HCAI5TML01 – Mathematics of Learning.

Week – 2: Lecture – 02

A Refresher on the Mathematics Behind Machine Learning.

Least Squares and Fundamental Theory of Linear Algebra.

Siman Giri

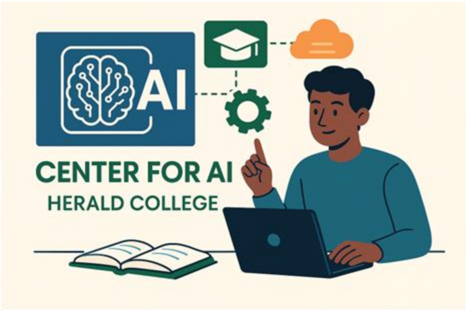


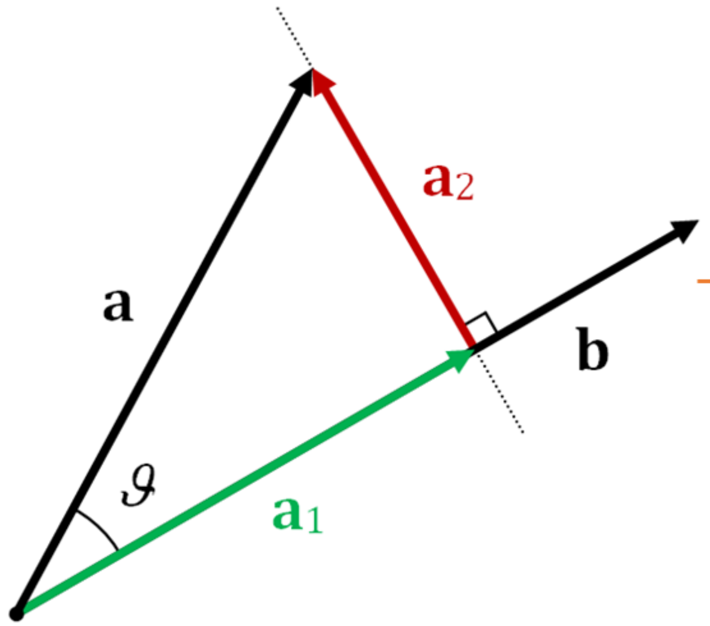
image generated via copilot.

HERALD COLLEGE ing CENTER FOR AI.

1. Some More Definitions.

1.1 Introduction: Vector projection.

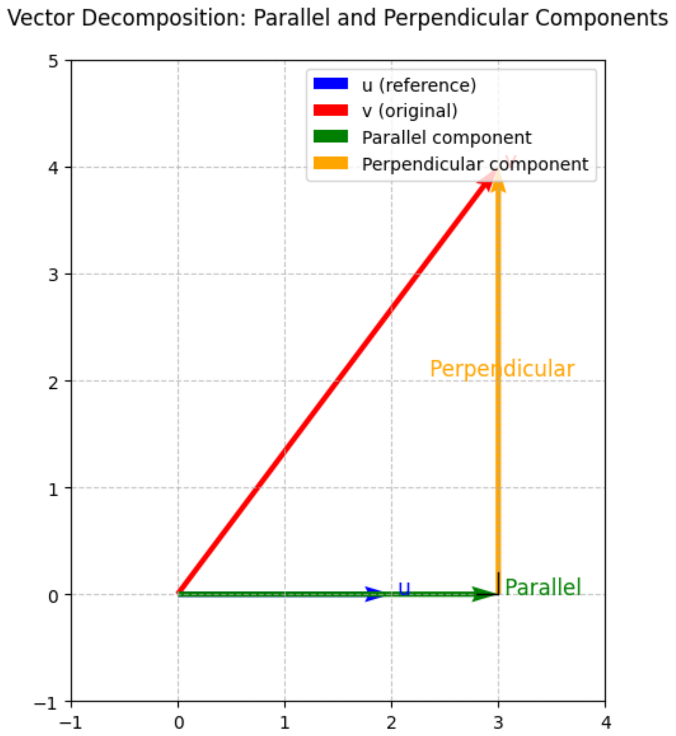
- Vector projection is a fundamental concept in linear algebra that decomposes
  - a vector into components **parallel and perpendicular** to another vector or subspace.



- In this example we **project vector a onto vector b,**
- Then we can indeed **decompose a** into:
  - A **parallel component aligned with b**
  - A **perpendicular component orthogonal to b.**
- Mathematically, this is expressed as:
  - $\mathbf{a} = \mathrm{proj}_{\mathbf{b}}\mathbf{a} + \mathrm{perp}_{\mathbf{b}}\mathbf{a}$
    - $\mathrm{proj}_{\mathbf{b}}\mathbf{a}$ → **the projection a onto b.**
    - $\mathrm{perp}_{\mathbf{b}}\mathbf{a}$ → **is the rejection of a from b (perpendicular to b)**

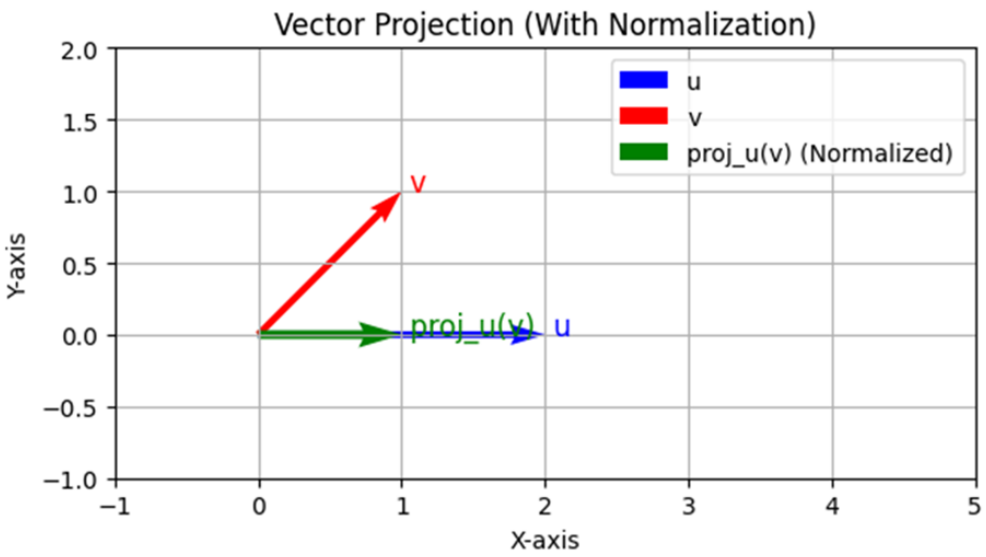
1.2 Projection on to vector.

- **Definition:**
  - Given two vector **u and v,** the projection of **v onto u (denoted $\mathbf{proj_u v}$)**
    - is a **component of v that lies in the direction of u.**
    - The $\mathbf{proj_u v}$ is commuted as:
      - $\mathbf{proj_u v} = \left(\frac{\mathbf{u \cdot v}}{\mathbf{u \cdot u}}\right)\mathbf{u} = \left(\frac{\mathbf{u \cdot v}}{\|\mathbf{u}\|^2}\right)\mathbf{u}$
  - **Key components:**
    - **Dot product** $(\mathbf{u} \cdot \mathbf{v})$:
      - Measures how **much v aligns with u** (the “overlap”).
      - If u and v are orthogonal, $\mathbf{u} \cdot \mathbf{v} = \mathbf{0}$ **(no projection).**
    - **Normalization** $(\mathbf{u} \cdot \mathbf{u} == \|\mathbf{u}\|^2)$:
      - This is the scaled magnitude of u.
      - Purpose: scales the projection to the unit length of u,
        - ensuring the result is **proportional to u’s direction** without **distorting its length**.
      - This ensures the projection depends only on the *direction* of **u**, not its magnitude.



1.2.1 Example: with and without Normalizations.

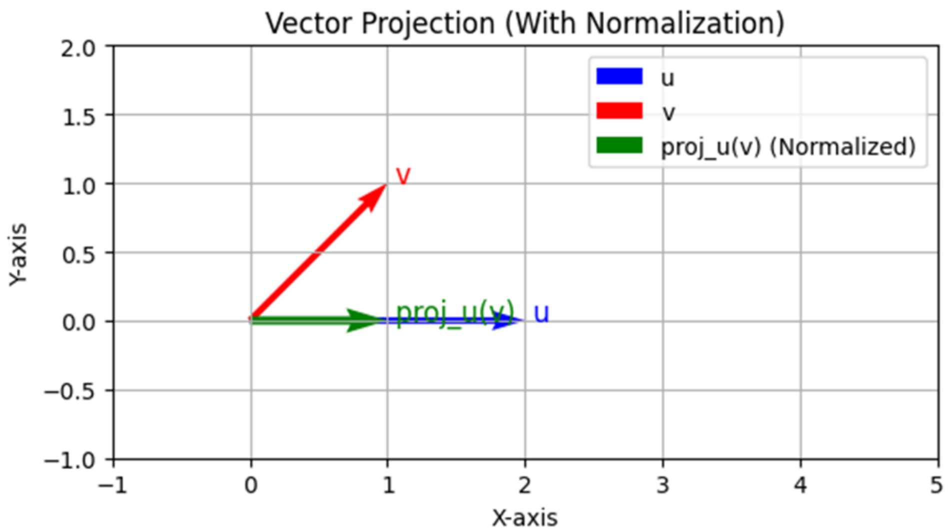
- Let's use the same vectors u and v but omit normalization to see how the projection becomes distorted.
  - Given vectors: $\mathbf{u} = \begin{matrix}2\\0\end{matrix}; \mathbf{v} = \begin{matrix}1\\1\end{matrix}$
  1. **Correct Projection with Normalization:**
     - $\mathbf{u} \cdot \mathbf{v} = [2 \quad 0] \cdot \begin{bmatrix}1\\1\end{bmatrix} = 2 \times 1 + 0 \times 1 = 2$
     - $\mathbf{u} \cdot \mathbf{u} = 2^2 + 0^2 = 4$
     - Projection:
       - $\mathbf{proj_u v} = \left(\frac{\mathbf{u \cdot v}}{\mathbf{u \cdot u}}\right)\mathbf{u} = \left(\frac{2}{4}\right)\begin{bmatrix}2\\0\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix}$

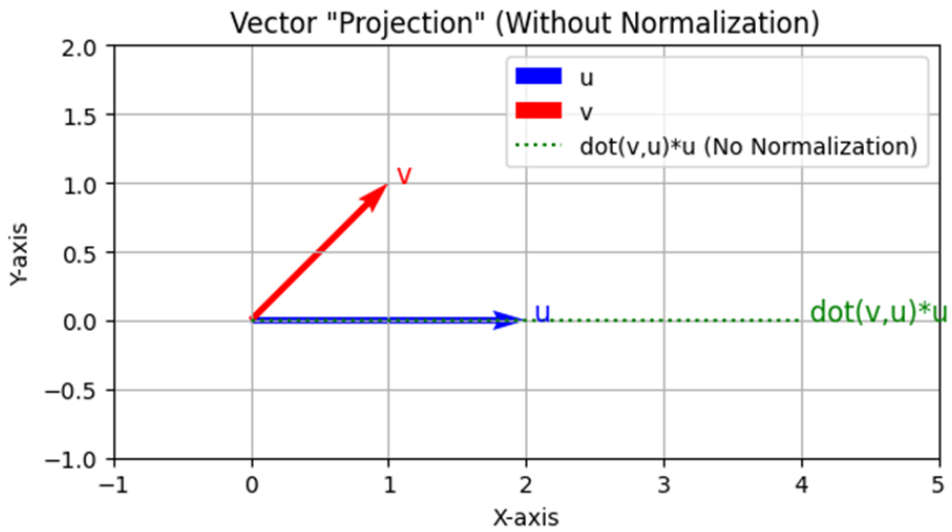


- The result is a vector in the direction of u with length scaled to match the "shadow" of v onto u.

1.2.1 Example: with and without Normalizations.

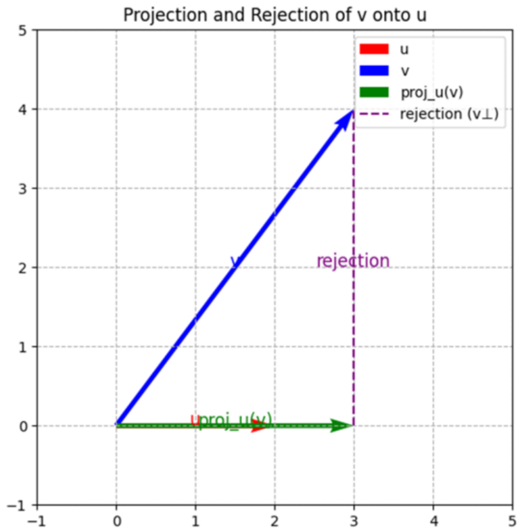
2. **Distorted Projection (without Normalization):**
   - If we **omit normalization**, the projection becomes:
     - **Distored Projection** $= (\mathbf{u} \cdot \mathbf{v})(\mathbf{u}) = 2\begin{bmatrix} \mathbf{2} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{4} \\ \mathbf{0} \end{bmatrix}$
       - Without dividing by $\mathbf{u} \cdot \mathbf{u} == \|\mathbf{u}\|^2$, the result, **over scales by the length of u.**
     - Length of **distorted projection**: 4 (2× longer than u itself!).





1.2.2 Orthogonal Projection (Rejection):

- When you **project v onto u**, you're extracting the **part of v that aligns with u** (the **projection**).
  - The **remaining part**—the **rejection**—is literally
    - **"rejected"** from u because **it has no component** in the **direction of u**.
- The perpendicular (rejection) component of v from u is:
  - $\mathbf{rej_u v = v - proj_u v}$
- **Key Property:**
  - The original vector v is the sum of its projection and rejection:
    - $\mathbf{v = proj_u v + rej_u v}$
  - The rejection component is orthogonal to u.
  - To verify orthogonality, take the dot product with u:
    - $\mathbf{rej_u v \cdot u = (v - proj_u v) \cdot u = v \cdot u - \left(\frac{u \cdot v}{u \cdot u}\right)(u \cdot u) = (v \cdot u) - (u \cdot v) = 0}$ {**Dot Product Commutativity**}
  - This confirms $\mathbf{rej_u v}$ is orthogonal to u.



1.3 Projection onto Subspace.

- **Idea:**
  - Instead of projecting **onto a single vector**, we can project onto a **subspace** (e.g., a plane or hyperplane).
- **Using Matrix Projection {only true for $\mathbf{b} \in \mathbf{C}(\mathbf{A})$} :**
  - Given a matrix $\mathbf{A} \in \mathbb{R}^{\mathbf{m}\times\mathbf{n}}$ whose columns form a basis for a subspace (full column rank matrix),
    - the projection of **b onto C(A) (column space of A)** is:
      - $\mathbf{proj_{C(A)}b = A\left(A^T A\right)^{-1} A^T b.}$
- **Special Case: Least Square Solution,**
  - If $\mathbf{b} \notin \mathbf{C}(\mathbf{A})$, this projection gives the best approximation or least square solution to $\mathbf{Ax = b.}$

1.3.1 Derivation: $\mathrm{proj}_{C(A)}b = A\left(A^T A\right)^{-1} A^T b$.

- Derive the projection of a **vector** $\mathbf{b} \in \mathbb{R}^{\mathbf{m}}$ onto the column space $\mathbf{C(A)} \exists\ \mathbf{A} \in \mathbb{R}^{\mathbf{m\times n}}$:
  - $\mathbf{proj_{C(A)} b = A\left(A^T A\right)^{-1} A^T b}$
- **Step 1 Definition of Projection onto a subspace:**
  - We want to find a vector $\mathbf{p} \in \mathbf{C(A)}$ (the projection of $\mathbf{b}$) such that the residual $\mathbf{b - p}$ is **orthogonal** to $\mathbf{C(A)}$.
  - Mathematically:
    - $\mathbf{p = A\,\hat{x}\ \left(since\ p\ is\ in\ C(A)\right)}$
  - And
    - $\mathbf{A^T(b - A\,\hat{x}) = 0\ (orthogonality\ condition)}$.

1.3.1 Derivation: $\text{proj}_{C(A)}b = A\left(A^T A\right)^{-1} A^T b.$

- **Step 2: Solve for $\hat{\mathbf{x}}$:**
  - From the orthogonality condition:
    - $\mathbf{A^T(b - A\,\hat{x}) = 0}$
    - $\mathbf{A^T b - A^T A\,\hat{x} = 0}$
    - $\boxed{\mathbf{A^T A\,\hat{x} = A^T b}}$
  - This is called **Normal Equation.**
- **Step 3: Solve the Normal Equation:**
  - Assuming A has full column rank (columns are linearly independent), **$\mathbf{A^T A}$ is invertible**, and:
    - $\mathbf{\hat{x} = \left(A^T A\right)^{-1} A^T b}$

1.3.1 Derivation: $\text{proj}_{C(A)}b = A\left(A^T A\right)^{-1} A^T b.$

- **Optional:**
  - **Step 4: Projection Matrix:**
    - The *projection p of a vector b onto the column space C(A)* is:
      - $\mathbf{p} = \overbrace{\mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T}\,\mathbf{b} = \mathbf{P}\mathbf{b}.$
    - Here:
      - $\mathbf{P} = \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T$ is the **projection matrix**, when **P multiplies b**, it returns the **projected vector p**.
  - **Step 4: Verification (Orthogonality Check):**
    - The residual $b-p$ must be orthogonal to $C(A)$:
      - $\mathbf{A}^T(\mathbf{b}-\mathbf{p}) = \mathbf{A}^T\mathbf{b} - \mathbf{A}^T\mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b} = \mathbf{A}^T\mathbf{b} - \mathbf{I}\mathbf{A}^T\mathbf{b} = \mathbf{A}^T\mathbf{b} - \mathbf{A}^T\mathbf{b} = \mathbf{0}.$

2. Solving Linear Regression.

{ With Normal and Least Squares.}

2.1 With Normal Equation.

- The **Normal Equation** is a fundamental result in linear regression
  - that provides a **closed-form solution** for finding the optimal parameters θ
    - that minimize the **sum of squared errors** (SSE) in linear regression.

Problem Setup: For Linear Regression Problem

Given:
A design matrix

$$\mathbf{X} \in \mathbb{R}^{\mathbf{m} \times \mathbf{n}}$$

(with m examples and n features, including a bias term if applicable).
A target vector $\mathbf{y} \in \mathbb{R}^{\mathbf{m}}$.
A parameter vector $\theta \in \mathbb{R}^{\mathbf{n}}$ (the weights we want to estimate).

Model:
The linear regression model is given by:

$$\mathbf{y} = \mathbf{X}\theta + \varepsilon$$

where ε is the error term.

2.1.1 Objective.

Objective: Minimize Sum of Squared Errors (SSE)

Objective:
Minimize the sum of squared errors (SSE):

$$\mathbf{J}(\theta) = \|\mathbf{y} - \mathbf{X}\theta\|^{\mathbf{2}} = (\mathbf{y} - \mathbf{X}\theta)^{\top}(\mathbf{y} - \mathbf{X}\theta)$$

2.1.2 Derivation of Normal Equation.

Goal: Find Optimal θ Minimizing the Cost Function

Goal: Find the optimal θ that minimizes the cost function $\mathbf{J}(\theta)$.
Start with the cost function:

$$\mathbf{J}(\theta) = \|\mathbf{y} - \mathbf{X}\theta\|^2 = (\mathbf{y} - \mathbf{X}\theta)^\top (\mathbf{y} - \mathbf{X}\theta)$$

Expand the expression:

$$\mathbf{J}(\theta) = \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X}\theta - \theta^\top \mathbf{X}^\top \mathbf{y} + \theta^\top \mathbf{X}^\top \mathbf{X}\theta$$

Note: $\mathbf{y}^\top \mathbf{X}\theta$ is a scalar, so

$$\mathbf{y}^\top \mathbf{X}\theta = \theta^\top \mathbf{X}^\top \mathbf{y}$$

Simplify:

$$\mathbf{J}(\theta) = \mathbf{y}^\top \mathbf{y} - 2\theta^\top \mathbf{X}^\top \mathbf{y} + \theta^\top \mathbf{X}^\top \mathbf{X}\theta$$

Next, take the gradient with respect to θ.

2.1.2 Derivation of Normal Equation.

Gradient Calculation of the Least Squares Cost Function

1. Cost Function Definition

The least squares cost function is:

$$\mathbf{J}(\theta) = \mathbf{y}^\top \mathbf{y} - \mathbf{2}\theta^\top \mathbf{X}^\top \mathbf{y} + \theta^\top \mathbf{X}^\top \mathbf{X}\theta$$

2. Term-by-Term Gradient Calculation

We compute the gradient $\nabla_\theta J(\theta)$ by differentiating each term separately.

- **Term 1: $\mathbf{y}^\top \mathbf{y}$**
  This term does not depend on θ.
  Its gradient is zero:
  $$\nabla_\theta(\mathbf{y}^\top \mathbf{y}) = \mathbf{0}$$
- **Term 2:**
  $$-\mathbf{2}\theta^\top \mathbf{X}^\top \mathbf{y}$$
  This is a linear term in θ.

  Using the identity
  $$\nabla_\theta(\theta^\top \mathbf{a}) = \mathbf{a}$$
  where $a = X^\top y$, we get
  $$\nabla_\theta(-\mathbf{2}\theta^\top \mathbf{X}^\top \mathbf{y}) = -\mathbf{2}\mathbf{X}^\top \mathbf{y}$$

2.1.2 Derivation of Normal Equation.

Gradient Calculation (Continued)

- **Term 3:**

$$\theta^\top \mathbf{X}^\top \mathbf{X}\theta$$

This is a quadratic term in θ.

Using the identity

$$\nabla_\theta(\theta^\top \mathbf{A}\theta) = (\mathbf{A} + \mathbf{A}^\top)\theta$$

and noting that $X^\top X$ is symmetric (i.e., $A^\top = A$), we have:

$$\nabla_\theta(\theta^\top \mathbf{X}^\top \mathbf{X}\theta) = \mathbf{2X}^\top \mathbf{X}\theta$$

2.1.2 Derivation of Normal Equation.

Combined Gradient of the Cost Function

Adding the gradients of all three terms, we get:

$$\nabla_\theta \mathbf{J}(\theta) = \mathbf{0} - \mathbf{2X}^\top \mathbf{y} + \mathbf{2X}^\top \mathbf{X}\theta$$

Simplifying:

$$\nabla_\theta \mathbf{J}(\theta) = -\mathbf{2X}^\top \mathbf{y} + \mathbf{2X}^\top \mathbf{X}\theta$$

2.1.2 Derivation of Normal Equation.

Setting Gradient to Zero and Deriving Normal Equation

To find the optimal θ, set the gradient to zero:

$$-2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\theta = \mathbf{0}$$

Divide both sides by 2:

$$-\mathbf{X}^\top \mathbf{y} + \mathbf{X}^\top \mathbf{X}\theta = \mathbf{0}$$

Rearranged, this gives the **Normal Equation**:

$$\mathbf{X}^\top \mathbf{X}\theta = \mathbf{X}^\top \mathbf{y}$$

- **Why this works?**
  - The **cost function** $\mathbf{J}(\boldsymbol{\theta})$ **is convex (bowl-shaped)**, so **the gradient zero-point gives the global minimum**.
  - The solution $\boldsymbol{\theta} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}$ is the **least squares estimator**.

2.1.3 Intuition Behind Normal Equation.

Geometric Intuition Behind the Normal Equation

Geometric Insight:
The normal equation arises from the fact that the optimal θ minimizes the distance between the observed vector y and the predicted vector $X\theta$, which lies in the column space of X.
This means the residual vector $r = y - X\theta$ must be **orthogonal to the column space** of X.
Mathematically, this orthogonality condition is expressed as:

$$X^\top (y - X\theta) = 0$$

This condition ensures that $X\theta$ is the orthogonal projection of y onto the column space of X, and the residual lies in the orthogonal complement (i.e., the left null space of X).

2.1.4 Interpretation of the Solution.

- The solution **θ minimizes** the **least-squares error**.
  - $\mathbf{X^TX}$ must be invertible (i.e., **X must have full column rank**).
  - If not, **regularization** (**like Ridge Regression**) can be used.
- **Advantage:** Direct solution (**no iterative optimization needed**).
- **Disadvantage:** Computationally expensive for large $\mathbf{n}$ $\left(\textbf{since } (\mathbf{X^TX})^{-1} \textbf{ is } O(\mathbf{n^3})\right)$.

2.2 What is the Least Square Solution?

- The least squares method solves the overdetermined system
- $X\theta \approx y$; $X \in \mathbb{R}^{m \times n}$ & $m > n$ by minimizing the sum of squared residuals:
- $J(\theta) = \|y - X\theta\|^2$
- The minimizer θ^* is called the least square solution and is:
- $\theta^* = (X^T X)^{-1} X^T y$
- Key Notes:
- **Uniqueness**: If X has full column rank (rank(X) = n), $X^T X$ is invertible, and θ^* is unique.
- **Degenerate case:** If rank(X) < n
  - infinitely many solutions exist (use pseudoinverses or regularization).

2.2.1 Why call it Least Squares?

1. **Minimizes Squared Error**:
   - $\boldsymbol{\theta}^*$ minimizes $\|\mathbf{y} = \mathbf{X}\boldsymbol{\theta}\|^2$ the sum of squared vertical distances (residuals) between data points and the model predictions.
2. **Geometric Interpretation**:
   - Projects $\mathbf{y}$ onto the **column space of X**, giving the **closest point** $\mathbf{X}\boldsymbol{\theta}^*$ in the subspace.
3. **Statistical Justification**:
   - Under Gaussian noise assumptions, $\boldsymbol{\theta}^*$ **is the maximum likelihood estimator**.

2.2.1 Existence and Uniqueness of Least Squares Solution

Conditions for Existence and Uniqueness of Least Squares Solution

Least Squares Solution:

$$\hat{\theta}_{\text{LS}} = (X^\top X)^{-1} X^\top y$$

Condition for Existence and Uniqueness:

- $X^\top X$ must be **invertible**.
- This requires that X has **full column rank**, i.e., $\text{rank}(X) = n$.
- Geometrically: the columns of X must be **linearly independent**.
- If $X^\top X$ is not invertible (i.e., singular or ill-conditioned), the solution:

$$\hat{\theta}_{\text{LS}} = (X^\top X)^{-1} X^\top y$$

  does not exist in the usual sense.

In such cases, we need a remedy — this is where *regularization* comes in.

2.2.2 Regularized Least Squares

Regularization: Ridge Regression (L2 Regularization)

Motivation: When $X^\top X$ is not invertible or when we want to control model complexity, we add a penalty on the size of the coefficients.

Modified Objective:

$$J_{\text{ridge}}(\theta) = \|y - X\theta\|^2 + \lambda\|\theta\|^2$$

where $\lambda > 0$ is the regularization parameter.

Ridge Solution:

$$\hat{\theta}_{\text{ridge}} = (X^\top X + \lambda I)^{-1} X^\top y$$

Why This Helps:

- $X^\top X + \lambda I$ is always invertible when $\lambda > 0$, even if $X^\top X$ is not.
- Helps prevent overfitting by penalizing large weights.
- Especially useful in high-dimensional or multicollinear data settings.

2.2.3 When Solution exist but is Not Unique.

What Does a Unique Solution Mean?

A **unique solution** is one where there is exactly *one* parameter vector θ that minimizes the residual error.

- No other distinct vector $\theta' \neq \theta$ achieves the same minimal error.
- Ensures stability and interpretability of the model.
- Occurs when the design matrix X has full column rank and $X^\top X$ is invertible.
- Geometrically, the projection of y onto the column space of X corresponds to exactly one $\hat{\theta}$.

2.2.3 When Solution exist but is Not Unique.

When Least Squares Solution Exists but Is Not Unique

Scenario:
A least squares solution always exists, but it may not be unique when:

$$\text{rank}(X) < n \quad \text{(i.e., } X \text{ does not have full column rank)}$$

Implications:

- $X^\top X$ is not invertible (singular matrix).
- There are infinitely many solutions θ that minimize $\|y - X\theta\|^2$.
- The system is underdetermined: multiple θ produce the same projection $X\theta$.

Selecting a Unique Solution:
To obtain a unique solution, we can use:

$$\hat{\theta} = X^+ y \quad \text{(minimum norm solution via Moore–Penrose pseudoinverse)}$$

or apply regularization:

$$\hat{\theta}_{\text{ridge}} = (X^\top X + \lambda I)^{-1} X^\top y \quad \text{(ridge regression with } \lambda > 0\text{)}$$

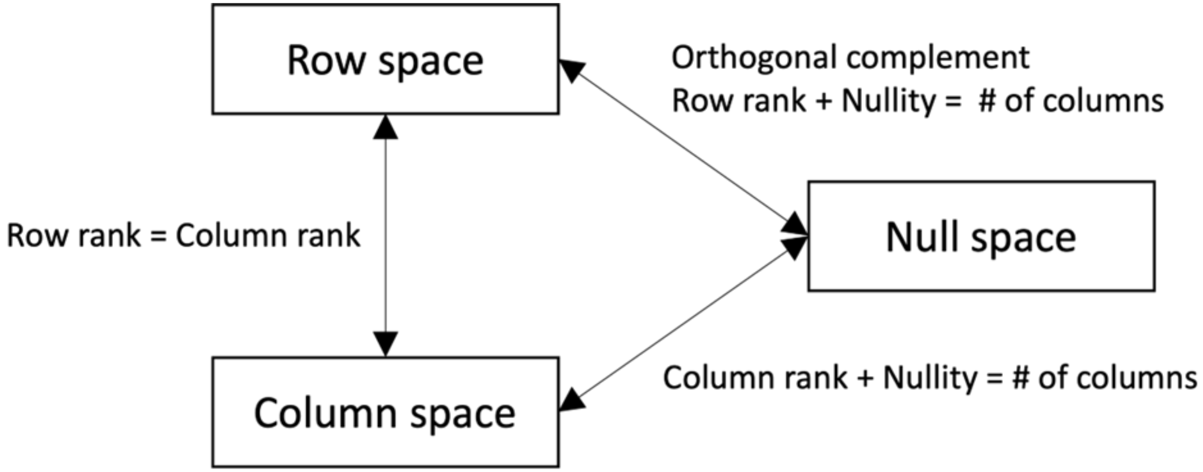
Both methods produce a unique $\hat{\theta}$ that minimizes the residual error.

3. Fundamental Theorem of Linear Algebra.

{ Combining all the fundamental subspaces of Matrix.}

3.1 What is FTLA?

- The **Fundamental Theorem of Linear Algebra (FTLA)** summarizes the relationships between the four fundamental subspaces associated with a matrix.



3.2 Domain vs Range Space.

- In Linear Algebra, if $A \in \mathbb{R}^{m \times n}$, we think of A as a linear transformation:
- $A: \mathbb{R}^n \to \mathbb{R}^m$
- So:
  - Domain → $\mathbb{R}^n$: the space where the input vectors x live.
  - Range/Co-domain → $\mathbb{R}^m$: the space where the output vectors Ax live.
- Fundamental Subspaces in Each space:
- Domain Space $\mathbb{R}^n$
- This is the space of input vectors x, and it contains:
- Row space: $C(A^T) \subseteq \mathbb{R}^n$
- Null space: $\mathcal{N}(A) \subseteq \mathbb{R}^n$
- Together:
  - $\mathbb{R}^{\mathbf{n}} = \mathbf{C}\left(\mathbf{A}^{\mathbf{T}}\right) \oplus \mathbf{N}(\mathbf{A})$

3.2.1 Fundamental Subspaces in Each space:

- **Domain Space $\mathbb{R}^{\mathbf{n}}$:**
  - This is the space of input vectors x, and it contains:
    - Row space: $\mathbf{C}\left(\mathbf{A}^{\mathbf{T}}\right) \subseteq \mathbb{R}^{\mathbf{n}}$
    - Null space: $\boldsymbol{\mathcal{N}}(\mathbf{A}) \subseteq \mathbb{R}^{\mathbf{n}}$
  - Together:
    - $\mathbb{R}^{\mathbf{n}} = \mathbf{C}\left(\mathbf{A}^{\mathbf{T}}\right) \oplus \boldsymbol{\mathcal{N}}(\mathbf{A})$
- **Range/Co-domain Space $\mathbb{R}^{\mathbf{m}}$:**
  - This is the space of **output vectors y = Ax** and it contains:
    - Column Space: $\mathbf{C}(\mathbf{A}) \subseteq \mathbb{R}^{\mathbf{m}}$
    - Left Null Space: $\boldsymbol{\mathcal{N}}\left(\mathbf{A}^{\mathbf{T}}\right) \subseteq \mathbb{R}^{\mathbf{n}}$
  - Together:
    - $\mathbb{R}^{\mathbf{n}} = \mathbf{C}(A) \oplus \boldsymbol{\mathcal{N}}\left(\mathbf{A}^{\mathbf{T}}\right)$

3.3 The Three Standard Statements of the FTLA

- **Statement 1:**
  - The Column Space and Left Null Space Are Orthogonal Complements in $\mathbb{R}^m$:
    - $\mathbf{C(A) \perp \mathcal{N}\left(A^T\right) \text{ and } \mathbb{R}^n = C(A) \oplus \mathcal{N}\left(A^T\right)}$
- **Statement 2:**
  - The Row Space and Null Space Are Orthogonal Complements in $\mathbb{R}^n$:
    - $\mathbf{C\left(A^T\right) \perp \mathcal{N}(A) \text{ and } \mathbb{R}^n = C(A) \oplus \mathcal{N}(A)}$
- **Statement 3:**
  - The Dimensions of the Four Subspaces Are Related by Rank:
    - $\mathbf{dim\big(C(A)\big) = rank(A)}$
    - $\mathbf{dim\left(C\left(A^T\right)\right) = rank(A)}$
    - $\mathbf{dim\big(\mathcal{N}(A)\big) = n - rank(A)}$
    - $\mathbf{dim\left(\mathcal{N}\left(A^T\right)\right) = m - rank(A)}$

Putting them Together.

Range Space and FTAL

Fundamental Theorem of Linear Algebra – Orthogonality and Decomposition (Range Space)

1. **Orthogonality:** $C(A) \perp N(A^\top)$
   - The column space $C(A) \subseteq \mathbb{R}^m$ is orthogonal to the left null space $N(A^\top) \subseteq \mathbb{R}^m$.
   - For $y_1 = Ax \in C(A)$ and $y_2 \in N(A^\top)$:

$$y_1^\top y_2 = x^\top A^\top y_2 = 0 \Rightarrow y_1 \perp y_2$$

2. **Direct Sum:** $\mathbb{R}^m = C(A) \oplus N(A^\top)$
   - Every vector $y \in \mathbb{R}^m$ can be uniquely decomposed as:

$$y = y_c + y_n, \quad \text{with } y_c \in C(A),\ y_n \in N(A^\top)$$

   - Since:

$$\dim(C(A)) + \dim(N(A^\top)) = \text{rank}(A) + (m - \text{rank}(A)) = m$$

   and $C(A) \cap N(A^\top) = \{0\}$

Column Space and FTAL

Fundamental Theorem of Linear Algebra – Orthogonality and Decomposition (Domain Space)

3. Orthogonality: $C(A^\top) \perp N(A)$

- The row space $C(A^\top) \subseteq \mathbb{R}^n$ is orthogonal to the null space $N(A) \subseteq \mathbb{R}^n$.
- For $x_1 \in C(A^\top)$ and $x_2 \in N(A)$, $Ax_2 = 0 \Rightarrow x_1^\top x_2 = 0$

4. Direct Sum: $\mathbb{R}^n = C(A^\top) \oplus N(A)$

- Every vector $x \in \mathbb{R}^n$ can be uniquely written as:

$$x = x_r + x_n, \quad \text{with } x_r \in C(A^\top),\ x_n \in N(A)$$

- Since:

$$\dim(C(A^\top)) + \dim(N(A)) = \operatorname{rank}(A) + (n - \operatorname{rank}(A)) = n$$

and $C(A^\top) \cap N(A) = \{0\}$

Thank You.