Day - 52, Jan-21, 2025 (Magh 08, 2081 BS).

→ Orthogonality and Least Squares:

→ Orthogonality refers to the relationship between two vectors.

→ Two vectors are orthogonal if their dot product is zero. They are perpendicular to eachother in vector space.

→ Least Square finds the best-fit solution by minimizing squared error, with residuals orthogonal to the prediction space.

$y = mx + c$ Commonly used in this equation –

→ In 2D space, vectors $u = [1, 0]$ and $v = [0, 1]$ are orthogonal because –

$$u \cdot v = [1, 0] \cdot [0, 1]$$

$$\Rightarrow 0 \ \#.$$

Inner Product

If u and v are $n \times 1$ matrices then $u^T v$ be 1×1 matrix where u^T be transpose of u, which is called the Inner product of u and v.

If $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ then the inner product of u and v is

$$u^T v = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

$$\Rightarrow u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

$$\boxed{u^T v \Rightarrow u \cdot v}$$

This means $u^T v$ is same as the dot product between two vectors u and v.

Definition (Scalar product or Dot Product)

Let $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$ then the scalar product of u and v is denoted by $u \cdot v$ and defined as:

$$u \cdot v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

This product is also known as Dot Product.

Example: Compute $u \cdot v$ and $v \cdot u$ when $u = (2, 4, 5)$ and $v = (-1, 3, -1)$.

Here,

$$u \cdot v = (2, 4, 5) \cdot (-1, 3, -1)$$

$$\Rightarrow (-2 + 12 - 5)$$

$$\Rightarrow 5$$

$$v \cdot u = (-1, 3, -1)(2, 4, 5)$$

$$\Rightarrow -2 + 12 - 5$$

$$\Rightarrow 5$$

So, $\boxed{u \cdot v = v \cdot u}$

Properties of Inner Product.

Theorem: Let u, v and w are vectors in R^n. Also, let c be a scalar then,

i) $u \cdot v = v \cdot u$

ii) $(u+v) \cdot w = u \cdot w + v \cdot w$

iii) $(cu) \cdot v = c(u \cdot v) = u(cv)$

iv) $u \cdot v \geq 0$ and $u \cdot u = 0$ if and only if $u = 0$.

Definition (Length of vector)

The length or norm of vector v is non-negative scalar $\|v\|$, is defined as

$$\|v\| \Rightarrow \sqrt{v \cdot v}$$

$$\Rightarrow \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$$

Where $v = (v_1, v_2 \dots v_n)$

Note that, this definition implies $\|v^2\| = v \cdot v$.

finding the length of vector $(4, 5, 6)$.

$$v = (4, 5, 6)$$

The length of v is

$$\|v\| \Rightarrow \sqrt{4^2 + 5^2 + 6^2}$$

$$\Rightarrow \sqrt{77}$$

Unit Vector: A vector having length 1 is called unit vector.

Normally, if we divide the non-zero vector by its length then we get a new vector called unit vector and its direction is same as to given vector. Mathematically, if v be a vector in $\mathbb{R}^n$ then the unit vector is,

$$\frac{v}{\|v\|}.$$

(Normalization of a vector)

Let v be a vector in $\mathbb{R}^n$. Set $u = \frac{v}{\|v\|}$ then process creating u is called normalizing v.

17 Definition (Distance between two Vectors):

Let u and v are in R^n, then the distance between 'u' and 'v' is the length between them it is denoted by dis(u, v) and define as,

$$\boxed{\text{dis}(u, v) = \|u - v\|}$$

Q. If $u = (2, 3)$ and $v = (3, -1)$ then find the distance between them.

$u = (2, 3)$

$v = (3, -1)$

$u - v = (2, 3) - (3, -1)$ $\quad \boxed{u - v = (-1, 4)}$

Now, distance between u and v is,

$$||u - v|| \Rightarrow \sqrt{(-1)^2 + (4)^2}$$

$$\Rightarrow \sqrt{1+16}$$

$$\Rightarrow \sqrt{17}$$

Definition (Orthogonal)

Two vectors u and v in R^n are orthogonal to each other if $u \cdot v = 0$.

Definition (Orthogonal Complements)

The set of all orthogonal vectors to a subspace W of R^n, is called orthogonal complements of W. It is denoted by $W^{\perp}$ and read as

'W perpendicular' or 'W perp'.

Orthogonal Sets:

A set of vectors $\{u_1, u_2 \ldots u_p\}$ in R^n, is said to be an orthogonal set if $u_i \cdot u_j = 0$ for $i \neq j$ for $i, j = 1, 2 \ldots p$

Example given $\{u_1, u_2, u_3\}$ we can test if they are orthogonal set or not by

$u_1 \cdot u_2 = 0$? or not

$u_2 \cdot u_3 = 0$? or not

Theorem: If $S = \{u_1 \ldots u_p\}$ is an orthogonal set of non-zero vectors in R^n then S is linearly Independent and hence is a basis for the Subspace Spanned by S.

So, we have $S = \{u_1 \dots u_p\}$ is an orthogonal set of non-zero vectors in $\mathbb{R}^n$. Then

$$u_i \cdot u_j = 0 \quad \text{for } i \neq j \quad \text{and } i, j = 1 \dots p.$$

Definition (Orthogonal Basis):

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

Theorem:

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

$$y = c_1 u_1 + \dots \dots + c_p u_p$$

are given by,

$$c_j = \frac{y \cdot u_j}{u_j \cdot u_j}$$

$$\Rightarrow \frac{y \cdot u_j}{\|u_j\|^2} \quad \text{for } j = 1, 2, \ldots, p.$$

Orthogonal Projections:

The orthogonal projection of a straight line or a curve in R^2 is a straight line.

Definition (Orthogonal Projection):

Let x and y are two vectors in R^n. Then the orthogonal projection of y on x is αx.

Orthogonal Projection is finding the closest point on a subspace to a given point.

An orthogonal Basis consists of vectors that are all perpendicular (dot product = 0) to each other and span the vector space.

A Orthogonal set is a collection of vectors in which every pair of distinct vectors is $\perp$ (their dot product is zero).

Normalizing vector means Scaling it so that its length (or magnitude) becomes 1, while maintaining its direction.

Gram-Schmidt Process:

A method for orthogonalizing a set of vectors in a vector space, meaning it converts linearly independent vectors into a set of orthogonal vectors that span the same distance.

Theorem (The Gram-Schmidt Process):

Given a basis $\{x_1, \ldots x_p\}$ for a subspace W of $\mathbb{R}^n$,

Define,

$$v_1 = x_1$$

$$v_2 = x_2 - \left(\frac{x_2 \cdot v_1}{v_1 \cdot v_1}\right) \cdot v_1$$

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

$$v_p = x_p - \left(\sum_{n=1}^{p-1} \frac{x_p \cdot v_n}{v_n \cdot v_n} \cdot v_n \right).$$

Then $\{v_1, \ldots v_p\}$ is an orthogonal basis for W.

In addition, $\text{Span}\{v_1, \ldots, v_k\} = \text{Span}\{x_1, \ldots, x_k\}$

for $1 \leq k \leq p$.

The Least Squares Problems:

First we assume that Ax as an approximation to b. Then $\|Ax - b\|$ be the smaller distance between Ax and b. The least-square problem is to find an x that makes $\|Ax - b\|$ as small as possible

Definition (Least-Square Solution):

If A is $m \times n$ and b is in $\mathbb{R}^m$ then a least-squares solution of $Ax = b$ is an $\hat{x}$ in $\mathbb{R}^n$ such that

$$\|b - A\hat{x}\| \le \|b - Ax\|$$

for all x in $\mathbb{R}^n$.

Solution of Least Square Problem $Ax = b$

Suppose $\hat{x}$ satisfies $A\hat{x} = b$ by the orthogonal Decomposition Theorem, the projection $\hat{b}$ has the property that $b - \hat{b}$ is orthogonal to Col A, so $b - A\hat{x}$ is orthogonal to each column of A. If a_j is any column of A, then $a_j \cdot (b - A\hat{x}) = 0$ and $a_j^T (b - A\hat{x}) = 0$. Since each a_j^T is a row of A^T

Thus, $$A^T(b - A\hat{x}) = 0 \quad \text{— eqn (P)}$$

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

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

These calculations show that each least-squares solution of $Ax = b$ satisfies the equation

$$A^T A x = A^T b \quad \text{—— eqn (ii)}$$

The matrix equation represents a system of equations called the normal equations for $Ax = b$. A solution of eqn (i) is often denoted by $\hat{x}$

Theorem: The matrix $A^T A$ is invertible if and only if the column A are linearly Independent. In this case, the equation $Ax = b$ has only one least-squares solution $\hat{x}$, and it is given by

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

Applications to Linear Models:

Solving Least Squares Problems help to analyze and understand the relationships among several quantities.

Instead of $Ax = b$ we write $X\beta = y$

where X as a design matrix

β as parameter vector

y as observation vector

If the data points were on the line, the parameters β_0 and β_1 would satisfy the equations -

Predicted y-value = Observed y-value

$\beta_0 + \beta_1 (x_1) = y_1$

$\beta_0 + \beta_1 (x_2) = y_2$

$\vdots \qquad \vdots$

$\beta_0 + \beta_1 x_n = y_n$

Simply we can write this system as,

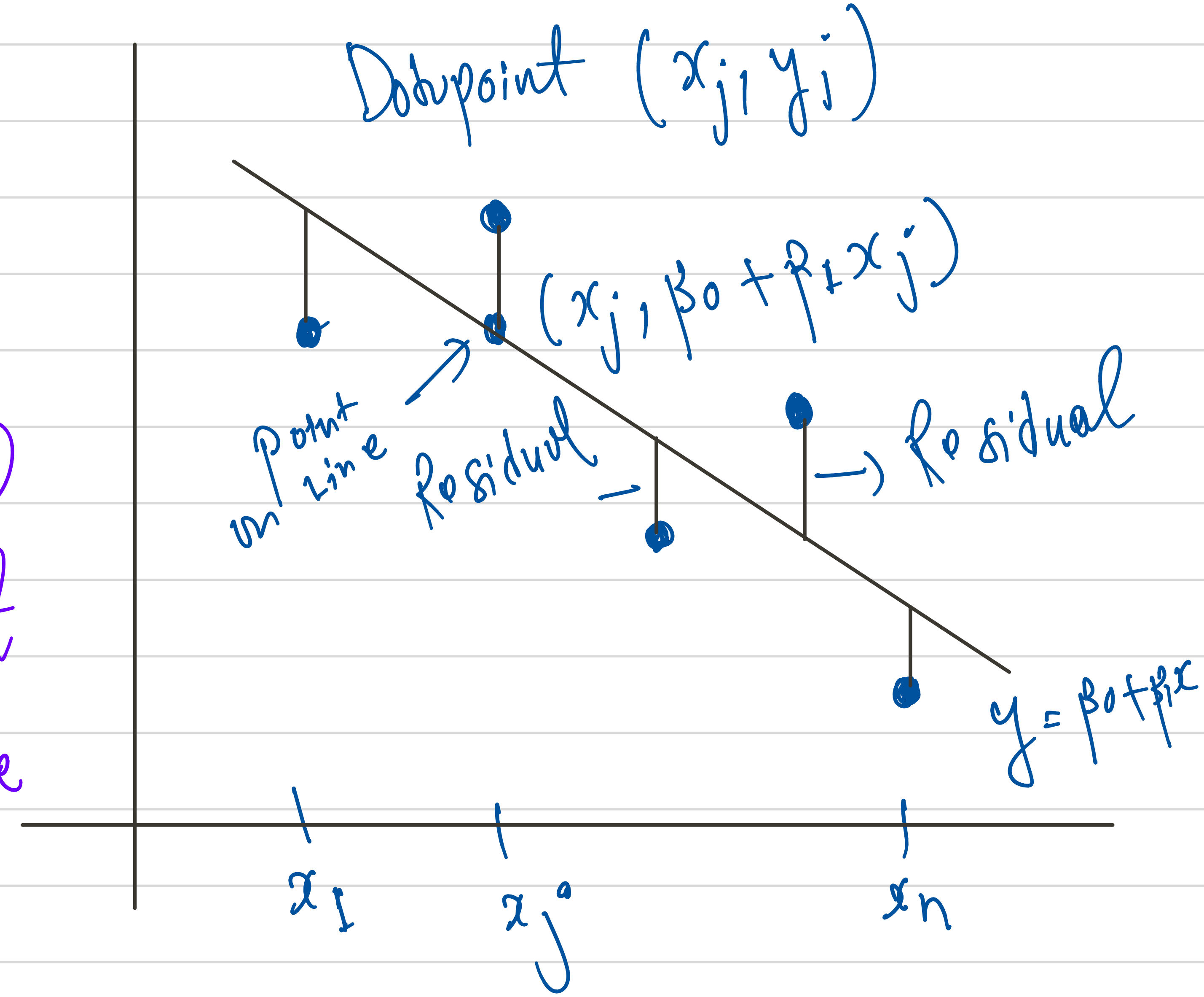
$X\beta = y$, where $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$, $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

Least - Squares Lines:

Suppose β_0 and β_1 are fixed and consider a line

$$y = \beta_0 + \beta_1 x \quad \text{—— eqn (i)}$$

as in figure. Let (x_j, y_j) be a point. Then $(x_j, \beta_0 + \beta_1 x_j)$ be a point on the line eqn (i).



In this case y_j be observed value of y and $\beta_0 + \beta_1 x_j$ be predicted y-value. Then residual is the difference between observed y-value and predicted y-value.

The least-squares line is the line (1) that minimizes the sum of the squares of the residuals.

It is also called a line of Regression of y on x. Here β_0, β_1 are called regression coefficients.

Note: The least-squares lines assumed that the errors in the data, to be only y-coordinates.

The General Linear Model:

Let $$y = \beta_0 + \beta_1 x \quad \text{—— eqn (P)}$$

be the least-squares line. If 'ε' be a residual error defined by $\varepsilon = y - X\beta$. Then eqn (i) becomes as

$$y = X\beta + \varepsilon \quad \text{—— eqn (i)}$$ called as linear model.

The least squares solution '$\hat{\beta}$ cap' is a solution of the normal equations,

$$X^T X\beta = X^T y.$$

Least Squares fitting to Other Curves:

Let the points $(x_1, y_1) \ldots (x_n, y_n)$ lie on a scatter plot that do not lie close to any line. def,

$$y = \sum_{k=0}^{n} \beta_k f_k(x) \quad \text{—— eqn (i)}$$

where $f_0, \dots f_n$ are functions and $\beta_0, \dots \beta_n$ are parameters.
The line eqn (i) is the general form of Least-Squares' line -

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 .$$

So for given eqn $y = \beta X + \varepsilon$

$$\rightarrow y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ x_n & x_n & x_n^2 \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Inner Product Space

Built from vector space + Inner Product -

① Vector Space (a set of vector where you can perform scalar multiplication and vector addition)

② A function that assigns scalar value to every pair of vectors which satisfies specific properties is called an Inner Product.

Definition : An Inner product on a vector space V is a function that to each pair of vectors u and v in V associates a real number $\langle u, v \rangle$ and satisfies, for all u, v, w in V and all scalar c

i) $\langle u, v \rangle = \langle u, v \rangle$

ii) $\langle u+v, w \rangle = \langle u, w \rangle + \langle u, w \rangle$

iii) $\langle cu, v \rangle = c \langle u, v \rangle$

iv) $\langle u, v \rangle \geq 0$ and $\langle u, v \rangle = 0$ if and only if $u = 0$.

[A vector space with an inner product is called an inner product space.]

Definition (Length or norm)

Let V be an inner product space and let $v \in V$. Then the length or norm of the vector v is a scalar, is defined as

$\|v\| \Rightarrow \sqrt{\langle v, v \rangle}$ or equivalently $\|v\|^2 = \langle v, v \rangle$

Definition:

Let V be an inner Product Space and let u, v in V then the distance between u and v is denoted by $\|u-v\|$ is defined as

$$\|u-v\| \Rightarrow \sqrt{\langle u-v, u-v \rangle}.$$

Definition (Orthogonal)

Let V be an inner product Space and u, v in V. Then the vectors u and v are orthogonal if $\langle u, v \rangle = 0$.