mathbf{u}.\ \mathbf{v} = \mathbf{u}^T \mathbf{v} = [1\ 3\ {-2}]\begin{bmatrix} 2 \\ 6 \\ 5 \end{bmatrix}$

$= 1(2) + 3(6) + (-2)(5) = 10$. $\therefore \mathbf{u}.\ \mathbf{v} = 10 \Rightarrow |\mathbf{u}.\ \mathbf{v}| = 10$

$||\mathbf{u}||=\sqrt{(1)^2+(3)^2+(-2)^2}=\sqrt{1+9+4}=\sqrt{14},$
$||\mathbf{v}||=\sqrt{(2)^2+(6)^2+(5)^2}=\sqrt{4+36+25}=\sqrt{65}$
Hence $||\mathbf{u}||\,||\mathbf{v}||=\sqrt{14}\sqrt{65}=\sqrt{910}$
Hence $|\mathbf{u}.\mathbf{v}|^2<||\mathbf{u}||\,||\mathbf{v}||^2$. $\therefore|\mathbf{u}.\mathbf{v}|<||\mathbf{u}||\,||\mathbf{v}||$

(12)Triangle Inequality

Theorem: If **u, v** are two vectors in R^n then $||\mathbf{u}+\mathbf{v}|| \leq ||\mathbf{u}||+||\mathbf{v}||$

Example: Consider the two vectors $\boldsymbol{u}=\begin{bmatrix}2\\3\\0\end{bmatrix}$,**v**=$\begin{bmatrix}-4\\2\\4\end{bmatrix}$. Here **u**+**v**=$\begin{bmatrix}-2\\5\\4\end{bmatrix}$,
$||\mathbf{u}+\mathbf{v}||=\sqrt{(-2)^2+(5)^2+(4)^2}=\sqrt{45}$, $||\mathbf{u}||=\sqrt{(2)^2+(3)^2+(0)^2}=\sqrt{13}$
$||\mathbf{v}||=\sqrt{(-4)^2+(2)^2+(4)^2}=\sqrt{36}=6$
As $\sqrt{45}<\sqrt{13}+6$,we get $||\mathbf{u}+\mathbf{v}||<||\mathbf{u}||+||\mathbf{v}||$

(13) Pythagoras theorem

Theorem: If **u, v** are two vectors in an inner product space R^n then **u, v** are orthogonal if and only if $||\mathbf{u}+\mathbf{v}||^2=||\mathbf{u}||^2+||\mathbf{v}||^2$

Example: Consider two orthogonal vectors $\boldsymbol{u}=\begin{bmatrix}-3\\4\\1\end{bmatrix}$,**v**=$\begin{bmatrix}2\\3\\-6\end{bmatrix}$in R^3.

Here **u** + **v**=$\begin{bmatrix}-1\\7\\-5\end{bmatrix}$.Also if $||\mathbf{u}+\mathbf{v}||=\sqrt{(-1)^2+(7)^2+(-5)^2}=\sqrt{1+49+25}=\sqrt{75}$
$\therefore||\mathbf{u}+\mathbf{v}||^2=75$. $||\mathbf{u}||=\sqrt{(-3)^2+(4)^2+(1)^2}=\sqrt{26}$,$||\,||\mathbf{v}||=\sqrt{(2)^2+(3)^2+(-6)^2}=\sqrt{49}$
Hence $||\mathbf{u}||^2+||\mathbf{v}||^2=75$. $\therefore||\mathbf{u}+\mathbf{v}||^2=||\mathbf{u}||^2+||\mathbf{v}||^2$

(14) Orthogonal Projection

For a non-zero vector **u** in R^n , we can write a vector **y** in W,where W is a subspace of R^n as sum of two vectors such that one is multiple of **u** and other is orthogonal to **u** i.e.**y**=**ŷ** + **z**.Here we say **y** is decomposed in to two vectors such that one is multiple of **u** and other is orthogonal to **z**.
Here **ŷ** is multiple of **u, z** is orthogonal to **u**. We can find **z** by using **z=y- ŷ**
Here **ŷ** is defined as orthogonal projection of **y** on to **u**, where $\hat{\mathbf{y}}=\left(\frac{\boldsymbol{y.u}}{\boldsymbol{u,u}}\right)\mathbf{u}$
We can take **ŷ** as **ŷ=k u** where k is a scalar**. y=ŷ + z⇒z=y- ŷ**.

As **z** is orthogonal to **u**,**z.u**=o. **z.u**=0⇒ (**y**- **ŷ**). **u** =0⇒**y. u- ŷ. u**=0⇒**y.u**=**ŷ. u**⇒**y. u**= (k **u**). **u**

⇒**y. u**=k (**u. u**) ⇒ k=$\frac{\boldsymbol{y.u}}{\boldsymbol{u.u}}$ .But **ŷ=k u**. $\therefore\hat{\mathbf{y}}=\left(\frac{\boldsymbol{y.u}}{\boldsymbol{u,u}}\right)\mathbf{u}$

Note: **ŷ**. (**y**- **ŷ**) = **ŷ**. (**z**)= (k **u**). **z**=k (**u. z**) =k (0) =0. Hence **ŷ**, **y**-**ŷ** are orthogonal vectors.

Example: Consider the vector **u**= $\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$ in R^3. Let us find the orthogonal projection of **y**= $\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ on to u.

$\mathbf{y}=\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}=\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} \Rightarrow \mathbf{y.\ u}=\mathbf{y}^{\mathbf{T}}\mathbf{u}=[2\ 1\ 1]\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$ =2(1) +1(3) +1(4) =2+3+4=9. $\therefore$**y. u**=9

$\mathbf{u}=\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} \Rightarrow \mathbf{u.\ u}=\mathbf{u}^{\mathbf{T}}\mathbf{u}=[1\ 3\ 4]\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$ =1(1) +3(3) +4(4) =1+9+16=26. $\therefore$**u. u**=26

$\hat{\mathbf{y}} = (\frac{y.u}{u,u})\ \mathbf{u} \Rightarrow \hat{\mathbf{y}}=\frac{9}{26}\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{9}{26} \\ \frac{27}{26} \\ \frac{36}{26} \end{bmatrix} \Rightarrow \hat{\mathbf{y}}=\begin{bmatrix} \frac{9}{26} \\ \frac{27}{26} \\ \frac{36}{26} \end{bmatrix}$ is orthogonal projection of **y** on **u**.

$\mathbf{z}=\mathbf{y}-\ \hat{\mathbf{y}} \Rightarrow \mathbf{z}=\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{9}{26} \\ \frac{27}{26} \\ \frac{36}{26} \end{bmatrix} = \begin{bmatrix} \frac{43}{26} \\ \frac{23}{26} \\ \frac{22}{26} \end{bmatrix} \Rightarrow \mathbf{z}=\begin{bmatrix} \frac{43}{26} \\ \frac{-1}{26} \\ \frac{-10}{26} \end{bmatrix}$.We observe that $\mathbf{u.z}=\mathbf{u}^{\mathbf{T}}\ \mathbf{z}=[1\ \ 3\ \ 4]\begin{bmatrix} \frac{43}{26} \\ \frac{-1}{26} \\ \frac{-10}{26} \end{bmatrix}$ $=1(\frac{43}{26})+3(\frac{-1}{26})+4(\frac{-10}{26}) = \frac{43-3-40}{26}$=0.Hence **u. z**=0⇒**u,z** are orthogonal to each other and $\mathbf{y}=\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}=\begin{bmatrix} \frac{9}{26} \\ \frac{27}{26} \\ \frac{36}{26} \end{bmatrix}+\begin{bmatrix} \frac{43}{26} \\ \frac{-1}{26} \\ \frac{-10}{26} \end{bmatrix}$.

Hence **y** is written as sum of two vectors **ŷ, z** where **z** is orthogonal to **u**.

Here observe that $\hat{\mathbf{y}}=\begin{bmatrix} \frac{9}{26} \\ \frac{27}{26} \\ \frac{36}{26} \end{bmatrix}$,$\mathbf{y}-\ \hat{\mathbf{y}} =\begin{bmatrix} \frac{43}{26} \\ \frac{-1}{26} \\ \frac{-10}{26} \end{bmatrix}$ are orthogonal.

(15) The Orthogonal Decomposition Theorem

Theorem: Let W be a subspace of R^n . Then each **y** in R^n can be written uniquely in the form
y= ŷ + zwhere **ŷ** is in W and **z** is in W⊥. In fact, if {$\mathbf{u}_1, \mathbf{u}_2$ .... $\mathbf{u}_n$} is any orthogonal basis of W then $\hat{\mathbf{y}}=(\frac{y.u_1}{u_1.u_1})u_1+\ (\frac{y.u_2}{u_2.u_2})u_2+\ (\frac{y.u_3}{u_3.u_3})u_3+\ldots\ldots\ldots\ldots+(\frac{y.u_n}{u_n.u_n})\mathbf{u_n}$ and **z=y- ŷ**.(W⊥contains the vectors which are orthogonal to every vector of W).Here ŷ is defined as orthogonal proj

(16) Orthonormal column of a matrix: For an m × n matrix, if a column of the matrix is orthonormalvector, then we say the column of the matrix as orthonormal column.

Theorem: An m ×n matrix U has orthonormal columns if and only if $U^TU= I$.
Theorem: Let U be an m×n matrix with orthonormal columns, and let x and y be in R^n . Then
a) |Ux||=||x||b) (U x). (U y) =x. y c) (U x). (U y) =0 if and only if x. y=0
Note: An orthogonal matrix is a square invertible matrix U such that $U^{-1}=U^T$

(17) The Orthogonal Decomposition Theorem

Theorem: Let W be a subspace of R^n . Then each y in R^n can be written uniquely in the form $y = \hat{y} + z$ where $\hat{y}$ is in W and z is in $W^{\perp}$. In fact, if $\{u_1, u_2 \ldots\ldots\ldots u_n\}$ is any orthogonal basis of W then $\hat{y} = (\frac{y.u_1}{u_1.u_1})u_1 + (\frac{y.u_2}{u_2.u_2})u_2 + (\frac{y.u_3}{u_3.u_3})u_3 + \ldots\ldots\ldots\ldots\ldots\ldots + (\frac{y.u_n}{u_n.u_n})u_n$ and $z = y - \hat{y}$.($W^{\perp}$contains the vectors which are orthogonal to every vector of W).Here $\hat{y}$ is defined as orthogonal projection of y on to W.

(18) The Best Approximation Theorem

Theorem: Let W be a subspace of R^n , let y be any vector in R^n, and let $\hat{y}$ be the orthogonal projection of y onto W.Then $\hat{\mathbf{y}}$ is the closest point in W to **y**, in the sense that $||\mathbf{y} - \hat{\mathbf{y}}|| <$
$||\mathbf{y} - \mathbf{v}||$for all v in W distinct from $\hat{\mathbf{y}}$.

Theorem: If $\{\mathbf{u_1}, \mathbf{u_2} \ldots\ldots\ldots \mathbf{u_n}\}$ is an orthonormal basis for a subspace W of R^n, then projection of **y** on to W $= (\mathbf{y}.\boldsymbol{u_1})\ \boldsymbol{u_1} + (\mathbf{y}.\boldsymbol{u_2})\ \boldsymbol{u_2} + (\mathbf{y}.\boldsymbol{u_3})\ \boldsymbol{u_3} + \ldots\ldots\ldots + (\mathbf{y}.\boldsymbol{u_n})\ \boldsymbol{u_n}$.If $U = \{\mathbf{u_1}, \mathbf{u_2} \ldots\ldots\ldots \mathbf{u_n}\}$ then projection of **y** on to $W = U\, U^T y$ for all **y** in R^n.

(19) Gram – Schmidt orthogonalization process

This Process is used to construct orthogonal basis from a given basis of a subspace of a vector space.

Theorem: Given a basis$B = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}, \ldots\ldots\ldots \boldsymbol{u_n}\}$for a non-zero subspace W of R^n.

Define,$\mathbf{v_1} = \mathbf{u_1}$

$$\boldsymbol{v_2} = \boldsymbol{u_2} - (\frac{\boldsymbol{u_2}.\boldsymbol{v_1}}{\boldsymbol{v_1}.\boldsymbol{v_1}})$$

$$\boldsymbol{v_3} = \boldsymbol{u_3} - (\frac{\boldsymbol{u_3}.\boldsymbol{v_1}}{\boldsymbol{v_1}.\boldsymbol{v_1}})\boldsymbol{v_1} - (\frac{\boldsymbol{u_3}.\boldsymbol{v_2}}{\boldsymbol{v_2}.\boldsymbol{v_2}})\boldsymbol{v_2}$$

…………………………

…………………………

…………………………

…………………………

$$\boldsymbol{v_n} = \boldsymbol{u_n} - (\frac{\boldsymbol{u_n}.\boldsymbol{v_1}}{\boldsymbol{v_1}.\boldsymbol{v_1}})\boldsymbol{v_1} - (\frac{\boldsymbol{u_n}.\boldsymbol{v_2}}{\boldsymbol{v_2}.\boldsymbol{v_2}})\boldsymbol{v_2} - \ldots\ldots\ldots - (\frac{\boldsymbol{u_n}.\boldsymbol{v_{n-1}}}{\boldsymbol{v_2}.\boldsymbol{v_{n-1}}})\boldsymbol{v_{n-1}}$$

then $B_1 = \{\mathbf{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \ldots\ldots\ldots, \boldsymbol{v_n}\}$ is an orthogonal basis of R^n.

(20) Orthonormal basis

Definition: A set $B = \{\propto_1, \propto_2, \propto_3, \ldots\ldots\ldots, \propto_n\}$ of vectors in R^n is said to be orthonormal basis if 1) $B = \{\propto_1, \propto_2, \propto_3, \ldots\ldots\ldots, \propto_n\}$ is a basis of R^n 2) $B = \{\propto_1, \propto_2, \propto_3, \ldots\ldots\ldots, \propto_n\}$ is orthonormal set in R^n

Example: If B= { $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ }then u is an orthonormal basis of R^2, because

1)B is basis of R^n , 2)Bis an orthonormal set.

(∵ The Vectors $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{-1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ are orthogonal and unit vectors)

(∵ The Vectors $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{-1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ are Linearly independent vectors and $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{-1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ spans R^2)

(21) The QR Factorization Theorem

Theorem: If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as A=QR, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

Here we find the matrix Q by using Gram – Schmidt orthogonalization process and normalize the columns of Q. As the columns of the matrix are orthonormal, $Q^TQ=I$. $A=QR \Rightarrow Q^{-1}A=R$
Hence, we can find R by using $R= Q^{-1}A$.

MAIN RESULT

(22) Least Square Lines.

In this section we find solution x of the equation Ax=B,where A is an $m\times n$ *matrix, x is* $n \times 1$ *matrix and B is* $m\times 1$ *matrix*.
The general least-squares problem is to find an x that makes $||B - A \times ||$ as small as possible.

Theorem: The set of least-squares solutions of $A\times= B$ coincides with the nonempty set of solutions of the normal equations $A^TA\times=A^Tb$.

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

a) The equation Ax = B has a unique least-squares solution for each b in R^m.

b) The columns of A are linearly independent.

c) The matrix A^tA is invertible. When these statements are true, the least-squares solution $\hat{x}$ is given by $\hat{x}=(A^TA)^{-1} A^T. B$

Theorem: Given an $m \times n$ matrix A with linearly independent columns, let A= QR be a QR factorization of A. Then, for each B in R^m, the equation Ax =B has a unique least-squares solution, given by $\hat{x}= R^{-1} Q^T B$.

(23) APPLICATIONS TO LINEAR MODELS

A common task in science and engineering is to analyze and understand relationships among several quantities that vary. This section describes a variety of situations in which data are used to build or verify a formula that predicts the value of one variable as a function of other variables. In each case, the problem will amount to solving a least squares problem.

For easy application of the discussion to real problems that you may encounter later in your career, we choose notation that is commonly used in the statistical analysis of scientific and engineering data. Instead of Ax=B, we write X β= y and refer to X as the design matrix, βas the parameter vector, and y as the observation vector.

The simplest relation between two variables x and y is the linear equation $y=\beta_0+\beta_1 x$Experimental data often produce points $(x_1,y_1),(x_2, y_2), (x_3, y_3)$ .... (x_n, y_n) that, when graphed seem to lie close to a line. We want to determine the parameters β_0 and β_1 that make the line as "close" to the points as possible. The least-squares line is the line $y= \beta_0+ \beta_1 x$ that minimizes the sum of the squares of the residuals. This line is also called a line of regression of y on x, because any errors in the data are assumed to be only in the y-coordinates. The coefficientsβ_0,β_1 of the line are called (linear) regression coefficients. If the data points were on the line, the parameters $\beta_0,\beta_1$and would satisfy the following equations.

Predicted y-value	**observed y-value**
$\beta_0+\beta_1 x_1$	$= y_1$
$\beta_0+\beta_1 x_2$	$= y_2$
$\beta_0+\beta_1 x_3$	$= y_3$
....	......
....	......
....	......
$\beta_0+\beta_1 x_n$	$= y_n$

we can write this system equations as X β=Y, where $X=\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ - & - \\ - & - \\ - & - \\ 1 & x_n \end{bmatrix}, \beta=\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, Y=\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ - \\ - \\ - \\ - \\ y_n \end{bmatrix}$

By least square method we can fit the line to the given data points

Example: Let us find the equation $y=\beta_0+\beta_1 x$of the least squares line that best fits the data points (2,1), (5,2), (7,3) and (8,3)

Let y= $y=\beta_0+\beta_1 x$ be the least squares line that best fits the data points (2,1), (5,2), (7,3)

and (8,3).

For least squares line,we find the solution β of x β=y. The normal equation is $x^T x \beta= x^T y$

Here $x=\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, y=\begin{bmatrix} 1 \\ 2 \\ 3 \\ 5 \end{bmatrix}$

The normal equation is$x^T x \beta= x^T y$

$$x^T x=\begin{bmatrix}1&1&1&1\\2&5&7&8\end{bmatrix}\begin{bmatrix}1&2\\1&5\\1&7\\1&8\end{bmatrix}=\begin{pmatrix}4&22\\22&142\end{pmatrix},\ x^T y=\begin{bmatrix}1&1&1&1\\2&5&7&8\end{bmatrix}\begin{bmatrix}1\\2\\3\\5\end{bmatrix}=\begin{bmatrix}9\\57\end{bmatrix}$$

$$x^T y=\begin{bmatrix}1&1&1&1\\2&5&7&8\end{bmatrix}\begin{bmatrix}1\\2\\3\\5\end{bmatrix}=\begin{bmatrix}9\\57\end{bmatrix}.\ x^T x\ \beta= x^T y\Rightarrow\begin{pmatrix}4&22\\22&142\end{pmatrix}\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}=\begin{bmatrix}9\\57\end{bmatrix}$$

$$\begin{pmatrix}4&22\\22&142\end{pmatrix}\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}=\begin{bmatrix}9\\57\end{bmatrix}\Rightarrow\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}=\begin{pmatrix}4&22\\22&142\end{pmatrix}^{-1}\begin{bmatrix}9\\57\end{bmatrix}=\frac{1}{568-484}\begin{pmatrix}142&-22\\-22&4\end{pmatrix}\begin{bmatrix}9\\57\end{bmatrix}$$

$$=\frac{1}{84}\begin{bmatrix}278-1254\\-198+228\end{bmatrix}=\frac{1}{84}\begin{bmatrix}24\\30\end{bmatrix}\Rightarrow\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}=\begin{bmatrix}\frac{24}{84}\\\frac{30}{84}\end{bmatrix}=\begin{bmatrix}\frac{2}{7}\\\frac{5}{14}\end{bmatrix}\Rightarrow\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}=\begin{bmatrix}\frac{2}{7}\\\frac{5}{14}\end{bmatrix}$$

Hence $\beta_0=\frac{2}{7}, \beta_1=\frac{5}{14}$.

Hence the least square line best fits the data points (2,1), (5,2), (7,3) and (8,3) is $y=\frac{2}{7}+\frac{5}{14}x$.

The General Linear Model

In some applications, it is necessary to fit data points with a curve other than a straight line. In the examples that follow, the matrix equation is still X β=y, but the specific form of Xchanges from one problem to the next. Statisticians usually introduce a residual vector ε,defined by ε=y-Xβ and write y= Xβ+ε.Any equation of this form is referred to as a linear model. Once X and y are determined, the goal is to minimize the length of ε, which amounts to finding a least-squares solution of Xβ =y in each case, the least-squares solution $\hat{B}$ is a solution of the normal equations $X^T X\beta=X^T y$.

Least-Squares Fitting of Other Curves

When data points $(x_1, y_1), (x_2, y_2), (x_3, y_3) \ldots\ldots (x_n, y_n)$ on a scatter plot do not lie close to any line, it may be appropriate to postulate some other functional relationship between x and y. The next two examples show how to fit data by curves that have the general form $y=\beta_0 f_0(x)+\beta_1 f_1(x)+\beta_2 f_2(x)+\cdots+\beta_n f_n(x)$,where $f_0, f_1, f_2, \ldots\ldots f_n$are known functions and $\beta_0, \beta_1, \beta_2, \ldots\ldots \beta_n$are parameters that must be determined. As we will see, equation $y=\beta_0 f_0(x)+\beta_1 f_1(x)+\beta_2 f_2(x)+\cdots+\beta_n f_n(x)$,describes a linear model because it is linear in the unknown parameters. For a particular value of x, it gives a predicted, or "fitted," value of y. The difference between the observed value and the predicted value is the residual. The parameters must be determined so as to minimize the sum of the squares of the residuals.

Example: Let the given data points $(x_1, y_1), (x_2, y_2), (x_3, y_3) \ldots\ldots (x_n, y_n)$ lie on a parabola other than a straight line. Let the x-coordinate denotes the production level for a company denotes the average cost per unit of operating at a level of x units per day.Then a typical average cost curve looks like a parabola that opens upward.

Let us approximate the data by an equation of the form $y=\beta_0+\beta_1 x+\beta_2 x^2+\varepsilon_1$,where ε_1 is the residual vector which is the difference between observed value and predicted value.

As $y=\beta_0+\beta_1 x+\beta_2 x^2+\varepsilon_1$ *passes through the points* $(x_1, y_1), (x_2, y_2), (x_3, y_3) \ldots (x_n, y_n)$,we have $y_1=\beta_0+\beta_1 x_1+\beta_2 x_1^2+\varepsilon_1$

$y_2 = \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \varepsilon_2$

$y_3 = \beta_0 + \beta_1 x_3 + \beta_2 x_3^2 + \varepsilon_3$

…………………………

…………………………

…………………………

$y_n = \beta_0 + \beta_1 x_n + \beta_2 x_n^2 + \varepsilon_n$

We can write the system of equations in the form Y=Xβ+ ε

Here $Y=\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ - \\ - \\ - \\ y_n \end{bmatrix}$, $X=\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ - & - & - \\ - & - & - \\ - & - & - \\ 1 & x_n & x_n^2 \end{bmatrix}$, $\beta=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$, $\varepsilon=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ - \\ - \\ - \\ - \\ \varepsilon_n \end{bmatrix}$.

By using normal equations, minimizing ε we solve the system of equations.

In order to approximate the given points (x_1, y_1), (x_2, y_2), (x_3, y_3) ……… (x_n, y_n) to a cubic equation of the form $y=\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon_1$, where ε_1 is the residual vector which is the difference between observed value and predicted value, we follow the above procedure.

(24) Some more Applications of Inner Product.

1**)**Wecan approximate a line or second degree or …. or an nth degree polynomial of exponential, logarithms and trigonometric functions by using the least square method.

2) We can approximate a Fourier function to a continuous function on a closed interval spanned by a non-empty set with respect to a vector spaceby using the least square method.

References

1. Linear algebra and its applications – David.c.Lay
2. B. Ulrychova – several world applications of linear algebraic tools
3. Loet Leidesdorff and Liwen Vaughan – co – occurrence of matrices and their applications in information
4. Matrix Analysis and Applied Linear Algebra by Carl. D. Meyer
5. Linear Algebra: A Modern Introduction by David Poole
6. Applied Linear Algebra and Matrix Analysis by Thomas S. Shores
7. Elementary Linear Algebra with Applications by Stanley I. Grossman
8. Introduction to Linear Algebra by Gilbert Strang
9. Theory and Problems of Linear Algebra by Seymour Lipschutz
10. Linear Algebra by Kenneth Hoffman and Ray Kunze
11. Linear Algebra by Friede burg, Insel, Spence
12. Advanced Linear Algebra by Steven Roman
13. Linear Algebra in Data Science by Peter Zizler, Roberta La Haye
14. A Course in Linear Algebra by Raju.k. George, Abhijith Ajaya kumar