VISPAD INSTITUTE OF TECHNOLOGY

Nea onnim no sua a ohu,

System of Equations

Fall 2025

(October – December Virtual Internship)

At the end of the session the students or candidates should be able to understand and work with:

- Linear systems in machine learning
- Introduction to Gaussian Elimination
- The Matrix Inverse
- Solving $Ax = b$ using the Inverse
- Numerical Stability and Alternatives

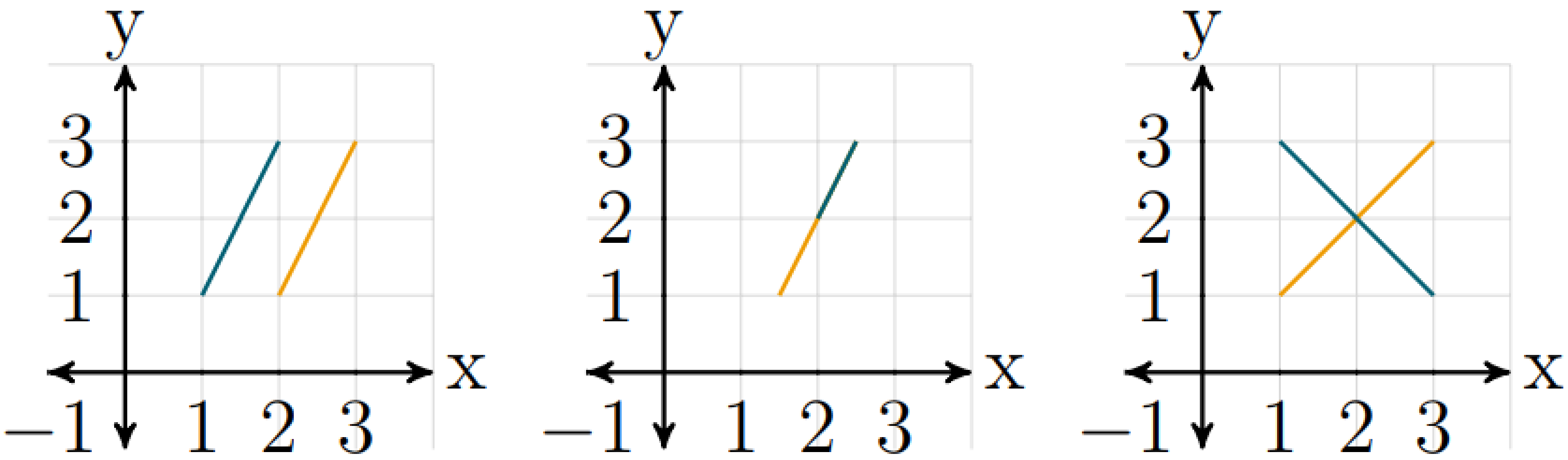
Linear Systems in Machine Learning Models

- ❑Vectors represent individual data points or features.
- ❑Matrices organize entire datasets or represent linear transformation.
- ❑These concepts are fundamental to a common task in machine learning: ***finding the optimal parameters for a model***

- A **system** is a collection of components that accepts an input and produces an output.
- A **linear system** will then be a set linear equations that receives a vector and produces one, several, or no solutions.

$$\begin{cases} ax_1 + bx_2 = c \\ dx_1 + ex_2 = f \end{cases}$$

- Elements a, b, c, d, e, f are scalars that belong to $\mathbb{R}$ and define the system by characterizing two equations which are lines.
- The solution is defined by x_1 and x_2, which can be unique, nonexistent, or non-unique.



- If the values are such that the lines are parallel, then the system has no solutions.
- If the lines land on top of each other, then the system has an infinite number of solutions.

- If the lines intersect each other, this reflects the case where a system has a unique solution.

$$A.\vec{x} = \vec{b}$$

- A represent a matrix whose elements are the scalars a, b, c, d.
- The vector $\vec{x}$ is the solution we are looking for(if it exists)
- $\vec{b}$ is another vector.

- If A represent a linear transformation, then we want to find a vector $\vec{x}$, that when we transformed by A, lands on $\vec{b}$.
- $\vec{b}$ is where the solution has to land after being transformed by A.

$$\begin{pmatrix} a & b \\ d & e \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c \\ f \end{pmatrix}$$

- For two vectors to intercept somewhere in space, they need to be **linearly independent.**
- The **determinant** of a matrix that has these same vectors for columns cannot be zero.
- For a system to have unique solution, the determinant of the matrix that represent the system must be non-zero.

- Consider a system defined as:

$$\begin{cases} -5x_1 + 2x_2 = -2 \\ -4x_1 + x_2 = -4 \end{cases}$$

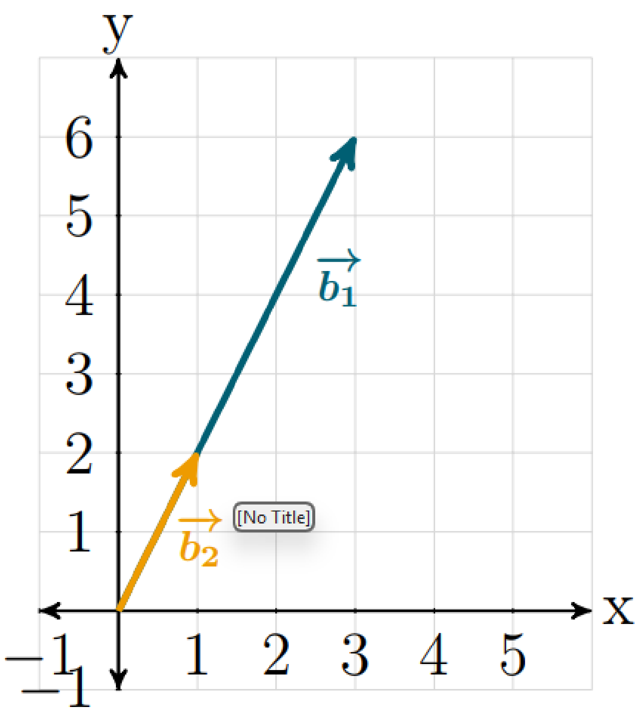
- Let's verify if there is a solution for the system: $A = \begin{pmatrix} -5 & 2 \\ -4 & 1 \end{pmatrix}$. The determine of A is calculated as:

$$\begin{vmatrix} -5 & 2 \\ -4 & 1 \end{vmatrix} = -5 \cdot 1 - 2 \cdot (-4) = -5 + 8 = 3$$

- Consider another matrix $B = \begin{pmatrix} 3 & 1 \\ 6 & 2 \end{pmatrix}$. The determine is 0

$$\begin{vmatrix} 3 & 1 \\ 6 & 2 \end{vmatrix} = 3 \cdot 2 - 6 \cdot 1 = 0$$

- The vectors are linearly dependent.
- The angle formed by them is 0°



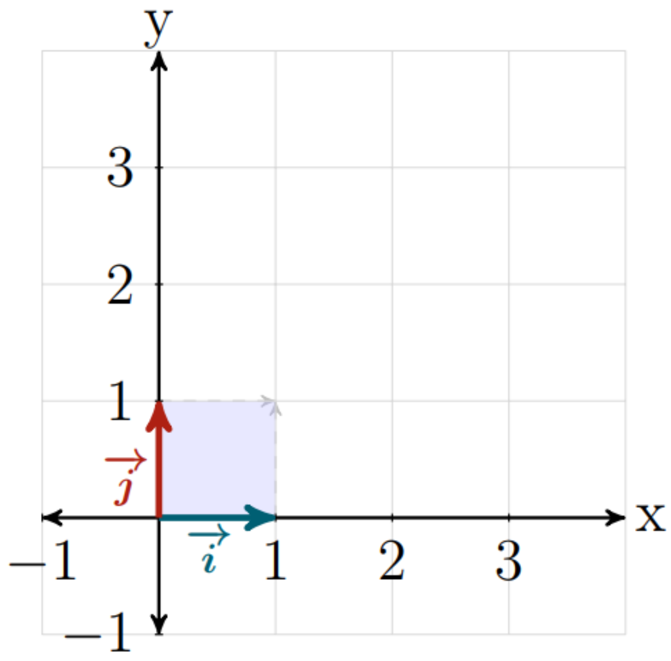
- The equation below represents the system in question:

$$\overbrace{\begin{pmatrix} -5 & 2 \\ -4 & 1 \end{pmatrix}}^{A} \cdot \overbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}^{\vec{x}} = \overbrace{\begin{pmatrix} -2 \\ -4 \end{pmatrix}}^{\vec{b}}$$

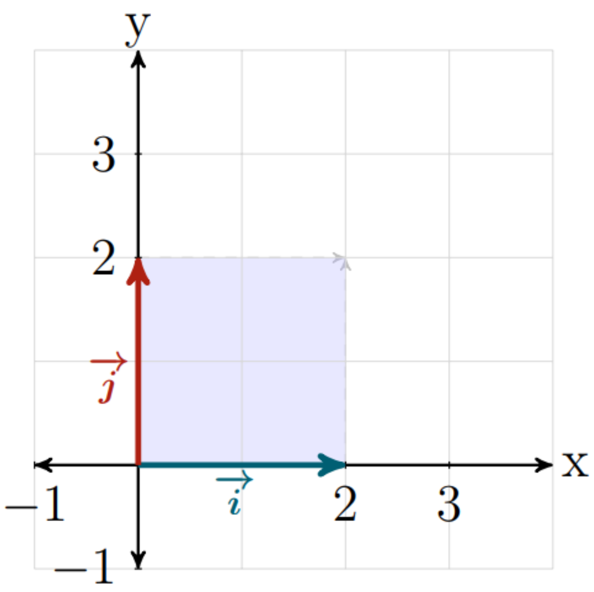
- A is a matrix that represent a linear transformation, and $\vec{x}$ is a vector.
- $A \cdot \vec{x}$ will map $\vec{x}$ into a new vector.

- The resultant recast vector of that transformation has to be equal to another vector, $\vec{b}$

❑Consider the standard basis $\vec{i} = (1,0)^T$ and $\vec{j} = (0,1)^T$ which form a square:

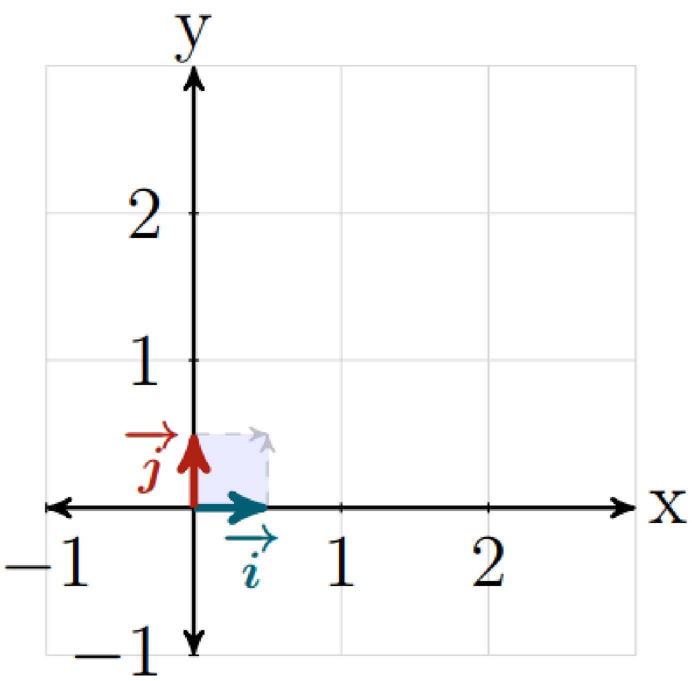


- Let's define a linear transformation represented by a matrix: $L = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.
- Applying the mapping L to $\vec{\imath}$ and $\vec{\jmath}$ will result into two new vectors, $\overrightarrow{i^*}$ and $\overrightarrow{j^*}$, with values $(2, 0)^T$ and $(0, 2)^T$ respectively.



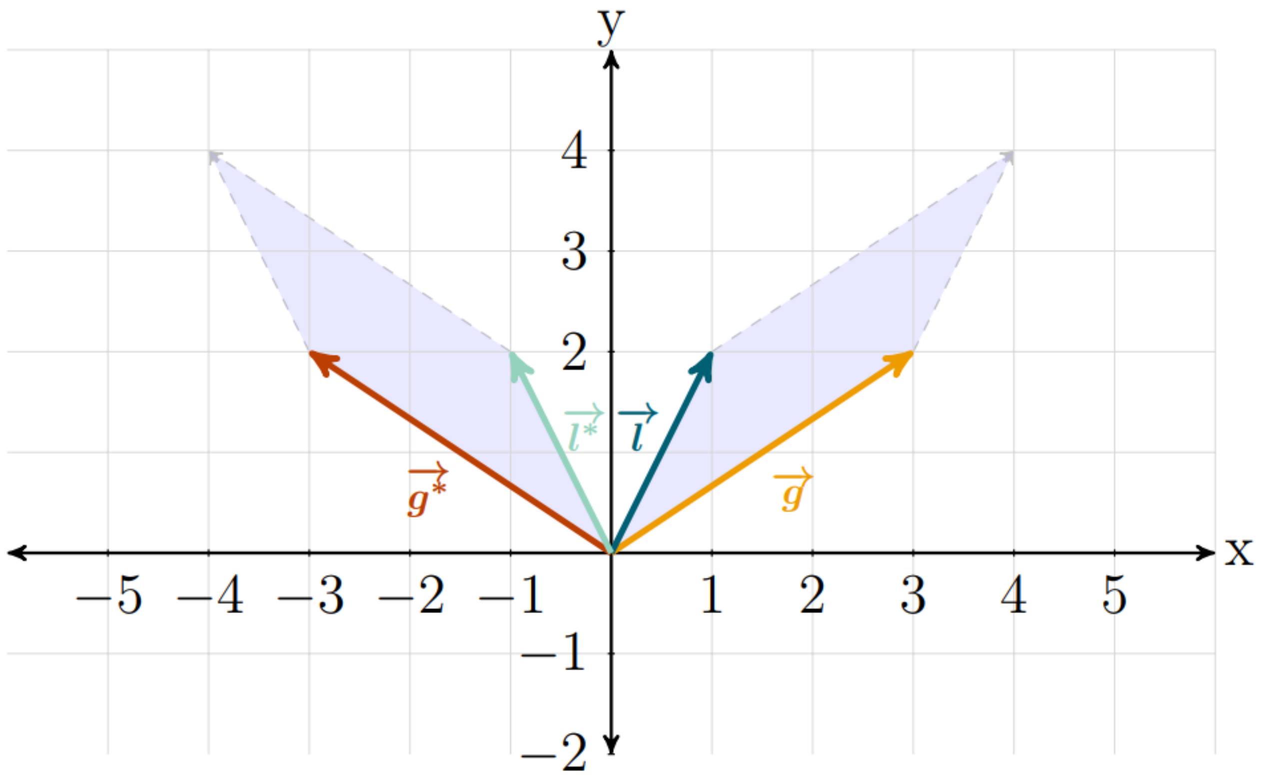
- The area of square formed by the basis vectors quadruples in size when transformed by L.
- Every shape that L modifies will be scaled by a factor of four.

- Consider the following linear transformation H defined by: $H = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$
- If we apply H to the vectors $\vec{i}$ and $\vec{j}$ we will end up with $\overrightarrow{i^{**}}$ and $\overrightarrow{j\ast\ast}$ values $\left(\frac{1}{2}, 0\right)^T$ and $\left(0, \frac{1}{2}\right)^T$:

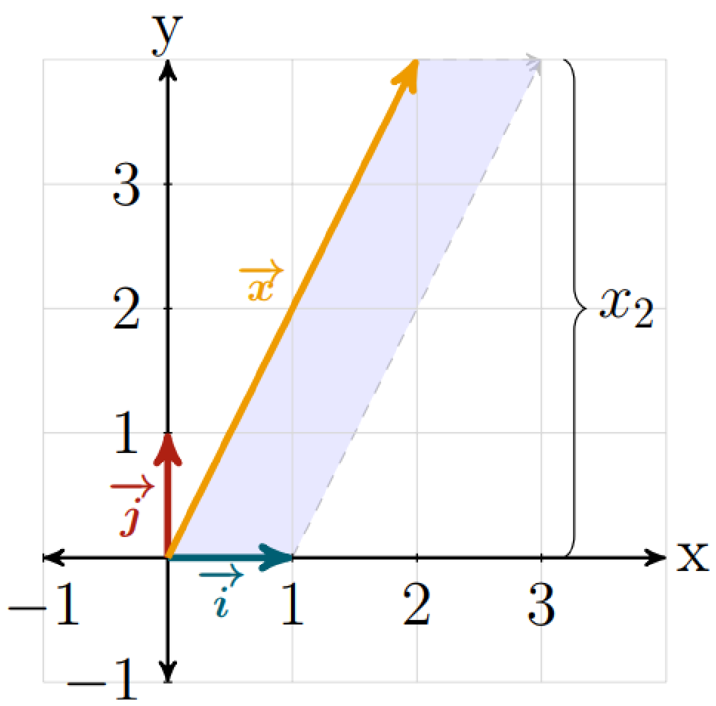


- The scalar that represents the change in area size due to the transformation in called the **determinant.**

❑Consider the transformation $Z = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ and two new vectors: $\vec{g} = \binom{3}{2}$, $l = \binom{1}{2}$



- Let’s consider the plot:



- The vector $\vec{x}$ represent any vector that verifies a system of equations

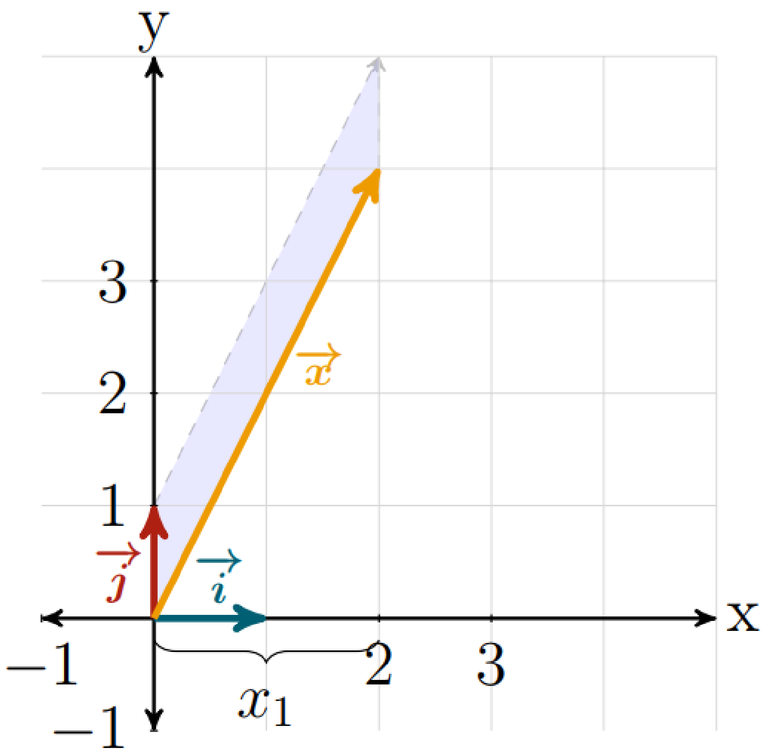
- The blue parallelogram has an area equal to

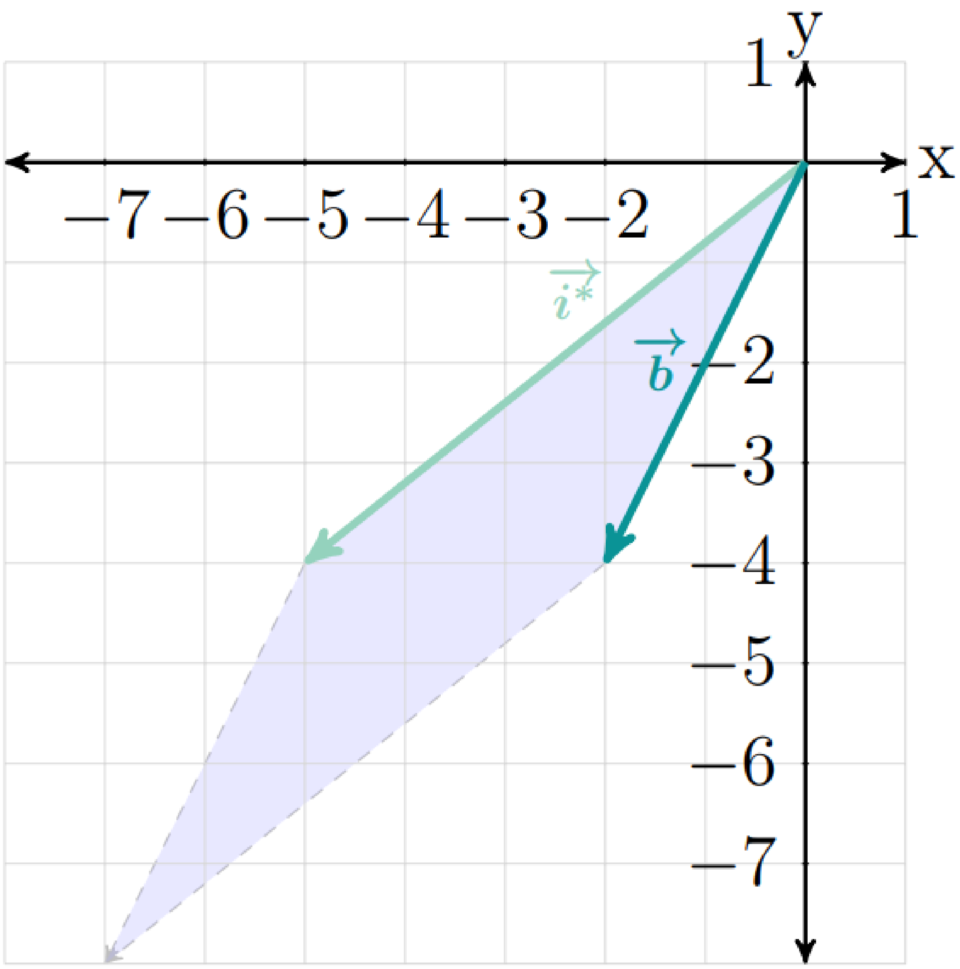
  1. x_2 $(area\ is\ equal\ to\ x_2)$.

- The area is called **signed area** because x_2 can be negative.

❑The signed area of the blue parallelogram will be

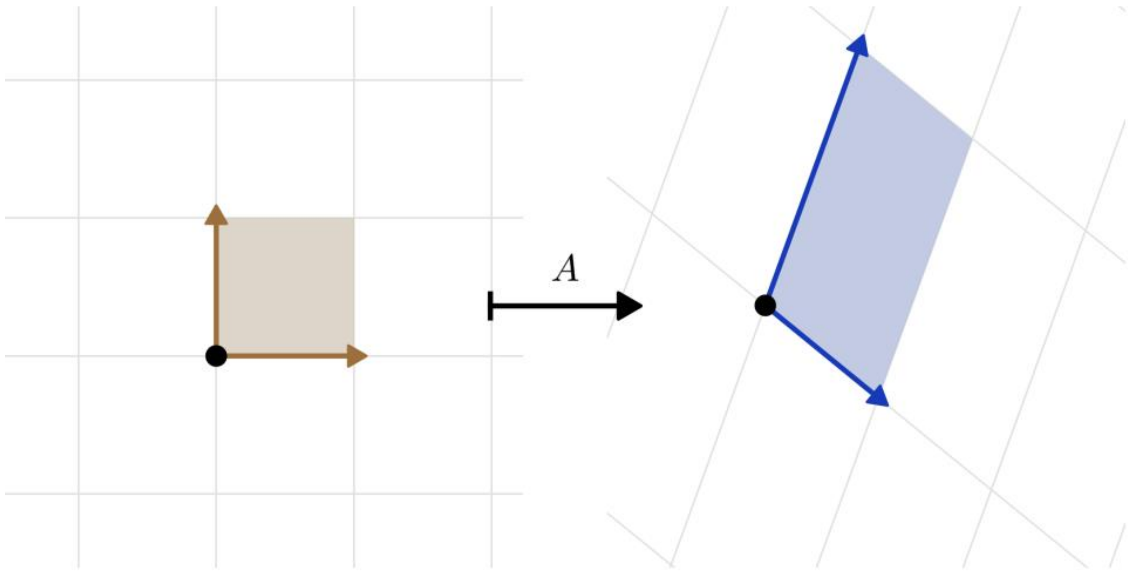
1. x_1, resulting in x_1.



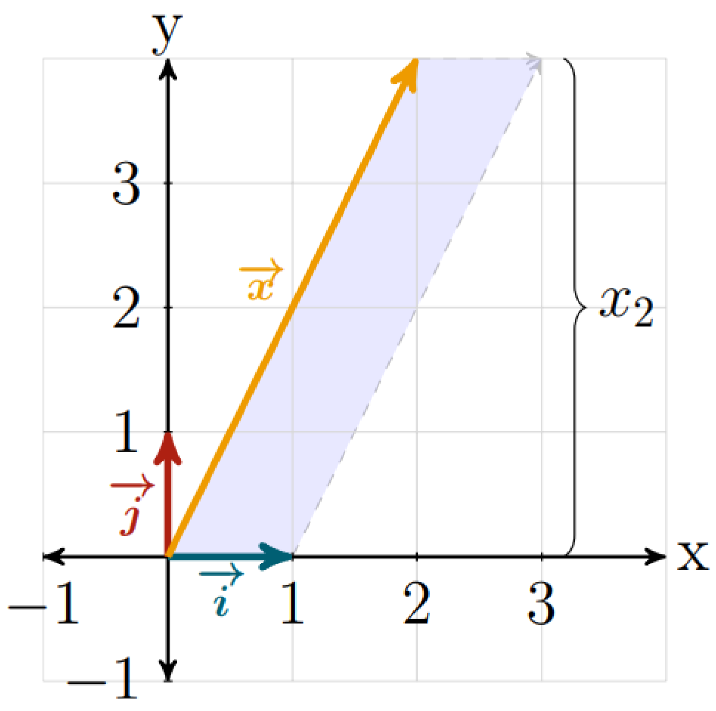


y
x
1
−7 −6 −5 −4 −3 −2
1
$\overrightarrow{i^*}$
$\overrightarrow{b}$
−2
−3
−4
−5
−6
−7

Consider any linear transformation A, mapping the unit square to parallelogram.



The area of this parallelogram describes how A scales the unit square.



If the area of the original parallelogram was x_2, this new shape will have an area of $\det(A) \cdot x_2$:

$$Area = \det(A) \cdot x_2$$

$$x_2 = \frac{Area}{\det(A)}$$

- We know the coordinates of vector $\vec{b}$ and we know where it lands, so we can create a new matrix:

$$\begin{pmatrix} -5 & -2 \\ -4 & -4 \end{pmatrix}$$

- The determinant of this matrix will represent the area of the transformed parallelogram:

$$x_2 = \frac{\begin{vmatrix} -5 & -2 \\ -4 & -4 \end{vmatrix}}{\begin{vmatrix} -5 & 2 \\ -4 & 1 \end{vmatrix}} = \frac{(-5)\cdot(-4) - (-2)\cdot(-4)}{(-5)\cdot 1 - 2\cdot(-4)} = \frac{12}{3} = 4$$

$$x_1 = \frac{\begin{vmatrix} -2 & 2 \\ -4 & 1 \end{vmatrix}}{\begin{vmatrix} -5 & 2 \\ -4 & 1 \end{vmatrix}} = \frac{(-2)\cdot 1 - (-2)\cdot(-4)}{3} = \frac{6}{3} = 2$$

The solution for our system is the vector $\vec{x} = (2,4)^T$

Cramer's Rule

- ❑Consider a linear system represented as follows:

$$A\vec{x} = \vec{b}$$

- ❑Where A is a squared matrix with size $n \times n, \vec{x} = (x_1, x_2, \dots, x_n)^T$ and $\vec{b} = (b_1, b_2, \dots, b_n)$.

- We can define the values of the vector $\vec{x}$ by: $x_i = \frac{det(A_i)}{det(A)}$
- The value $\det(A_i)$ is the matrix formed by replacing the i_{th} column of A with the column vector $\vec{b}$.

❑Next, for the verification part, we have to transform $(2,4)^T$ with A and to check if we get $\vec{b}$:

$$\begin{pmatrix} -5 & 2 \\ -4 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} -10+8 \\ -8+4 \end{pmatrix} = \begin{pmatrix} -2 \\ -4 \end{pmatrix}$$

Linear Regression

- Linear regression is a method for modeling the relationship between two scalar values: the input variable x and the output variable y
- The model assumes that y is a linear function or weighted sum of the input variable.

$$y = f(x)$$

$$y = b_0 + b_1 \times x_1$$

- The model can also be used to model an output variable given multiple input variable called multivariate linear regression.

$$y = b_0 + (b_1 \times x_1) + (b_2 \times x_2) + \cdots$$

- The objective of creating a linear regression model is to find the values for the coefficient values (b) that minimize the error in the prediction of the output variable y.

Matrix Formulation of Linear Regression

- Linear regression can be stated using Matrix notation:

$$Ax = b$$

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{32} & a_{3,3} \\ a_{4,1} & a_{4,2} & a_{4,3} \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix}$$

- The problem becomes a system of linear equations where the x vector values are unknown.
- The solution can be determined using linear least square.

$$\|A.x - b\|^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j} \cdot \left(b_j - y_i\right)^2$$

- This formulation has a unique solution as long as the input columns are independent.
- We cannot always get the error $e = b - Ax$ down to zero.
- When e is zero, x is an exact solution to $Ax = b$.
- When the length of e is as small as possible, $\hat{x}$ is a least squares solution.

- In matrix notation, this problem is formulated using the so-named **Normal equation**:

$$X^T X\theta = X^T y$$

Where:

- X is the matrix of features (often called the designed matrix). Each row represent a data point, and each column represent a feature.

A column of ones is added to X to account for the intercept term.

- θ is the vector of model coefficients we want to find.
- y is the vector of actual target values.
- X^T is the transpose of the feature X.

- The **normal equation** is familiar to $Ax = b$, where $A = X^T X$, $x = \theta$, and $b = X^T y$.

Example

- Suppose we have data points:

$$(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$$

- We want a line:

$$y = \theta_0 + \theta_1 \times x_1$$

- We write in a matrix form:

$$y = X\theta$$

Where:

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}, \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}, y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

- ❑ Column 1 represents the constant term.
- ❑ Column X represent the x-values
- ❑ θ contains the unknown parameters we want.

- In linear algebra, a regression is model is written as $X\theta = y$ where each column of X is a vector in $\mathbb{R}^m$.
- For simple linear algebra: $y = \theta_0 + \theta_1 x$, we write

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}$$

- The constant term (first column) is essential because:
  - Without the constant, the model is forced to pass through the origin.
- The model becomes $y = \theta_1 x$ if we remove the constant term.

- The prediction vector becomes $y = \theta_1(x_1, x_2, \ldots, m)^T$.
- This means:
  - The line must pass through $(0, 0)$
  - The column space of X is only the line spanned by $(x_1, x_2, \ldots, m)^T$

- We can only slide along one direction – we cannot shift up or down.
- Most real data does not go through the origin, so the fit becomes poor.

- Each column of Z is a vector in $\mathbb{R}^m$:
  - $v_1 = (1, 1, \dots, 1)^T$
  - $v_2 = (x_1, x_2, \dots, x_m)^T$
- These two vector span a plane inside $\mathbb{R}^m$
- Any prediction vector $\hat{y}$ we produce must lie in that plane: $\hat{y} = X\theta = \theta_0 v_1 + \theta_1 v_2$
- NB: Actual data vector y may not lie in that plane.

- To find the point in the plane (column space) closest to y, we project y into the column space of X.
- In other words, that point is the orthogonal projection of y onto the column space of X.

$$\hat{y} = proj_{Col(X)}(y)$$

The error vector: $r = y - \hat{y}$ must be orthogonal to the

- This gives the condition: $X^T(y - X\theta) = 0$
- From the orthogonality condition: $X^T y - X^T X\theta = 0$
- We rearrange: $X^T X\theta = X^T y$
- If the columns of X are linearly independent, then $X^T X$ is invertible: $\theta = (X^T X)^{-1} X^T y$

- Given: $(0, 1), (1, 2), (2, 2.5)$
- $X =??, y =??$
- Compute: $X^T X =??, X^T y =??$
- Solve: $\theta = (X^T X)^{-1} X^T y =???$

What is the best-fitting line: $\hat{y} = \_ + \_x$


THANK YOU
BEARY MUCH