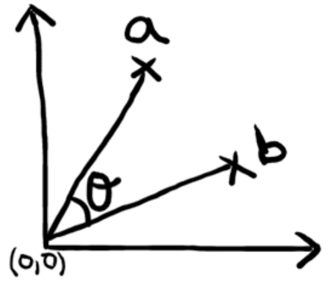
Fundamentals of Machine Learning

LINEAR ALGEBRA

Angles and Orthogonality

Angle between two vectors $\vec{a}$ and $\vec{b}$ can be computed using the formula

$$\cos\theta = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|.\|\vec{b}\|}$$



Orthogonal Vectors

Two given nonzero vectors $\vec{v} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\vec{u} = \begin{bmatrix} c \\ d \end{bmatrix}$ are called orthogonal vectors if and only if their dot product is zero

i.e $\vec{v}\cdot\vec{u} = 0$

$\vec{v}\cdot\vec{u} = 0$ implies $\cos\theta = 0$, hence $\theta = 90°$.

What will be the angle θ

Vectors $\vec{v}$ *and* $\vec{u}$ *are perpendicular to each other*

Vector $\vec{v}$ *is orthogonal (at right angles) to* $\vec{u}$ *if and only if* $\vec{v}.\vec{u} = 0$

Exercise

1. Find the angle between $u = \langle -3, 4 \rangle$ and $v = \langle 5, 12 \rangle$.
2. Determine whether the vectors $v = \langle 3, 4 \rangle$ and $u = \langle -4, 3 \rangle$ are orthogonal or not

Vector space V

Vector Space V is set of all linear combination of its elements(Vector).

1. If x and y $\epsilon\, V$, $then\ x + y\ \epsilon\, V$
2. If x ϵ V and αx ϵ V for any scalar α
3. There exist 0 ϵ V then x+0=x for any x ϵ V

Definition – Linear combination

We say that v is a linear combination of $v_1, v_2, \ldots, v_n$, if there exist scalars $x_1, x_2, \ldots, x_n$ such that $v = x_1v_1 + x_2v_2 + \ldots + x_nv_n$.

Let a $=\begin{bmatrix}1\\2\end{bmatrix}$ and b = $\begin{bmatrix}3\\4\end{bmatrix}$ then

2.a +3.b is a linear combination of a and b

$Example:$ $2\begin{bmatrix}1\\2\end{bmatrix} + 3\begin{bmatrix}3\\4\end{bmatrix}$ is a linear combination of vectors $\begin{bmatrix}1\\2\end{bmatrix}$ and $\begin{bmatrix}3\\4\end{bmatrix}$

Theorem– Expressing a vector as a linear combination

Let B denote the matrix whose columns are the vectors $v_1, v_2, \ldots, v_n$. Expressing $v = x_1v_1+x_2v_2+\ldots+x_nv_n$ as a linear combination of the given vectors is then equivalent to solving the linear system $Bx = v$.

Defintion : - Span of set of vectors

Given a set of vectors, V= { $v_1, v_2 \ldots v_k$}, the set of all linear combination of V is called the span

Suppose V= {v_1,v_2} where $v_1=\begin{bmatrix}1\\3\end{bmatrix}$ and $v_2=\begin{bmatrix}2\\5\end{bmatrix}$, then

span {v_1,v_2}$=C_1.\begin{bmatrix}1\\3\end{bmatrix}+C_2.\begin{bmatrix}2\\5\end{bmatrix}$, where $C_1, C_2 \epsilon R^n$.

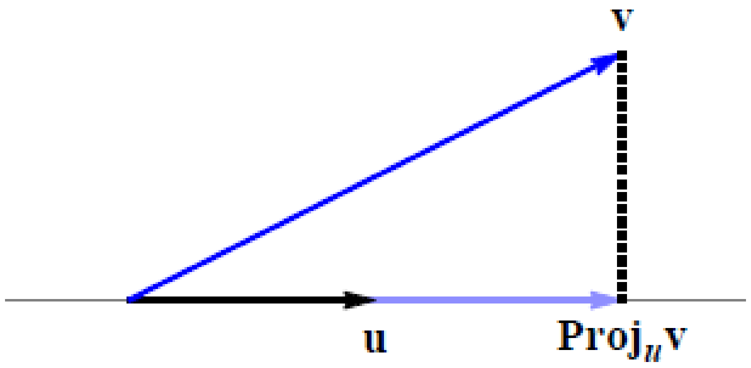
For example, let $C_1=1$, and $C_2=4$

$$v_3=1.\begin{bmatrix}1\\3\end{bmatrix}+4.\begin{bmatrix}2\\5\end{bmatrix} = \begin{bmatrix}1\\3\end{bmatrix}+\begin{bmatrix}8\\20\end{bmatrix}=\begin{bmatrix}1+8\\3+20\end{bmatrix}=\begin{bmatrix}9\\23\end{bmatrix}$$

v_3 will be in the span of {v_1,v_2}

Show that the vector $\begin{bmatrix} 20 \\ 4 \end{bmatrix}$ belongs to span of $\left\{\begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}\right\}$

Orthogonal Projections



The orthogonal projection of $\vec{v}$ onto $\vec{u}$ gives the component vector $\mathbf{Proj_u}\vec{v}$ of $\vec{v}$ in the direction of $\vec{u}$.

$$Proju^{\vec{v}} = \frac{\vec{u}.\vec{v}}{||u||.||u||}\vec{u}$$

Example : Find the orthogonal projection of v onto u.

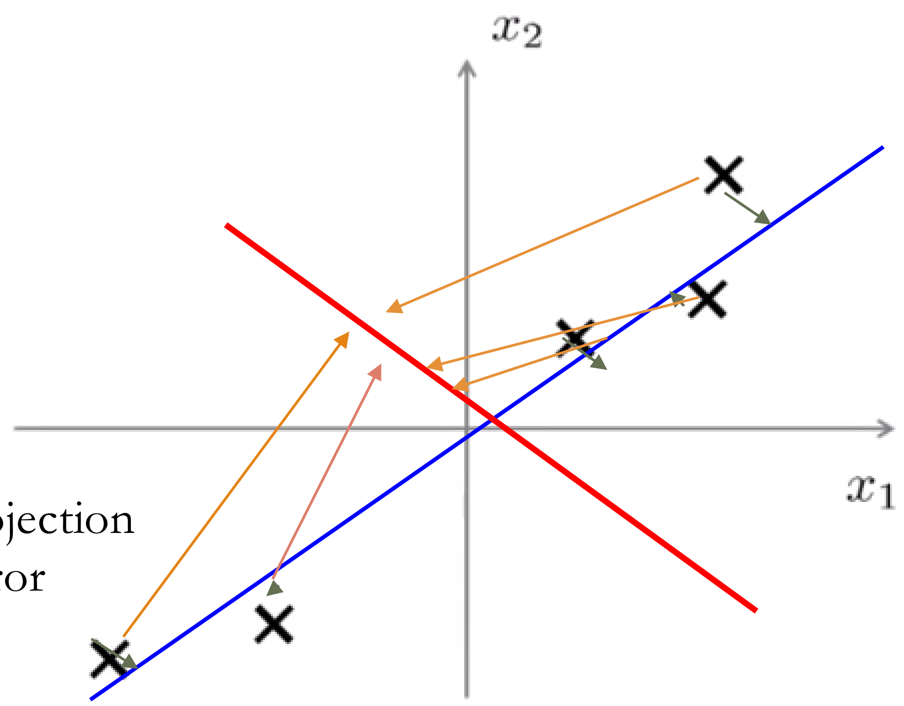
u=$\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ and v=$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

Solution:

$$Proju^{\vec{v}} = \frac{\vec{u}.\vec{v}}{||u||.|u|}\vec{u}$$

$$= \frac{(-1 \text{ x } 1 + 3 \text{ x } -1)}{\sqrt{(-1^2)+(3^2)} \ .\sqrt{(-1^2)+(3^2)}} \ . \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \frac{-4}{10}. \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \frac{-2}{5}\begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2/5 \\ -6/5 \end{bmatrix}$$

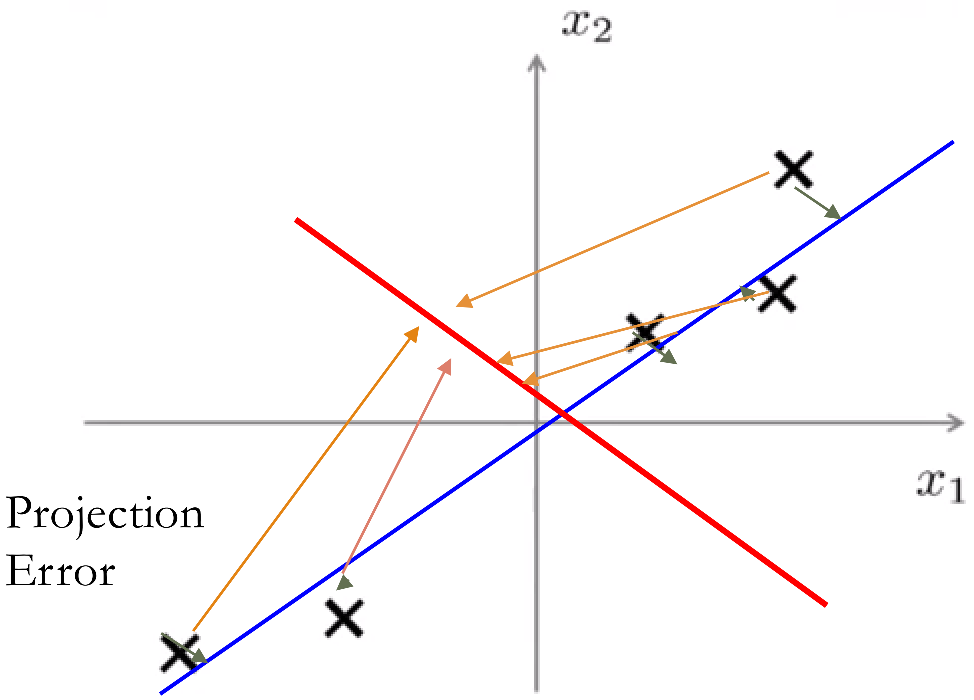
Principle Component Analysis(PCA)-Dimension reduction



Projection
Error

Principle Component Analysis(PCA)-Dimension reduction

Reduce from 2-dimension to 1-dimension: Find a direction (a vector $u^{(1)} \in \mathbb{R}^n$) onto which to project the data so as to minimize the projection error.



VECTOR SPACE MODEL IN SEARCH ENGINES:VSM

Consider the database of articles for a magazine called "*Sports*" . Assume, articles are distinguishable by only three keywords:

- Cricket, Hockey, Tennis.
- One article (document 1) talks only about Cricket.
- Another article (document 2) discusses only Hockey.
- A third article (document 3) is 30% about Tennis and 70% about Hockey. All divisions of article space among the three keywords are possible.

Vectors are used to describe how a vector space search engine processes a query in the *Sports* database.

To store data about the division of topics discussed in the particular ***Sports*** document number 1, use the column vector , $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

where the first entry of the vector represents the percentage of the article devoted to CRICKET, the second corresponds to Hockey, and the third to Tennis.

Similarly, document number 3 has an associated vector of $\begin{pmatrix} 0 \\ .70 \\ .30 \end{pmatrix}$.

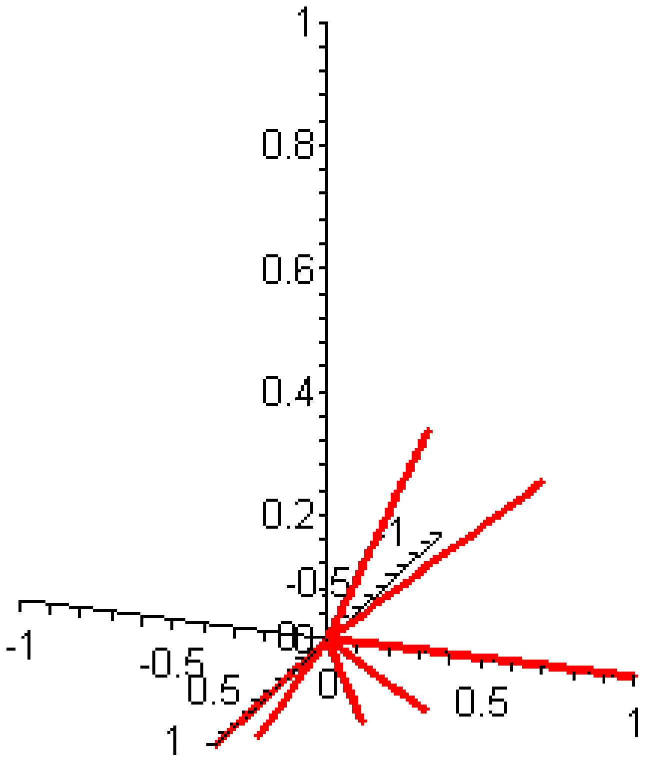
All of the document vector information can be stored together in a *matrix*.
Suppose there are only 7 documents labeled d_1, d_2, d_3, d_4, d_5, d_6, d_7, in the ***Sports*** database and only the three keywords (Cricket, Hockey, Tennis) to distinguish article topics. Then, the matrix **A** capturing the data for ***Sports*** documents is

A =

	D1	D2	D3	D4	D5	D6	D7
Cricket	1	0	0	0.20	0.65	0.50	0.90
Hockey	0	1	0.70	0.40	0.35	0.50	0.10
Tennis	0	0	0.30	0.40	0	0	0

- **A** is a 3-by-7 matrix.
- If there were, say, 20 keywords and 1000 documents, then **A** would be a 20-by-1000 matrix.
- **A** is called a ***term-by-document matrix*** because the rows correspond to terms (keywords) in the database and the columns correspond to documents.

- Since there are only three terms in our ***Sports*** example, we can plot each column (document) vector in the matrix **A** in three-dimensional space.
- A static plot of the seven document vectors is shown in Figure 1



Suppose a user of this ***Sports!*** search engine wants to find all articles related to a query of **Tennis**. Then our query vector is

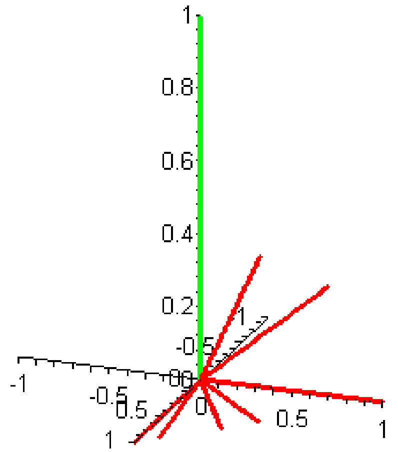
$$q = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

- **Now the problem is mathematical:** Find which documents represented by the vectors (**d_1**, **d_2**, **d_3**, **d_4**, **d_5**, **d_6**, **d_7**) are "closest" to this query vector **q**.

- You might try to inspect the 3-D plot of the document vectors and the query vector visually (Figure 2) to see which document vectors are "closest."

- This visual inspection can be difficult, since the display of a 3-D plot must be done on a 2-D screen --

- **Can you guess which document vectors are closest to the query vector? Why did you make that choice?**



Another Solution to find the articles closest to the query

- .
- The angle of the vector with the query vector tells us that document 4's vector is closest to the query vector.
- For each document in the *Sports!* database, the angle between the document vector and the query vector, to capture the physical "distance" of that document from the query.
- The document vector whose direction is closest to the query vector's direction (i.e., for which the angle is smallest) is the best choice, yielding the document most closely related to the query.

Angle between the vectors

We can compute the cosine of the angle θ between the nonzero vectors **x** and **y** by

$$\cos\theta = \frac{x^T y}{||x||_2 ||y||_2},$$

where $||\mathbf{x}||_2$ and $||\mathbf{y}||_2$ represent the Euclidean norms (lengths) of the vectors **x** and **y**, respectively. Thus, the cosine of the angle between document vector $\mathbf{d}_3$ and query vector **q** is

$$\cos\theta = \frac{d_3^T q}{||d_3||_2 ||q||_2} = \frac{.3}{.7616} = .394.$$

Cosines of the angles for each ***Sports*** document vector with respect to **query 1** in Table 1.

- If the cosine of an angle is 1, then the angle between the document vector and the query vector measures 0°, meaning the document vector and the query vector point in the same direction.
- A cosine measure of 0 means the document is unrelated to the query, i.e., the vectors are perpendicular.
- Thus, a cosine measure close to 1 means that the document is closely related to the query.

Table 1. Cosines of angles between Sports! document vectors and query vector 1

Document Vector	Cosine of angle between d(i) and q
d1	0
d2	0
d3	0.394
d4	0.667
d5	0
d6	0
D7	0

Class exercise

We now consider a second query, this time desiring articles with an equal mix of information about Cricket and Hockey so the new query vector is

$$\begin{pmatrix} .50 \\ .50 \\ 0 \end{pmatrix}.$$

Show the cosines of the angles between the document vectors and this query vector and identify the documents closest to the query

Solution

Document vector	cosine of angle between d_i and q
d_1	.707
d_2	.707
d_3	.650
d_4	.707
d_5	.958
d_6	1.000
d_7	.781

- In practice, a search engine system must designate a *cutoff value*.

- Only those documents whose angles with the query vector are above this cutoff value are reported in the list of retrieved documents viewed by the user.

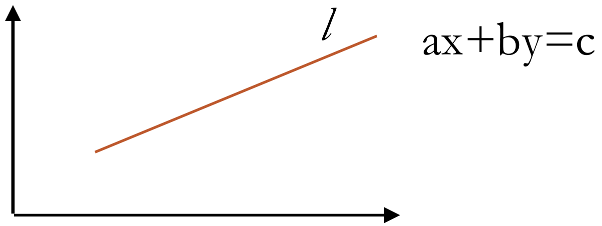
- If, for example, the cutoff value for query 2 is 0.9, then only $\mathbf{d}_5$ and $\mathbf{d}_6$ are returned to the user as relevant (see Table 2 again). If, however, a cutoff value of 0.7 is selected, all but $\mathbf{d}_3$ will be returned as relevant. Thus, choosing an appropriate cutoff value is not a trivial task and clearly affects both the precision and recall of the search engine.

System of Linear equation of two variables

A **linear equation** is a polynomial equation in which the unknown variables have a degree of one.

Example : $3x+2y=9$

Geometrically, a linear equation with two variable (i.e.,$ax+by=c$) is represents a line. A linear equation with n variables represents a hyperplane.

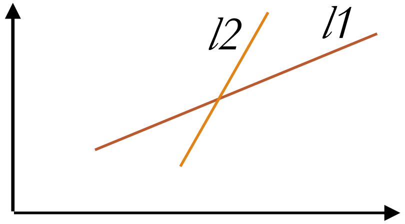


A system of linear equations is a set of **two or more linear equations** with the same variables.

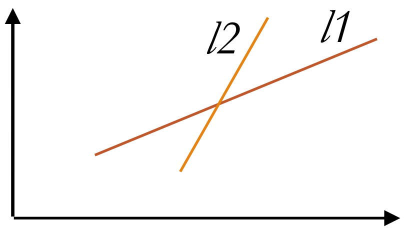
Example:

$$a_1x_1+b_1x_2=c_1$$

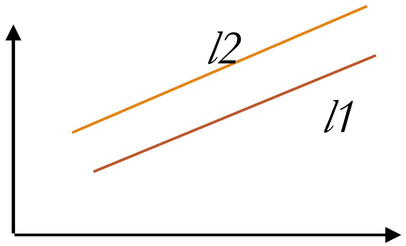
$$a_2x_2+b_2x_2=c_2$$



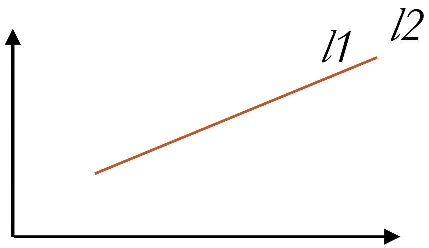
Solution to System of linear equation



If two lines interest at one common point,
then there is a unique solution



If no common point,
then there is no solution



If the graph of two lines coincides, then infinite number of solutions

System of linear equations of n-variables

A linear equation of n-variable is of the form

$a_1x_1+a_2x_2+\ldots+a_nx_n=b$, where $x_1,x_2,..x_n$ are variables(unknowns) and $a_1,a_2\ldots a_n$ are coefficients ϵ R. Geometrically, it denotes a hyperplane.

A **system of linear equations** (or linear system) is a finite collection of linear equations in same variables. Each equation is a hyperplane and if they have a common point, then there will be a solution. For instance, a linear system of m equations in n variables $x_1, x_2, \ldots, x_n$ can be written as

Example :

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots a_{1nxn} = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + am_2x_2 + \cdots amnxn = bm \end{cases}$$

Note : Any system of linear equations has one of the following exclusive conclusions.

(a) No solution.
(b) Unique solution.
(c) Infinitely many solutions

A linear system is said to be **consistent** if it has **at least one solution**; and is said to be **inconsistent** if it has **no solution**.

Representation of system of linear equations in matrix form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots a_{1nxn} = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n = b_2 \\ \quad \cdots \\ a_{m1}x_1 + am_2x_2 + \cdots amnxn = bm \end{cases}$$

Can be represented as augmented matrix

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

$where \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ is the coefficient matrix of linear system

Example: Represent the following linear system of equations in matrix form

$$\begin{cases} x_1 + x_2 - 2x_3 = 1 \\ 2x_1 - 3x_2 + x_3 = -8 \\ 3x_1 + x_2 + 4x_3 = 7 \end{cases}$$

$$\left[\begin{array}{ccc|c} 1 & 1 & -2 & 1 \\ 2 & -3 & 1 & -8 \\ 3 & 1 & 4 & 7 \end{array}\right]$$

Solving Linear Equations

Consider the linear system

$$3x + 4y = 5$$

$$2x - y = 0.$$

A system of linear equations can be compactly represented in their matrix form as **Ax = b;** where A is the coefficient matrix of size m x n .The **m and n** are **number of equations** and **variables** respectively , b denotes the value in the R.H.S

For example, compact form of above linear equation is:

$$\begin{bmatrix} 3 & 4 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

Every linear system may have only one of three possible number of solutions:

1.The system has a ***single unique*** *solution.*

2.The system has ***infinitely many*** *solutions.*

3.The system has ***no solution***

m=n	If number of equations is equal to number of variables, then unique solutions
m>n	No solution exists
m<n	More than one solution exists

Solving a System of Linear Equations Using Matrices

Step 1 : Represent the linear system of equations in augmented matrix form .To create the augmented matrix, add the constant matrix as the last column of the coefficient matrix
For example

$$\begin{cases} 3x - 2y + z = 5 \\ x + 3y - z = 0 \\ -x + \quad 4z = 11 \end{cases}$$

Augmented matrix

$$\left[\begin{array}{ccc|c} 3 & -2 & 1 & 5 \\ 1 & 3 & -1 & 0 \\ -1 & 0 & 4 & 11 \end{array}\right]$$

Symbol	Description
$\mathbf{R}_i + \propto \mathbf{R}_j \rightarrow \mathbf{R}_i$	**Change the *i*th row by adding ∝times row *j* to it. Then, put the result back in row *i*.**
$\propto \mathbf{R}_i$	**Multiply the *i*th row by ∝.**
$\mathbf{R}_i \leftrightarrow \mathbf{R}_j$	**Interchange the *i*th and *j*th rows.**

Step 2 : Reduce into Row-Echelon form **by applying elementary row operations. Three basic row operations are :**

1. interchange two rows (swapping)
2. Multiply one of the rows by a nonzero constant
3. Add a multiple of one row to another row.

Step 3: the matrix in row-echelon form is solved by back-substitution. (**Gauss elimination method**)

Row Echelon Form (REF)

A matrix is in **row echelon form** (ref) when it satisfies the following conditions.

- The first non-zero element in each row, called the **leading entry**, is 1.
- Each leading entry is in a column to the right of the leading entry in the previous row.
- Rows with all zero elements, if any, are below rows having a non-zero element.

Example:

$$A=\begin{bmatrix}1 & 2 & 3 & 4\\ 0 & 0 & 1 & 3\\ 0 & 0 & 0 & 1\end{bmatrix} \qquad B=\begin{bmatrix}1 & 2 & 3 & 4\\ 0 & 0 & 1 & 3\\ 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0\end{bmatrix} \qquad C=\begin{bmatrix}1 & 3\\ 0 & 1\\ 0 & 0\end{bmatrix}$$

Example : Solution of Linear equations using Gauss elimination method

$$\begin{cases} 4x + 8y - 4z = 4 \\ 3x + 8y + 5z = -11 \\ -2x + y + 12z = -17 \end{cases}$$

Step 1: Construct augmented matrix

$$\left[\begin{array}{ccc|c} 4 & 8 & -4 & 4 \\ 3 & 8 & 5 & -11 \\ -2 & 1 & 12 & -17 \end{array}\right]$$

Step 2: convert into Row-echelon form (REF form)

$$\xrightarrow{\frac{1}{4}(R1)} \left[\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 3 & 8 & 5 & -11 \\ -2 & 1 & 12 & -17 \end{array}\right]$$

$$\xrightarrow[R_3+2R_1 \to R_3]{R_2-3R_1 \to R_2} \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 2 & 8 & -14 \\ 0 & 5 & 10 & -15 \end{bmatrix}$$

$$\xrightarrow{\frac{1}{2}R_2} \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 1 & 4 & -7 \\ 0 & 5 & 10 & -15 \end{bmatrix}$$

$$\xrightarrow{R_3-5R_2 \to R_3} \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 1 & 4 & -7 \\ 0 & 0 & -10 & 20 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 1 & 4 & -7 \\ 0 & 0 & -10 & 20 \end{bmatrix} \xrightarrow{-\frac{1}{10}R_3} \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 1 & 4 & -7 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

The matrix is now in row-echelon form

Solve using back-substitution. The last matrix represents the following equations

z=-2
Y+4z= -7
X+2y-z=1

Using back-substitution, we obtain the solution as **(-3,1,-2)**

A linear equation may have one of the following possibilities

One solution
No solution
Infinitely many solutions

Case 1 : One Solution -if each variable in the row-echelon form is a leading variable, the system has exactly one solution. Then find the solution by using back-substitution or other methods

$$\begin{bmatrix} 1 & 6 & -1 & 3 \\ 0 & 1 & 2 & -2 \\ 0 & 0 & 1 & 8 \end{bmatrix}$$

Each variable is a leading variable.

Case 2 : No Solution : If the row-echelon form contains row that represents the equation **0 =** $\boldsymbol{c}$**,** where $\boldsymbol{c}$ **is not zero**, the system has no solution.

A system with no solution is called **inconsistent.**

$$\underbrace{\begin{bmatrix} 1 & 2 & 5 & 7 \\ 0 & 1 & 3 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{Last equation says } 0 = 1.}$$

Case 3: Infinitely many solution - If the variables in the row-echelon form are not all leading variables, and if the system is not inconsistent, it has infinitely many solutions.

The system is called dependent.

$$\begin{bmatrix} 1 & 2 & -3 & 1 \\ 0 & 1 & 5 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

z is not a leading variable.

Exercise

1. Determine whether each matrix is in row-echelon form. If it is, determine whether the matrix is in reduced row-echelon form.

a. $$\begin{bmatrix} 1 & 2 & -1 & 4 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

b. $$\begin{bmatrix} 1 & 2 & -1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & -4 \end{bmatrix}$$

c. $$\begin{bmatrix} 1 & -5 & 2 & -1 & 3 \\ 0 & 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

d. $$\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Exercises

2. Write the augmented matrix for the system of linear equations.

$$x + 3y = 9$$
$$-y + 4z = -2$$
$$x - 5z = 0$$

What is the dimension of the augmented matrix?

3. Solve the following system of linear equations using matrices.

$$-x-2y+z=-12$$

$$x+3y=2$$

$$y-2z=0$$