ES 691
Mathematics for Machine Learning

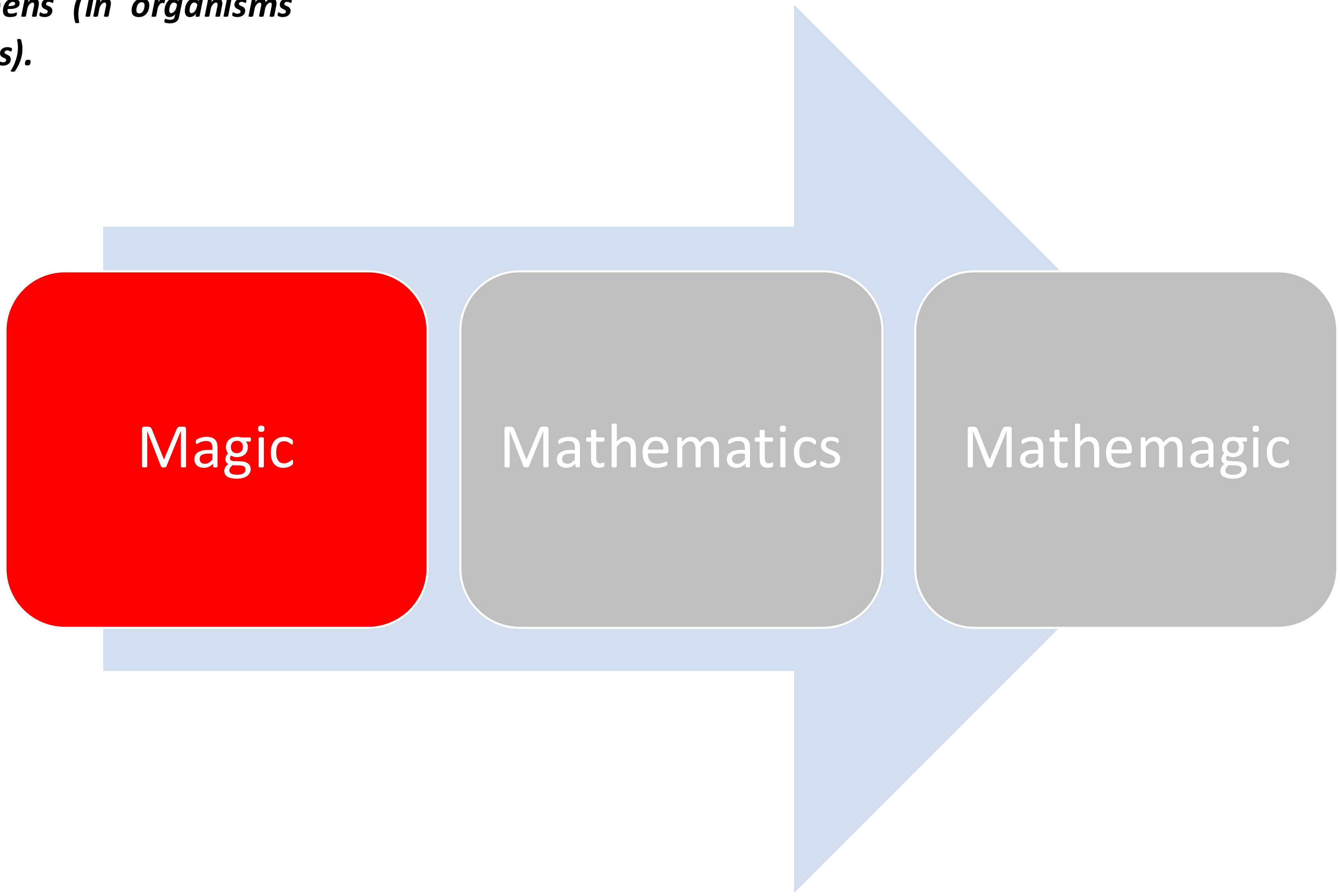
with

Dr. Naveed R. Butt

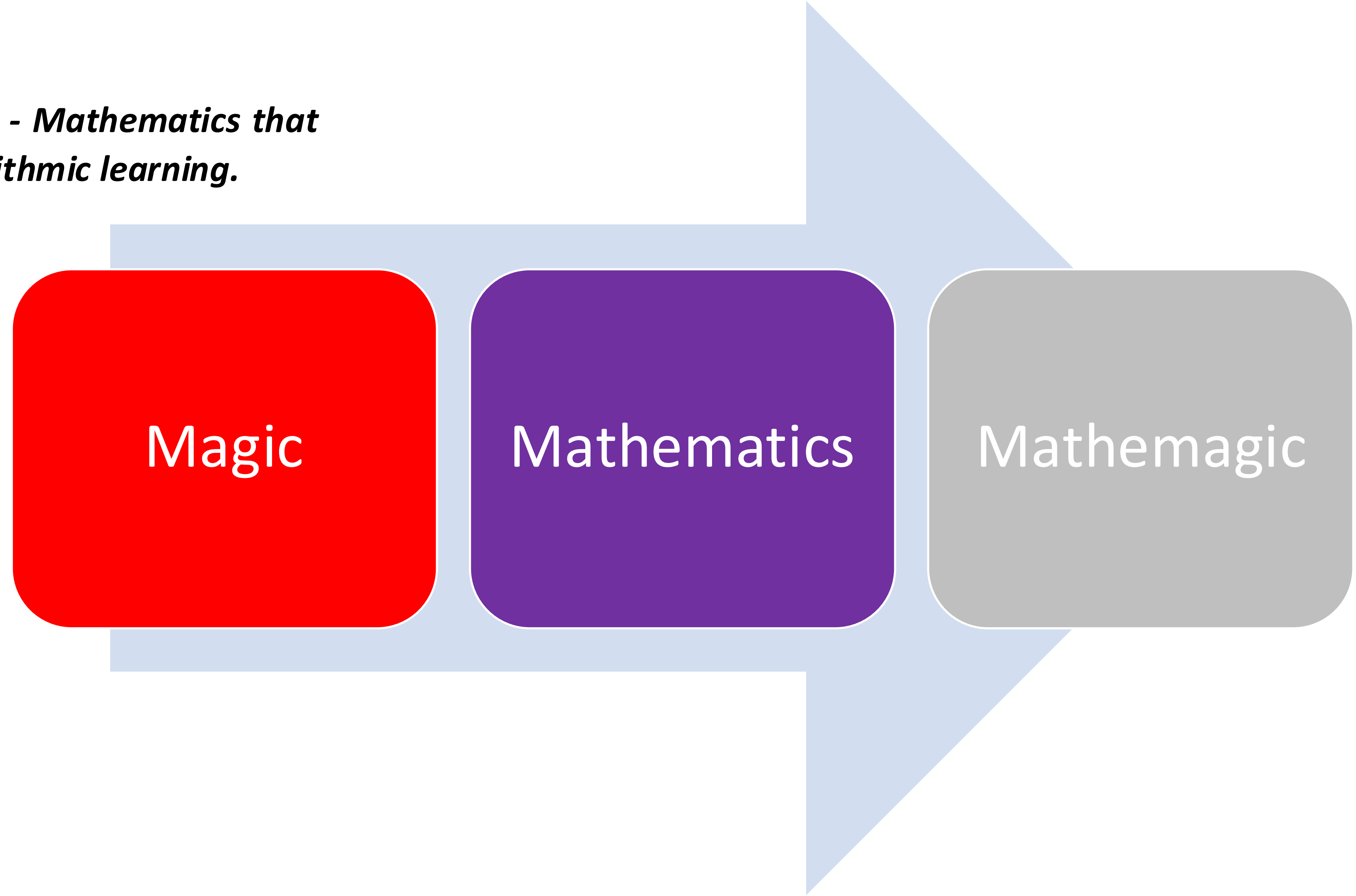
@

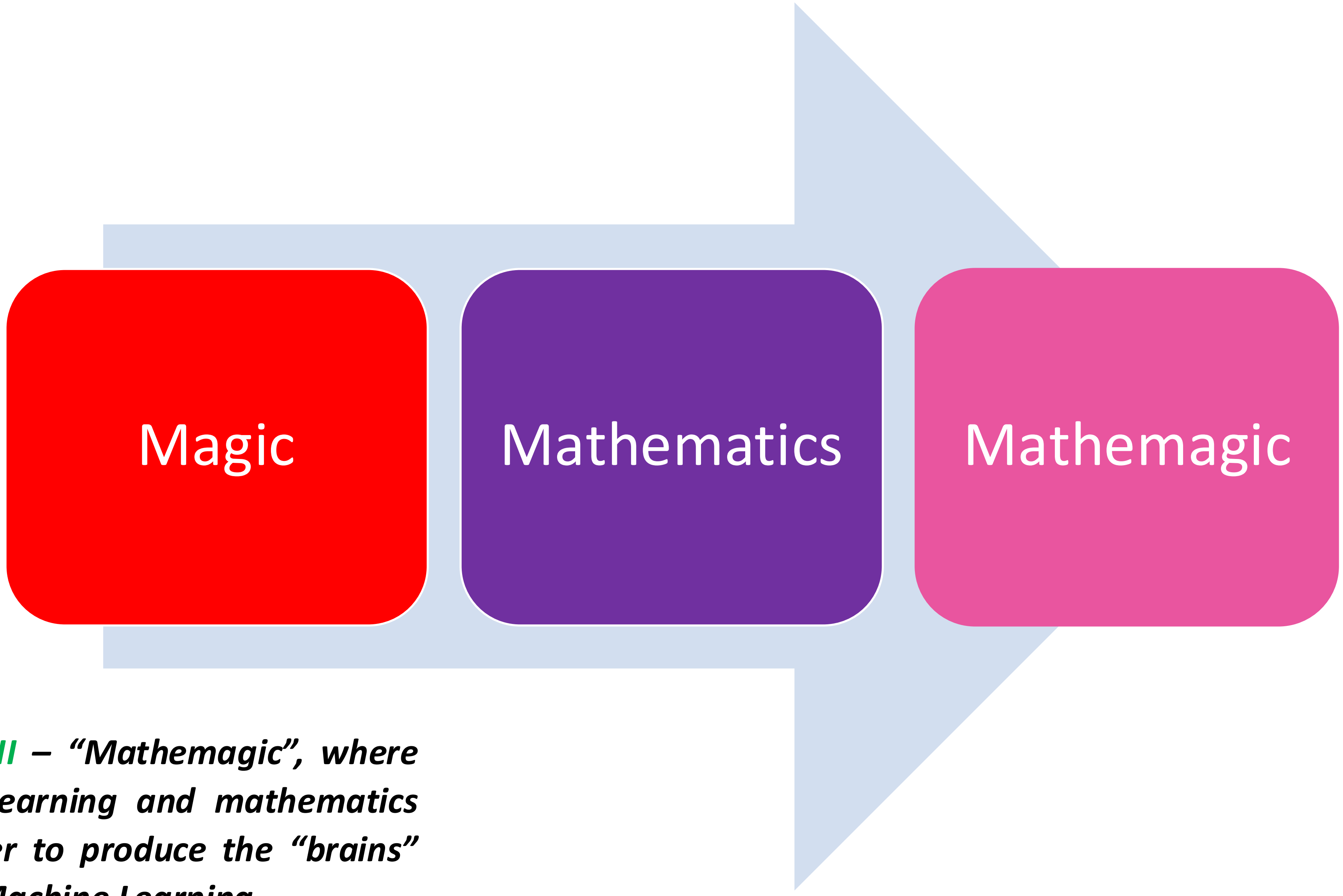
GIKI - FES

Course Part I* - *Magic of how learning happens (in organisms and algorithms).



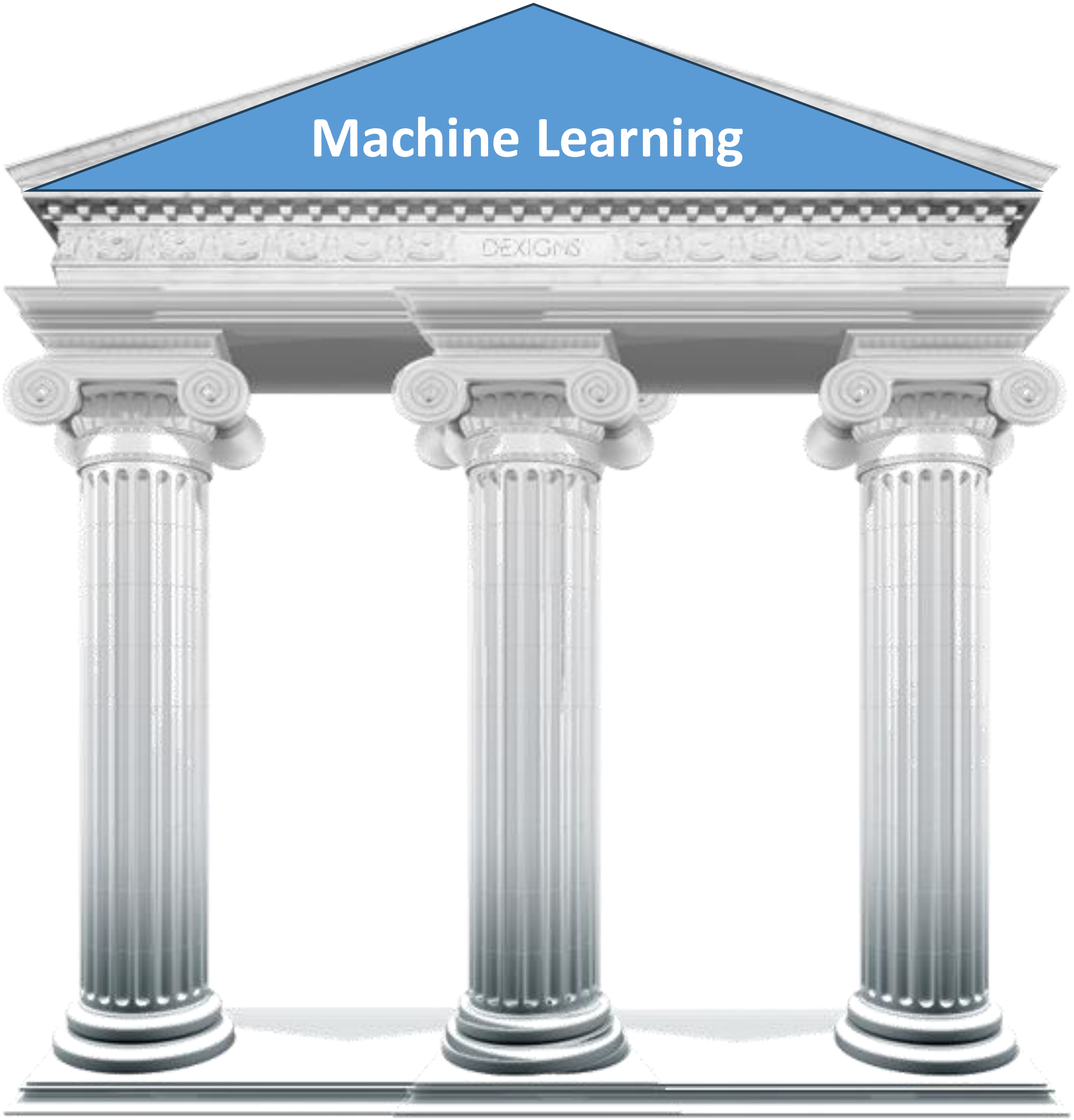
***Course Part II** - Mathematics that enables algorithmic learning.*



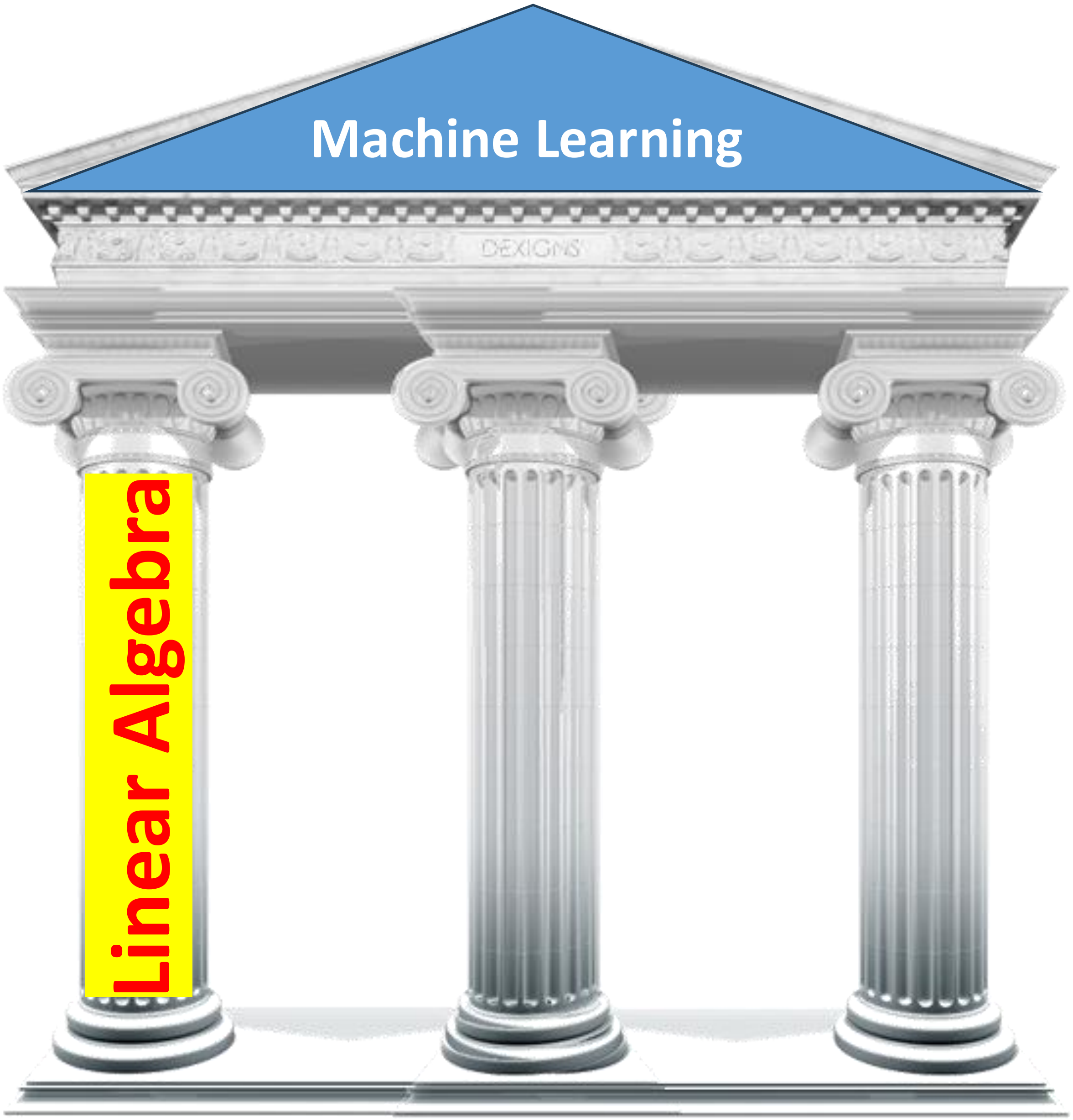


***Course Part III* – "Mathemagic", where concepts of learning and mathematics come together to produce the "brains" of AI, called Machine Learning.**

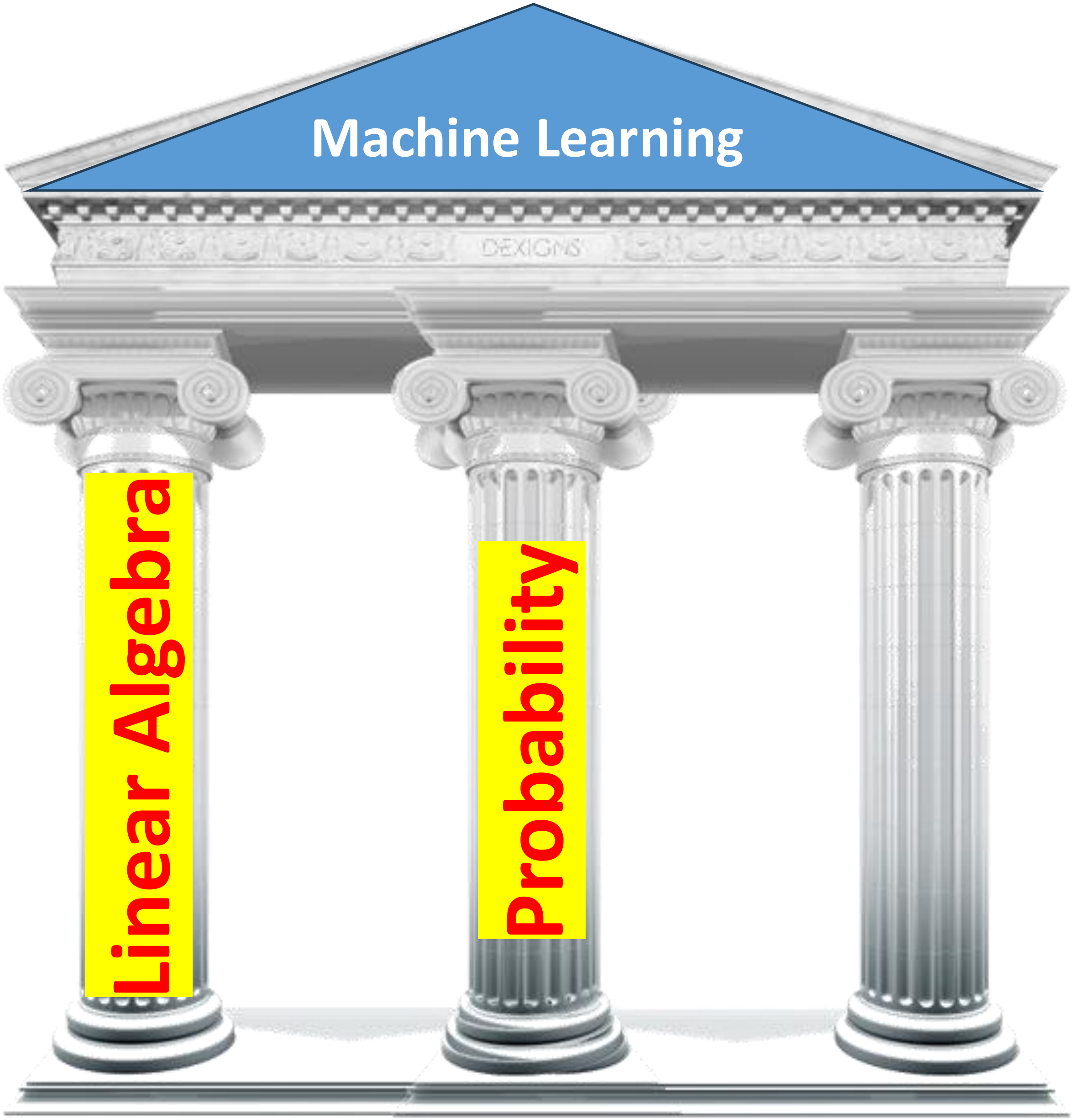
Broadly Speaking…



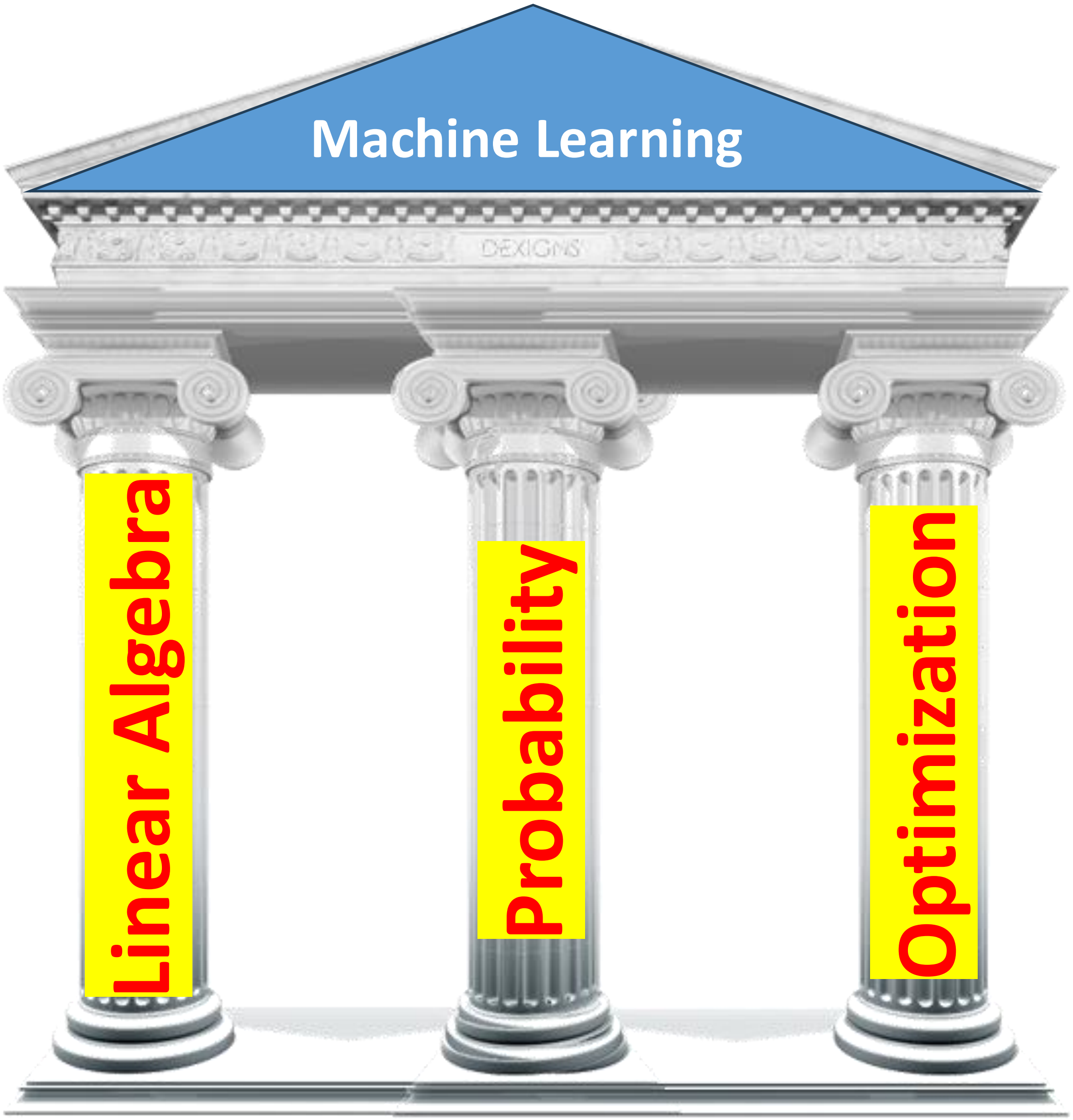
Broadly Speaking…



Broadly Speaking…



Broadly Speaking…

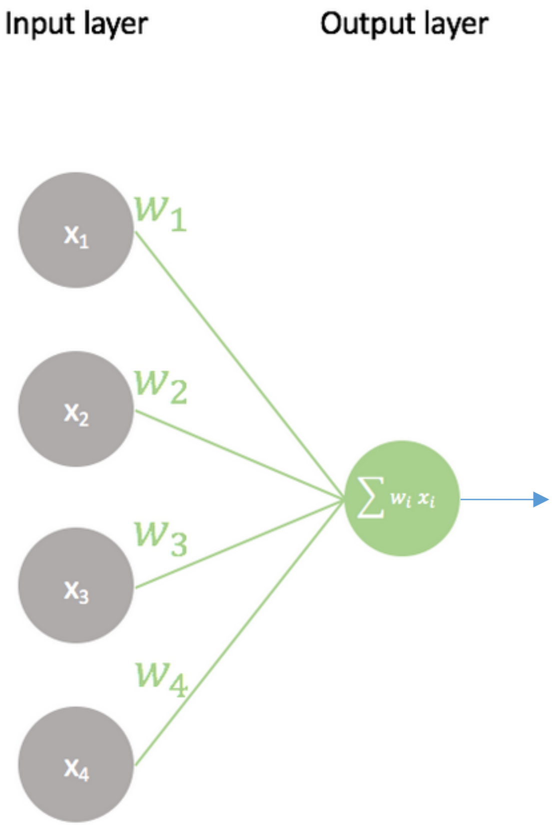


How?

A *Very Very* Brief Reply…

A lot of ML Deals With “Linear Combinations”.

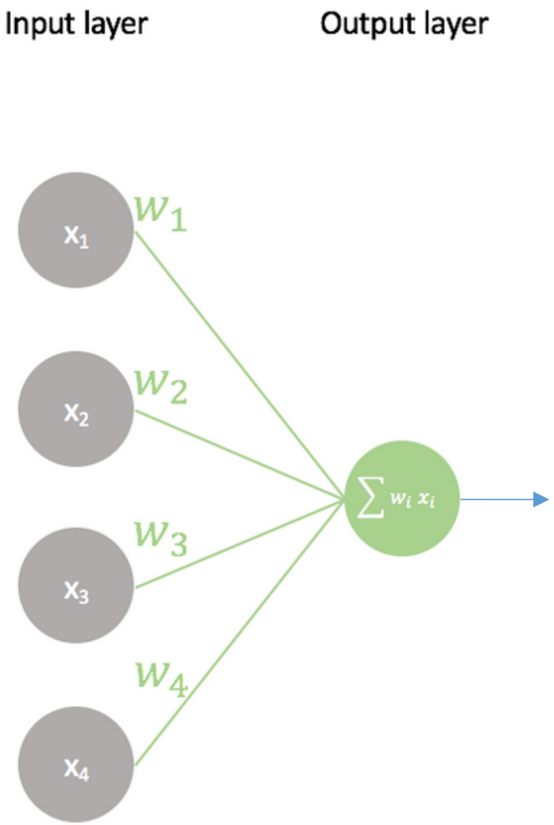
A lot of ML Deals With "Linear Combinations".



Perceptron Unit

$$f_w(x) = \begin{Bmatrix} \sum w_i x_i \geq \theta \rightarrow \text{neuron fires} \\ \sum w_i x_i < \theta \rightarrow \text{neuron doesn't fire} \end{Bmatrix}$$

A lot of ML Deals With "Linear Combinations".



Perceptron Unit

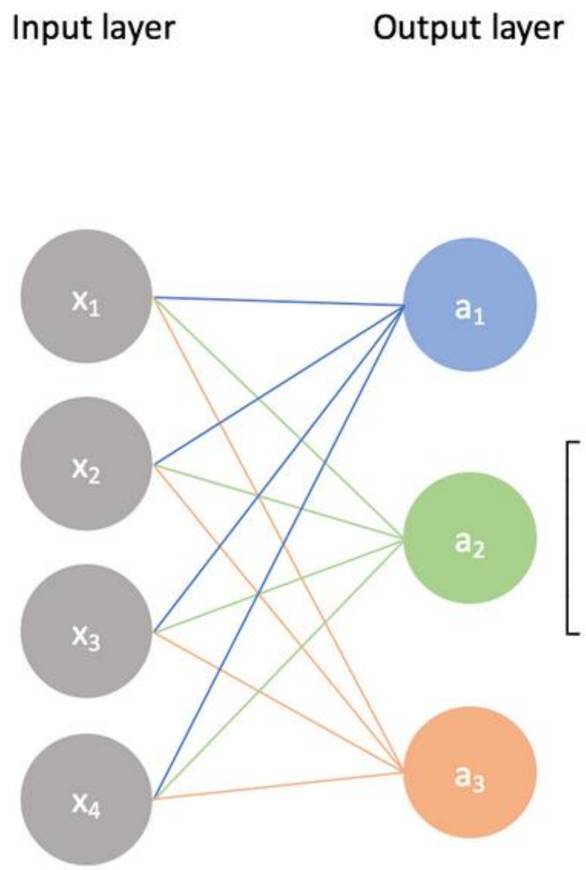
$$f_w(x) = \begin{cases} \sum w_i x_i \geq \theta \rightarrow \text{neuron fires} \\ \sum w_i x_i < \theta \rightarrow \text{neuron doesn't fire} \end{cases}$$

$\boldsymbol{w}^T \boldsymbol{x}$

Can be written more succinctly as an inner product of two vectors.

"Systems of Linear Equations" Naturally Show Up…

“Systems of Linear Equations” Naturally Show Up…

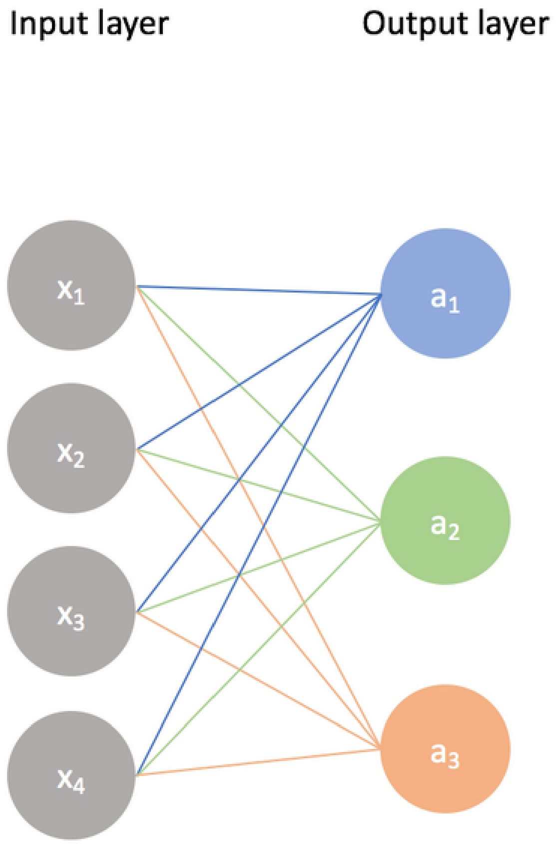


A simple neural network

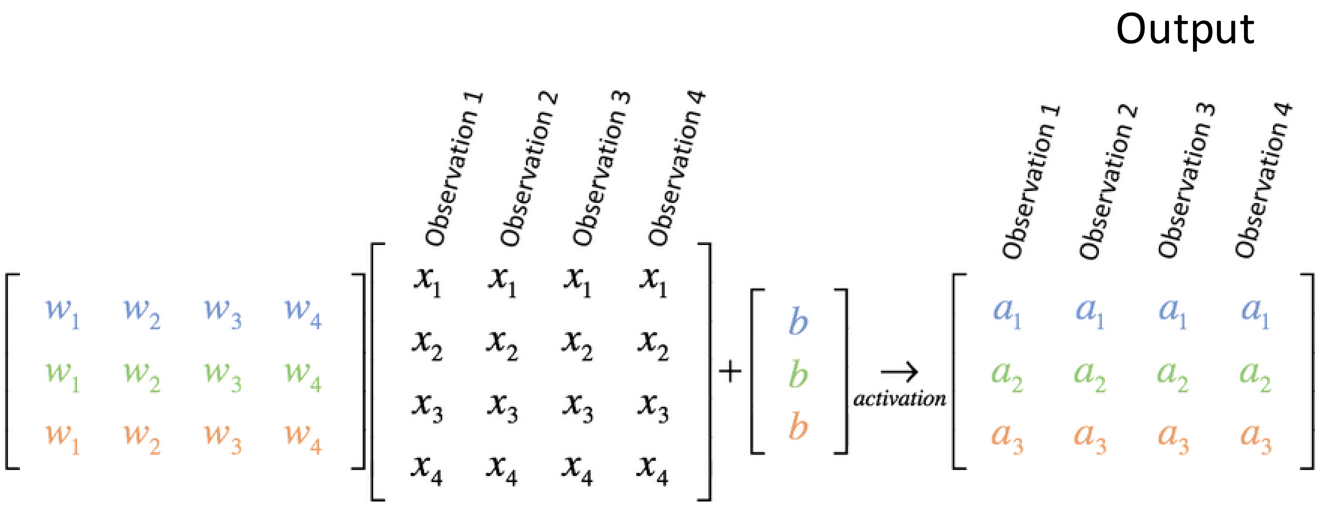
$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b \\ b \\ b \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \end{bmatrix} \xrightarrow[\text{activation}]{} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Nice Illustrations by Jeremy Jordan.

"Systems of Linear Equations" Naturally Show Up…



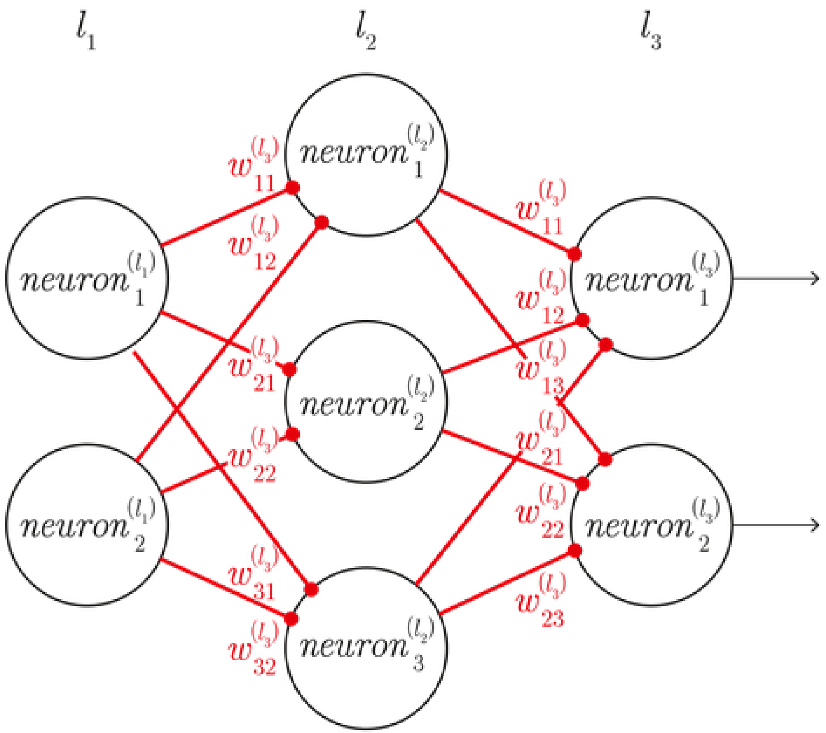
Using multiple observations



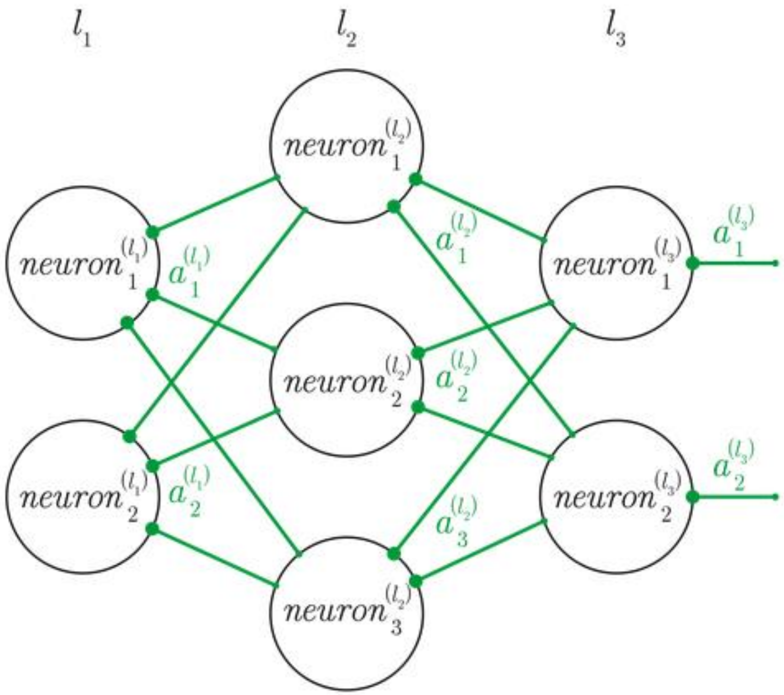
Nice Illustrations by Jeremy Jordan.

Lots of Parameters to Handle…

Lots of Parameters to Handle…

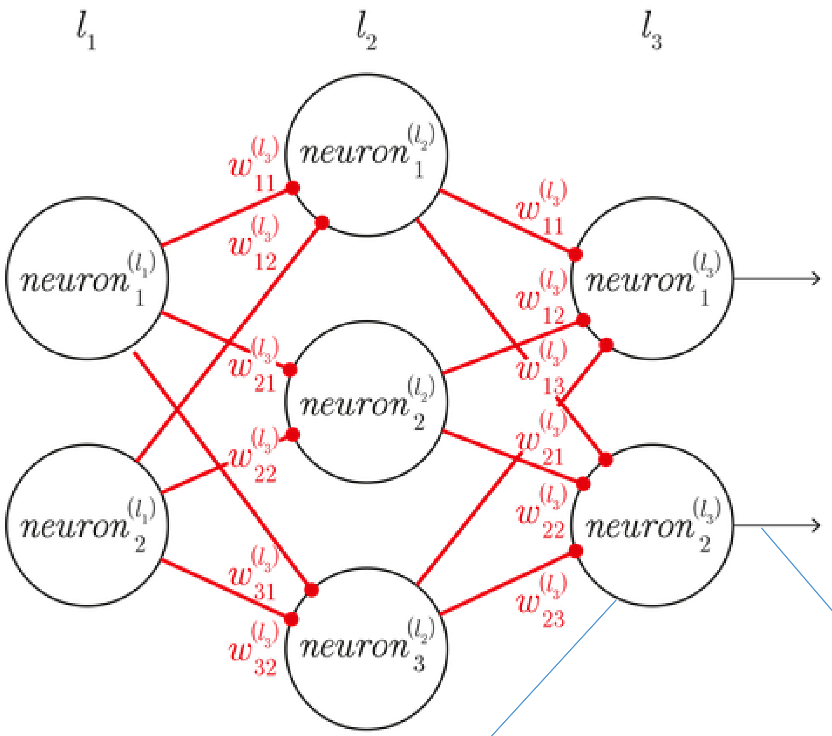


Activations

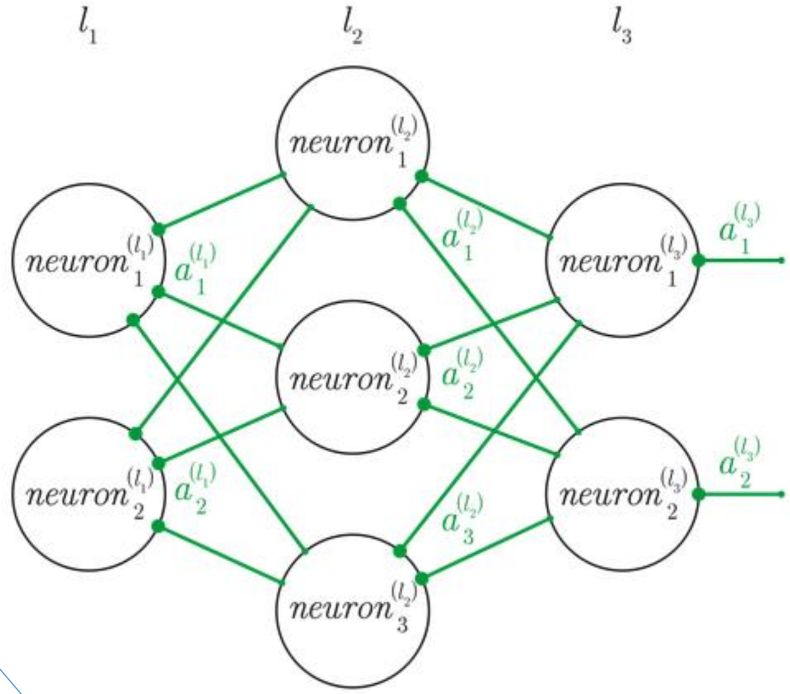


Nice Illustrations by Nikola Marincic.

Lots of Parameters to Handle…



Activations

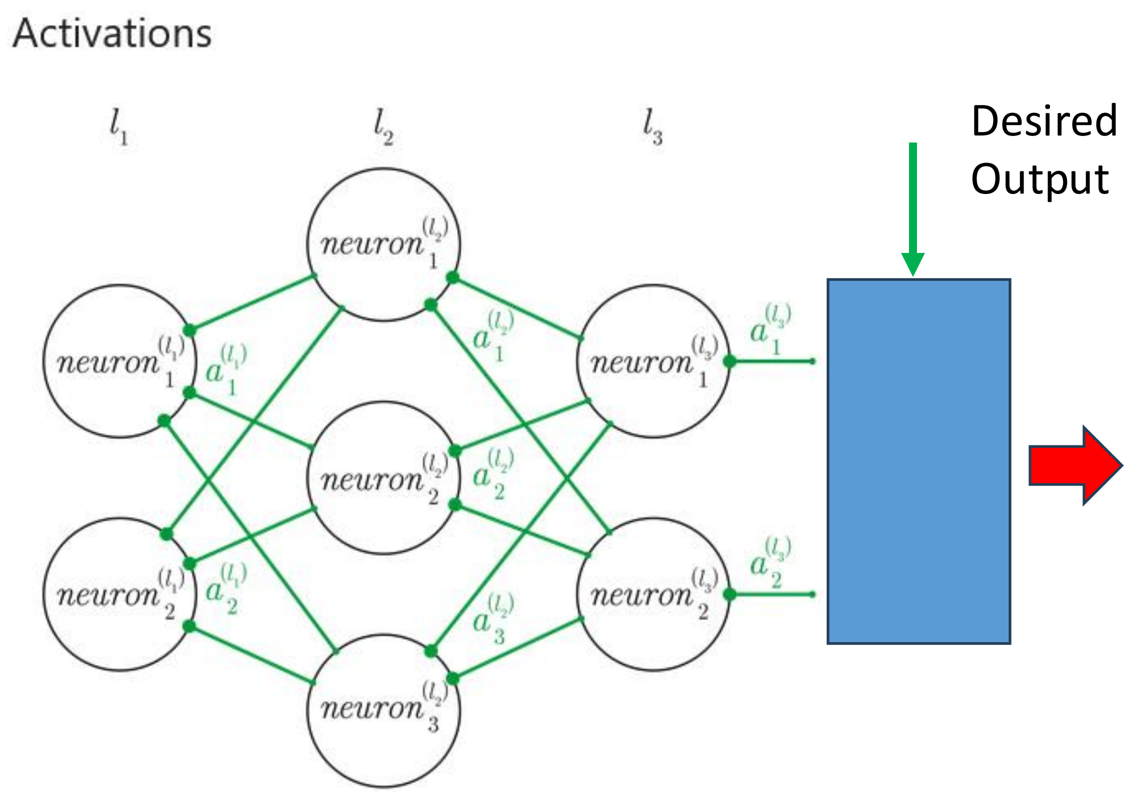


$$z^{(l_3)} = \underset{\textbf{2x3}}{\begin{bmatrix} w_{11}^{(l_3)} & w_{12}^{(l_3)} & w_{13}^{(l_3)} \\ w_{21}^{(l_3)} & w_{22}^{(l_3)} & w_{23}^{(l_3)} \end{bmatrix}} \times \underset{\textbf{3x1}}{\begin{bmatrix} a_1^{(l_2)} \\ a_2^{(l_2)} \\ a_3^{(l_2)} \end{bmatrix}} + \underset{\textbf{2x1}}{\begin{bmatrix} b_1^{(l_3)} \\ b_2^{(l_3)} \end{bmatrix}}$$

$$a^{(l_3)} \equiv \begin{bmatrix} a_1^{(l_3)} \\ a_2^{(l_3)} \end{bmatrix} = \sigma(z^{(l_3)}) \equiv \begin{bmatrix} \sigma(z_1^{(l_3)}) \\ \sigma(z_2^{(l_3)}) \end{bmatrix}$$

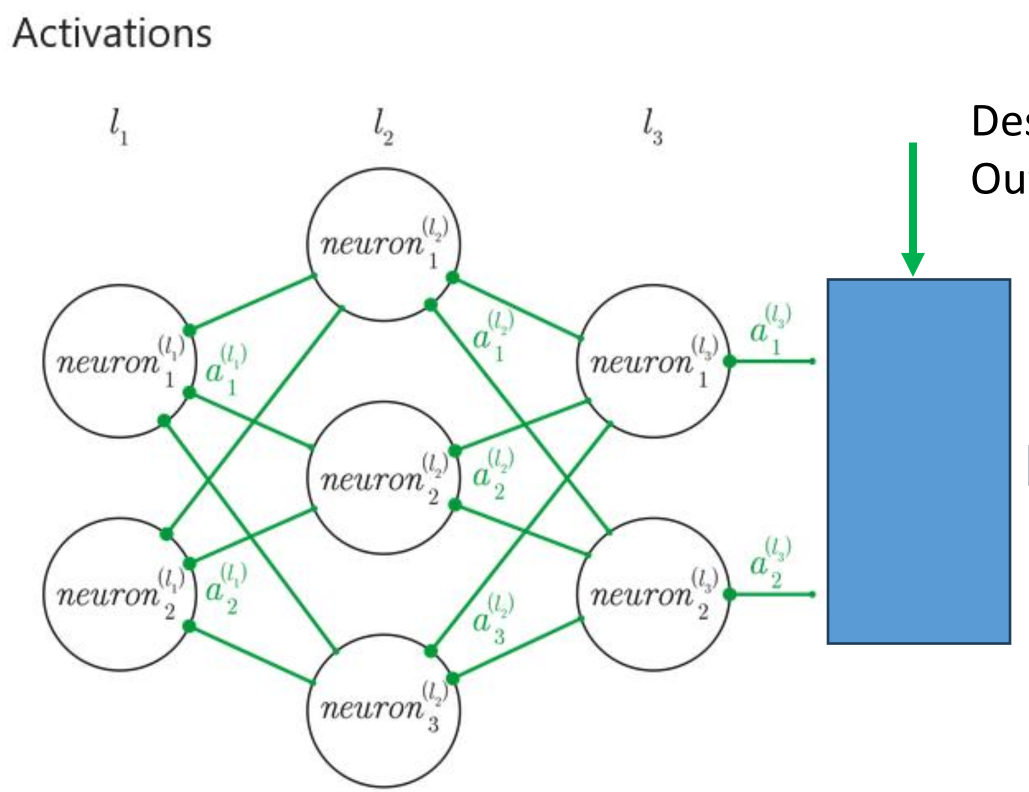
Cost Functions Will Often be Defined on Vectors…

Cost Functions Will Often be Defined on Vectors…



$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

Cost Functions Will Often be Defined on Vectors…



Desired Output

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

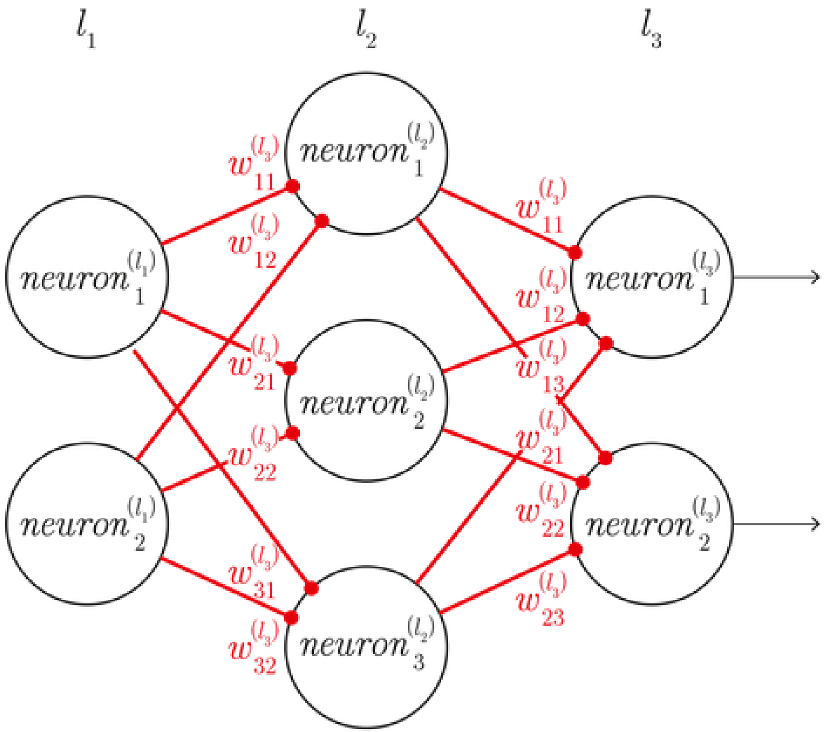
$$C_x = \frac{1}{2} \| y - a^{(l_3)} \|^2$$

Vectors

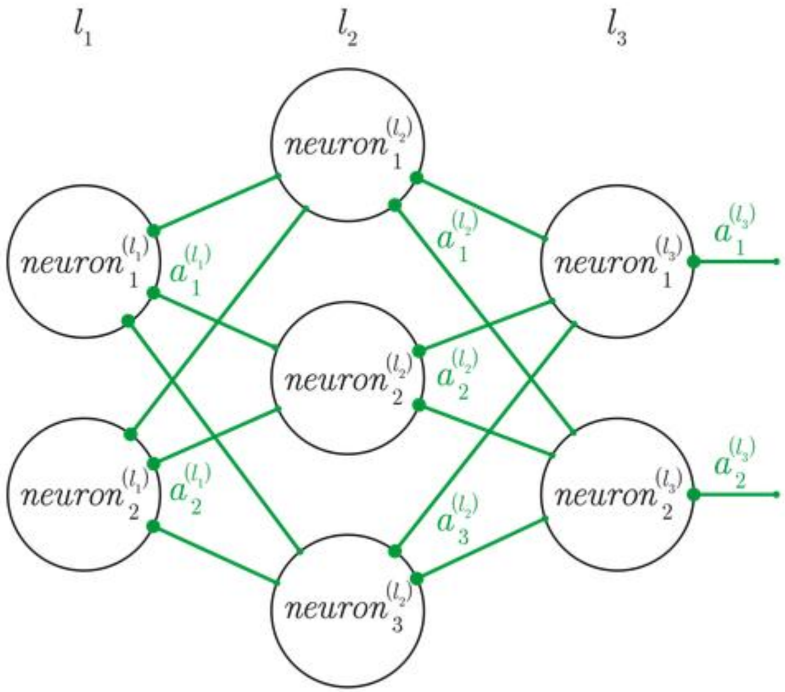
Optimization of Parameters w.r.t Cost Function Will be Required

Optimization of Parameters w.r.t Cost Function Will be Required

$$\frac{\partial C}{\partial a^{(l_3)}} = (a^{(l_3)} - y)$$



Activations

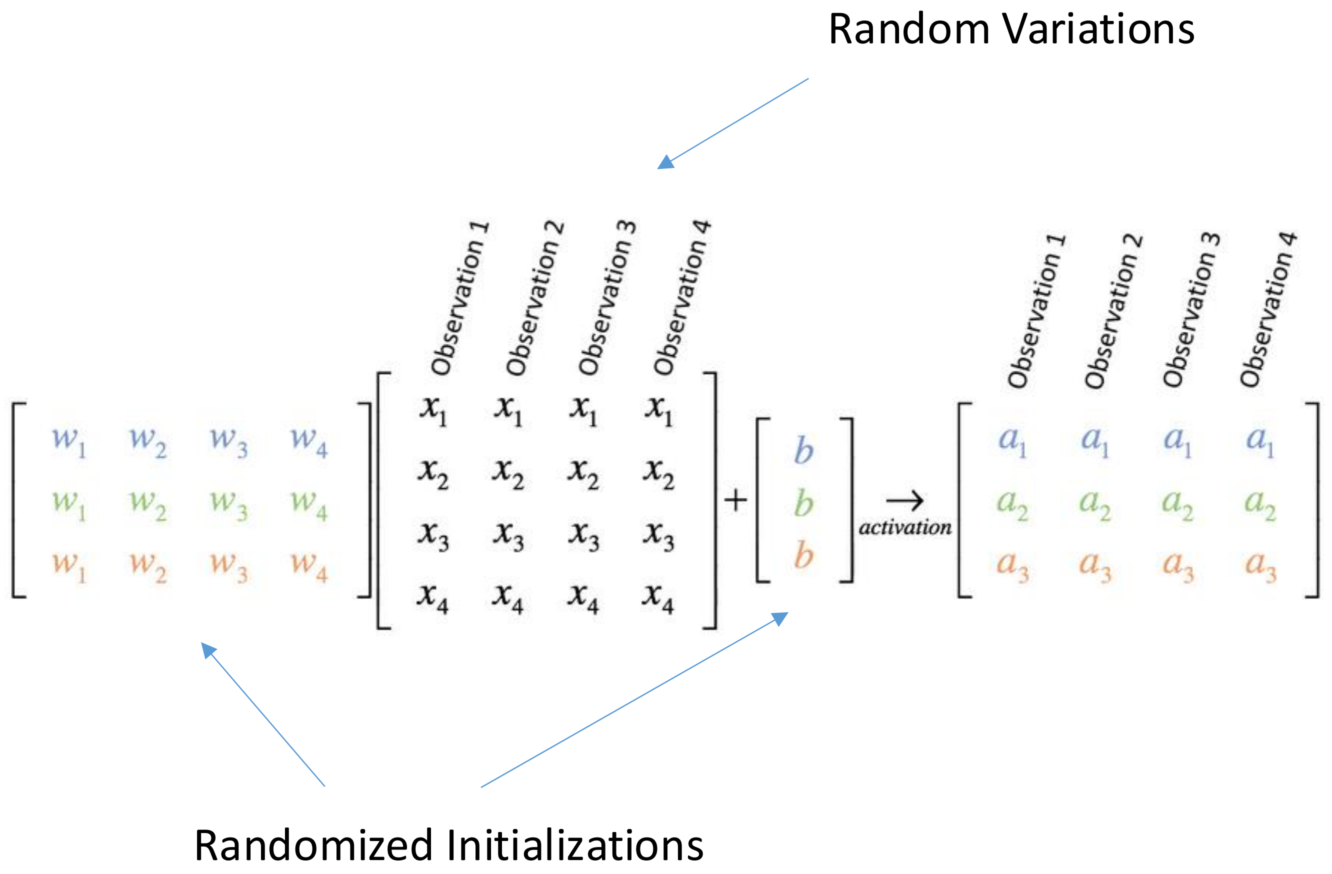


Probabilities Will be Needed to Handle/Analyze Randomness in Data, Weight Initializations, and Model Behavior…

Probabilities Will be Needed to Handle/Analyze Randomness in Data, Weight Initializations, and Model Behavior…

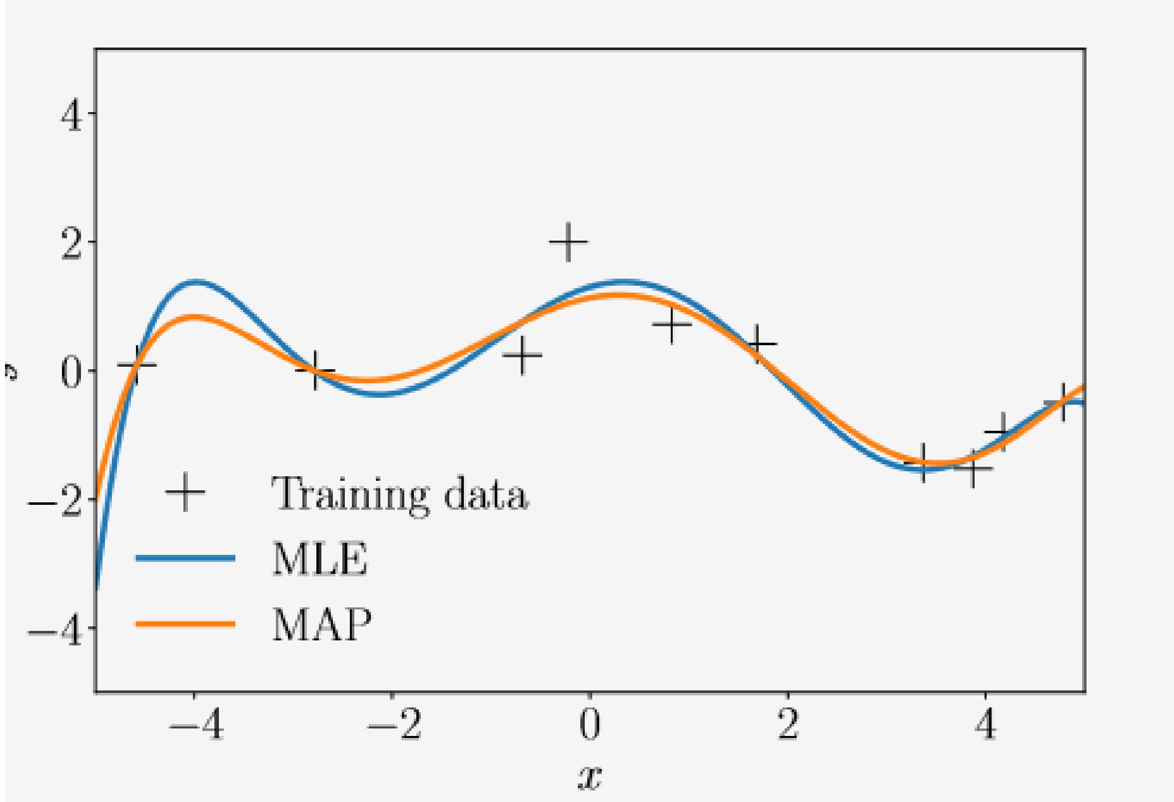
$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \end{bmatrix} \begin{array}{c} \begin{array}{cccc} \text{Observation 1} & \text{Observation 2} & \text{Observation 3} & \text{Observation 4} \end{array} \\ \begin{bmatrix} x_1 & x_1 & x_1 & x_1 \\ x_2 & x_2 & x_2 & x_2 \\ x_3 & x_3 & x_3 & x_3 \\ x_4 & x_4 & x_4 & x_4 \end{bmatrix} \end{array} + \begin{bmatrix} b \\ b \\ b \end{bmatrix} \xrightarrow[\text{activation}]{} \begin{array}{c} \begin{array}{cccc} \text{Observation 1} & \text{Observation 2} & \text{Observation 3} & \text{Observation 4} \end{array} \\ \begin{bmatrix} a_1 & a_1 & a_1 & a_1 \\ a_2 & a_2 & a_2 & a_2 \\ a_3 & a_3 & a_3 & a_3 \end{bmatrix} \end{array}$$

Probabilities Will be Needed to Handle/Analyze Randomness in Data, Weight Initializations, and Model Behavior…



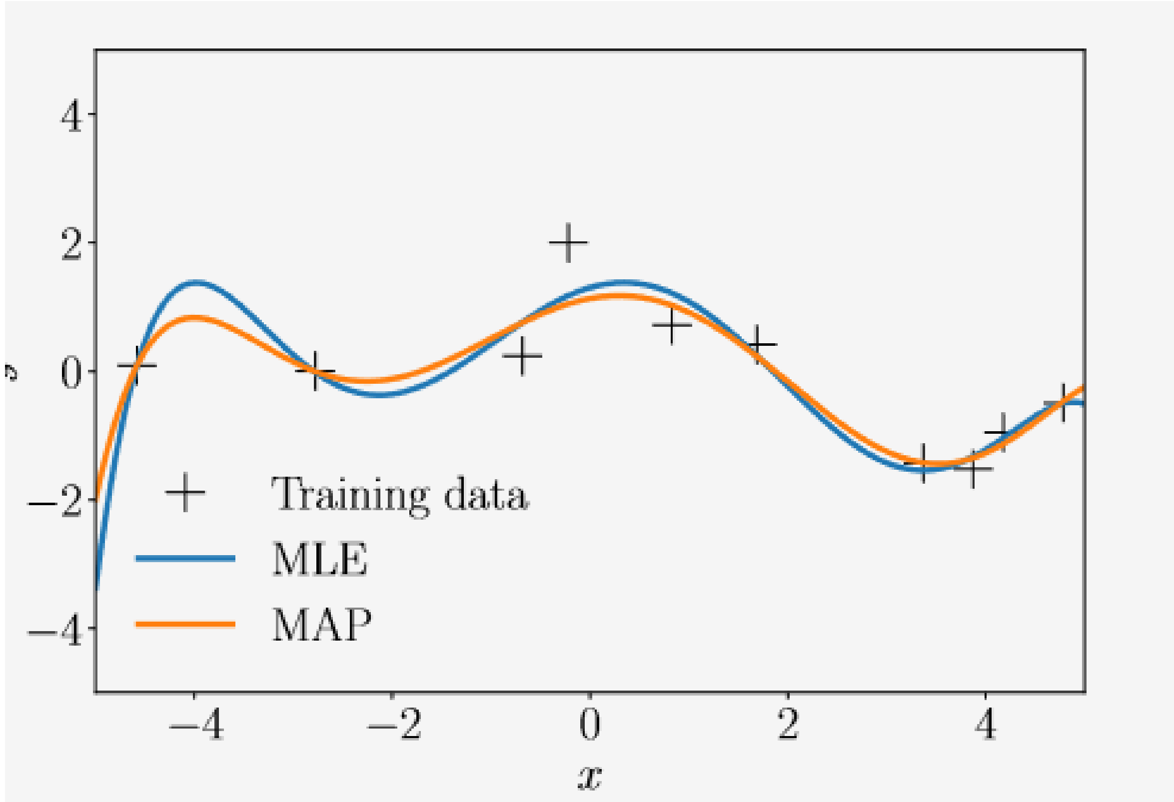
We Will Also Adopt Probabilistic Formulations in Various Contexts to Handle Uncertainties

We Will Also Adopt Probabilistic Formulations in Various Contexts to Handle Uncertainties



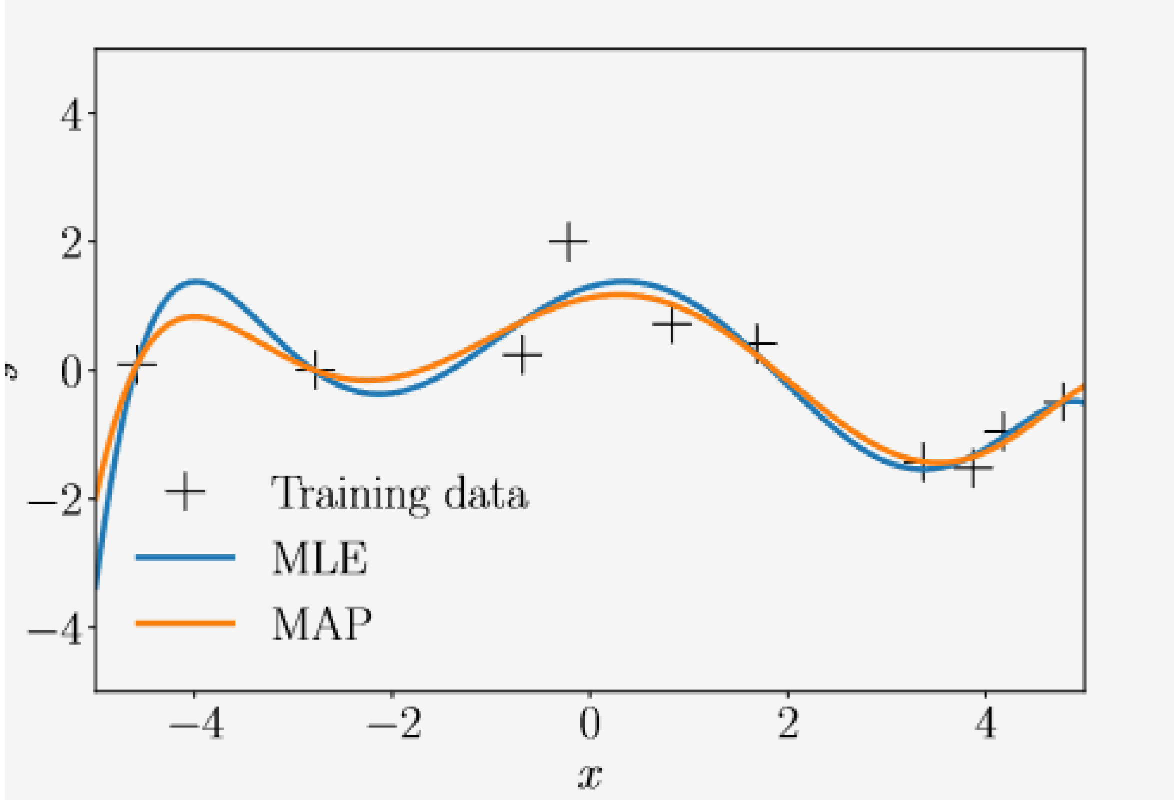
We Will Also Adopt Probabilistic Formulations in Various Contexts to Handle Uncertainties

$$\boldsymbol{\theta}_{\mathrm{ML}} \in \arg\max_{\boldsymbol{\theta}} p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta})$$



We Will Also Adopt Probabilistic Formulations in Various Contexts to Handle Uncertainties

$$\boldsymbol{\theta}_{\mathrm{ML}} \in \arg\max_{\boldsymbol{\theta}} p(\mathcal{Y} \,|\, \mathcal{X}, \boldsymbol{\theta})$$



Which will often again lead us to *Linear Algebra* *and* *Optimization*

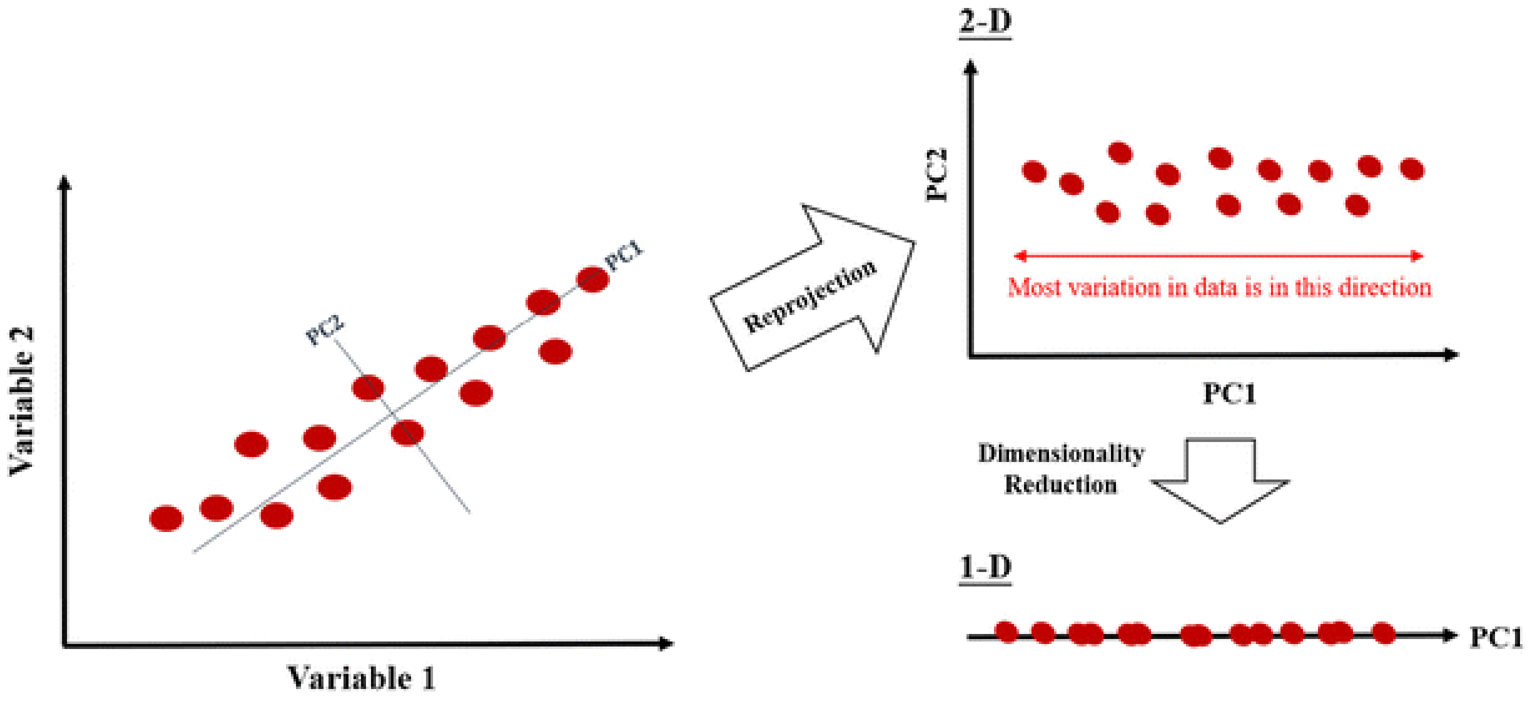
$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\boldsymbol{\theta}} = \mathbf{0}^\top \overset{(9.11c)}{\iff} \boldsymbol{\theta}_{\mathrm{ML}}^\top \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{y}^\top \boldsymbol{X}$$

$$\iff \boldsymbol{\theta}_{\mathrm{ML}}^\top = \boldsymbol{y}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$

$$\iff \boldsymbol{\theta}_{\mathrm{ML}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} \,.$$

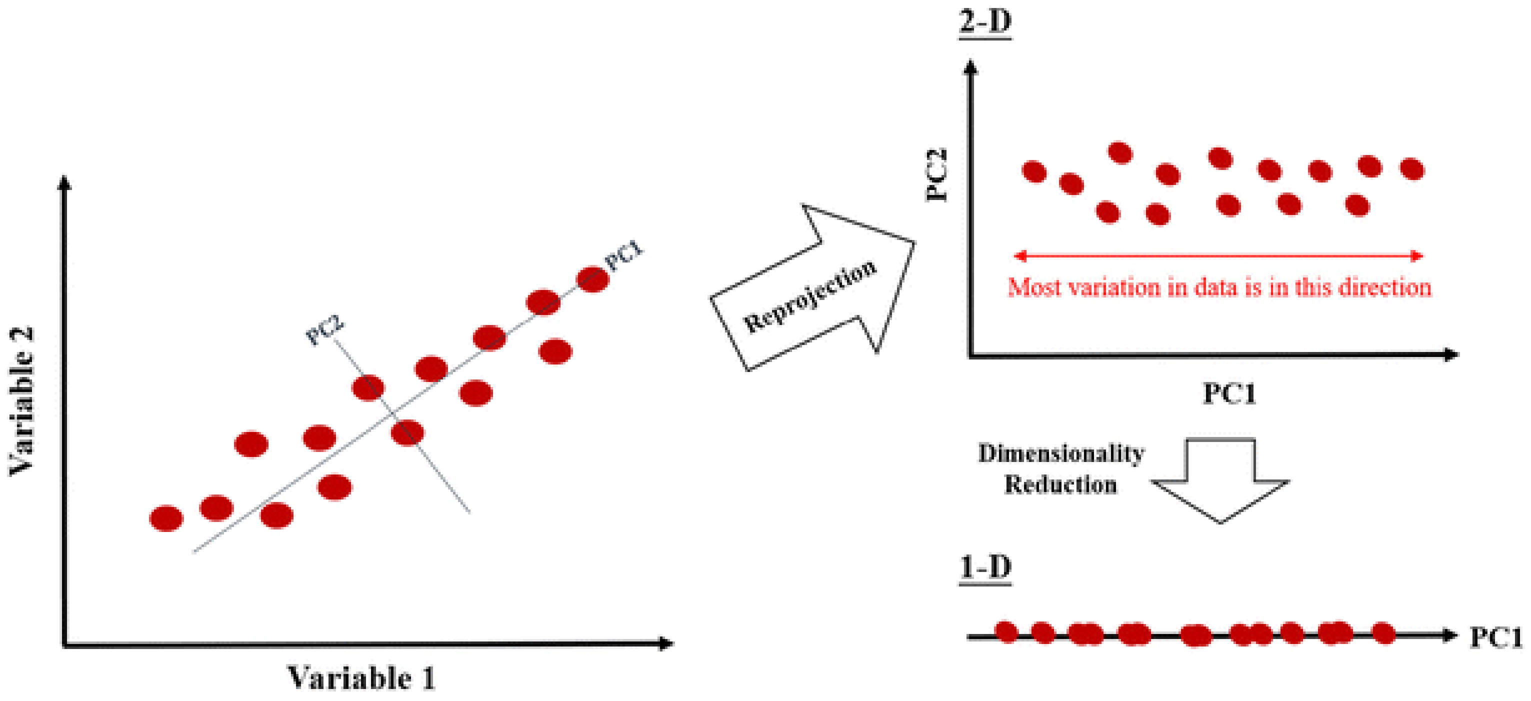
In Reducing Dimensions, it Would Help to Know How to Smartly Decompose Matrices…

In Reducing Dimensions, it Would Help to Know How to Smartly Decompose Matrices…



In Reducing Dimensions, it Would Help to Know How to Smartly Decompose Matrices…

$$Z = XW = U\,\Sigma V^T W = U\,\Sigma,$$



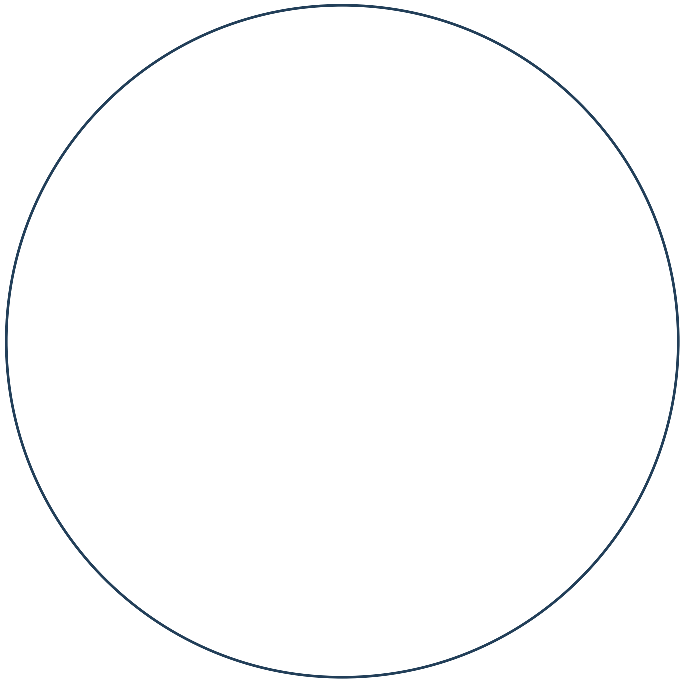
The above examples are just a small sample. In fact, it will help immensely with **smart** problem formulation, implementation, and analysis to know the three fields well.

Poetry of Mathematics…

Poetry of Mathematics…

Just like poetry, mathematics creates its own worlds, with various “entities” and their “rules of engagement”…

World of Poetry



Poetry of Mathematics…

Just like poetry, mathematics creates its own worlds, with various “entities” and their “rules of engagement”…

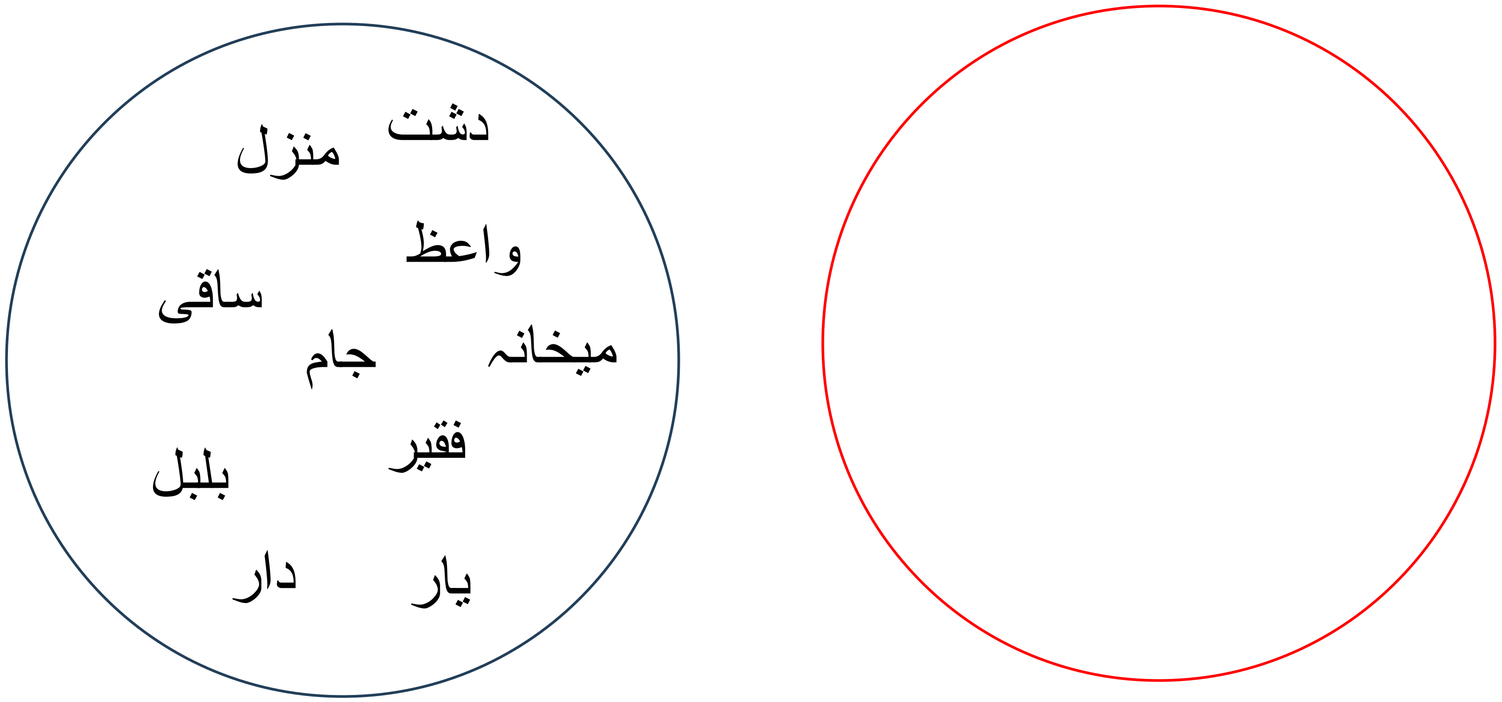
World of Poetry

Poetry of Mathematics…

Just like poetry, mathematics creates its own worlds, with various "entities" and their "rules of engagement"…

World of Poetry

World of Mathematics



Poetry of Mathematics…

Just like poetry, mathematics creates its own worlds, with various “entities” and their “rules of engagement”…

World of Poetry

دشت
منزل
واعظ
ساقی
جام
میخانہ
فقیر
بلبل
دار
یار

World of Mathematics

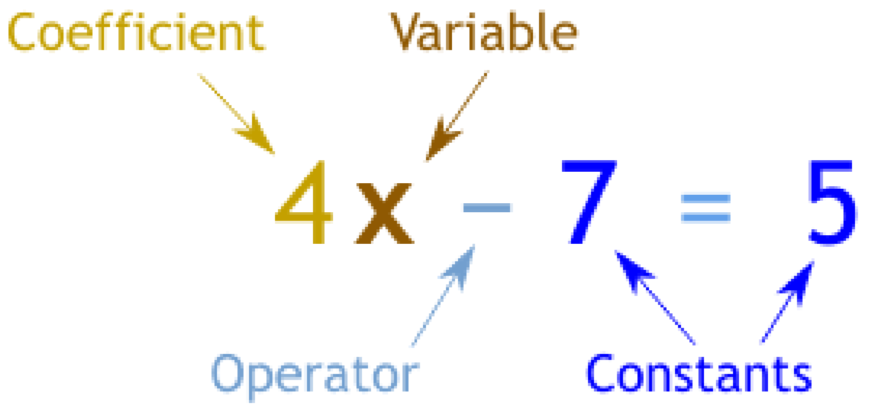
Sets
Numbers
Variables
Vectors
Matrices
Spaces
Theorems
Groups
Tensors
functions

Some Mathematical **Objects**

Poetry of Mathematics…

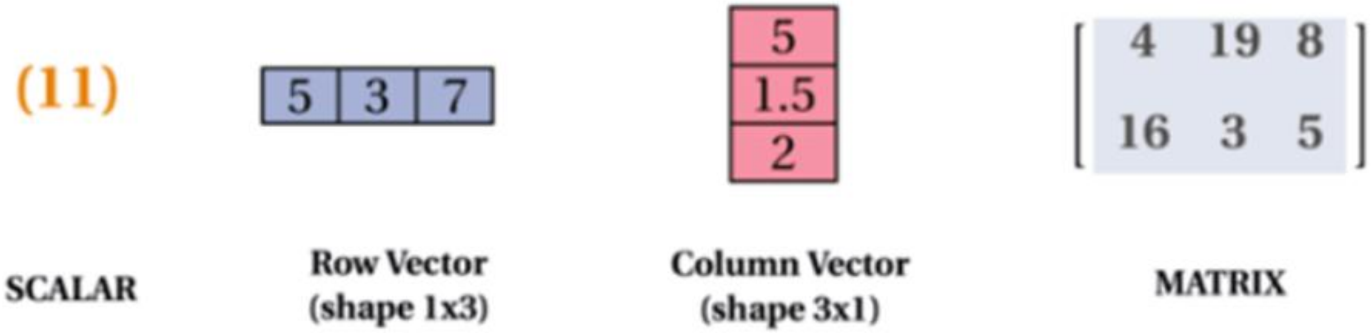
Just like poetry, mathematics creates its own worlds, with various “entities” and their “rules of engagement”…

Expressions and Some Objects in Them

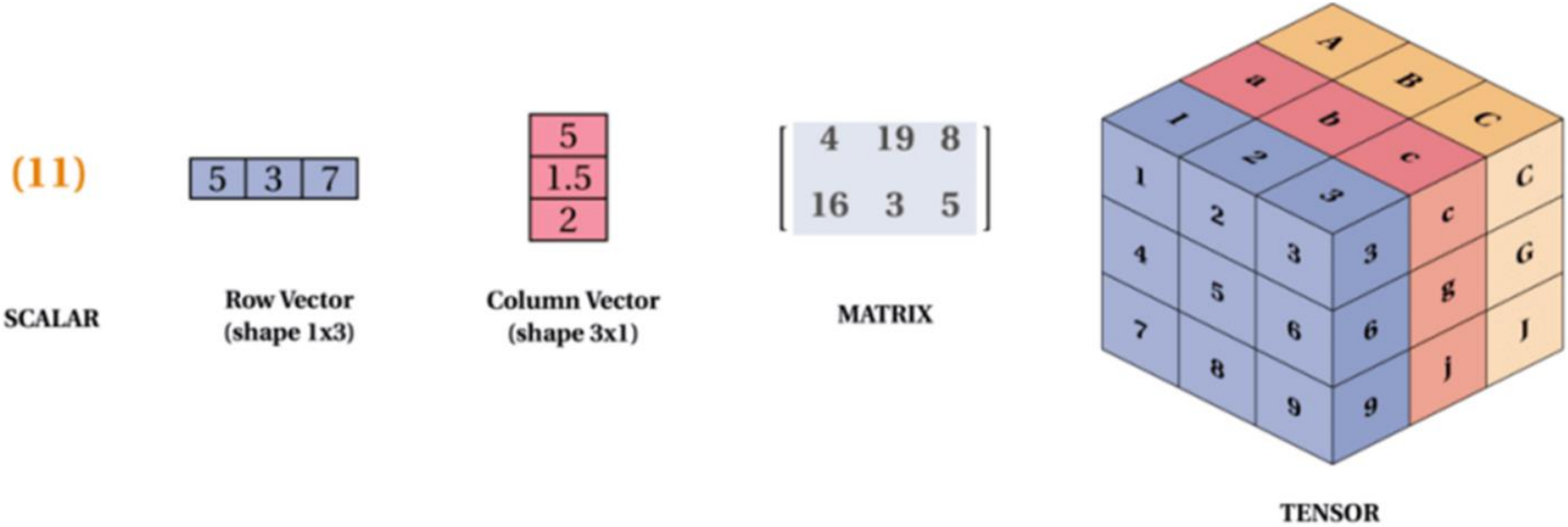


Beyond Single Variables…

Beyond Single Variables...

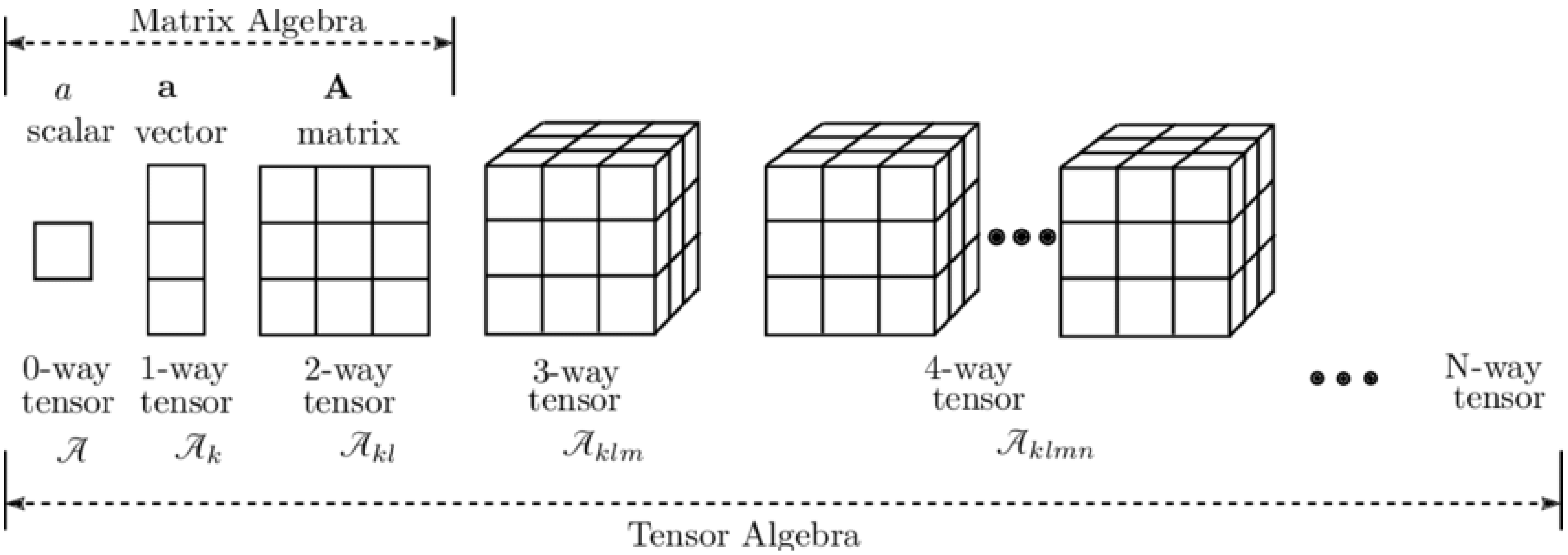


Beyond Single Variables...

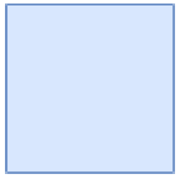
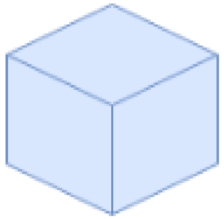


Poetry of Mathematics…

Just like poetry, mathematics creates its own worlds, with various "entities" and their "rules of engagement"…



Beyond Single Variables...

Type	Scalar	Vector	Matrix	Tensor
Definition	a single number	an array of numbers	2-D array of numbers	k-D array of numbers
Notation	x	$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$	$\boldsymbol{X} = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,n} \\ X_{2,1} & X_{2,2} & \dots & X_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ X_{m,1} & X_{m,2} & \dots & X_{m,n} \end{bmatrix}$	$\mathbf{X}$ $X_{i,j,k}$
Example	1.333	$\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ 9 \end{bmatrix}$	$\boldsymbol{X} = \begin{bmatrix} 1 & 2 & \dots & 4 \\ 5 & 6 & \dots & 8 \\ \vdots & \vdots & \vdots & \vdots \\ 13 & 14 & \dots & 16 \end{bmatrix}$	$x = \left[\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 10 & 20 & 30 \\ 50 & 60 \\ 80 & 90 \end{bmatrix} \begin{bmatrix} 100 & 200 & 300 \\ 00 & 600 \\ 00 & 900 \end{bmatrix} \right]$
Python code example	`x = np.array(1.333)`	`x = np.array([1,2,3, 4,5,6, 7,8,9])`	`x = np.array([[1,2,3,4], [5,6,7,8], [9,10,11,12], [13,14,15,16]])`	`x = np.array([[[1, 2, 3], [4, 5, 6], [7, 8, 9]], [[10, 20, 30], [40, 50, 60], [70, 80, 90]], [[100, 200, 300], [400, 500, 600], [700, 800, 900]]])`
Visualization				 3-D Tensor

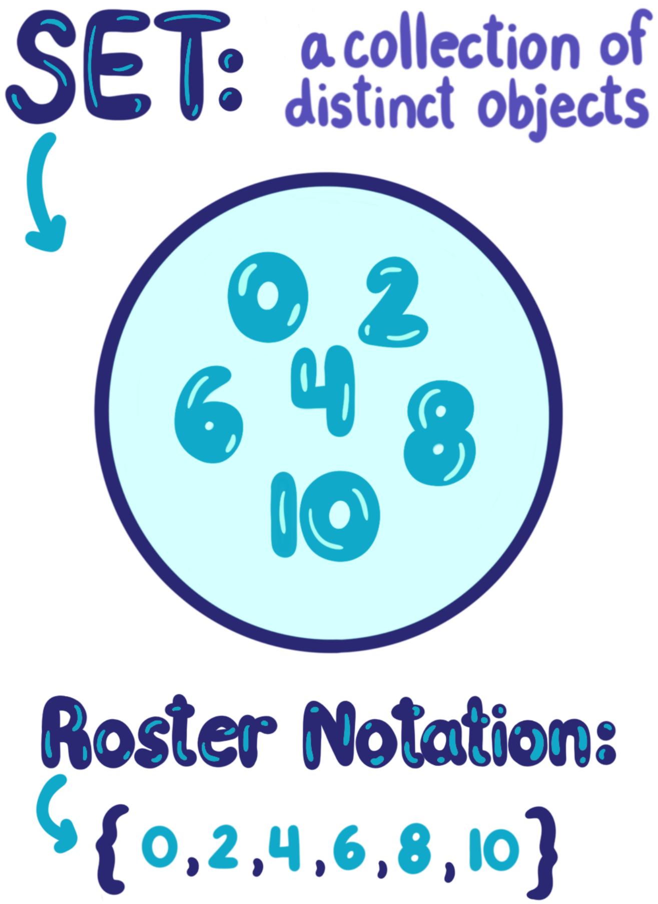
Set

Set

A set is a collection of well-defined and distinct objects. These objects can be concrete (numbers, letters) or abstract (ideas, concepts).

Set

A set is a collection of well-defined and distinct objects. These objects can be concrete (numbers, letters) or abstract (ideas, concepts).



Collections of Mathematical Objects...

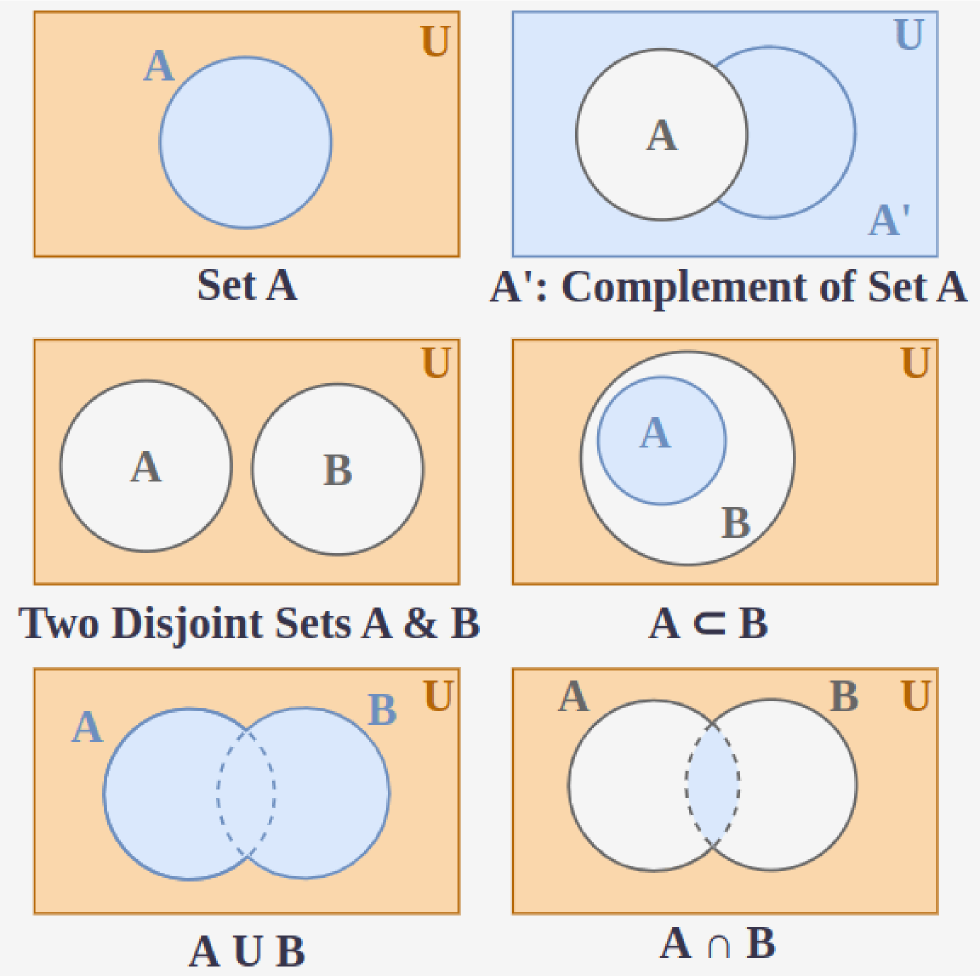
Set

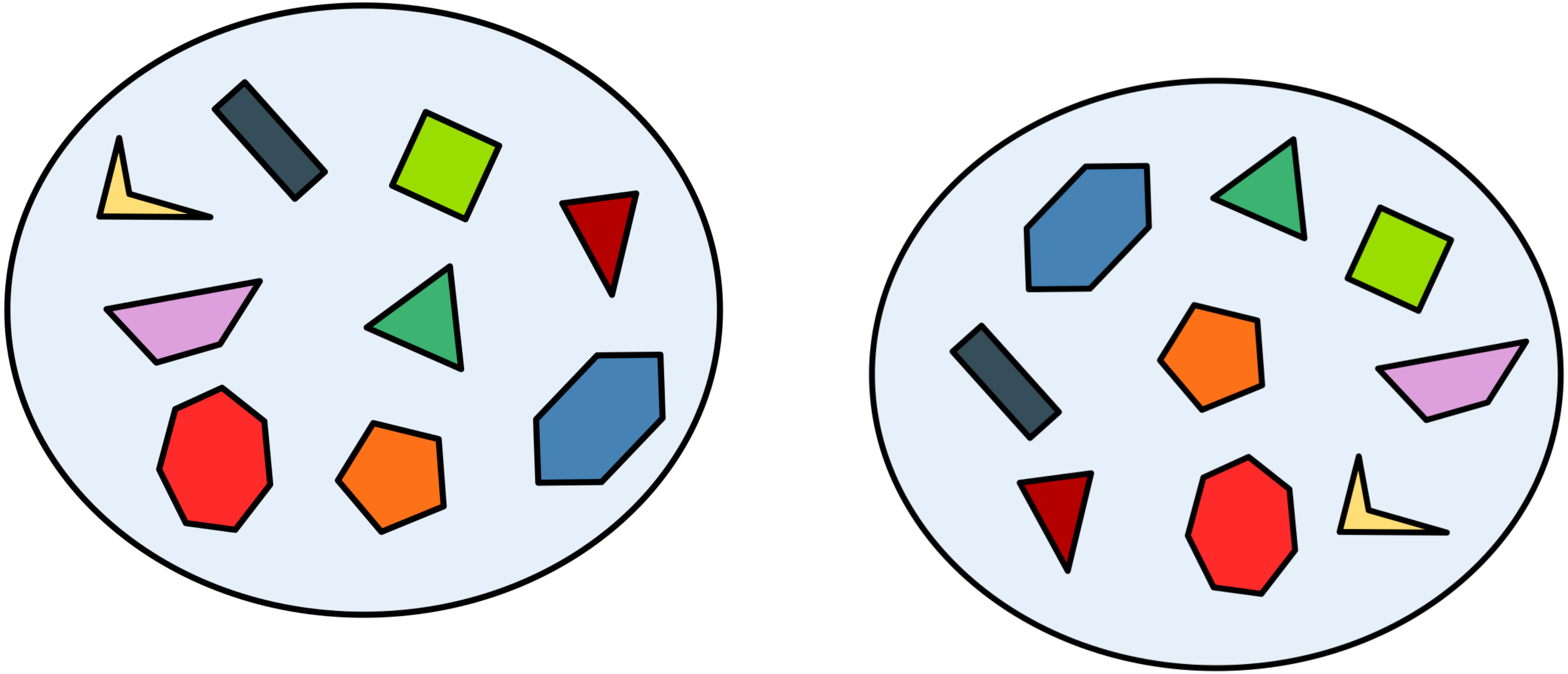
A set is a collection of well-defined and distinct objects. These objects can be concrete (numbers, letters) or abstract (ideas, concepts).

Symbol	Name	Example	Explanation
{ }	Set	$A = \{1,3\}$ $B = \{2,3,9\}$ $C = \{3,9\}$	Collection of objects
$\cap$	Intersect	$A \cap B = \{3\}$	Belong to both set A and set B
$\cup$	Union	$A \cup B = \{1,2,3,9\}$	Belong to set A or set B
$\subset$	Proper Subset	$\{1\} \subset A$ $C \subset B$	A set that is contained in another set
$\subseteq$	Subset	$\{1\} \subseteq A$ $\{1,3\} \subseteq A$	A set that is contained in or equal to another set
$\not\subset$	Not a Proper Subset	$\{1.3\} \not\subset A$	A set that is not contained in another set
$\supset$	Superset	$B \supset C$	Set B includes set C
$\in$	Is a member	$3 \in A$	3 is an element in set A
$\notin$	Is not a member	$4 \notin A$	4 is not an element in set A

Set

A set is a collection of well-defined and distinct objects. These objects can be concrete (numbers, letters) or abstract (ideas, concepts).

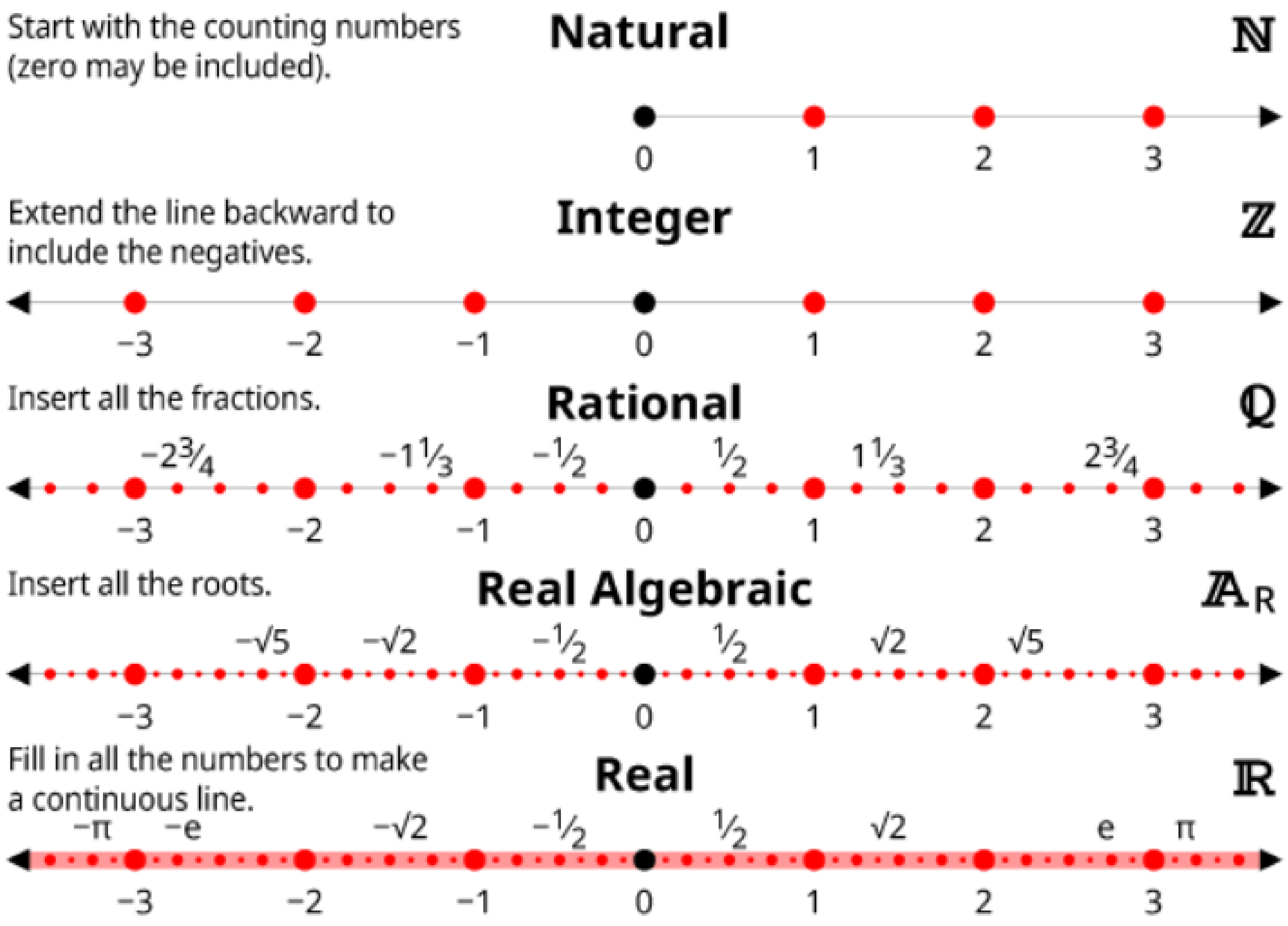




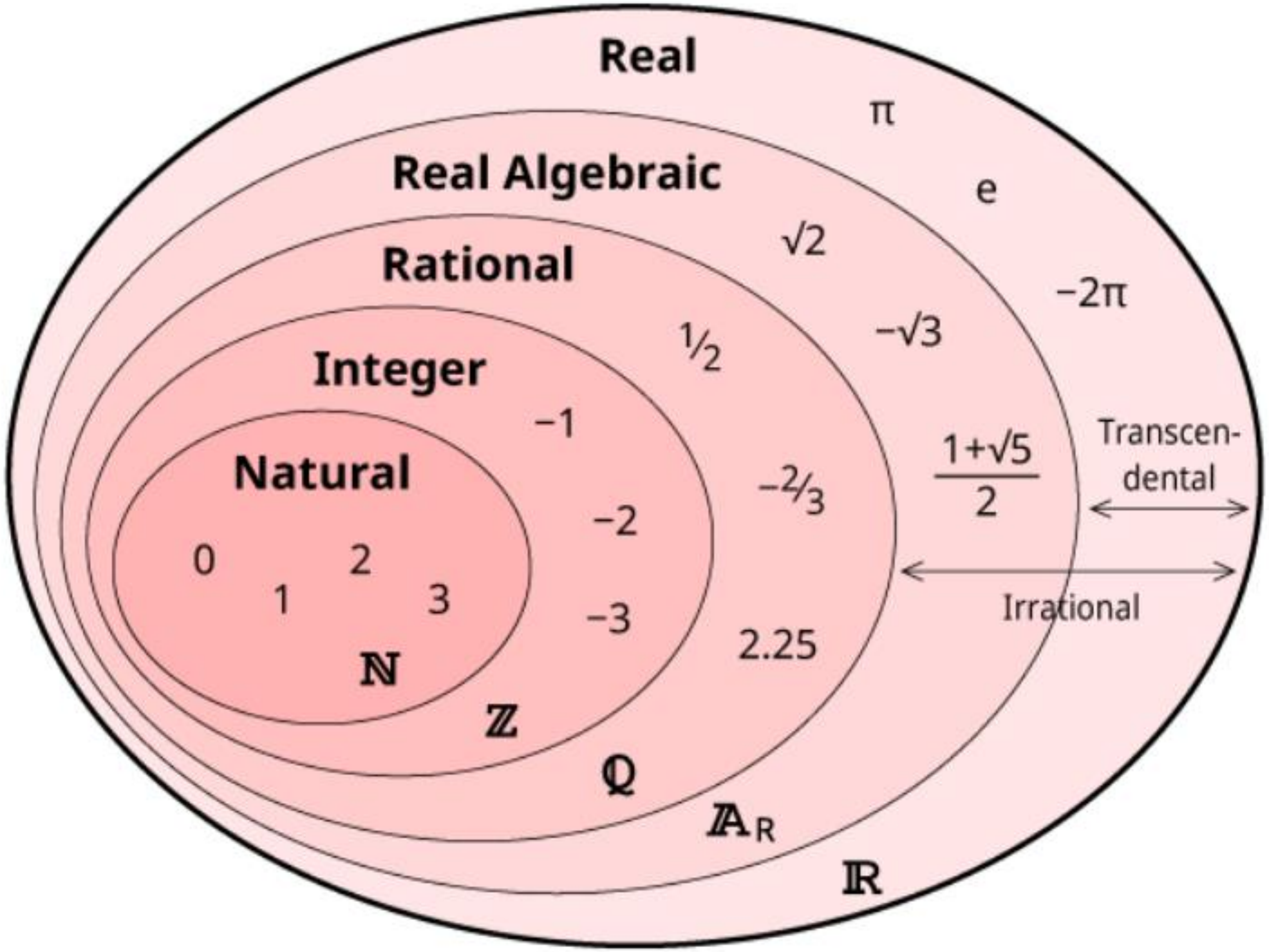
Q. Are these sets the same?

Real Number Line

Some familiar sets…



Some familiar sets…



$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{A}_{\mathrm{R}} \subset \mathbb{R}$$

Rational

$\mathbb{Q}$

Rational numbers are the ratios of integers, also called fractions, such as 1/2 = 0.5 or 1/3 = 0.333... Rational decimal expansions end or repeat. (Q is from quotient.)

Rational
$\mathbb{Q}$

Rational numbers are the ratios of integers, also called fractions, such as 1/2 = 0.5 or 1/3 = 0.333... Rational decimal expansions end or repeat. (Q is from quotient.)

Real Algebraic
$\mathbb{A}_{\mathrm{R}}$

Non-zero polynomials of finite degree with integer coefficients.

The real subset of the algebraic numbers: the real roots of polynomials. Real algebraic numbers may be rational or irrational. √2 = 1.41421... is irrational. Irrational decimal expansions neither end nor repeat.

Rational

$\mathbb{Q}$

Rational numbers are the ratios of integers, also called fractions, such as 1/2 = 0.5 or 1/3 = 0.333... Rational decimal expansions end or repeat. (Q is from quotient.)

Real Algebraic

$\mathbb{A}_{\mathrm{R}}$

Non-zero polynomials of finite degree with integer coefficients.

The real subset of the algebraic numbers: the real roots of polynomials. Real algebraic numbers may be rational or irrational. $\sqrt{2}$ = 1.41421... is irrational. Irrational decimal expansions neither end nor repeat.

Transcendental – Subset of Irrationals that are not algebraic.

Rational

$\mathbb{Q}$

Rational numbers are the ratios of integers, also called fractions, such as 1/2 = 0.5 or 1/3 = 0.333... Rational decimal expansions end or repeat. (Q is from quotient.)

Real Algebraic

$\mathbb{A}_{\mathrm{R}}$

Non-zero polynomials of finite degree with integer coefficients.

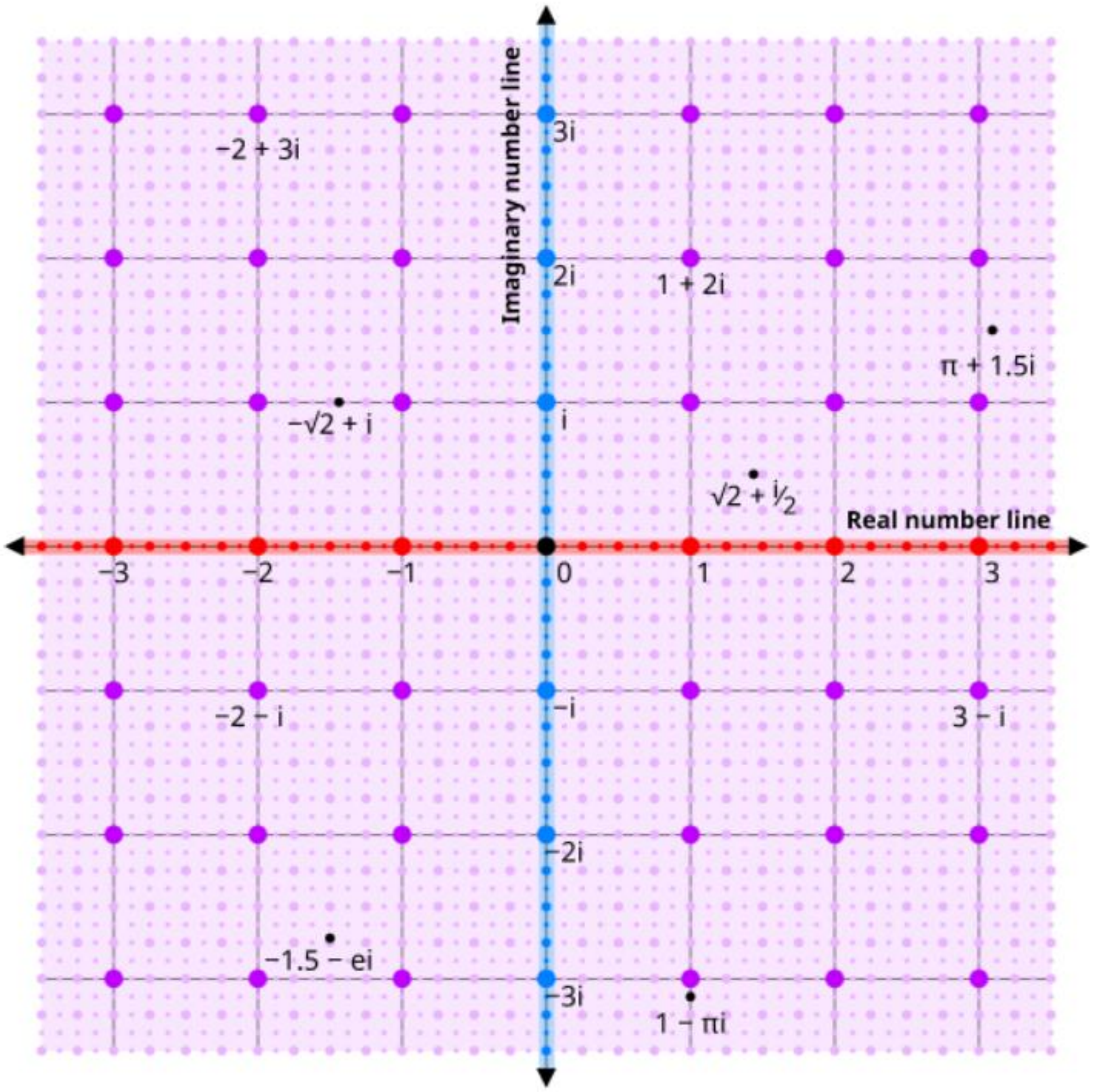
The real subset of the algebraic numbers: the real roots of polynomials. Real algebraic numbers may be rational or irrational. √2 = 1.41421... is irrational. Irrational decimal expansions neither end nor repeat.

Transcendental – Subset of Irrationals that are not algebraic.

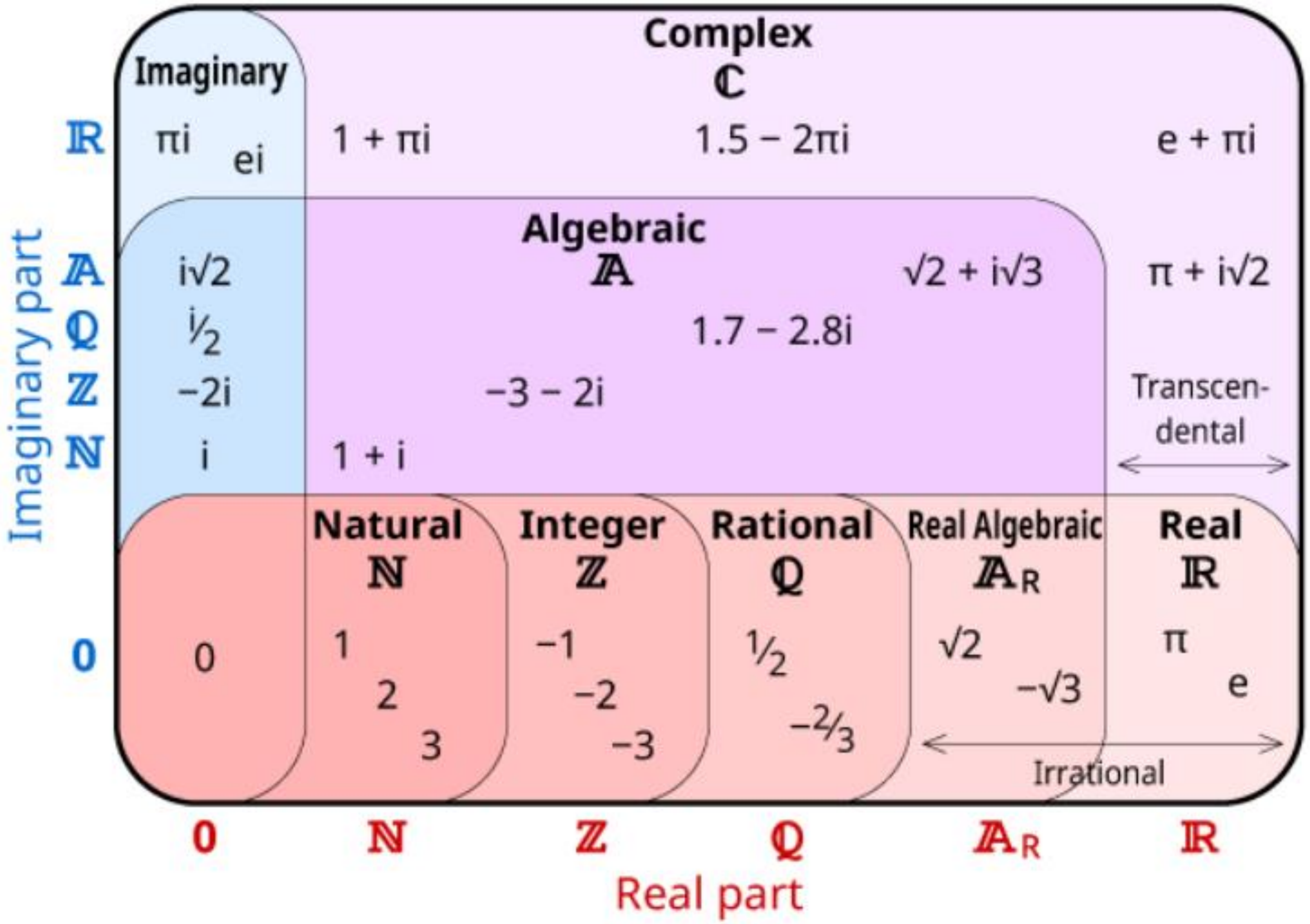
$$\sqrt{2}$$

Irrational, but not Transcendental
(as it is root of polynomial $x^2 - 2 = 0$)

Some familiar sets…



Some familiar sets…



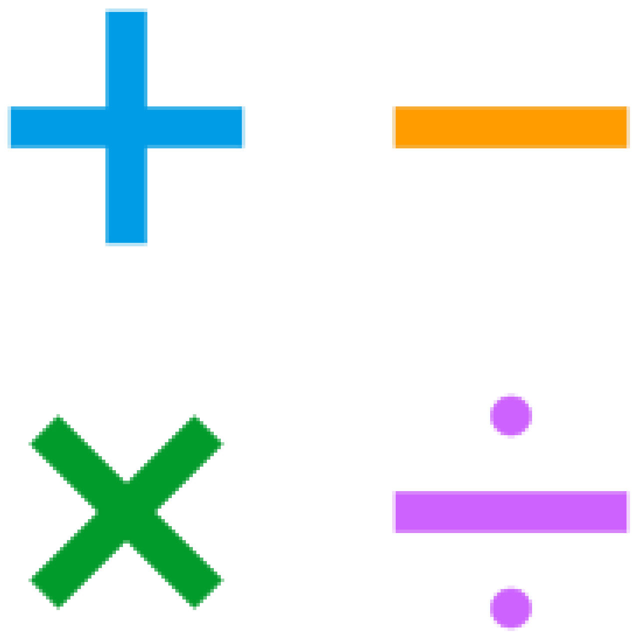
$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{A}_R \subset \mathbb{R} \subset \mathbb{C}$$

Predefined Interactions…

Operators

Predefined Interactions…

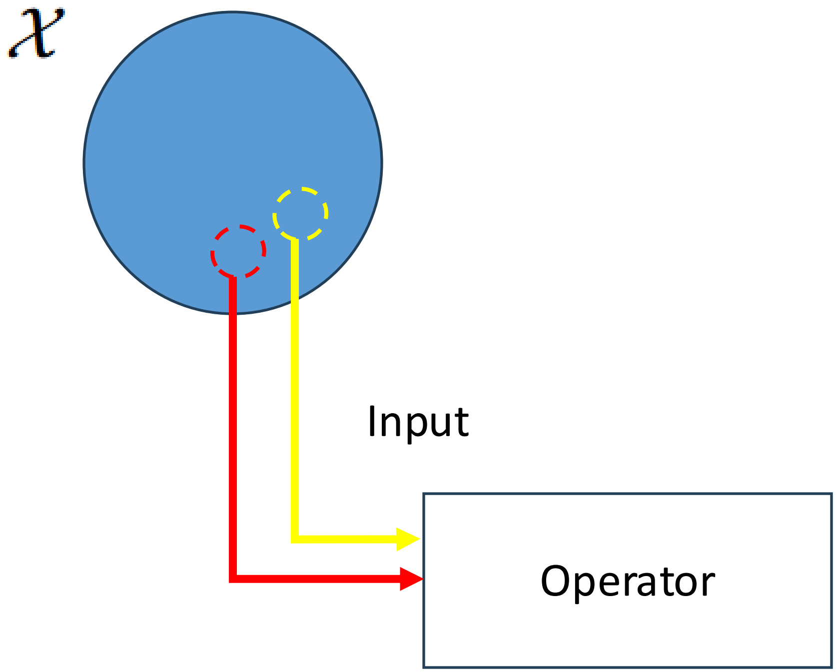
Operators



Most familiar operators…

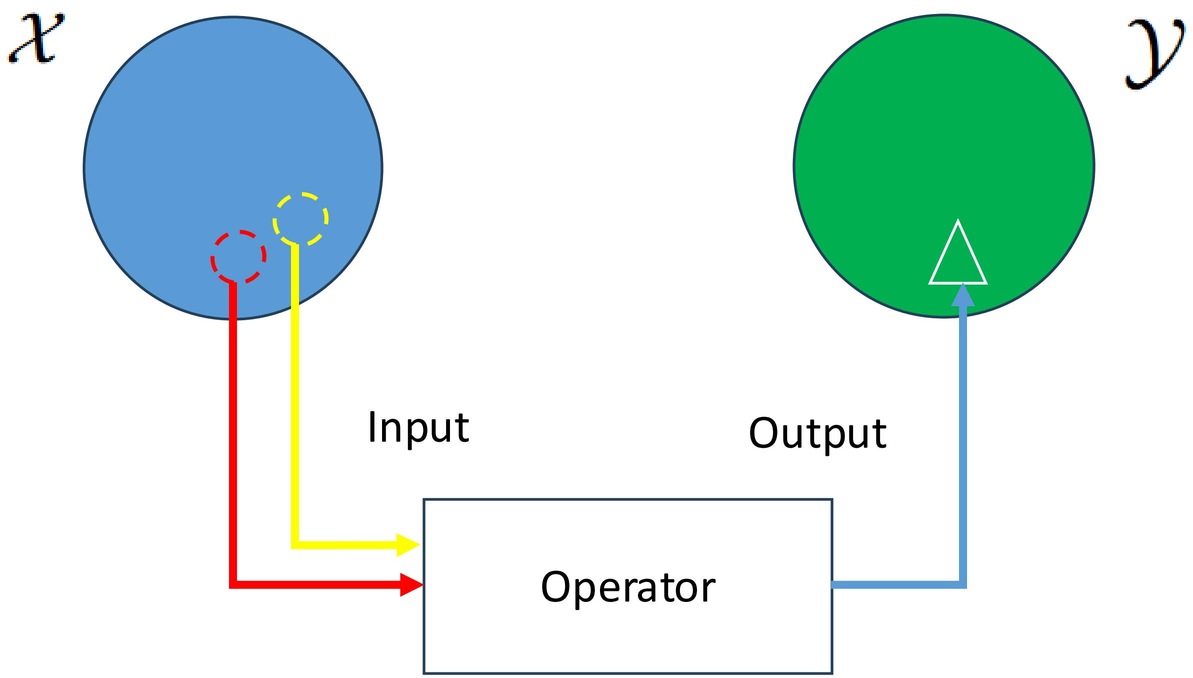
Predefined Interactions…

Operators



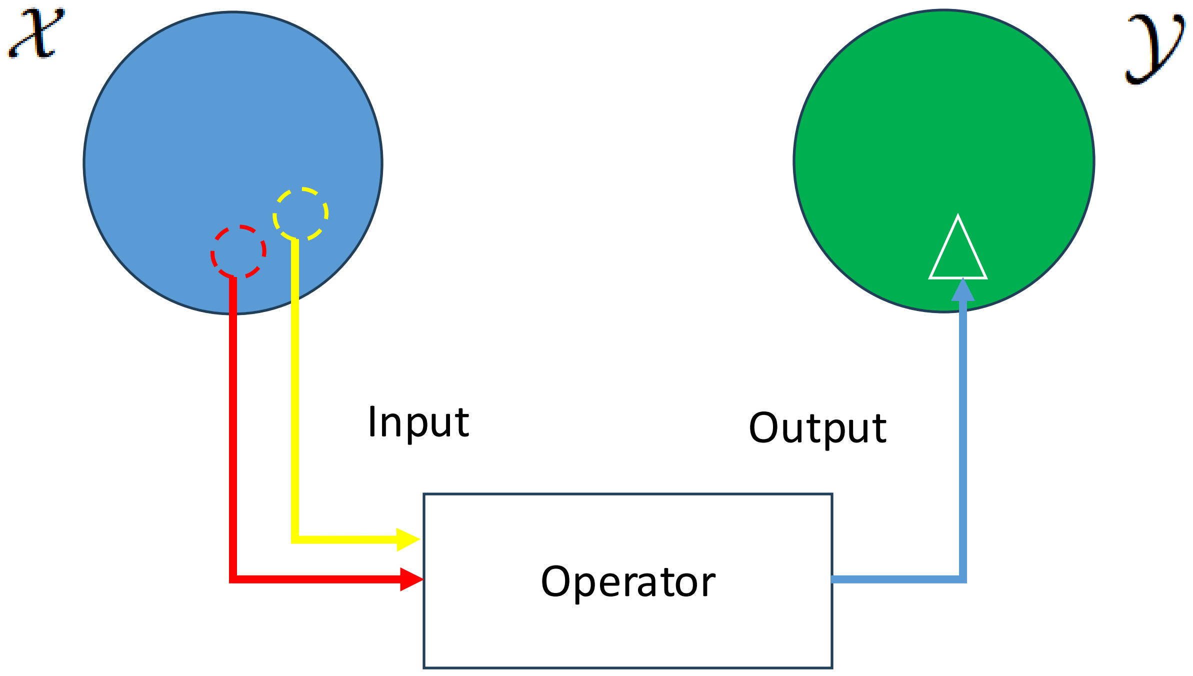
Predefined Interactions…

Operators



Predefined Interactions…

Operators

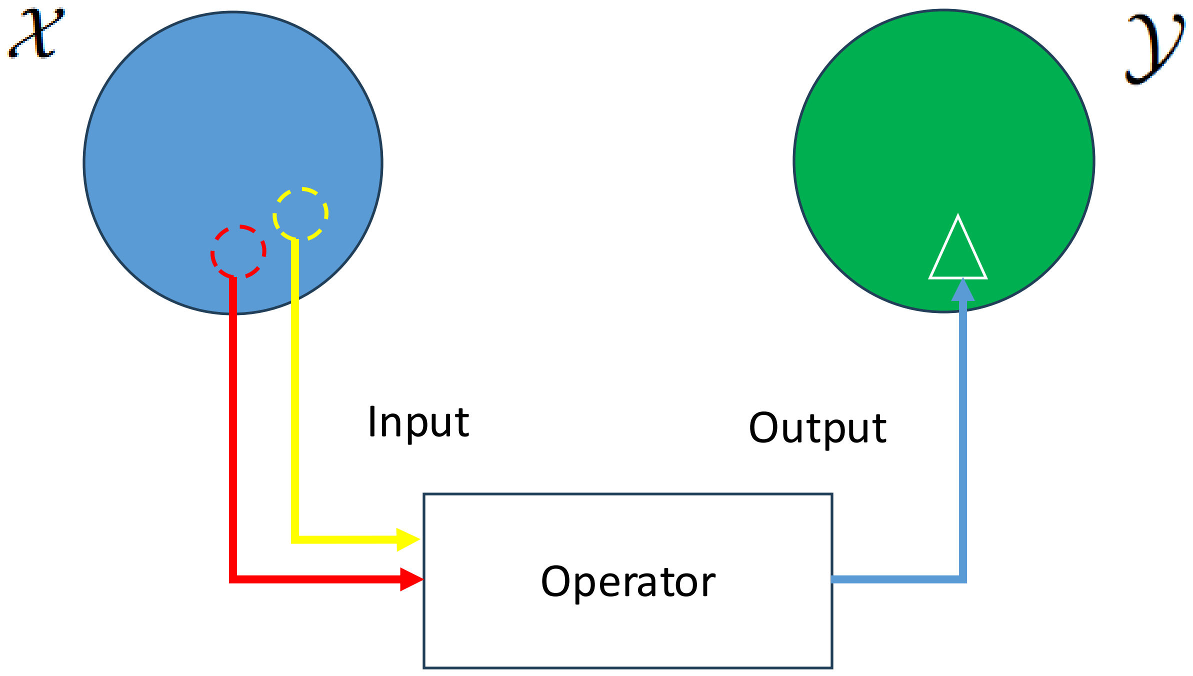


e.g., $x_1 + x_2 = y$

Predefined Interactions…

Operators

$+ : X \times X \to Y$



e.g., $x_1 + x_2 = y$

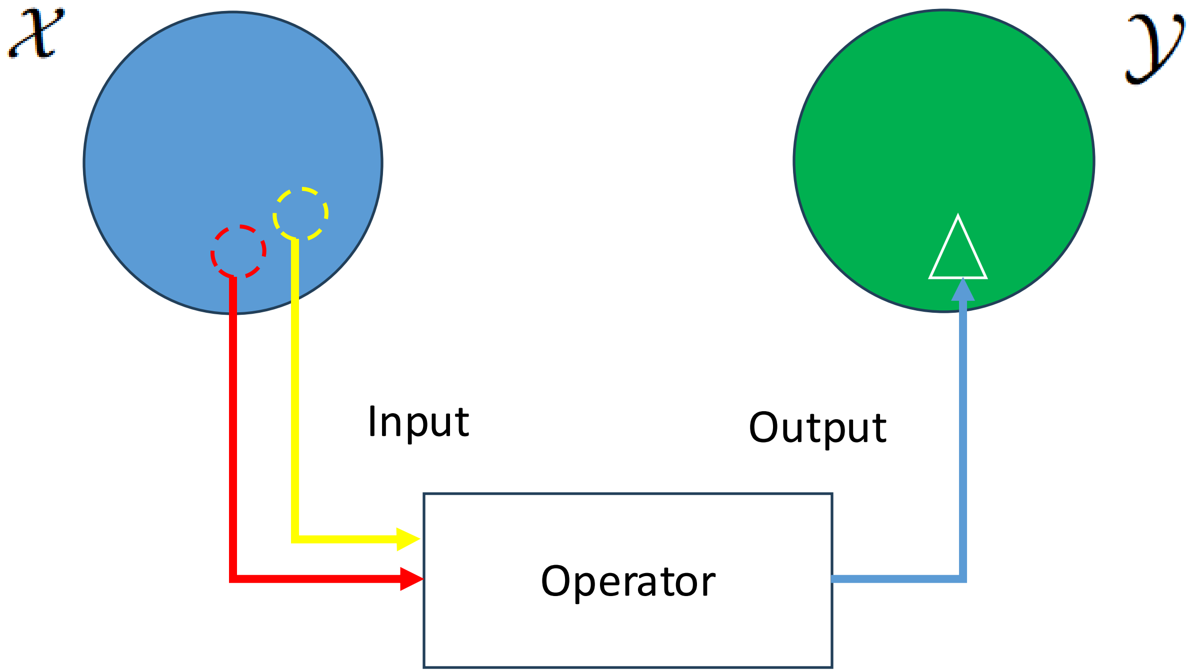
Predefined Interactions…

Operators

$+ : X \times X \rightarrow Y$

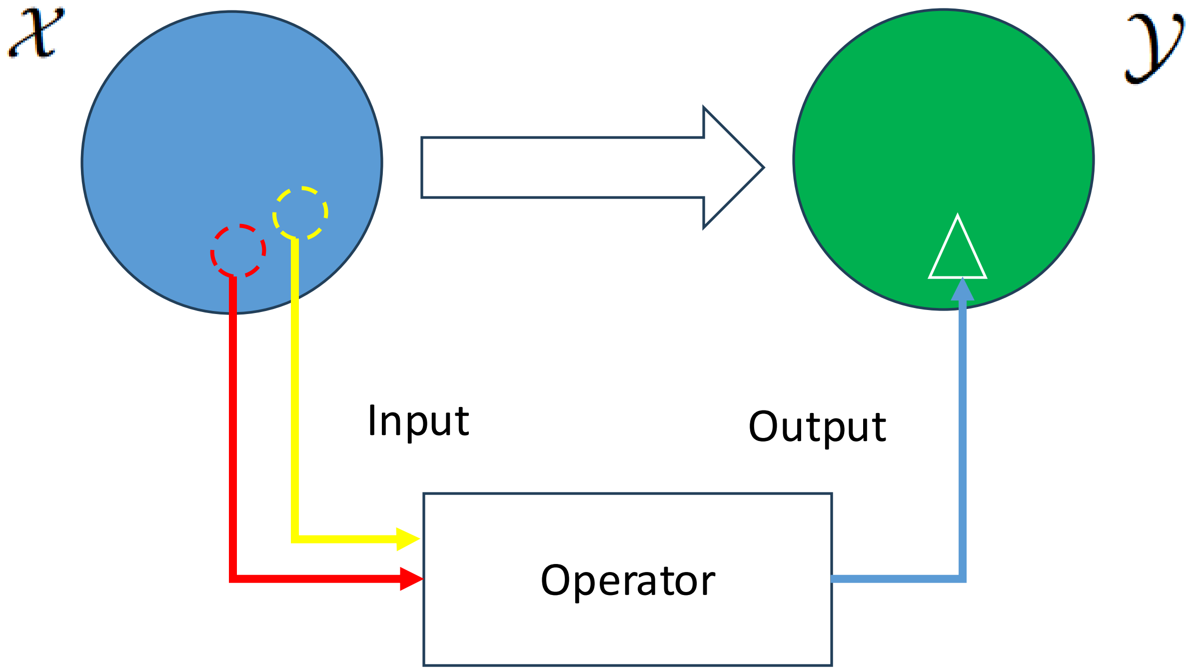
$\omega : X_1 \times \ldots \times X_n \rightarrow Y$

More general case.



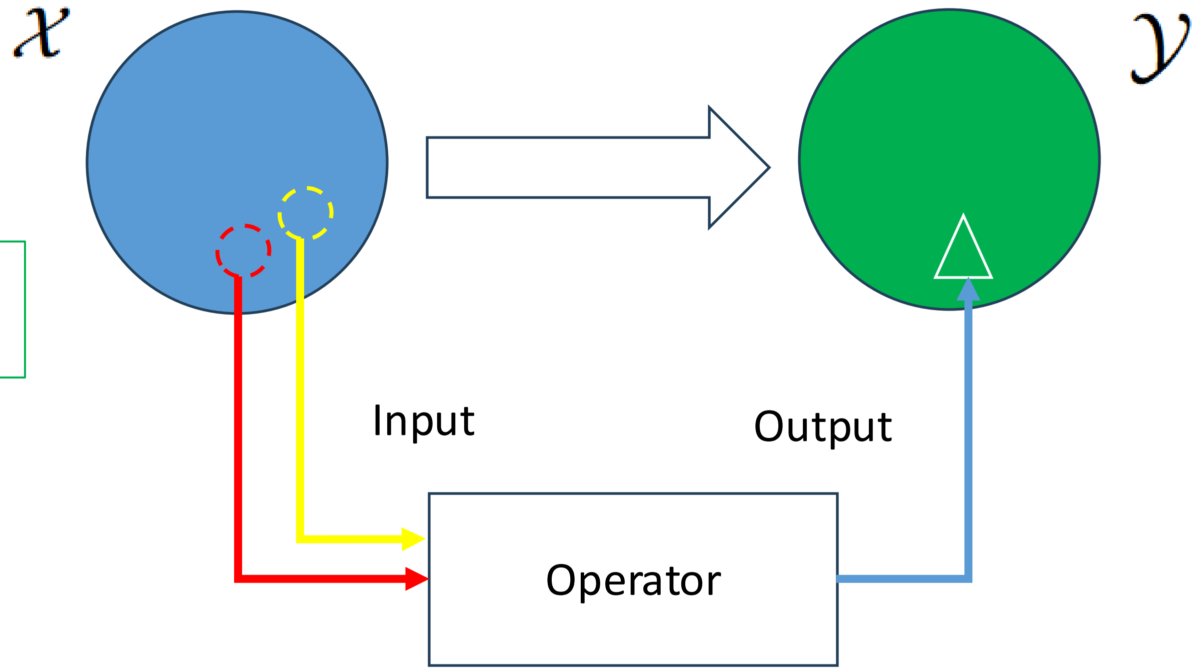
e.g., $x_1 + x_2 = y$

Operators may have some properties of interest and [analytical] use!

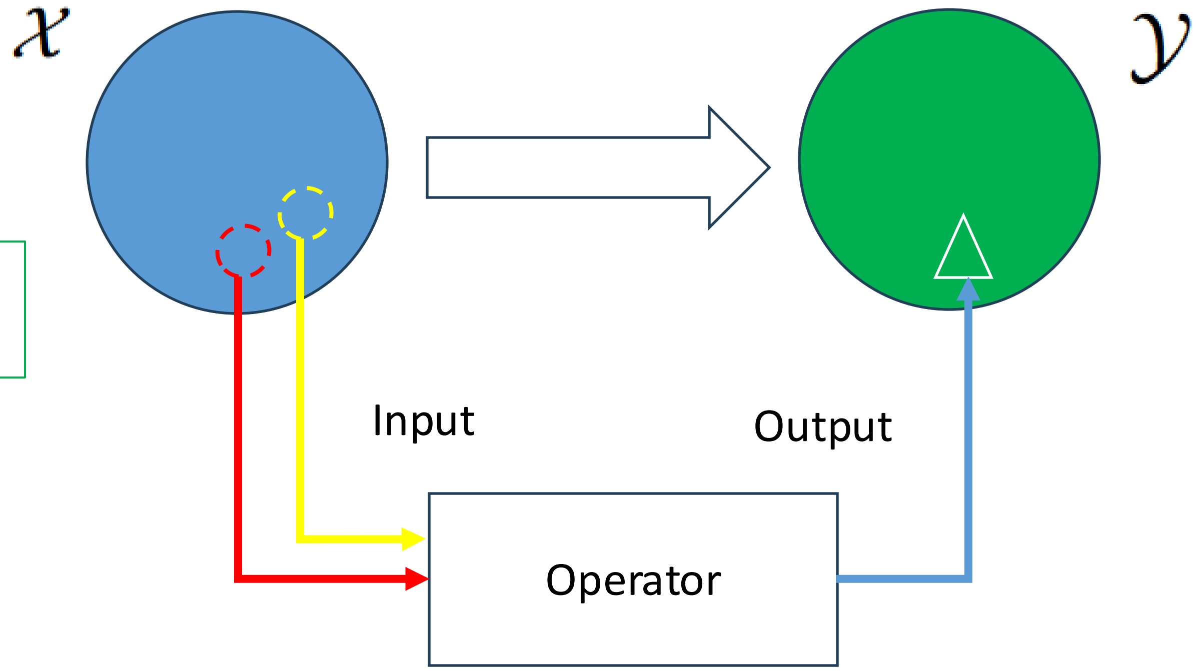


Operators may have some properties of interest and [analytical] use!

We know that $x_1 + x_2$, and $x_2 + x_1$ lead to the same point y (at least for numbers).



Operators may have some properties of interest and [analytical] use!



We know that $x_1 + x_2$, and $x_2 + x_1$ lead to the same point y (at least for numbers).

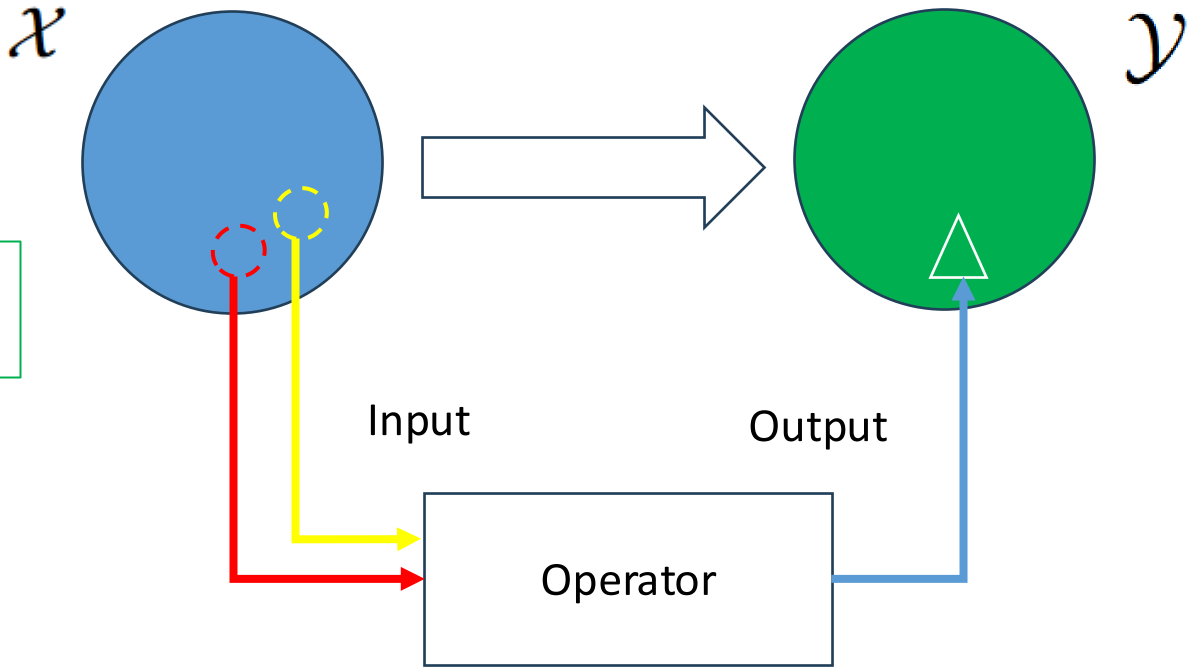
Addition is Commutative!

Operators may have some properties of interest and [analytical] use!

We know that $x_1 + x_2$, and $x_2 + x_1$ lead to the same point y (at least for numbers).

Addition is Commutative!

Subtraction is Not!

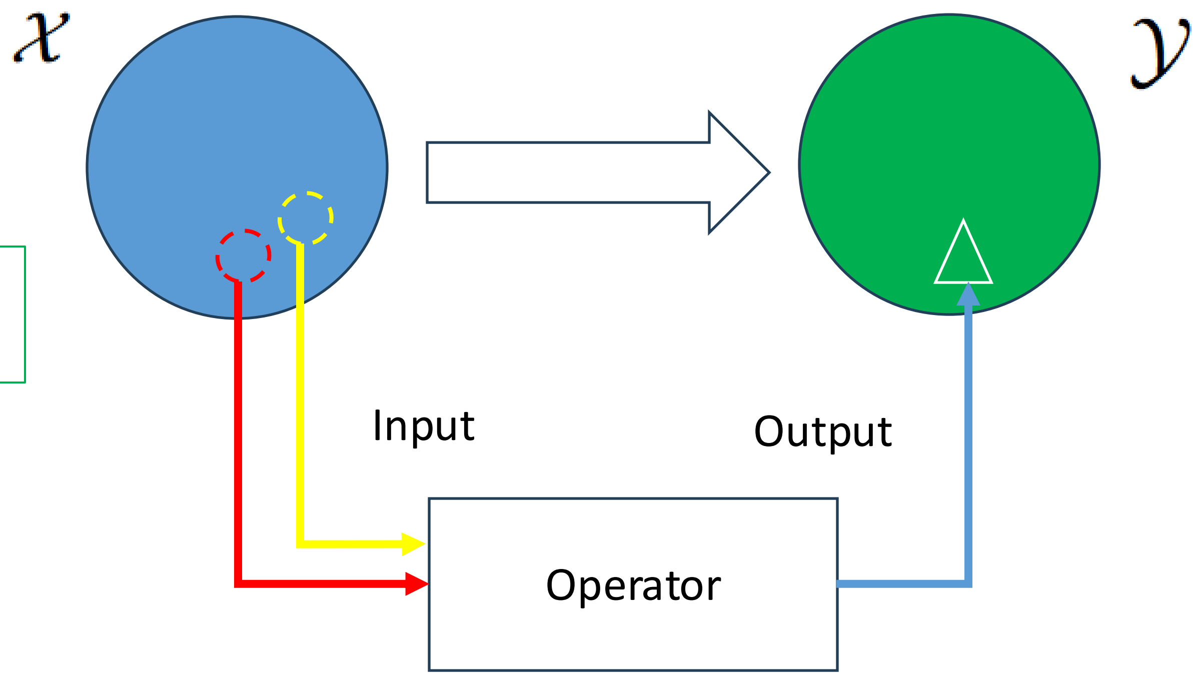


Operators may have some properties of interest and [analytical] use!

We know that $x_1 + x_2$, and $x_2 + x_1$ lead to the same point y (at least for numbers).

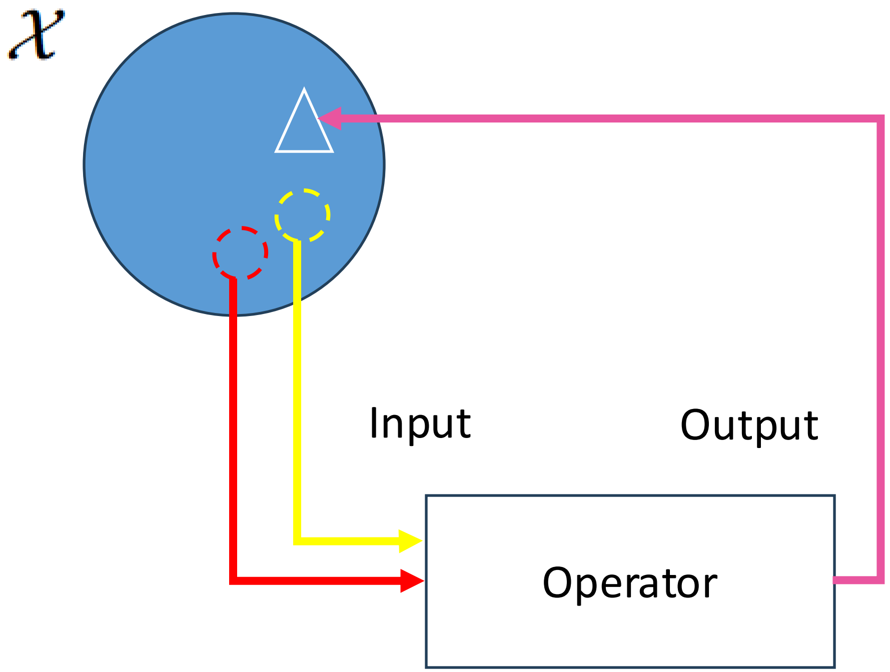
Addition is Commutative!

Subtraction is Not!

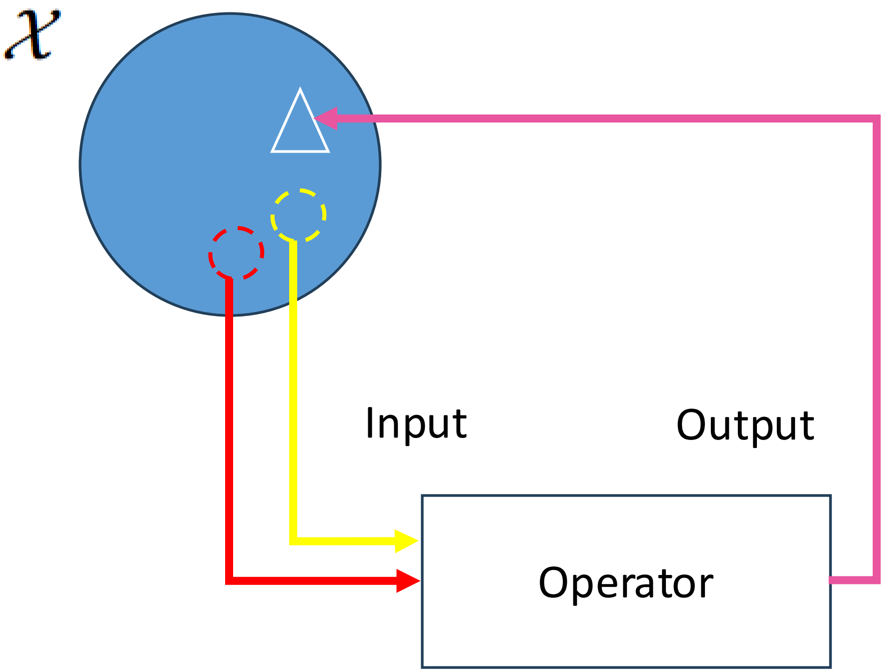


Property	Addition	Multiplication
Commutative	a + b = b + a	a × b = b × a
Associative	a + (b + c) = (a + b) + c	a × (b × c) = (a × b) × c
Distributive	a × (b + c) = a × b + a × c	

Taken together, Operators and Sets may have some properties of interest and [analytical] use!

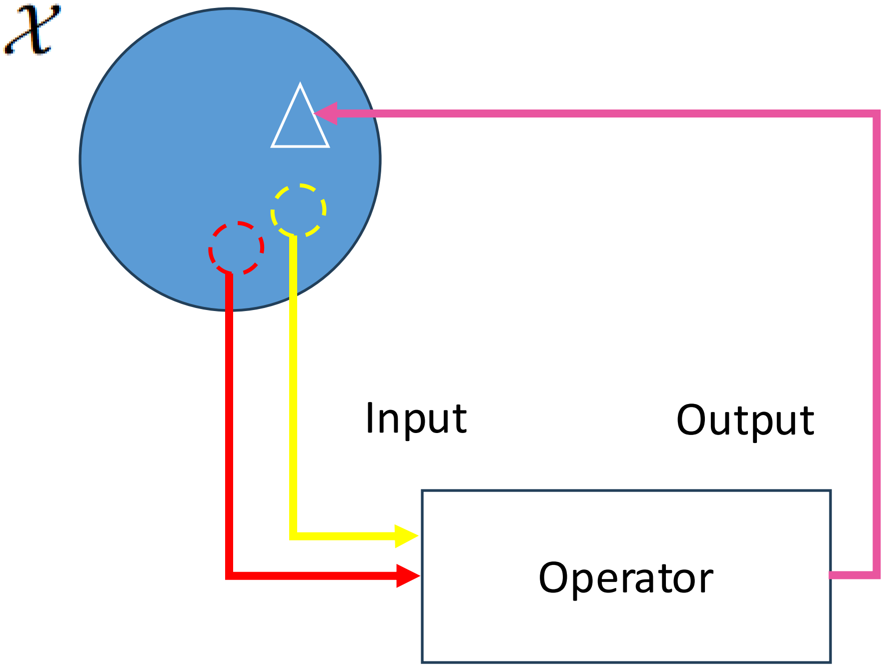


Taken together, Operators and Sets may have some properties of interest and [analytical] use!



Consider set X and an operator on it $\otimes$, i.e., $\{X, \otimes\}$

Taken together, Operators and Sets may have some properties of interest and [analytical] use!

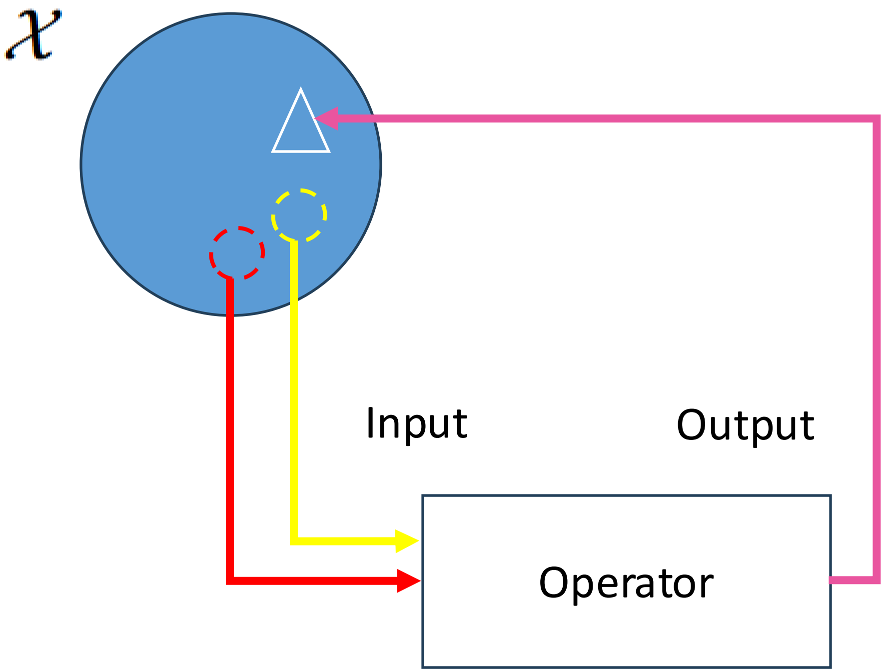


Consider set X and an operator on it $\otimes$, i.e., $\{X, \otimes\}$

Closure

$$\forall\, u, v \in X:$$
$$u \otimes v \in X$$

Taken together, Operators and Sets may have some properties of interest and [analytical] use!



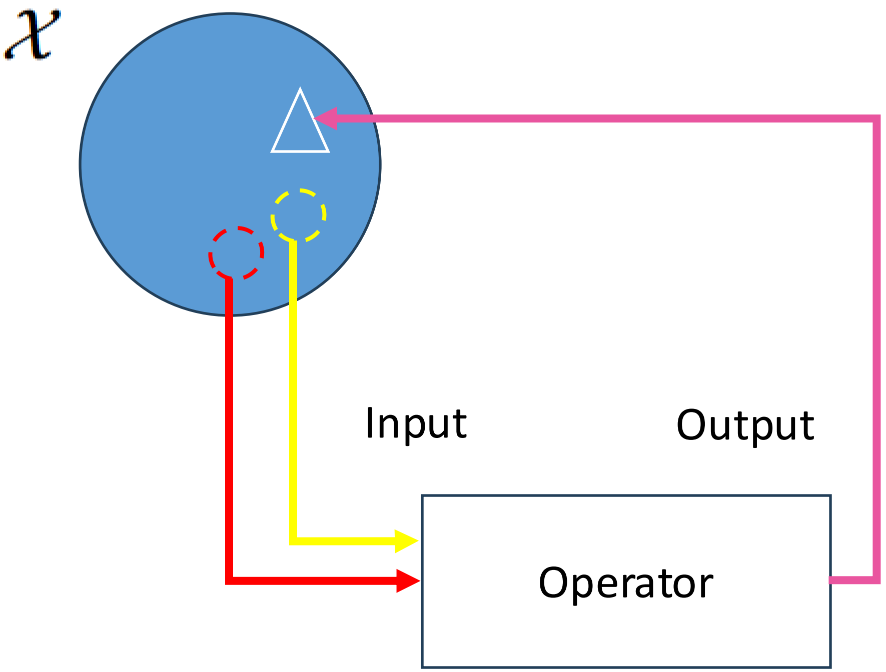
Consider set X and an operator on it $\otimes$, i.e., $\{X, \otimes\}$

Closure

$$\forall\, u, v \in X:$$
$$u \otimes v \in X$$

$$\forall\, x \in X$$
$$\exists\, e \in X:$$
$$x \otimes e = e \otimes x = x$$

Taken together, Operators and Sets may have some properties of interest and [analytical] use!



Consider set X and an operator on it $\otimes$, i.e., $\{X, \otimes\}$

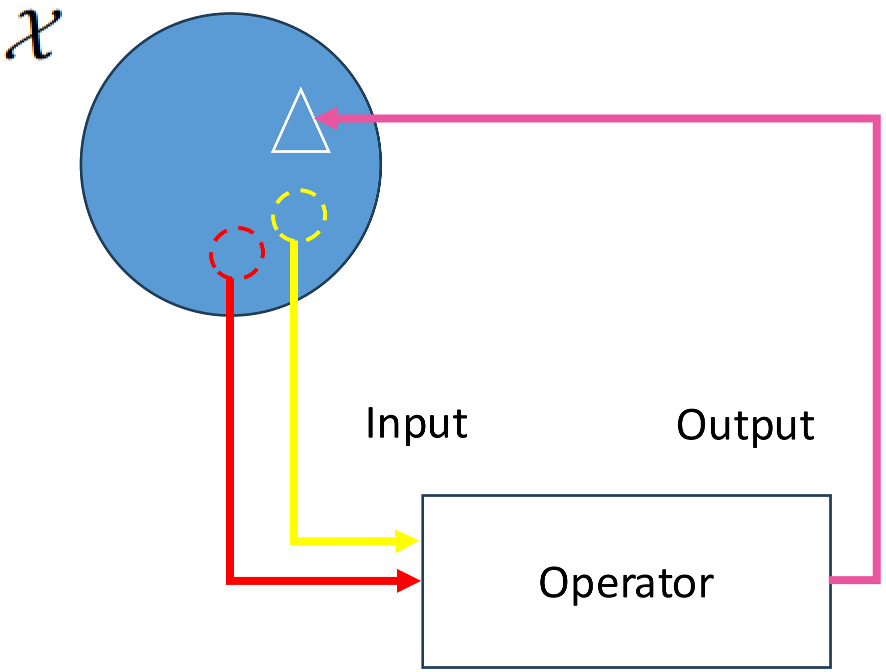
Closure

$$\forall\, u, v \in X:$$
$$u \otimes v \in X$$

Identity Element

$$\forall\, x \in X$$
$$\exists\, e \in X:$$
$$x \otimes e = e \otimes x = x$$

Taken together, Operators and Sets may have some properties of interest and [analytical] use!



Consider set X and an operator on it $\otimes$, i.e., $\{X, \otimes\}$

Closure

$$\forall\, u, v \in X: \\ u \otimes v \in X$$

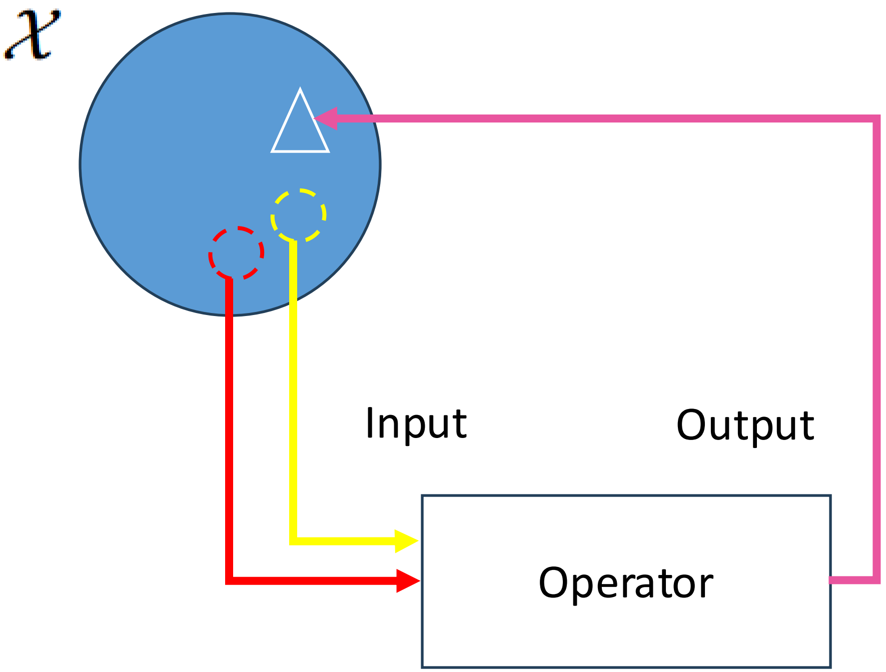
Identity Element

$$\forall\, x \in X \\ \exists\, e \in X: \\ x \otimes e = e \otimes x = x$$

$$\forall\, x \in X \\ \exists\, y \in X: \\ x \otimes y = y \otimes x = e \\ e = identity\ element$$

Taken together, Operators and Sets may have some properties of interest and [analytical] use!

Consider set X and an operator on it $\otimes$, i.e., $\{X, \otimes\}$



Closure

$$\forall\, u, v \in X:$$
$$u \otimes v \in X$$

Identity Element

$$\forall\, x \in X$$
$$\exists\, e \in X:$$
$$x \otimes e = e \otimes x = x$$

Inverse Element

$$\forall\, x \in X$$
$$\exists\, y \in X:$$
$$x \otimes y = y \otimes x = e$$
$$e = identity\ element$$

$Z = set\ of\ integers$

Closure Property

$Z = set\ of\ integers$

Closure Property

$$a + b \in Z$$
$$a - b \in Z$$
$$a \times b \in Z$$
$$a \div b \notin Z$$

$Z = set\ of\ integers$

Closure Property

$a + b \in Z$

$a - b \in Z$

$a \times b \in Z$

$a \div b \notin Z$

Identity Element for $\{\mathbf{Z}, +\}$

Identity Element for $\{\mathbf{Z}, \times\}$

Inverse Element for $\{Z, +\}$

Inverse Element for $\{\mathbf{Z}, \times\}$

$Z = set\ of\ integers$

Closure Property

$a + b \in Z$

$a - b \in Z$

$a \times b \in Z$

$a \div b \notin Z$

Identity Element for $\{\mathbf{Z}, +\}$

$\forall a \in Z$

$\exists\ e = 0 \in Z$:

$a + 0 = 0 + a = a$

Identity Element for $\{\mathbf{Z}, \times\}$

$\forall a \in Z$

$\exists\ e = 1 \in Z$:

$a \times e = e \times a = a$

Inverse Element for $\{\boldsymbol{Z}, +\}$

$\forall a \in Z$

$\exists\ y = -a \in Z$:

$a + y = y + a = 0$

Inverse Element for $\{\mathbf{Z}, \times\}$

for $a \in Z$

$\nexists\ \ y \in Z$:

$a \times y = y \times a = 1$

Why Are These Properties of Interest?

Why Are These Properties of Interest?

- A lot of mathematics deals with finding solutions (possibly under constraints), solution sets, and general proofs

Why Are These Properties of Interest?

- A lot of mathematics deals with finding solutions (possibly under constraints), solution sets, and general proofs

    - Closure, Identity, Inverse helpful as we can be sure that mappings can be reversed, and solution does not require expanding the set.

Why Are These Properties of Interest?

- A lot of mathematics deals with finding solutions (possibly under constraints), solution sets, and general proofs

    - Closure, Identity, Inverse helpful as we can be sure that mappings can be reversed, and solution does not require expanding the set.

    - They also help us understand nature and limitations of sets possibly leading to new math (e.g., need for complex numbers when trying to solve polynomials. Galois Theory another example.)

Is the square root of a real number always a real number?

Why Are These Properties of Interest?

- A lot of mathematics deals with finding solutions (possibly under constraints), solution sets, and general proofs

    - Closure, Identity, Inverse helpful as we can be sure that mappings can be reversed, and solution does not require expanding the set.

    - They also help us understand nature and limitations of sets possibly leading to new math (e.g., need for complex numbers when trying to solve polynomials. Galois Theory another example.)

Is the square root of a real number always a real number?

Non-Closure of $\{\mathbf{R}, \sqrt{\ }\}$

$$\exists\ a \in R$$
$$\sqrt{a} \notin R$$

Why Are These Properties of Interest?

- A lot of mathematics deals with finding solutions (possibly under constraints), solution sets, and general proofs
    - Closure, Identity, Inverse helpful as we can be sure that mappings can be reversed, and solution does not require expanding the set.
    - They also help us understand nature and limitations of sets possibly leading to new math (e.g., need for complex numbers when trying to solve polynomials. Galois Theory another example.)

Is the square root of a real number always a real number?

Non-Closure of $\{\mathbf{R}, \sqrt{\ }\}$

$$\exists\ a \in R$$
$$\sqrt{a} \notin R$$

No solution to the problem "**find Real number $\boldsymbol{x}$ such that $\boldsymbol{x^2 + 1 = 0}$**", since Real numbers are not closed under square root.

Why Are These Properties of Interest?

- A lot of mathematics deals with finding solutions (possibly under constraints), solution sets, and general proofs

  - Closure, Identity, Inverse helpful as we can be sure that mappings can be reversed, and solution does not require expanding the set.

  - They also help us understand nature and limitations of sets possibly leading to new math (e.g., need for complex numbers when trying to solve polynomials. Galois Theory another example.)

Is the square root of a real number always a real number?

Non-Closure of $\{\mathbf{R}, \sqrt{\ }\}$

$$\exists\ a \in R$$
$$\sqrt{a} \notin R$$

No solution to the problem "**find Real number $\boldsymbol{x}$ such that $\boldsymbol{x^2 + 1 = 0}$**", since Real numbers are not closed under square root.

Imagine the time before introduction of Set of Complex numbers!

Why Are These Properties of Interest? (Contd…)

- Analytical solutions and proofs become easier if expressions can be simplified
    - Commutative, Associative, Distributive properties come in handy.

Why Are These Properties of Interest? (Contd…)

- Analytical solutions and proofs become easier if expressions can be simplified

    - Commutative, Associative, Distributive properties come in handy.

Prove the useful identity

$$(a+b)(a-b) = a^2 - b^2$$

Why Are These Properties of Interest? (Contd…)

- Analytical solutions and proofs become easier if expressions can be simplified
    - Commutative, Associative, Distributive properties come in handy.

Prove the useful identity

$$(a+b)(a-b) = a^2 - b^2$$

$$(a+b)(a-b) = (a+b)a + (a+b)(-b) = a(a+b) + (-b)(a+b) = a^2 + ab - ba - b^2$$
$$= a^2 + ab - ab - b^2 = a^2 + 1ab - 1ab + b^2 = a^2 + (1-1)ab - b^2 = a^2 + 0ab - b^2 = a^2$$
$$+ 0 - b^2 = a^2 - b^2$$

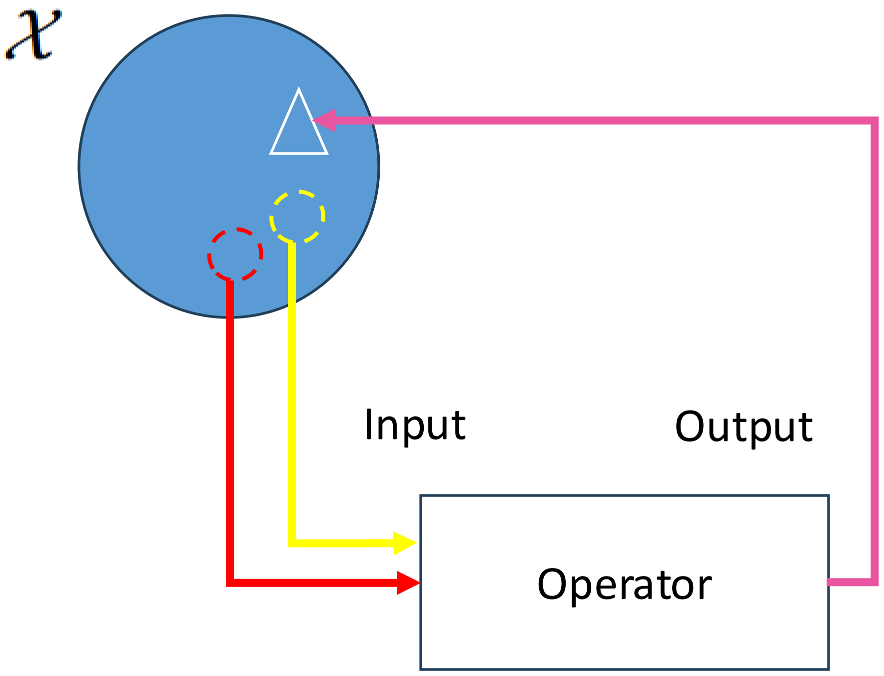
A Very Common Type of Operators

Binary Operations

A Very Common Type of Operators

Binary Operations

An operator that takes **two** elements of a set and produces another element of the same set.



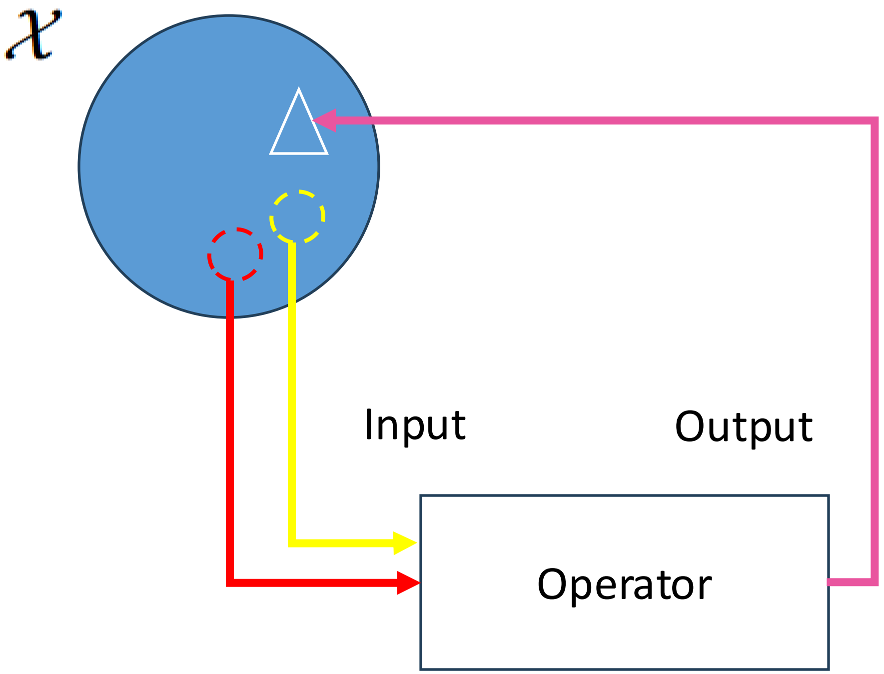
$f: X \times X \to X$

A Very Common Type of Operators

Binary Operations

An operator that takes **two** elements of a set and produces another element of the same set.

Note that this implies "X is closed under f"



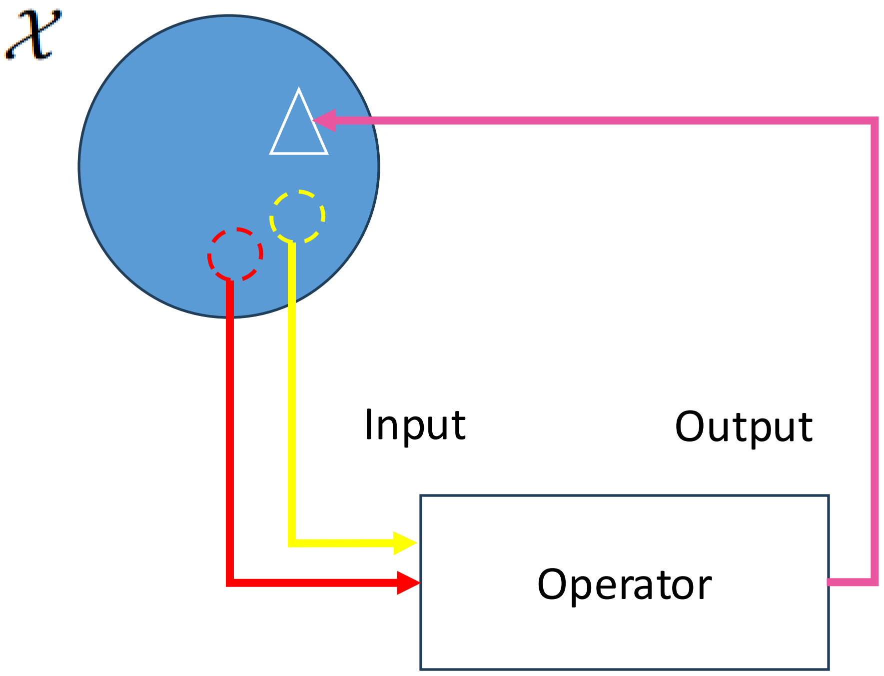
$f: X \times X \to X$

A Very Common Type of Operators

Binary Operations

An operator that takes **two** elements of a set and produces another element of the same set.

Note that this implies "X is closed under f"



$f: X \times X \to X$

HW:
What are nullary, unary, ternary, and n-ary operators?

Some Examples

On the set of real numbers $\mathbb{R}$, $f(a,b) = a + b$

Some Examples

On the set of real numbers $\mathbb{R}$, $f(a, b) = a + b$ is a binary operation

On the set $M(2, \mathbb{R})$ of 2×2 matrices with real entries, $f(A, B) = AB$

Some Examples

On the set of real numbers $\mathbb{R}$, $f(a, b) = a + b$ is a binary operation

On the set $M(2, \mathbb{R})$ of 2×2 matrices with real entries, $f(A, B) = AB$ is a binary operation since the product of two such matrices is a 2×2 matrix.

Some Examples

On the set of real numbers $\mathbb{R}$, $f(a, b) = a + b$ is a binary operation

On the set $M(2, \mathbb{R})$ of 2×2 matrices with real entries, $f(A, B) = AB$ is a binary operation since the product of two such matrices is a 2×2 matrix.

Division on real numbers $\{\boldsymbol{R}, /\}$

Some Examples

On the set of real numbers $\mathbb{R}$, $f(a, b) = a + b$ is a binary operation

On the set $M(2, \mathbb{R})$ of 2×2 matrices with real entries, $f(A, B) = AB$ is a binary operation since the product of two such matrices is a 2×2 matrix.

Division on real numbers $\{\boldsymbol{R}, /\}$

Division is a partial binary operation on set of Real Numbers since $f(a, b) = a/b$ satisfies $f: R \times R \to R$ except for $b = 0$ where it is undefined.

Multiverse & Mathematics

Multiverse & Mathematics

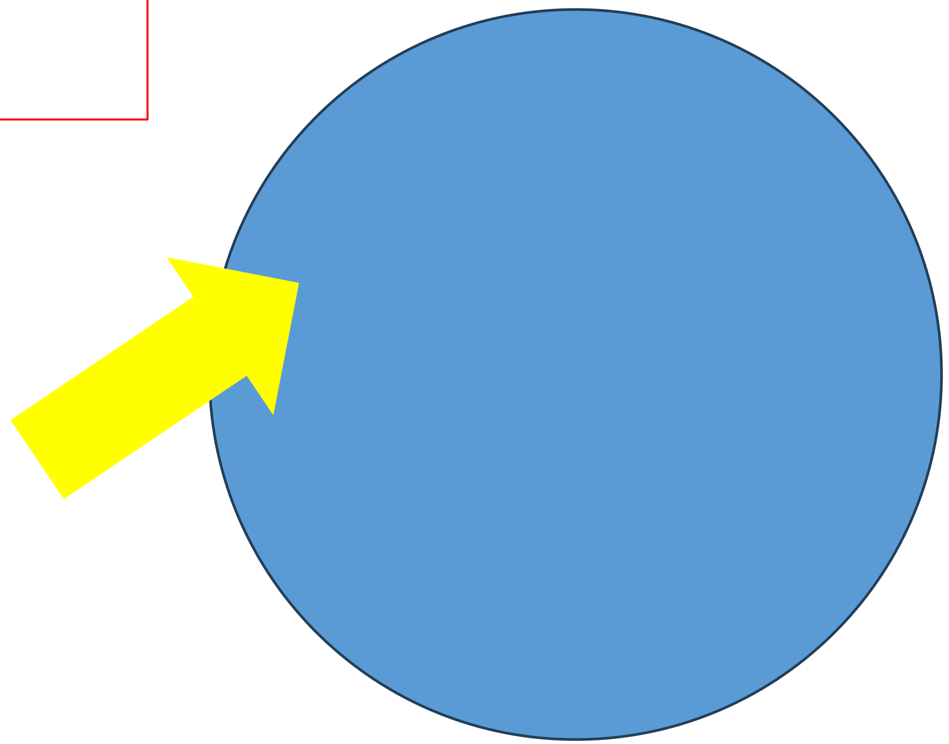
"Multiverse" – many universes, each with its own set of natural laws and objects.

Multiverse & Mathematics

"Multiverse" – many universes, each with its own set of natural laws and objects.

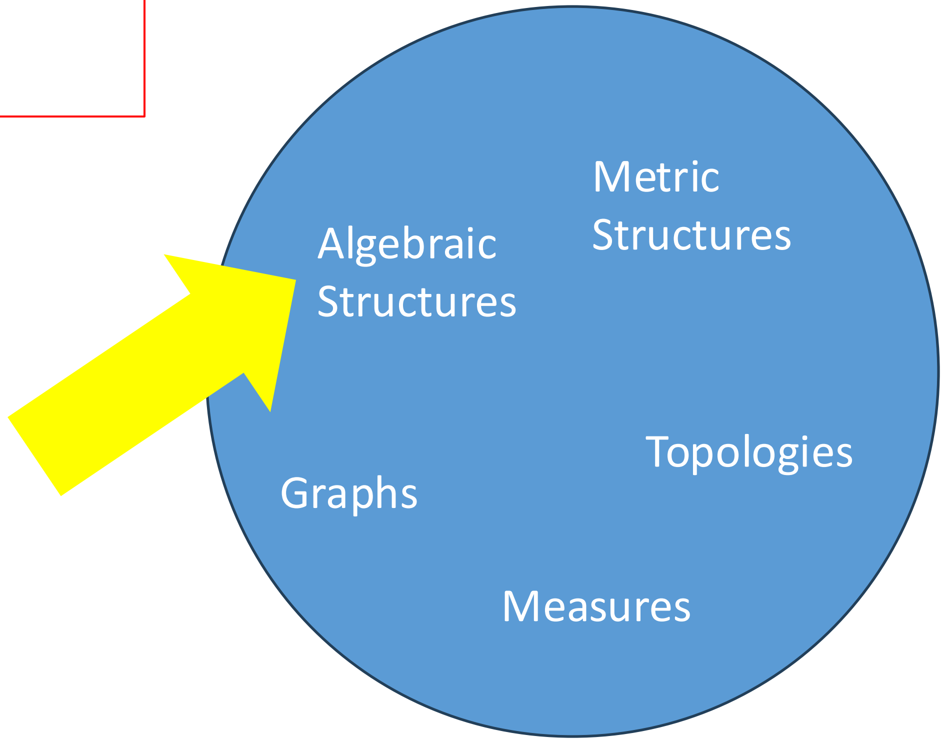
Mathematics also likes to create its own "universes", each with set[s] of objects and "laws" (axioms)

Some Mathematical Multiverses ("Structures")



Some Mathematical Multiverses (“Structures”)

Mathematical Structure = A Set of Objects with Some Features/Laws (operation, relation, distance etc.)



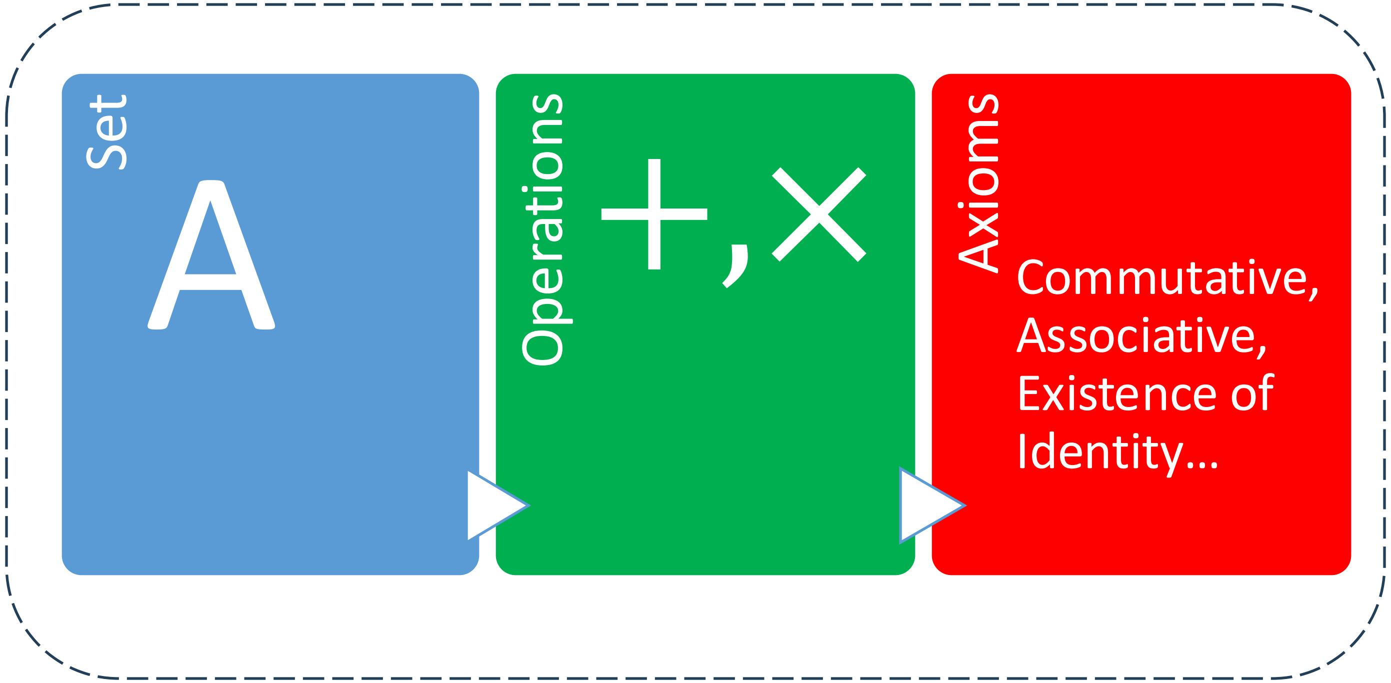
Algebraic Structure

Algebraic Structure

- A Non-Empty Set,
- With a Selection of Operations on the Set,
- And a Finite Set of Axioms (“laws”) that the Operations must follow.

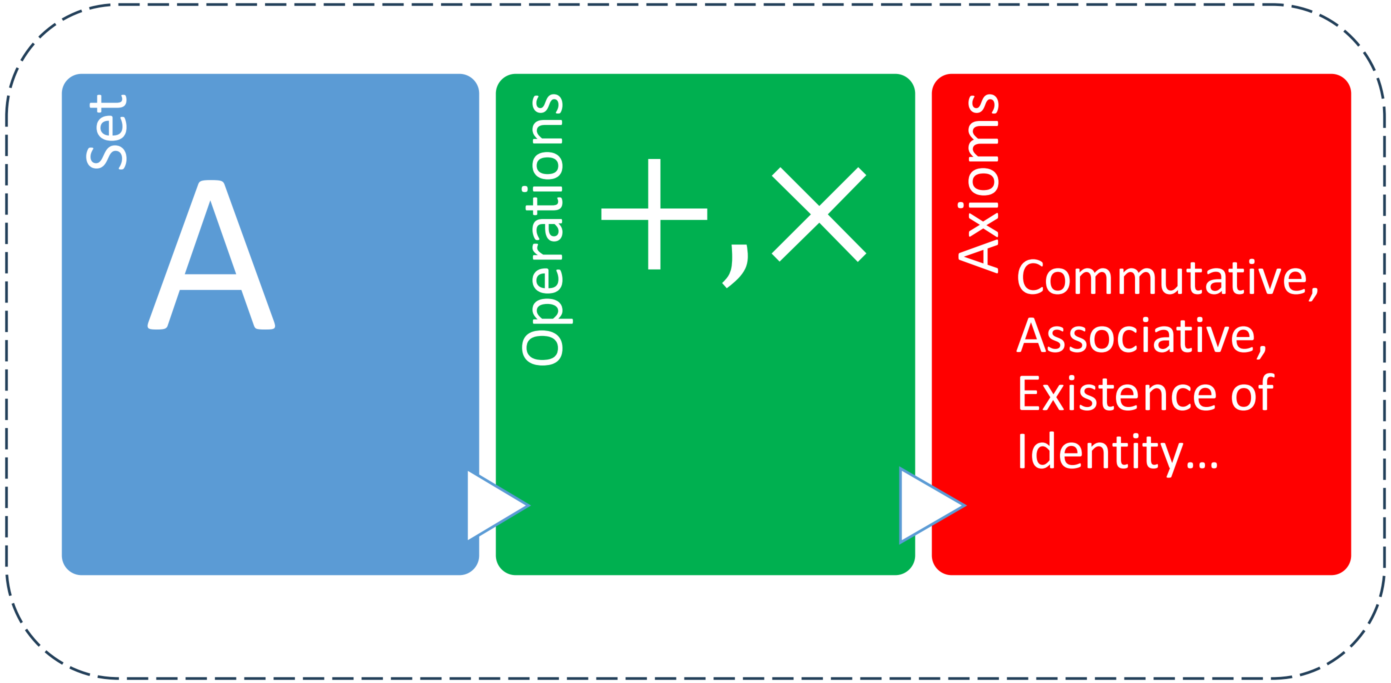
Algebraic Structure

- A Non-Empty Set,
- With a Selection of Operations on the Set,
- And a Finite Set of Axioms (“laws”) that the Operations must follow.

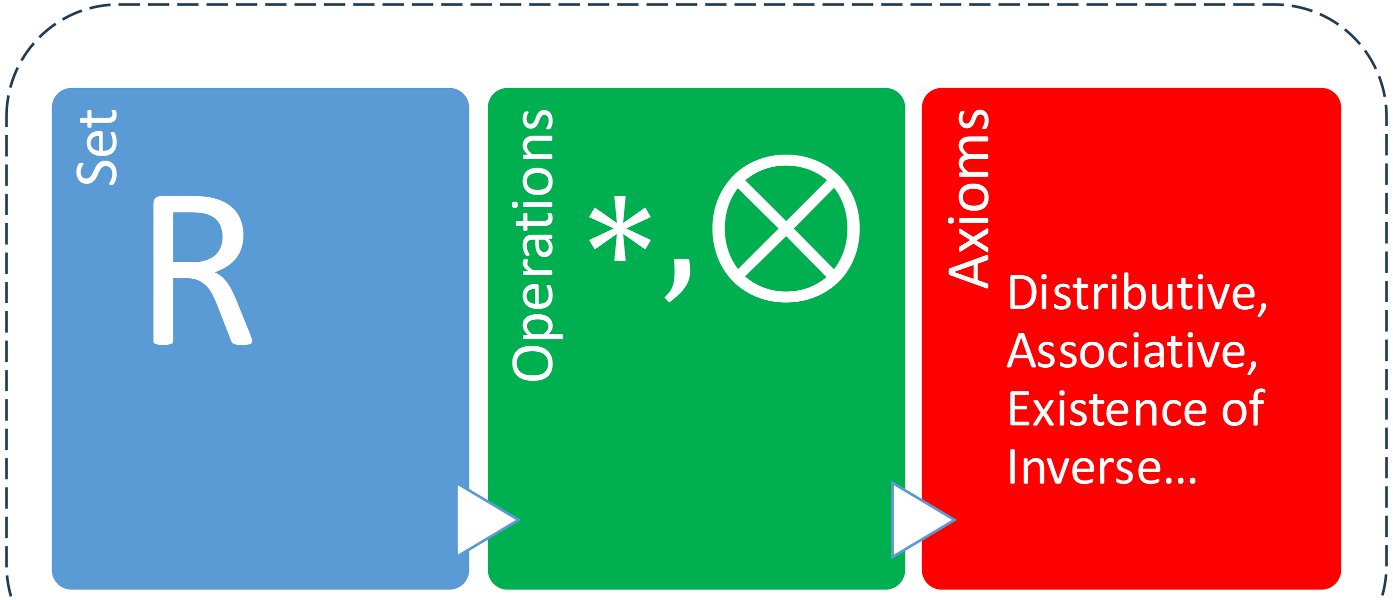


Algebraic Structure

- A Non-Empty Set,
- With a Selection of Operations on the Set,
- And a Finite Set of Axioms (“laws”) that the Operations must follow.



Each choice leads to a different structure (“universe”)



Algebraic Structure – Group

A Set (of objects) **with a Set** (of rules)…

Algebraic Structure – Group

A Set (of objects) with a Set (of rules)...

Axioms

Definition 2.7 (Group). Consider a set $\mathcal{G}$ and an operation $\otimes : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ defined on $\mathcal{G}$. Then $G := (\mathcal{G}, \otimes)$ is called a *group* if the following hold:

1. *Closure* of $\mathcal{G}$ under $\otimes$: $\forall x, y \in \mathcal{G} : x \otimes y \in \mathcal{G}$
2. *Associativity:* $\forall x, y, z \in \mathcal{G} : (x \otimes y) \otimes z = x \otimes (y \otimes z)$
3. *Neutral element:* $\exists e \in \mathcal{G}\, \forall x \in \mathcal{G} : x \otimes e = x$ and $e \otimes x = x$
4. *Inverse element:* $\forall x \in \mathcal{G}\, \exists y \in \mathcal{G} : x \otimes y = e$ and $y \otimes x = e$, where e is the neutral element. We often write x^{-1} to denote the inverse element of x.

Algebraic Structure – Group

A Set (of objects) with a Set (of rules)…

Axioms

Definition 2.7 (Group). Consider a set $\mathcal{G}$ and an operation $\otimes : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ defined on $\mathcal{G}$. Then $G := (\mathcal{G}, \otimes)$ is called a *group* if the following hold:

1. *Closure* of $\mathcal{G}$ under $\otimes$: $\forall x, y \in \mathcal{G} : x \otimes y \in \mathcal{G}$
2. *Associativity:* $\forall x, y, z \in \mathcal{G} : (x \otimes y) \otimes z = x \otimes (y \otimes z)$
3. *Neutral element:* $\exists e \in \mathcal{G} \, \forall x \in \mathcal{G} : x \otimes e = x$ and $e \otimes x = x$
4. *Inverse element:* $\forall x \in \mathcal{G} \, \exists y \in \mathcal{G} : x \otimes y = e$ and $y \otimes x = e$, where e is the neutral element. We often write x^{-1} to denote the inverse element of x.

If, in addition, Commutative Property also holds, then G is called an ……………….

$$\forall x, y \in \mathcal{G} : x \otimes y = y \otimes x,$$

Algebraic Structure – Group

A Set (of objects) with a Set (of rules)...

Axioms

Definition 2.7 (Group). Consider a set $\mathcal{G}$ and an operation $\otimes : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ defined on $\mathcal{G}$. Then $G := (\mathcal{G}, \otimes)$ is called a *group* if the following hold:

1. *Closure* of $\mathcal{G}$ under $\otimes$: $\forall x, y \in \mathcal{G} : x \otimes y \in \mathcal{G}$
2. *Associativity:* $\forall x, y, z \in \mathcal{G} : (x \otimes y) \otimes z = x \otimes (y \otimes z)$
3. *Neutral element:* $\exists e \in \mathcal{G}\, \forall x \in \mathcal{G} : x \otimes e = x$ and $e \otimes x = x$
4. *Inverse element:* $\forall x \in \mathcal{G}\, \exists y \in \mathcal{G} : x \otimes y = e$ and $y \otimes x = e$, where e is the neutral element. We often write x^{-1} to denote the inverse element of x.

If, in addition, Commutative Property also holds, then G is called an **Abelian Group**.

$$\forall x, y \in \mathcal{G} : x \otimes y = y \otimes x,$$

General Linear Group

$$(\boldsymbol{A}, \cdot)$$

Set of $n \times n$ matrices that are also invertible.

Matrix multiplication

General Linear Group

$$(\boldsymbol{A}, \cdot)$$

Set of $n \times n$ matrices that are also invertible.

Matrix multiplication

Definition 2.8 (General Linear Group). The set of regular (invertible) matrices $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is a group with respect to matrix multiplication as defined in (2.13) and is called *general linear group* $GL(n, \mathbb{R})$. However, since matrix multiplication is not commutative, the group is not Abelian.

General Linear Group

$$(\boldsymbol{A}\ ,\cdot)$$

Set of $n \times n$ matrices that are also invertible.

Matrix multiplication

Definition 2.8 (General Linear Group). The set of regular (invertible) matrices $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is a group with respect to matrix multiplication as defined in (2.13) and is called *general linear group* $GL(n, \mathbb{R})$. However, since matrix multiplication is not commutative, the group is not Abelian.

HW:
Show that $(\boldsymbol{A}\ ,\cdot)$ satisfies all four axioms of a Group.

The Algebraic Structure We Are Interested In…

Vector Space

A Set (of vectors) **with a Set** (of inner rules)
and a Set (of outer rules)…

But Let's Dial Back a Bit…

A Game of Knowns (and Unknowns)!

A Simple Life...

One Unknown, One Linear Equation

$$2x = 16$$

$$2x^2 = 128$$

A Simple Life…

One Unknown, One Linear Equation

$$2x = 16$$

Linear in the Unknown

$$2x^2 = 128$$

Nonlinear in the Unknown

A Simple Life…

One Unknown, One Linear Equation

$$2x = 16$$

Has one Unique Solution!

You need just one linear equation to solve for one unknown.

A Simple Life...

One Unknown, One Linear Equation

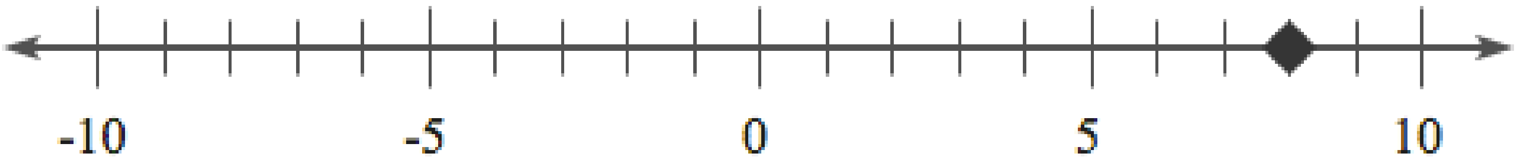
$$2x = 16$$

Has one Unique Solution!

You need just one linear equation to solve for one unknown.

Represents a single point on number line

$$2x = 16$$



One Unknown, No Equation…

Suppose I am asked to find x but no equation is given. Perhaps there are some constraints (e.g., inequality) available?

$$2x \leq 16$$

One Unknown, No Equation…

Suppose I am asked to find x but no equation is given. Perhaps there are some constraints (e.g., inequality) available?

$$2x \leq 16$$

$\boxed{\text{Underdetermined}}$

One Unknown, No Equation…

Suppose I am asked to find x but no equation is given. Perhaps there are some constraints (e.g., inequality) available?

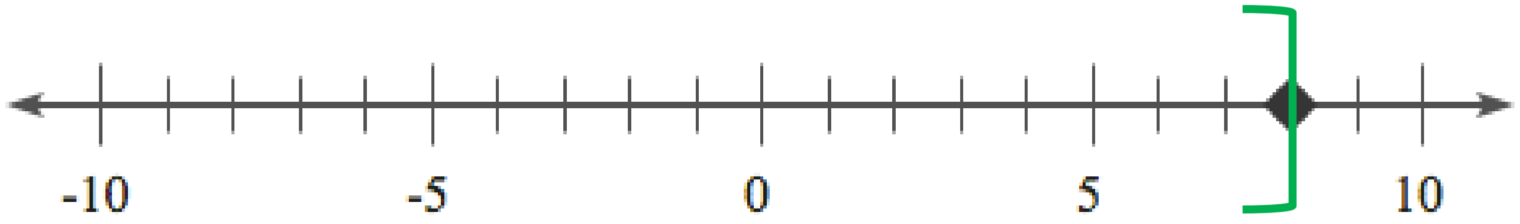
$$2x \leq 16$$

Underdetermined

Infinite Solutions!

Represents an interval on the number line

$$2x \leq 16$$



One Unknown, Two Equations, and **A Scam**…

$$2x = 16$$
$$4x = 32$$

One Unknown, Two Equations, and **A Scam**…

$$2x = 16$$
$$4x = 32$$

It "looks" like I've been given two equations, but clearly one is a scaled version of the other and **carries no new information independent of what the other already provides** (I've been scammed!)

Equations are scaled versions of each other ("Dependent" in 1D)

One Unknown, Two Equations, and **A Scam**…

Overdetermined
(But Consistent!)

$$
\begin{aligned}
2x &= 16 \\
4x &= 32
\end{aligned}
$$

It "looks" like I've been given two equations, but clearly one is a scaled version of the other and **carries no new information independent of what the other already provides** (I've been scammed!)

Equations are scaled versions of each other ("Dependent" in 1D)

One Unknown, Two Equations, and **A Scam**…

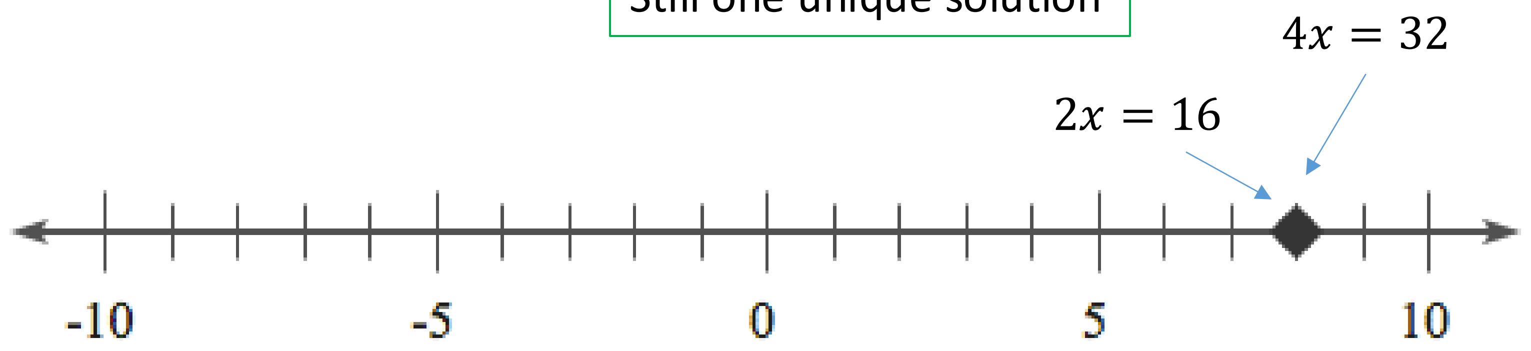
Overdetermined
(But Consistent!)

$$2x = 16$$
$$4x = 32$$

It "looks" like I've been given two equations, but clearly one is a scaled version of the other and **carries no new information independent of what the other already provides** (I've been scammed!)

Equations are scaled versions of each other ("Dependent" in 1D)

Still one unique solution



In fact, a unique point on the number line can only be represented by one distinct linear equation. Any other linear equation representing the same point would necessarily be a scaled version of that equation.

One Unknown, Two Equations, and A Big Fat Lie…

$$2x = 16$$
$$4x = 8$$

One Unknown, Two Equations, and A Big Fat Lie…

Overdetermined
(And Inconsistent!)

$$2x = 16$$
$$4x = 8$$

It “looks” like I’ve been given two equations, but clearly their statements don’t match (somebody’s lying!)

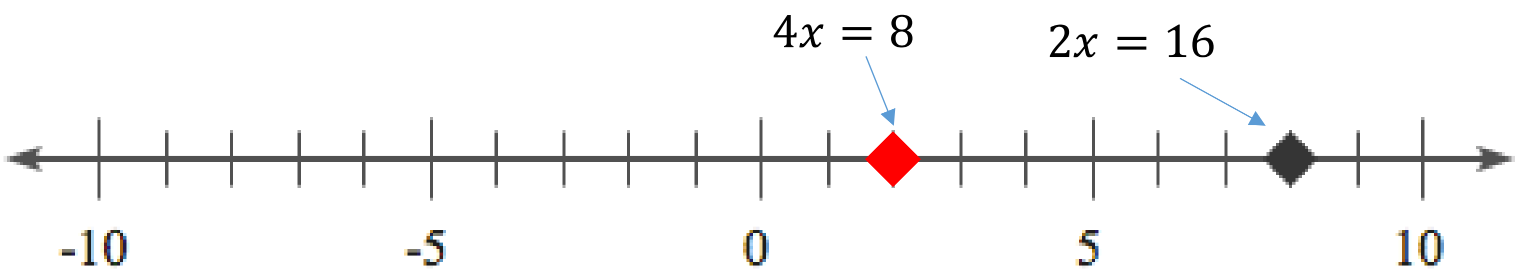
One Unknown, Two Equations, and A Big Fat Lie…

Overdetermined
(And Inconsistent!)

$$2x = 16$$
$$4x = 8$$

It "looks" like I've been given two equations, but clearly their statements don't match (somebody's lying!)

Which One to Pick? No Solution!



One Unknown, Two Equations, and A Big Fat Lie…

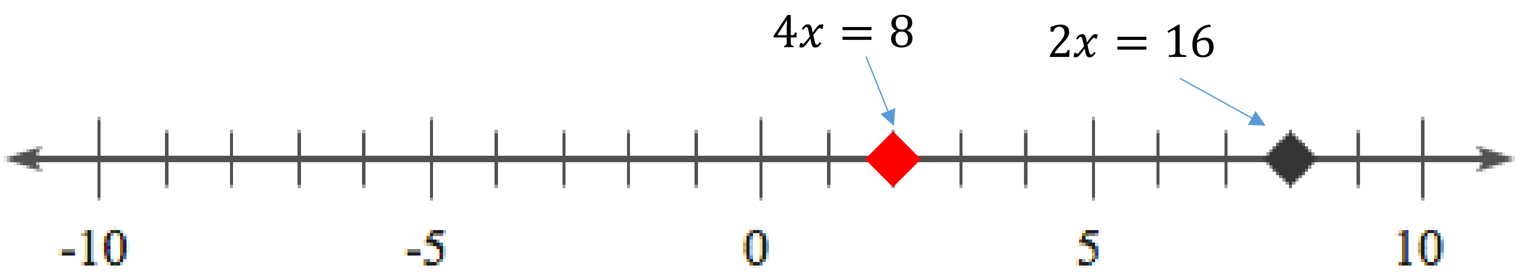
Overdetermined
(And Inconsistent!)

$$\begin{aligned} 2x &= 16 \\ 4x &= 8 \end{aligned}$$

It "looks" like I've been given two equations, but clearly their statements don't match (somebody's lying!)

Which One to Pick? No Solution!

Unless we agree on some sort of a rule to make a choice in such situations.



In fact, a unique point on the number line can only be represented by one distinct linear equation. Any other linear equation representing the same point would necessarily be a scaled version of that equation. Otherwise it represents a different point!

Surprise Surprise...

In real data applications, **we almost always work with overdetermined inconsistent equations**.

Surprise Surprise...

In real data applications, **we almost always work with overdetermined inconsistent equations**.

Say Hello to Randomness
(noise, measurement errors, variations, nature)

Surprise Surprise...

In real data applications, **we almost always work with overdetermined inconsistent equations**.

Overdetermined
(And Inconsistent!)

$$2x = 16$$
$$4.1x = 32$$
$$3.5x = 20.2$$

Say Hello to Randomness
(noise, measurement errors, variations, nature)

Surprise Surprise…

In real data applications, **we almost always work with overdetermined inconsistent equations**.

Overdetermined
(And Inconsistent!)

$$2x = 16$$
$$4.1x = 32$$
$$3.5x = 20.2$$

Say Hello to Randomness
(noise, measurement errors, variations, nature)

No Solution, but could we devise a strategy to pick a point these are "most likely" pointing to?

Surprise Surprise…

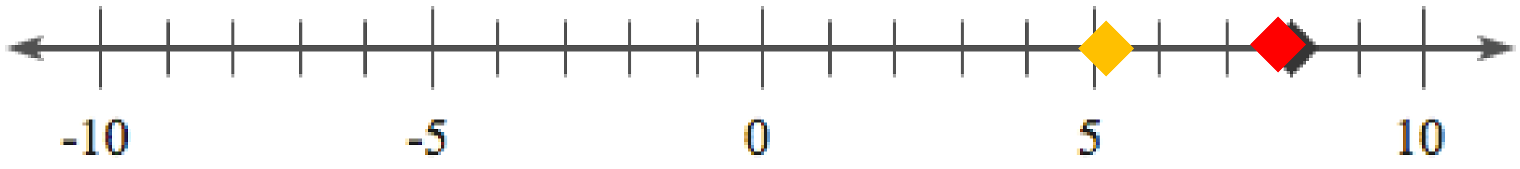
In real data applications, **we almost always work with overdetermined inconsistent equations**.

Overdetermined
(And Inconsistent!)

$$2x = 16$$
$$4.1x = 32$$
$$3.5x = 20.2$$

Say Hello to Randomness
(noise, measurement errors, variations, nature)

No Solution, but could we devise a strategy to **pick a point** these are "most likely" pointing to?



No Solution, but could we devise a strategy to **pick a point** these are “most likely” pointing to?

No Solution, but could we devise a strategy to **pick a point** these are "most likely" pointing to?

Let s_1, s_2, s_3 be the three points represented by the three equations.

No Solution, but could we devise a strategy to **pick a point** these are "most likely" pointing to?

Perhaps choose their Mean as the solution?

$$\hat{x} = \frac{s_1 + s_2 + s_3}{3}$$

Let s_1, s_2, s_3 be the three points represented by the three equations.

No Solution, but could we devise a strategy to **pick a point** these are “most likely” pointing to?

Perhaps choose their Mean as the solution?

$$\hat{x} = \frac{s_1 + s_2 + s_3}{3}$$

Let s_1, s_2, s_3 be the three points represented by the three equations.

Or perhaps something more elaborate?

$$\hat{x} = \arg\min_x [(x - s_1)^2 + (x - s_2)^2 + (x - s_3)^2]$$

No Solution, but could we devise a strategy to **pick a point** these are "most likely" pointing to?

Perhaps choose their Mean as the solution?

$$\hat{x} = \frac{s_1 + s_2 + s_3}{3}$$

Let s_1, s_2, s_3 be the three points represented by the three equations.

Or perhaps something more elaborate?

$$\hat{x} = \arg\min_x [(x - s_1)^2 + (x - s_2)^2 + (x - s_3)^2]$$

(Which, in this case, also leads to the Mean, btw.)

HW:
What does this formulation represent?

In real data applications, **we almost always work with overdetermined inconsistent equations**.

OK, I get that randomness may lead to inconsistent equations, but why "overdetermined"?

$$2x = 16$$
$$4.1x = 32$$
$$3.5x = 20.2$$

In real data applications, **we almost always work with overdetermined inconsistent equations**.

OK, I get that randomness may lead to inconsistent equations, but why "overdetermined"?

$$
\begin{aligned}
2x &= 16 \\
4.1x &= 32 \\
3.5x &= 20.2
\end{aligned}
$$

Since we know there will be randomness (leading to some inconsistencies) we do not want to rely on just one equation. This leads to overdetermined systems.

One Unknown, One Equation, And No Solution?

$$ax = b$$

Solvable if a and b known?

One Unknown, One Equation, And No Solution?

$$ax = b$$

Solvable if a and b known?

What if $a = 0$?

One Unknown, One Equation, And No Solution?

$$ax = b$$

Solvable if a and b known?

What if $a = 0$?

No Solution!

One Unknown, One Equation, And No Solution?

$$ax = b$$

Solvable if a and b known?

What if $a = 0$?

No Solution!

We've either been given an impossible equation

$$0x = 10$$

Or an equation with no new information

$$0x = 0$$

From Points to Lines…

From Points to Lines…

Two Unknowns, and Some Linear Equations

The above discussion nicely generalizes (mostly) to cases of two unknowns.

$$x + y = 5$$
$$\frac{1}{2}x - y = -2$$

Two unknowns, two linear equations, no scamming, no lying, one unique solution!

No linear dependencies

No inconsistencies

From Points to Lines…

Two Unknowns, and Some Linear Equations

The above discussion nicely generalizes (mostly) to cases of two unknowns.

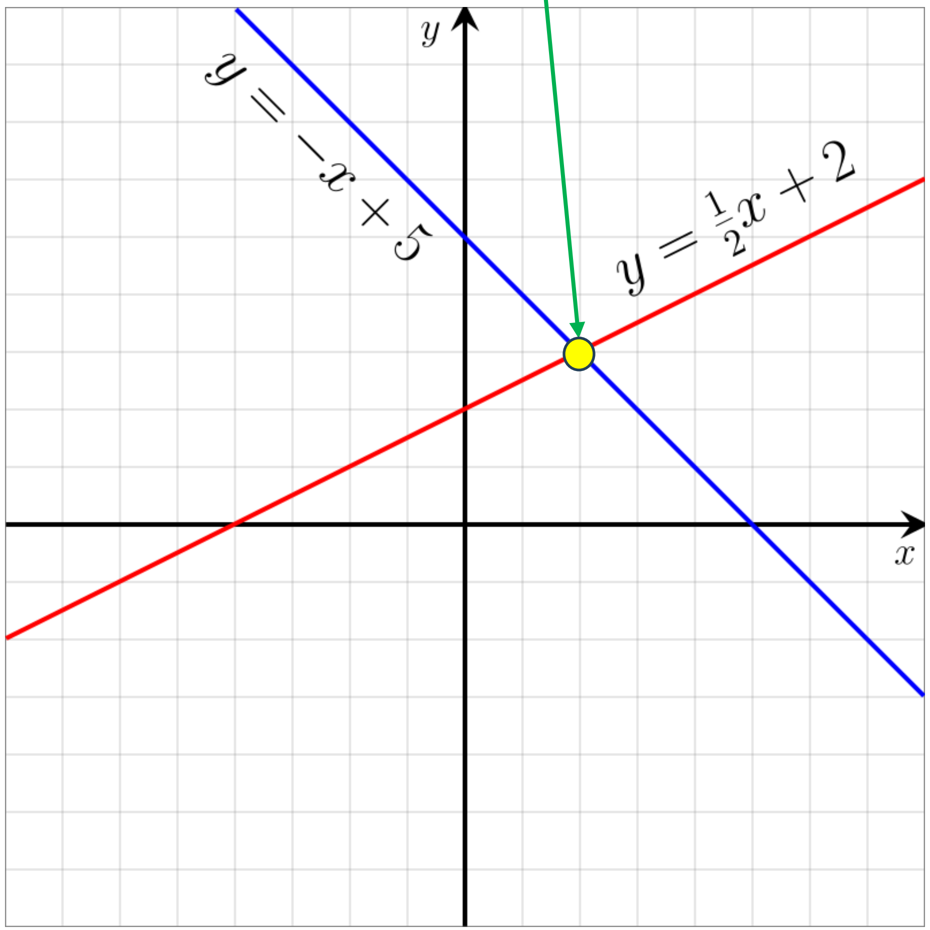
$$x + y = 5$$
$$\frac{1}{2}x - y = -2$$

Two unknowns, two linear equations, no scamming, no lying, one unique solution!

No linear dependencies

No inconsistencies

$(x, y) = (2, 3)$



Two Unknowns, One Equation…

Two Unknowns, One Equation…

Suppose I am asked to find x and y but am given only one linear equation.

$$x + y = 5$$

Two Unknowns, One Equation…

Suppose I am asked to find x and y but am given only one linear equation.

$$x + y = 5$$

Underdetermined

Infinite Solutions!

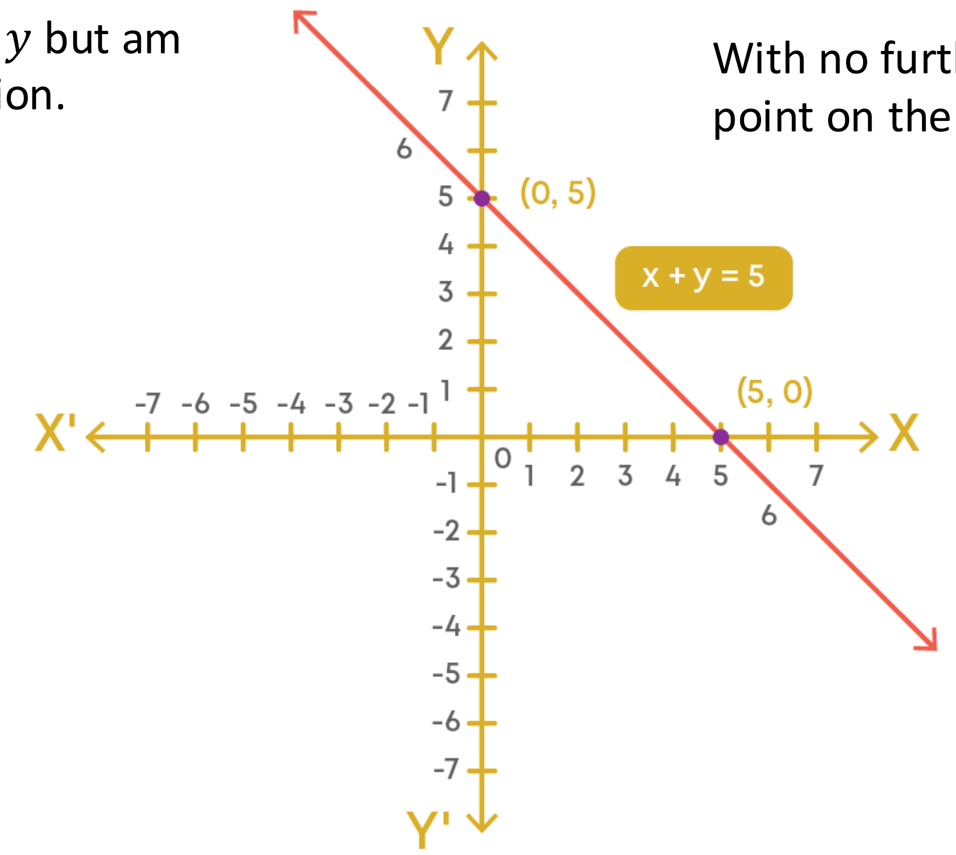
Two Unknowns, One Equation…

Suppose I am asked to find x and y but am given only one linear equation.

$$x + y = 5$$

Underdetermined

Infinite Solutions!



With no further information, every point on the line is a solution.

Two Unknowns, Three Equations, and A Scam…

$$x + 3y = 2$$
$$3x + 9y = 6$$
$$x - y = 0$$

Two Unknowns, Three Equations, and A Scam…

$$x + 3y = 2$$
$$3x + 9y = 6$$
$$x - y = 0$$

Apparently given three equations, but second is simply scaled version of first, and represents the exact same line!

Two Unknowns, Three Equations, and A Scam…

Overdetermined
(But Consistent!)

$$x + 3y = 2$$
$$3x + 9y = 6$$
$$x - y = 0$$

Apparently given three equations, but second is simply scaled version of first, and represents the exact same line!

Still one unique solution

Two Unknowns, Three Equations, and A Scam…

Overdetermined
(But Consistent!)

$$x + 3y = 2$$
$$3x + 9y = 6$$
$$x - y = 0$$

Apparently given three equations, but second is simply scaled version of first, and represents the exact same line!

Still one unique solution

Only two independent equations. So can we say that the **"true worth" (rank?)** of these equations is 2?

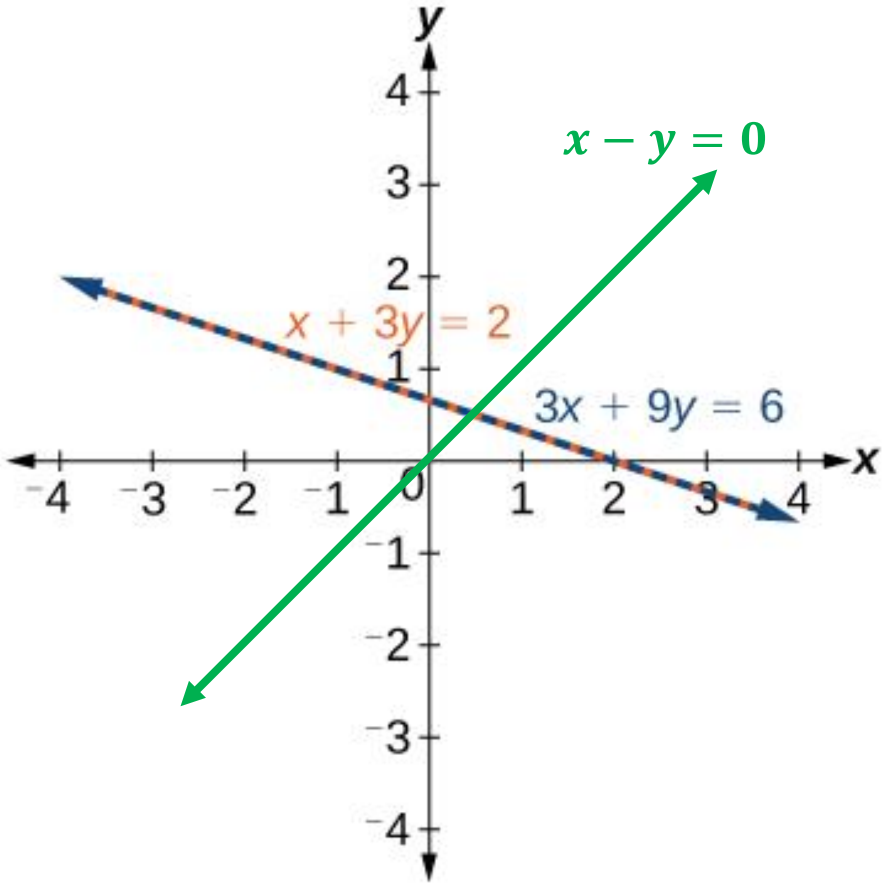
Two Unknowns, Three Equations, and A Scam…

Overdetermined
(But Consistent!)

$$x + 3y = 2$$
$$3x + 9y = 6$$
$$x - y = 0$$

Still one unique solution

Only two independent equations. So can we say that the **"true worth"** **(rank?)** of these equations is 2?



In fact, a unique point on the Cartesian Plane can only be intersection of two distinct lines (equations). Any other line passing through the same point would necessarily be a **simple scaled version** of one of those lines, or a **scaled and summed (linear combination)** version of the two lines.

Recap: Linear Combination (aka Superposition)

Recap: Linear Combination (aka Superposition)

Can be variables, vectors, matrices, equations, …

Given objects $e_1, e_2, \dots, e_k$,

Recap: Linear Combination (aka Superposition)

Can be variables, vectors, matrices, equations, …

Given objects $e_1, e_2, \dots, e_k$, every combination of the form

α_k Scalars

$$y = \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_k e_k$$

"Scaled and Summed"

Recap: Linear Combination (aka Superposition)

Can be variables, vectors, matrices, equations, …

Given objects $e_1, e_2, \dots, e_k$, every combination of the form

α_k Scalars

$$y = \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_k e_k$$

"Scaled and Summed"

is a **linear combination** of $e_1, e_2, \dots, e_k$.

Recap: Linear Combination (aka Superposition)

Can be variables, vectors, matrices, equations, …

Given objects $e_1, e_2, \dots, e_k$, every combination of the form

α_k Scalars

$$y = \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_k e_k$$

"Scaled and Summed"

is a **linear combination** of $e_1, e_2, \dots, e_k$.

Each choice of $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$ leads to a different linear combination.

Can you recall any linear combinations you have seen?

Two Unknowns, Three Equations, and A **Redundancy**…

$$\begin{aligned} x - 2y &= -1 \\ 3x + 5y &= 8 \\ 4x + 3y &= 7 \end{aligned}$$

Two Unknowns, Three Equations, and A **Redundancy**…

$$\begin{aligned} x - 2y &= -1 \\ 3x + 5y &= 8 \\ 4x + 3y &= 7 \end{aligned}$$

Each equation can be formed by scaling and summing the other two, thus it simply contains the same information as in the other two combined ("**Dependent**" in 2D)

Two Unknowns, Three Equations, and A **Redundancy**…

Overdetermined
(But Consistent!)

$$\begin{aligned} x - 2y &= -1 \\ 3x + 5y &= 8 \\ 4x + 3y &= 7 \end{aligned}$$

Each equation can be formed by scaling and summing the other two, thus it simply contains the same information as in the other two combined ("**Dependent**" in 2D)

Only two independent equations.

Two Unknowns, Three Equations, and A **Redundancy**…

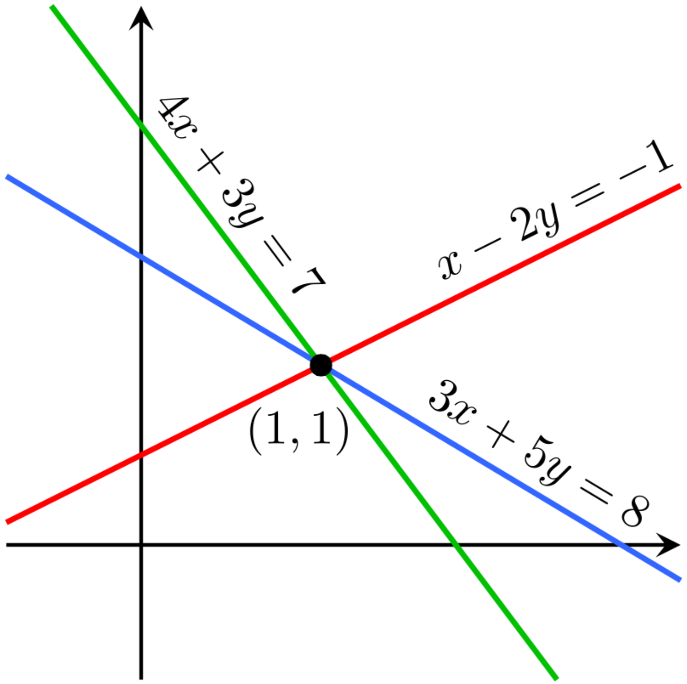
Overdetermined
(But Consistent!)

$$\begin{aligned} x - 2y &= -1 \\ 3x + 5y &= \quad 8 \\ 4x + 3y &= \quad 7 \end{aligned}$$

Each equation can be formed by scaling and summing the other two, thus it simply contains the same information as in the other two combined ("**Dependent**" in 2D)

Only two independent equations.

Still one unique solution



Two Unknowns, Three Equations, and A **Redundancy**…

Overdetermined
(But Consistent!)

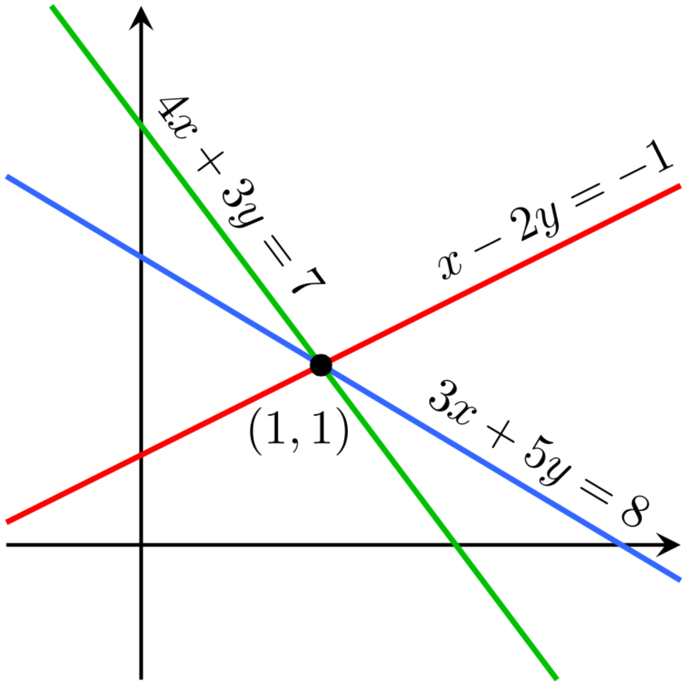
$$\begin{aligned} x - 2y &= -1 \\ 3x + 5y &= 8 \\ 4x + 3y &= 7 \end{aligned}$$

Each equation can be formed by scaling and summing the other two, thus it simply contains the same information as in the other two combined ("**Dependent**" in 2D)

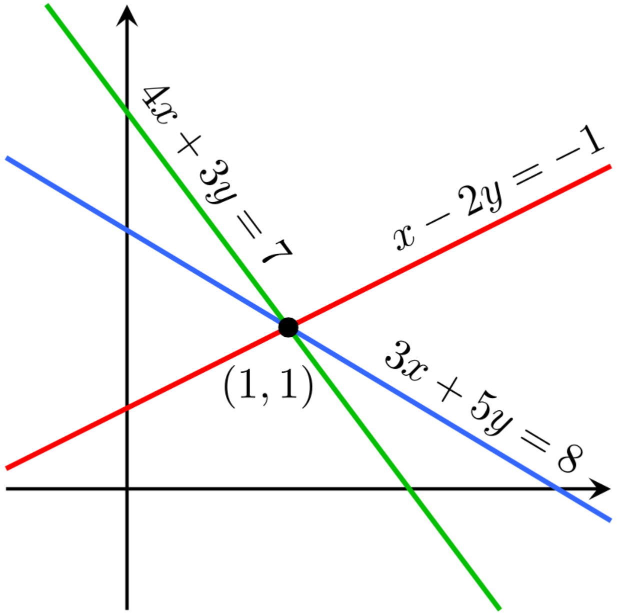
Only two independent equations.

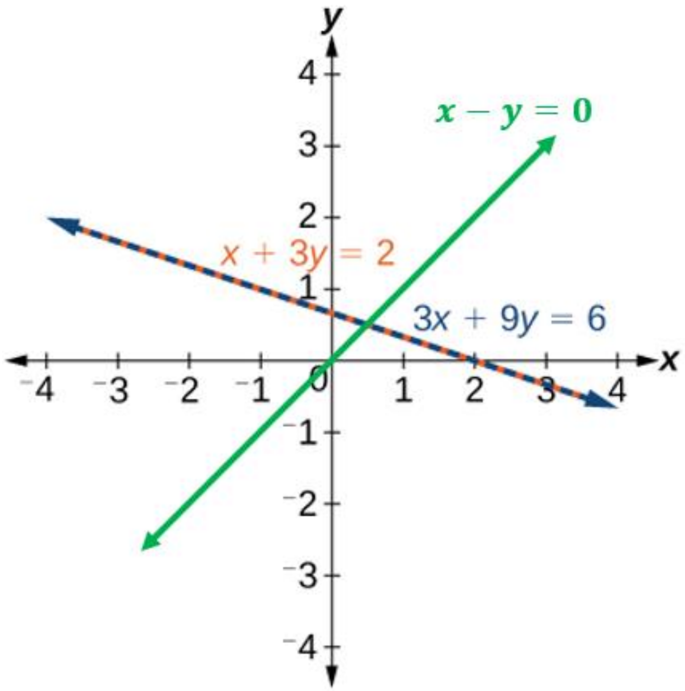
Still one unique solution

Why? **Because two independent equations are enough to solve uniquely for two unknowns**, *as long as any additional equations are not inconsistent with these*.



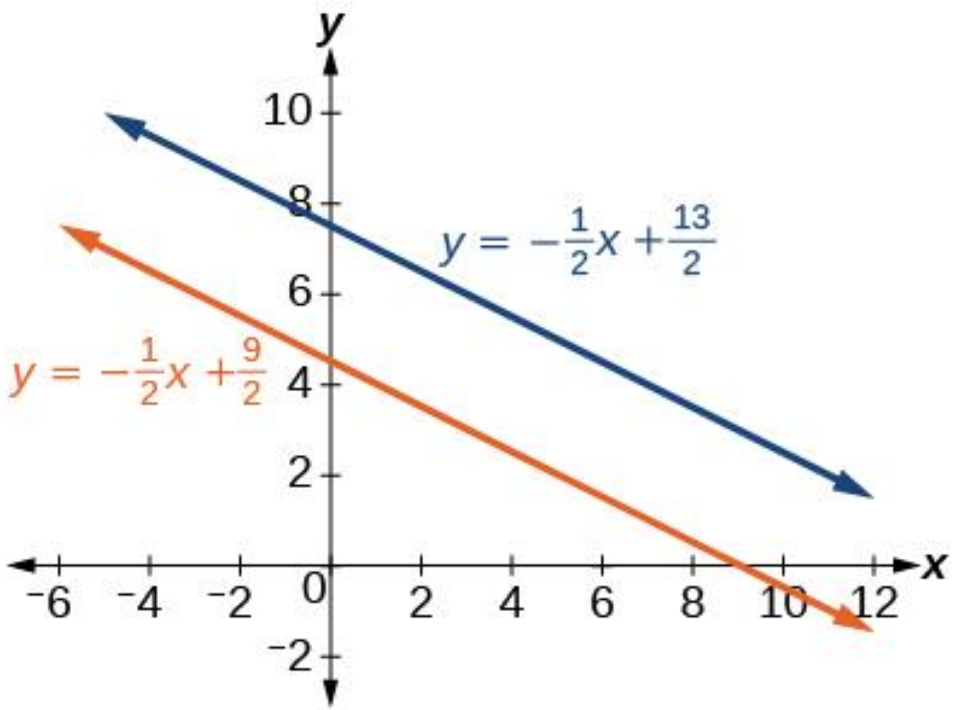
In fact, a unique point on the Cartesian Plane can only be intersection of two distinct lines (equations). Any other line passing through the same point would necessarily be a **simple scaled version** of one of those lines, or a **scaled and summed (linear combination)** version of the two lines.





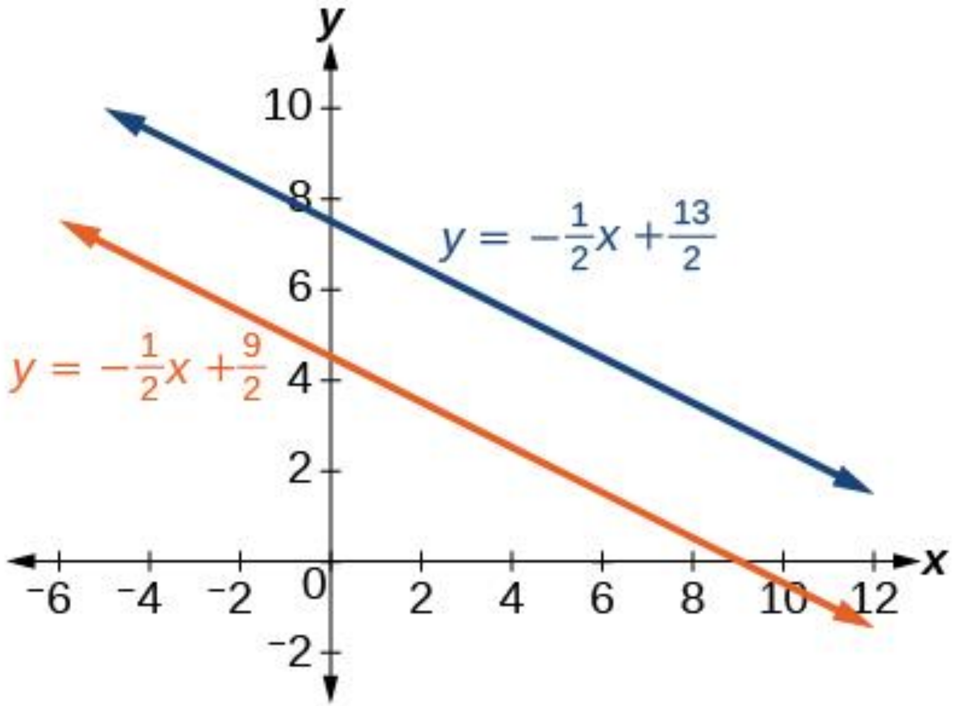
Two Unknowns, Two Equations, and A Big Fat Lie…

Two Unknowns, Two Equations, and A Big Fat Lie…



It "looks" like I've been given two equations, but clearly their statements don't match (somebody's lying!)

Two Unknowns, Two Equations, and A Big Fat Lie…

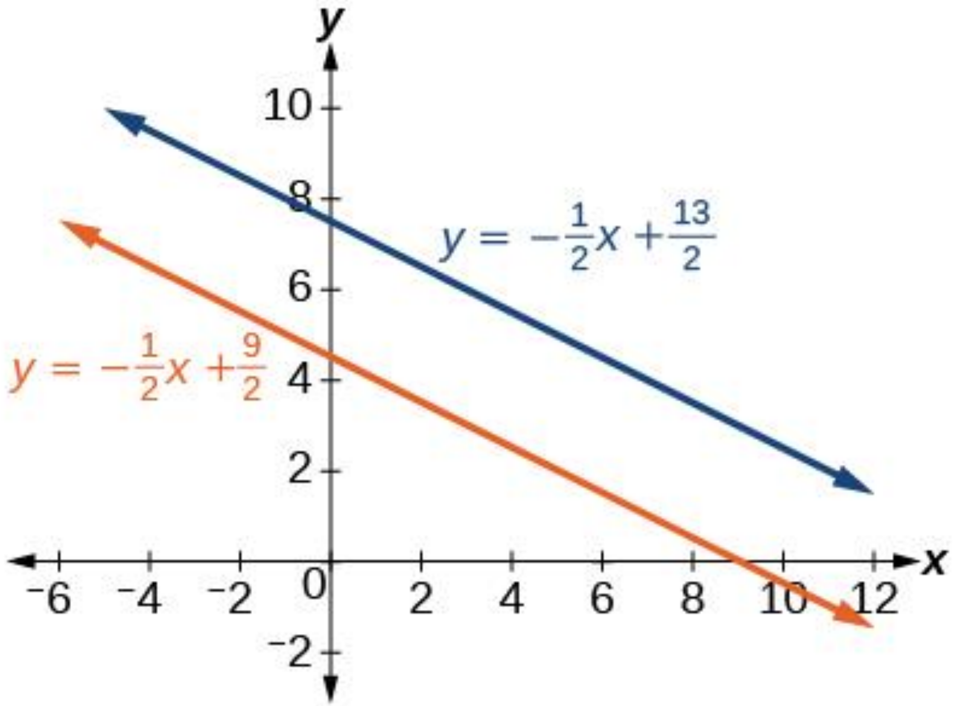


It "looks" like I've been given two equations, but clearly their statements don't match (somebody's lying!)

Inconsistent!

No Intersection (commonly satisfied) Point. No Solution!

Two Unknowns, Two Equations, and A Big Fat Lie…



It "looks" like I've been given two equations, but clearly their statements don't match (somebody's lying!)

HW:
Convince yourself that two mutually inconsistent linear equations in two variables will always lead to two parallel lines.

Inconsistent!

No Intersection (commonly satisfied) Point. No Solution!

Two Unknowns, Three Equations, and Another Lie…

Two Unknowns, Three Equations, and Another Lie…

Overdetermined and Inconsistent!

$$\begin{aligned} x + y &= 2 \\ 2x - y &= 0 \\ x - 2y &= 0 \end{aligned}$$

It "looks" like I've been given two equations, but clearly their statements don't match (somebody's lying!)

Two Unknowns, Three Equations, and Another Lie…

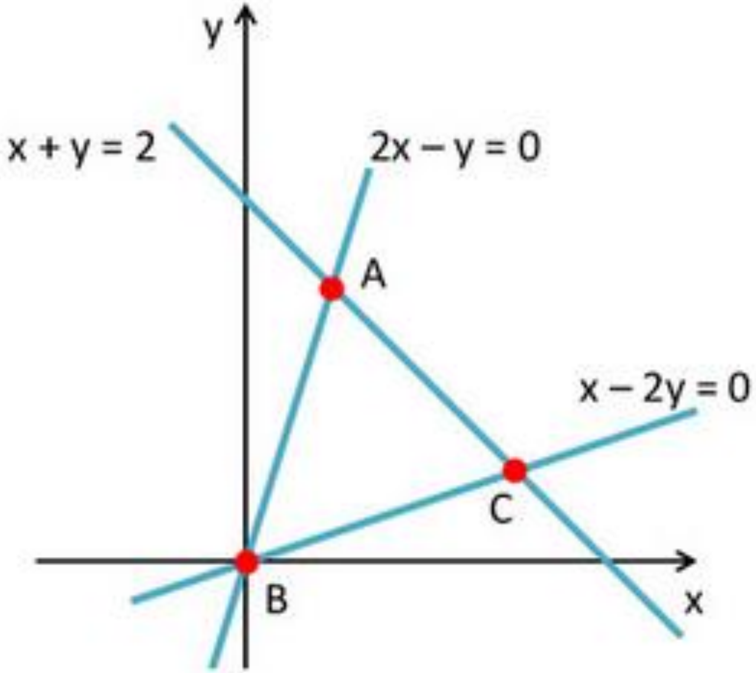
Overdetermined and Inconsistent!

$$\begin{aligned} x + y &= 2 \\ 2x - y &= 0 \\ x - 2y &= 0 \end{aligned}$$

It "looks" like I've been given two equations, but clearly their statements don't match (somebody's lying!)

No Intersection Point (commonly satisfied value).

No Solution!

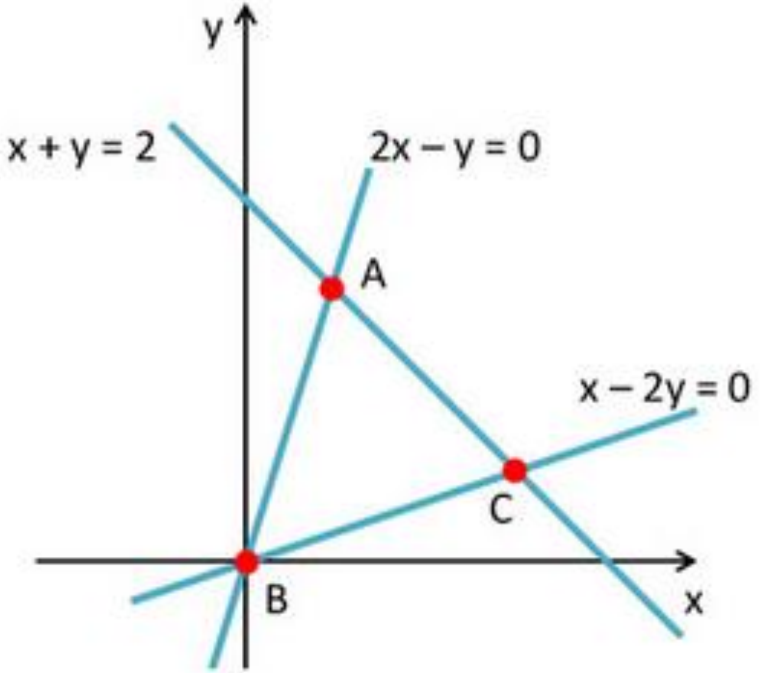


Two Unknowns, Three Equations, and Another Lie…

Overdetermined and Inconsistent!

$$\begin{aligned} x + y &= 2 \\ 2x - y &= 0 \\ x - 2y &= 0 \end{aligned}$$

It "looks" like I've been given two equations, but clearly their statements don't match (somebody's lying!)

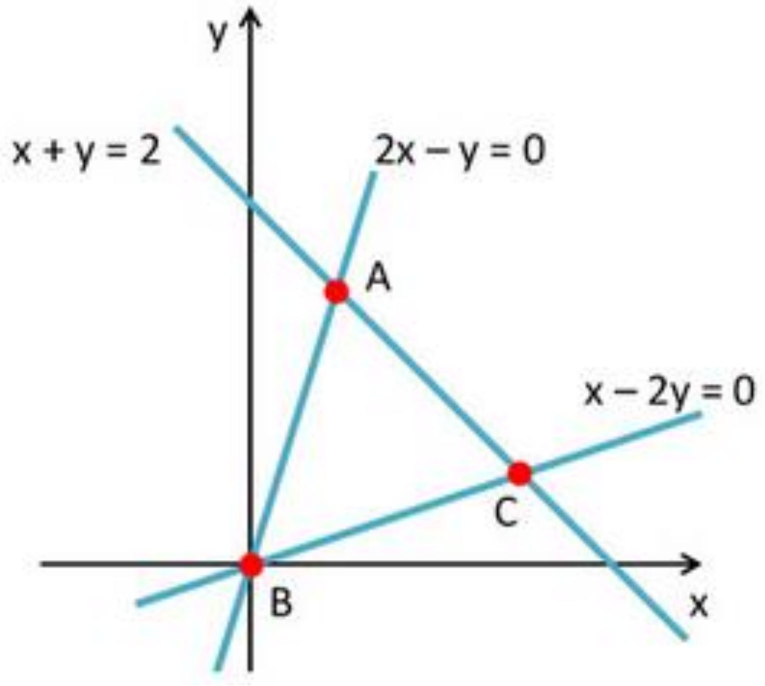


Q. Like the 1D case, is it possible to pick a solution that these equations are "most likely" pointing to?

No Intersection Point (commonly satisfied value).

No Solution!

Q. Like the 1D case, is it possible to pick a solution that these equations are "most likely" pointing to?

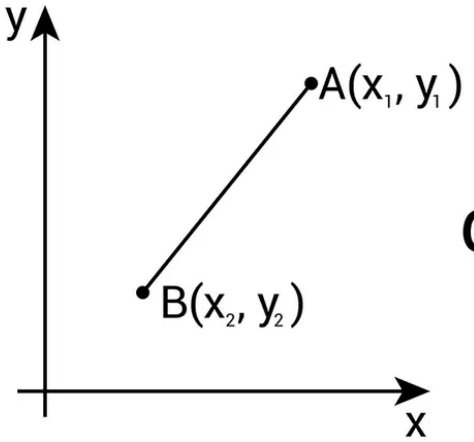


A "compromise" solution with "equal" [in]justice to all candidate points.

Perhaps choose a point that is at **smallest squared [Euclidean] distance** from the three points?

Perhaps choose a point that is at **smallest squared [Euclidean] distance** from the three points?

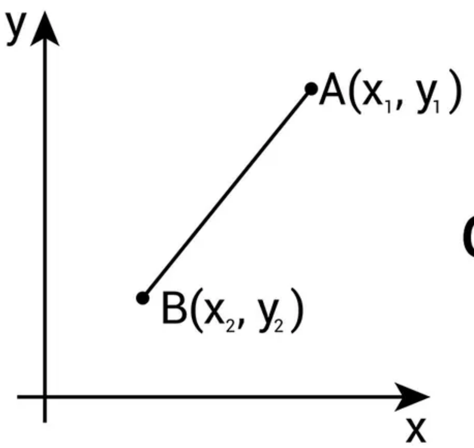
Perhaps choose a point that is at **smallest squared [Euclidean] distance** from the three points?



Euclidean Distance Between Two Points in a Plane

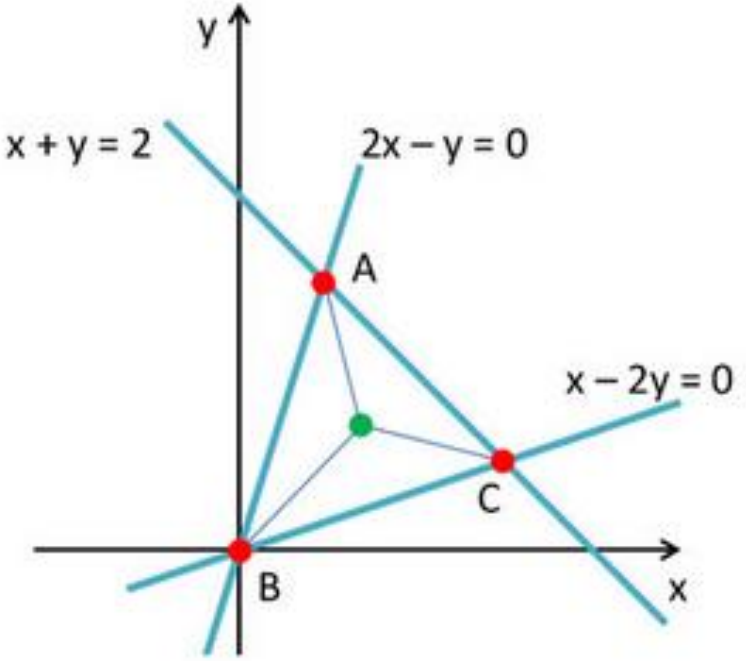
$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Perhaps choose a point that is at **smallest squared [Euclidean] distance** from the three points?



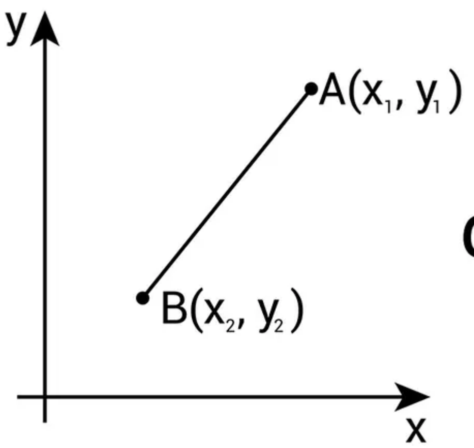
Euclidean Distance Between Two Points in a Plane

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$



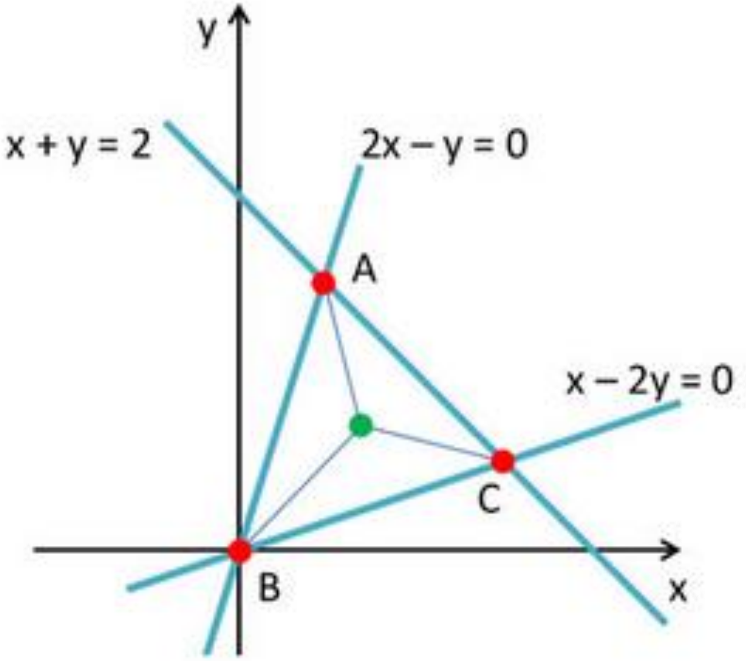
Let points with coordinates be $A(a_1, a_2), B(b_1, b_2), C(c_1, c_2)$

Perhaps choose a point that is at **smallest squared [Euclidean] distance** from the three points?



Euclidean Distance Between Two Points in a Plane

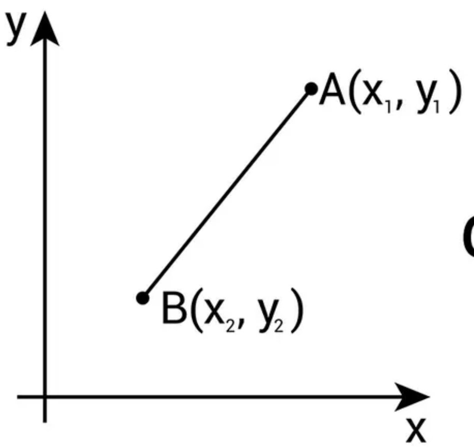
$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$



Let points with coordinates be
$A(a_1, a_2), B(b_1, b_2), C(c_1, c_2)$

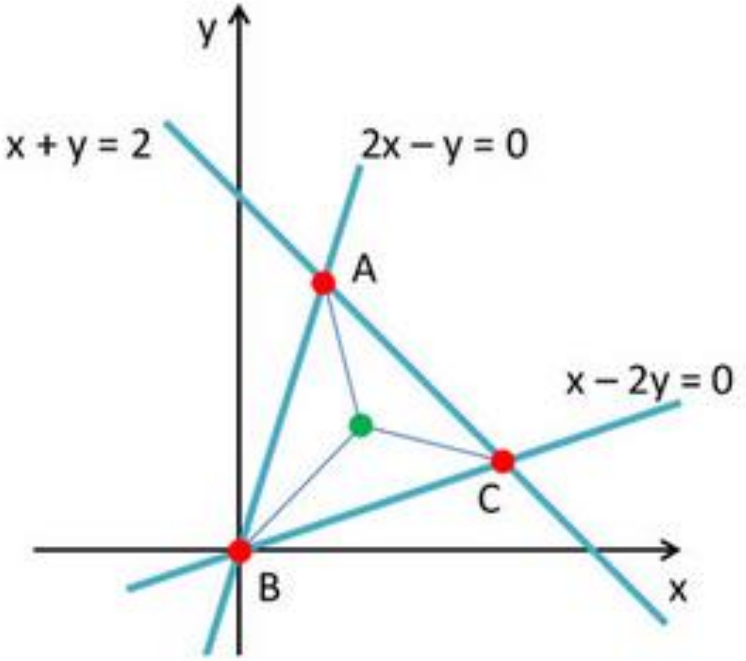
$$(\hat{x}, \hat{y}) = \arg\min_{x,y} \ (x - a_1)^2 + (y - a_2)^2 \\ +(x - b_1)^2 + (y - b_2)^2 \\ +(x - c_1)^2 + (y - c_2)^2$$

Perhaps choose a point that is at **smallest squared [Euclidean] distance** from the three points?



Euclidean Distance Between Two Points in a Plane

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$



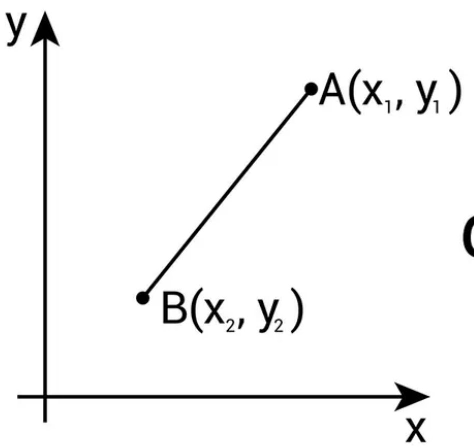
Let points with coordinates be
$A(a_1, a_2), B(b_1, b_2), C(c_1, c_2)$

$$(\hat{x}, \hat{y}) = \arg\min_{x,y} \; (x - a_1)^2 + (y - a_2)^2 + (x - b_1)^2 + (y - b_2)^2 + (x - c_1)^2 + (y - c_2)^2$$

"Least"

"Squares"

Perhaps choose a point that is at **smallest squared [Euclidean] distance** from the three points?



Euclidean Distance Between Two Points in a Plane

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Let points with coordinates be
$A(a_1, a_2), B(b_1, b_2), C(c_1, c_2)$

In real data applications, we almost always work with overdetermined inconsistent equations. And **techniques like these come in handy**!

$$(\hat{x}, \hat{y}) = \arg\min_{x,y} (x - a_1)^2 + (y - a_2)^2 + (x - b_1)^2 + (y - b_2)^2 + (x - c_1)^2 + (y - c_2)^2$$

"Least" "Squares"

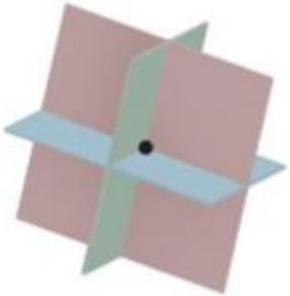
From Lines to Planes (Three Unknowns)…

Concepts we've already discussed easily extend to higher dimensions

From Lines to Planes (Three Unknowns)…

Concepts we've already discussed easily extend to higher dimensions

Exactly One Solution
The planes intersect in a single point, which is the solution of the system.



Infinitely Many Solutions
The planes intersect in a line. Every point on the line is a solution of the system.

The planes could also be the same plane. Every point in the plane is a solution of the system.

No Solution
There are no points in common with all three planes.

From Linear Equations to Matrices...

From Linear Equations to Matrices…

Smarter way of handling systems of linear equations!

From Linear Equations to Matrices...

Smarter way of handling systems of linear equations!

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \,, \end{aligned}$$

$$a_{ij} \in \mathbb{R}$$

From Linear Equations to Matrices…

Smarter way of handling systems of linear equations!

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \,, \end{aligned} \qquad a_{ij} \in \mathbb{R}$$

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

From Linear Equations to Matrices...

Smarter way of handling systems of linear equations!

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

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \, , \end{aligned}$$

$$a_{ij} \in \mathbb{R}$$

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

From Linear Equations to Matrices...

Smarter way of handling systems of linear equations!

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \qquad \boldsymbol{Ax} = \boldsymbol{b}.$$

$\boldsymbol{A} \in \mathbb{R}^{m \times n}$ $\qquad \mathbb{R}^{m \times n}$ is the set of all real-valued (m, n)-matrices

From Linear Equations to Matrices...

Smarter way of handling systems of linear equations!

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \qquad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}.$$

$\boldsymbol{A} \in \mathbb{R}^{m \times n}$ $\quad$ $\mathbb{R}^{m \times n}$ is the set of all real-valued (m, n)-matrices

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Btw, does this remind you of something?

From Linear Equations to Matrices…

Smarter way of handling systems of linear equations!

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \qquad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}.$$

$$\boldsymbol{A} \in \mathbb{R}^{m \times n} \qquad \mathbb{R}^{m \times n} \text{ is the set of all real-valued } (m, n)\text{-matrices}$$

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Interestingly, we are claiming here that vector $\boldsymbol{b}$ is made up of a linear combination of the columns of $\boldsymbol{A}$, and we have to find these linear "weights".

Btw, does this remind you of something?

Some Operations on Matrices…

Some Operations on Matrices…

$$\boldsymbol{A} + \boldsymbol{B} := \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Can’t add matrices of different dimensions!

Some Operations on Matrices…

$$\underset{\textbf{MATRIX 1}}{\begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix}} \times \underset{\textbf{MATRIX 1}}{\begin{bmatrix} 1 & 5 \\ 3 & 7 \end{bmatrix}} = \underset{\textbf{RESULT}}{\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}}$$

Some Operations on Matrices…

$$\underset{\text{MATRIX 1}}{\begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix}} \times \underset{\text{MATRIX 1}}{\begin{bmatrix} 1 & 5 \\ 3 & 7 \end{bmatrix}} = \underset{\text{RESULT}}{\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}}$$

$$\underbrace{A}_{n \times k} \underbrace{B}_{k \times m} = \underbrace{C}_{n \times m}$$

Matrix multiplication is not commutative!

$$AB \neq BA$$

Some Operations on Matrices…

$$\begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 5 \\ 3 & 7 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

MATRIX 1 MATRIX 1 RESULT

Does this remind you of something? Hint: Conv…

$$\underbrace{A}_{n\times k}\underbrace{B}_{k\times m} = \underbrace{C}_{n\times m}$$

Matrix multiplication is not commutative!

$$AB \neq BA$$

Identity Element

$$\boldsymbol{I}_n := \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

Some Properties of Interest…

- *Associativity:* (under matrix multiplication)

$$\forall \boldsymbol{A} \in \mathbb{R}^{m \times n}, \boldsymbol{B} \in \mathbb{R}^{n \times p}, \boldsymbol{C} \in \mathbb{R}^{p \times q} : (\boldsymbol{A}\boldsymbol{B})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{B}\boldsymbol{C})$$

Some Properties of Interest…

- *Associativity:* (under matrix multiplication)

$$\forall \boldsymbol{A} \in \mathbb{R}^{m \times n}, \boldsymbol{B} \in \mathbb{R}^{n \times p}, \boldsymbol{C} \in \mathbb{R}^{p \times q} : (\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$$

- *Distributivity:* (under matrix addition and multiplication)

$$\begin{aligned} \forall \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}, \boldsymbol{C}, \boldsymbol{D} \in \mathbb{R}^{n \times p} : (\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} &= \boldsymbol{AC} + \boldsymbol{BC} \\ \boldsymbol{A}(\boldsymbol{C} + \boldsymbol{D}) &= \boldsymbol{AC} + \boldsymbol{AD} \end{aligned}$$

Some Properties of Interest…

- Multiplication with the identity matrix:

$$\forall \boldsymbol{A} \in \mathbb{R}^{m \times n} : \boldsymbol{I}_m \boldsymbol{A} = \boldsymbol{A} \boldsymbol{I}_n = \boldsymbol{A}$$

Note that $\boldsymbol{I}_m \neq \boldsymbol{I}_n$ for $m \neq n$.

Inverse Element...

Definition 2.3 (Inverse). Consider a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. Let matrix $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ have the property that $\boldsymbol{AB} = \boldsymbol{I}_n = \boldsymbol{BA}$. $\boldsymbol{B}$ is called the *inverse* of $\boldsymbol{A}$ and denoted by $\boldsymbol{A}^{-1}$.

Arthur Cayley

Inverse Element…

Non-invertible matrices are called “**Singular**”.

Definition 2.3 (Inverse)**.** Consider a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. Let matrix $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ have the property that $\boldsymbol{AB} = \boldsymbol{I}_n = \boldsymbol{BA}$. $\boldsymbol{B}$ is called the *inverse* of $\boldsymbol{A}$ and denoted by $\boldsymbol{A}^{-1}$.

Arthur Cayley

Inverse Element...

Non-invertible matrices are called "**Singular**".

Definition 2.3 (Inverse). Consider a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. Let matrix $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ have the property that $\boldsymbol{AB} = \boldsymbol{I}_n = \boldsymbol{BA}$. $\boldsymbol{B}$ is called the *inverse* of $\boldsymbol{A}$ and denoted by $\boldsymbol{A}^{-1}$.

Arthur Cayley

Only square matrices can have inverses (though not all of them do).

Why?

Only square matrices can have inverses (though not all of them do).

Why?

- By definition, inverse requires existence of matrix $\boldsymbol{B}$, such that $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I_n}$.

- Since matrix multiplication is non-commutative and dimensions must conform, a non-square matrix $\boldsymbol{A}$ would necessarily lead to identities of two different dimensions (one for $\boldsymbol{AB}$ and one for $\boldsymbol{BA}$).

Non-square matrices may have a left and/or a right inverse, which will not be both same.

Right inverse: X is a *right inverse* of A if

$$AX = I$$

A is *right-invertible* if it has at least one right inverse

If $\boldsymbol{A}$ is $m \times n$, then $\boldsymbol{X}$ must be $n \times m$, and $I = I_m$

Left inverse: X is a *left inverse* of A if

$$XA = I$$

A is *left-invertible* if it has at least one left inverse

If $\boldsymbol{A}$ is $m \times n$, then $\boldsymbol{X}$ must be $n \times m$, and $I = I_n$

For the two RHS's to be Identity matrices of same dimensions, $\boldsymbol{A}$ must be square!

Can you prove this?

if A has a left **and** a right inverse, then they are equal and unique:

$$XA = I, \quad AY = I$$

- in this case, we call $X = Y$ the **inverse** of A (notation: A^{-1})
- A is *invertible* if its inverse exists

For the two RHS's to be Identity matrices
of same dimensions, $\boldsymbol{A}$ must be square!

Can you prove this?

if A has a left **and** a right inverse, then they are equal and unique:

$$XA = I, \quad AY = I \qquad \Longrightarrow \qquad X = X(AY) = (XA)Y = Y$$

- in this case, we call $X = Y$ the **inverse** of A (notation: A^{-1})
- A is *invertible* if its inverse exists

Transpose and Symmetry

Definition 2.4 (Transpose)**.** For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ the matrix $\boldsymbol{B} \in \mathbb{R}^{n \times m}$ with $b_{ij} = a_{ji}$ is called the *transpose* of $\boldsymbol{A}$. We write $\boldsymbol{B} = \boldsymbol{A}^\top$.

Definition 2.5 (Symmetric Matrix)**.** A matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is *symmetric* if $\boldsymbol{A} = \boldsymbol{A}^\top$.

Some Generally Useful Identities…

Assuming the inverses exist.

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I} = \boldsymbol{A}^{-1}\boldsymbol{A}$$

$$(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$

$$(\boldsymbol{A}+\boldsymbol{B})^{-1} \neq \boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}$$

$$(\boldsymbol{A}^{\top})^{\top} = \boldsymbol{A}$$

$$(\boldsymbol{A}\boldsymbol{B})^{\top} = \boldsymbol{B}^{\top}\boldsymbol{A}^{\top}$$

$$(\boldsymbol{A}+\boldsymbol{B})^{\top} = \boldsymbol{A}^{\top}+\boldsymbol{B}^{\top}$$

Between matrices themselves.

So far, we've looked at **Inner Operations/Properties**.

E.g., between matrices and scalars?

How about some **Outer Operations/Properties?**

Multiplication with Scalars…

Outer Operation

$$\text{If } A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}_{m \times n} \qquad kA = \begin{bmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \vdots & \vdots & & \vdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{bmatrix}_{m \times n}$$

$$k \in \boldsymbol{R}$$

Multiplication with Scalars...

Outer Operation

$\lambda, \psi \in \mathbb{R},$

- *Associativity:*
  $(\lambda\psi)\boldsymbol{C} = \lambda(\psi\boldsymbol{C}), \quad \boldsymbol{C} \in \mathbb{R}^{m\times n}$
- $\lambda(\boldsymbol{BC}) = (\lambda\boldsymbol{B})\boldsymbol{C} = \boldsymbol{B}(\lambda\boldsymbol{C}) = (\boldsymbol{BC})\lambda, \quad \boldsymbol{B} \in \mathbb{R}^{m\times n}, \boldsymbol{C} \in \mathbb{R}^{n\times k}.$

Multiplication with Scalars...

Outer Operation

$\lambda, \psi \in \mathbb{R}$,

- *Associativity:*
  $(\lambda\psi)\boldsymbol{C} = \lambda(\psi\boldsymbol{C}), \quad \boldsymbol{C} \in \mathbb{R}^{m\times n}$
- $\lambda(\boldsymbol{BC}) = (\lambda\boldsymbol{B})\boldsymbol{C} = \boldsymbol{B}(\lambda\boldsymbol{C}) = (\boldsymbol{BC})\lambda, \quad \boldsymbol{B} \in \mathbb{R}^{m\times n}, \boldsymbol{C} \in \mathbb{R}^{n\times k}$.
  Note that this allows us to move scalar values around.
- $(\lambda\boldsymbol{C})^\top = \boldsymbol{C}^\top\lambda^\top = \boldsymbol{C}^\top\lambda = \lambda\boldsymbol{C}^\top$ since $\lambda = \lambda^\top$ for all $\lambda \in \mathbb{R}$.

Multiplication with Scalars…

Outer Operation

$\lambda, \psi \in \mathbb{R}$,

- *Associativity:*
  $(\lambda\psi)\boldsymbol{C} = \lambda(\psi\boldsymbol{C}), \quad \boldsymbol{C} \in \mathbb{R}^{m\times n}$
- $\lambda(\boldsymbol{BC}) = (\lambda\boldsymbol{B})\boldsymbol{C} = \boldsymbol{B}(\lambda\boldsymbol{C}) = (\boldsymbol{BC})\lambda, \quad \boldsymbol{B} \in \mathbb{R}^{m\times n}, \boldsymbol{C} \in \mathbb{R}^{n\times k}$.
  Note that this allows us to move scalar values around.
- $(\lambda\boldsymbol{C})^\top = \boldsymbol{C}^\top\lambda^\top = \boldsymbol{C}^\top\lambda = \lambda\boldsymbol{C}^\top$ since $\lambda = \lambda^\top$ for all $\lambda \in \mathbb{R}$.
- *Distributivity:*
  $(\lambda + \psi)\boldsymbol{C} = \lambda\boldsymbol{C} + \psi\boldsymbol{C}, \quad \boldsymbol{C} \in \mathbb{R}^{m\times n}$
  $\lambda(\boldsymbol{B} + \boldsymbol{C}) = \lambda\boldsymbol{B} + \lambda\boldsymbol{C}, \quad \boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{m\times n}$

How Do We Solve Systems of Linear Equations?

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

How Do We Solve Systems of Linear Equations?

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

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}.$$

How Do We Solve Systems of Linear Equations?

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

$\boxed{\text{Hardly Ever!}}$

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}.$$

How Do We Solve Systems of Linear Equations?

$$Ax = b.$$

Hardly Ever!

$$x = A^{-1}b.$$

Why?

- As we saw, in practice, we rarely have square matrices (dealing instead with overdetermined inconsistent systems).

- Even for square matrices, inverse may not exist.

- Direct matrix inversion is computationally costly and prone to instability.

How Do We Solve Systems of Linear Equations?

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

Instead, many algorithmic and iterative procedures exist for

- Finding exact or approximate solutions where unique solutions exist.
- Finding a “best” candidate solution where many solutions exist.

$$\boldsymbol{x} \cancel{= \boldsymbol{A}}^{-1}\boldsymbol{b}.$$

How Do We Solve Systems of Linear Equations?

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

Instead, many algorithmic and iterative procedures exist for

- Finding exact or approximate solutions where unique solutions exist.
- Finding a “best” candidate solution where many solutions exist.

$$\boldsymbol{x} = \cancel{\boldsymbol{A}^{-1}}\boldsymbol{b}.$$

e.g.,

Transform problem to a function minimization!

Solve Ax=b

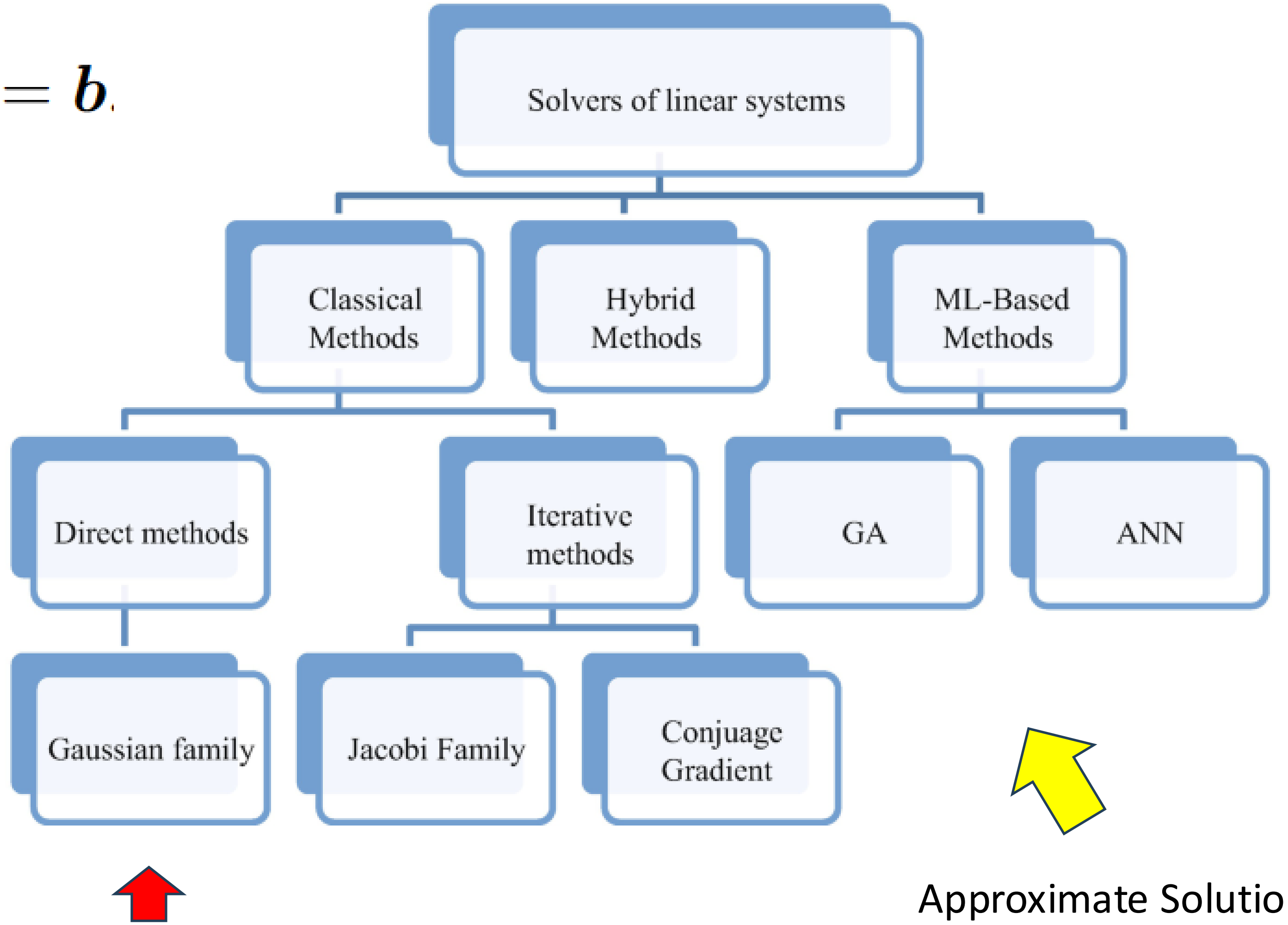
$\Rightarrow$ Minimize $f(x) = x^T A x - 2b^T x$

Scalar!

In fact, this aims to find the solution that minimizes the **total squared error** (*have we seen this before?)*

How Do We Solve Systems of Linear Equations?

$$Ax = b.$$



Approximate Solutions

Exact Solutions

Perhaps now we are ready!

Next time…

Questions?? Thoughts??