SIT787: Mathematics for Artificial Intelligence
Preliminaries and Calculus
Topic 1

Asef Nazari

School of Information Technology, Deakin University

Sets

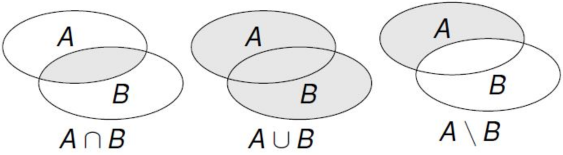
- A **set** is a collection of elements
  - $A = \{1, 5, -2, 7\} = \{1, 1, 5, -2, -2, 7, 7, 7\}$
  - $B = \{\text{monkey}, \text{money}, \text{key}\}$
- **Membership**
  - $1 \in A$ and $100 \notin A$
  - money $\in B$ and pool $\notin B$
- **Subsets**
  - $A \subset C$ if every element of A is in C
    Consider $A = \{1, 5, -2, 7\}$ and $C = \{1, 5, 8, -2, 7, 12\}$. We say $A \subset C$ but $C \not\subset A$.
- **Cardinality** of a set is the number of elements in that set
  - $|A| = 4$, $|B| = 3$, $|C| = 6$
- **Empty set**, $\emptyset$, has no member. $|\emptyset| = 0$.

- **Universal set U:** the set that contains all elements under consideration.
- **Set compliment:** The complement of A is the set of all elements in the universal set U, but not in A, and is denoted by A^c

$$A^c = \{x | x \in U \text{ and } x \notin A\}$$

Operations on Sets

- **Union** of A and B is $A \cup B$
  - $A \cup B = \{x | x \in A \text{ OR } x \in B\}$
- **Intersection** of A and B is $A \cap B$
  - $A \cap B = \{x | x \in A \text{ AND } x \in B\}$
- **Set difference** of A and B is $A \backslash B$ or $A - B$
  - $A \backslash B = \{x | x \in A \text{ AND } x \notin B\} = A \cap B^c = A - B$



- You search and learn **symmetric difference** between two sets

$$A \Delta B$$

Inclusion-exclusion Theorem

- **Inclusion-exclusion Theorem**

$$|A \cup B| = |A| + |B| - |A \cap B|$$

- Example: consider $A = \{1, 5, -7, 3\}$ and $B = \{1, -5, -7, 4\}$
  - Find $A \cup B$ and $A \cap B$
  - Find $|A|, |B|, |A \cup B|$ and $|A \cap B|$
  - Verify the Inclusion-exclusion Theorem.

Scalers

- Scalers are just numbers, like 1, $-\frac{1}{3}$, $\frac{\pi}{3}$
- Frequent sets of scalers
  - Natural Numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$ or the counting numbers
  - Integers $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ or

$$\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

  - Rational Numbers $\mathbb{Q}$ numbers of the form $\frac{m}{\mathbf{n}}$ where $n \neq 0$ and $m, n \in \mathbb{Z}$
    - number can be written as a Ratio of two integers
    - fractional, decimal $\frac{3}{2} = 1.5$ or $7 = \frac{7}{1}$
    - what about $\frac{1}{3} = 0.333333...$?
  - The decimal representation of a rational number always either terminates after a finite number of digits or begins to repeat the same digits over and over.
  - any repeating or terminating decimal represents a rational number.

Scalers

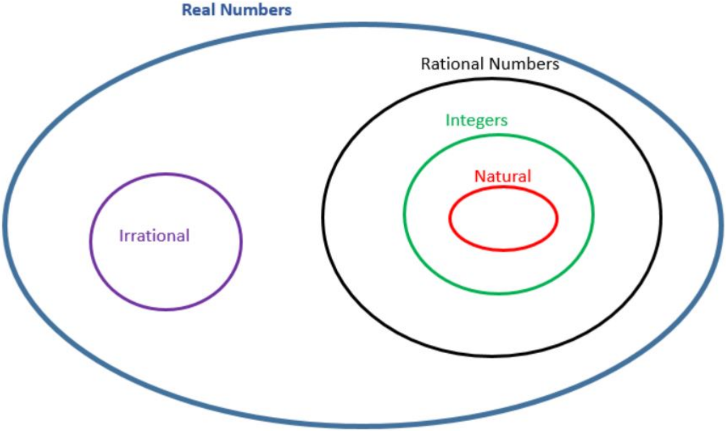
- Irrational Numbers $\mathbb{I}$:
  - numbers cannot be written as a ratio of two integers
  - $\pi = 3.14159265358979323846264338327 95...$
  - $e = 2.71828182845904523536028747135 27...$
  - Surds: $\sqrt{2}$
  - numbers non-terminating decimal with no pattern
- Real Numbers $\mathbb{R}$: all numbers mentioned above all together

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$$

$$\mathbb{I} \subset \mathbb{R}$$

- $\mathbb{Q}^c = \mathbb{I}$ and $\mathbb{I}^c = \mathbb{Q}$
- $\mathbb{Q} \cup \mathbb{I} = \mathbb{R}$

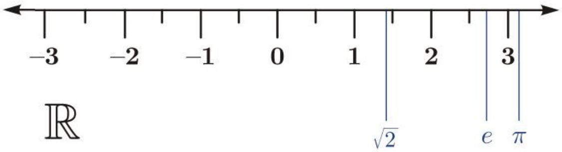
Scalers



Real Numbers
Rational Numbers
Integers
Natural
Irrational

Representing $\mathbb{R}$

- Real numbers can be shown on a line (two-sided arrow)
  - reference point as origin
  - Positive and negative direction
  - coordinate line or real line
  - The real numbers are ordered
- Concept of ∞ and $-\infty$
- If you have a real variable x, it could be everywheree on this two-sided arrow



Inequalities

- $x = y$, x is equal to y
- $x < y$, x is less than y
- $y < x$, x is greater than y
- $x \leq y$, x is less than or equal to y
- $x \geq y$, y is greater than or equal to x
- $x \ll y$, x is much, much less than y
- important considerations
  - If $x = y$, then for any real number a, $ax = ay$
  - if $x \leq y$
    - if a is a positive number, $ax \leq ay$
    - if a is a negative number, $ax \geq ay$

Intervals

- **Closed** intervals $[a, b] = \{x | a \leq x \leq b\}$
- **Open** intervals $(a, b) = \{x | a < x < b\}$
- **Half-open** intervals ,
  - $(a, b] = \{x | a < x \leq b\}$
  - $[a, b) = \{x | a \leq x < b\}$
- **Rays**
  - $[a, \infty) = \{x | x \geq a\}$
  - $(-\infty, a) = \{x | x < a\}$
  - similarly, (a, ∞), , $(-\infty, a]$

Summation and Product notations

Sigma Notation for Sums

- $\sum_{i=2}^{6} i = 2+3+4+5+6$
- $\sum_{i=2}^{6} i^2 = (2)^2+(3)^2+(4)^2+(5)^2+(6)^2 = 4+9+16+25+36$
- $\sum_{k=2}^{6} (2k+1) =$
  $(2(2)+1)+(2(3)+1)+(2(4)+1)+(2(5)+1)+(2(6)+1)$
- $\sum_{j=0}^{3} (\frac{j}{k}) = (\frac{0}{k}) + (\frac{1}{k}) + (\frac{2}{k}) + (\frac{3}{k})$
- $\sum_{j=0}^{3} 5 = 5+5+5+5 = (3-0+1)5$
- $\sum_{j=m}^{n} a = (n-m+1)a$
- Check that $\sum_{j=0}^{3} (j^2+2j) = \left(\sum_{j=0}^{3} j^2\right) + \left(2\sum_{j=0}^{3} j\right)$

Sigma Notation for Sums

- $\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$
- Suppose set $X = \{x_1, x_2, \ldots, x_n\}$ is given.
  - The average of the values of X

$$\bar{x} = \frac{\text{add all values}}{\text{divide by } |X|} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Notation for products

- $\prod_{i=1}^{n} x_i = x_1 x_2 \dots x_n$
- $\prod_{i=2}^{5} x_i = x_2 x_3 x_4 x_5$
- For a nonegative integer n, the factorial is defined as

$$n! = n(n-1)(n-2)\dots(3)(2)(1)$$

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120, \text{ also } 1! = 0! = 1$$

- We can represent factorial as

$$n! = \prod_{i=1}^{n} i$$

Coordinate Systems and Distance

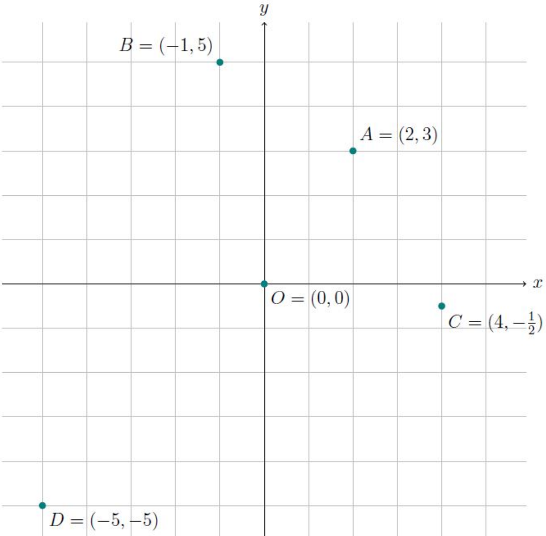
Coordinates of $\mathbb{R}^2$ (Cartesian Plane)

- Points on a line can be identified with real numbers
- Points on a plane can be identified by with ordered pairs of numbers

$$(x, y)$$

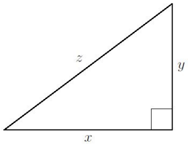
- axis and quadrants

Cartesian plane



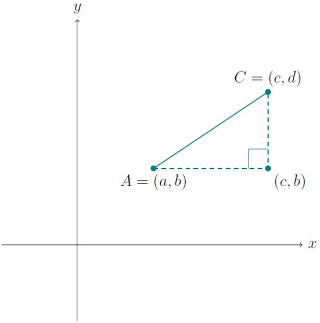
Distance

- Pythagoras' Theorem
- $z^2 = x^2 + y^2$



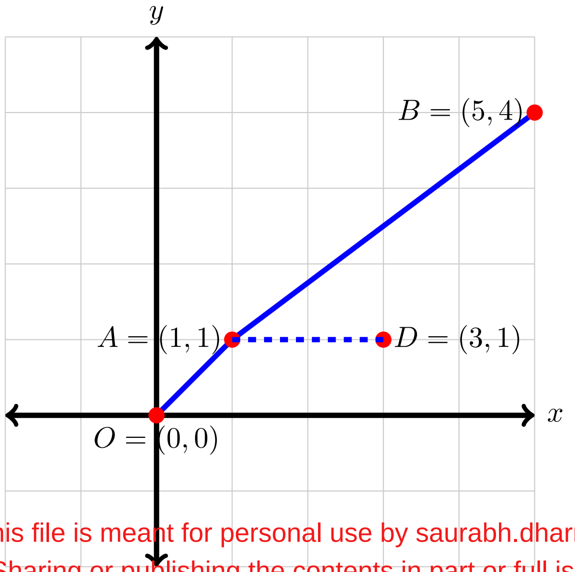
- Distance between two pints

$$\text{dist}(A, C) = \sqrt{(c-a)^2 + (b-d)^2}$$

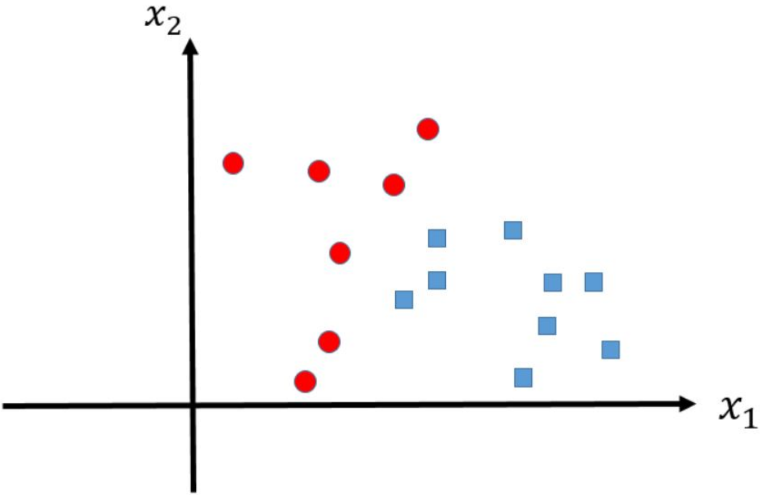


Distance applications: k Nearest Neighbours

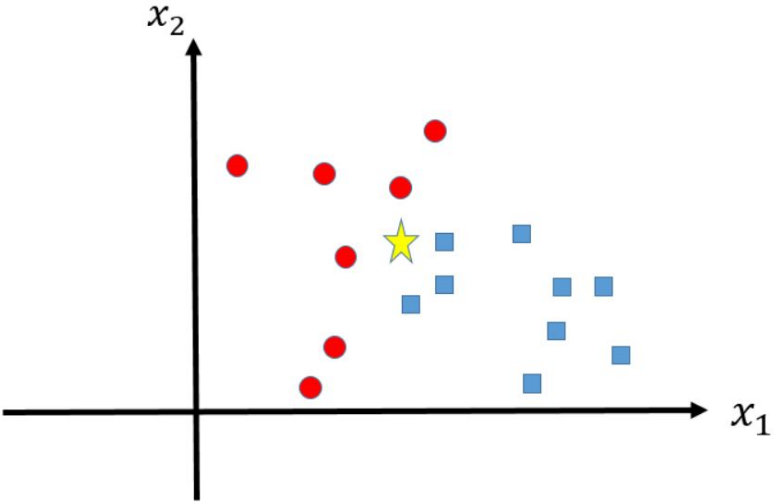
- dist(A, B), dist(A, O), dist(A, D)



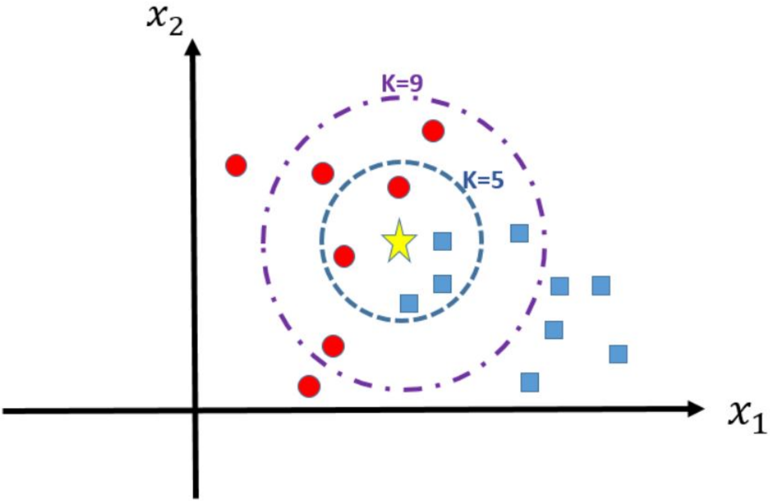
Distance applications: k Nearest Neighbours



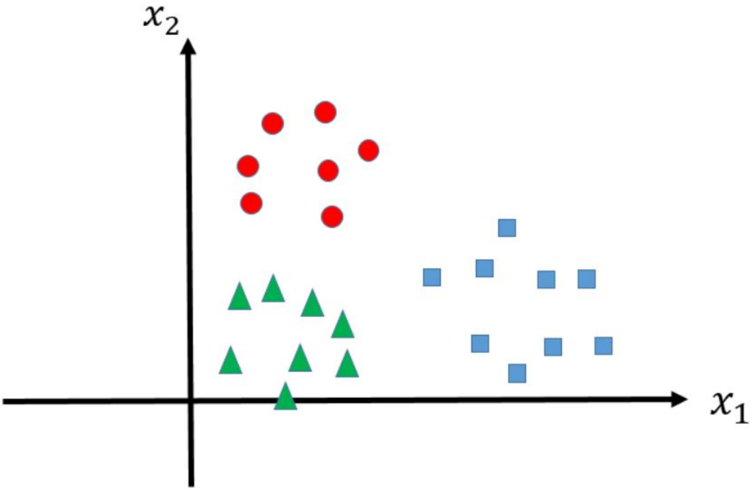
Distance applications: k Nearest Neighbours



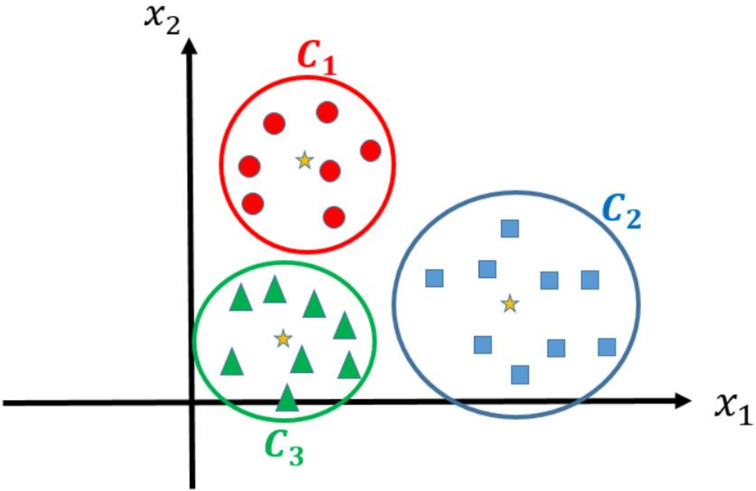
Distance applications: k Nearest Neighbours



Distance applications: Clustering



Distance applications: Clustering

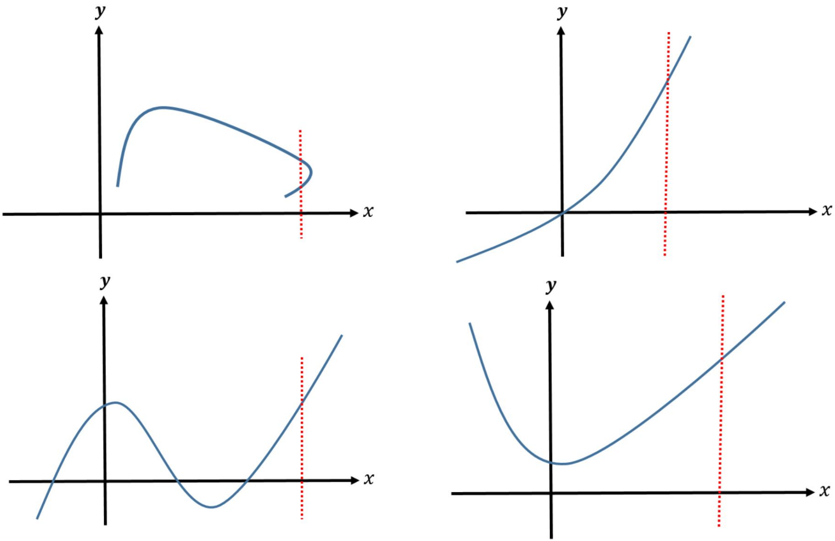


Functions

Functions

- Ordered pairs (x, y)
- Independent variable (shown generally by x) and dependent variable (shown generally by y)
- A function $f : A \to B$ transforms each elemnt $x \in A$ into $f(x) \in B$
- a particular ordered pair $(x, f(x))$
- visual detection of functions: vertical line test

Vertical line tests



Functions

- Let X and Y be sets.
- A function f from X to Y assigns (images) each element $x \in X$ to exactly one element in $y \in Y$.

$$f : X \to Y$$

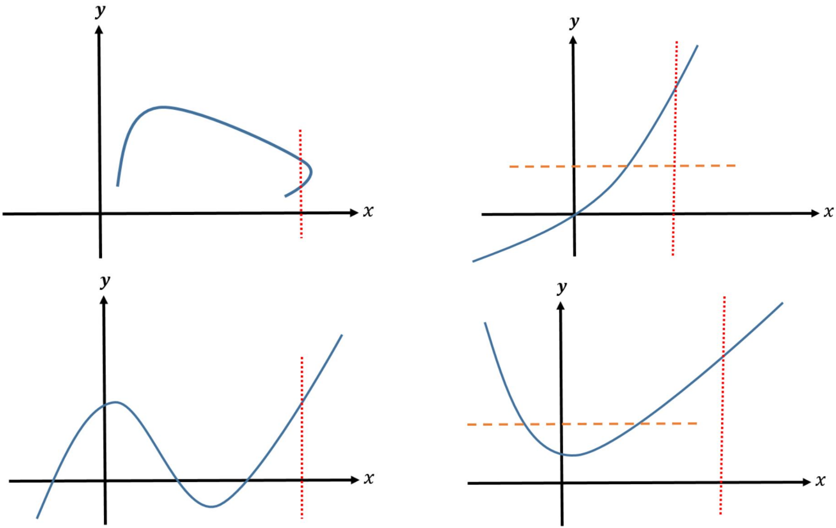
$$y = f(x)$$

- Functions are also called mappings or transformations.
- X is called domain
- Y is called co-domain
- The range of f is the set of all images of elements in X

Finding domain of functions

- Be carefull
  - Never devide by zero! There should not be zero in the denominator.
  - No negative number under the square root sign.
- Generally finding the domain of a function is easier than finding the range of that function.
- Injective or one-to-one
  For all $a, b \in X$, if $f(a) = f(b)$ then $a = b$

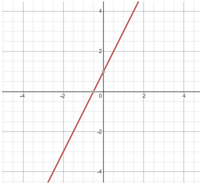
Vertical and horizontal line tests



Some Functions and their graphs

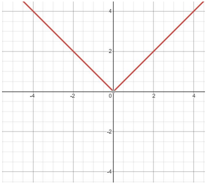
- A linear function: $f : \mathbb{R} \to \mathbb{R}$

$$f(x) = mx + b$$



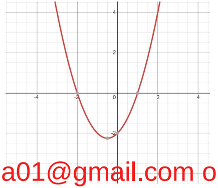
- Absolute value function: $f : \mathbb{R} \to \mathbb{R}$

$$f(x) = |x|$$

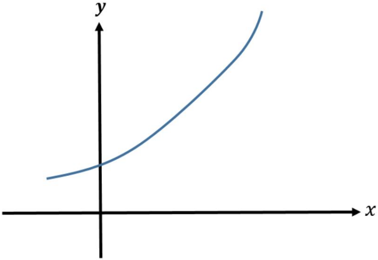


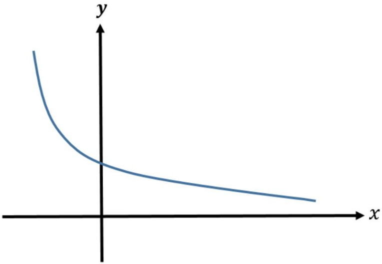
- Qusdratic functions $f : \mathbb{R} \to \mathbb{R}$

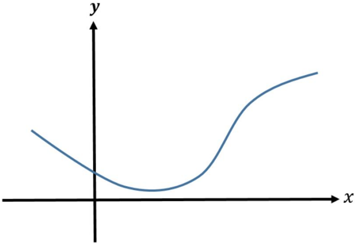
$$f(x) = ax^2 + bx + c$$



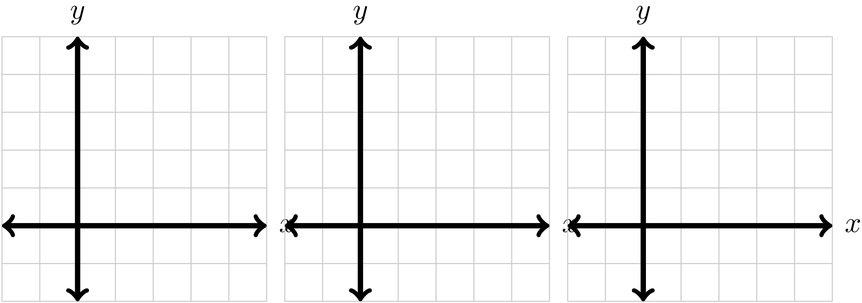
Types of functions: Increasing and Decreasing Functions







- $y = f(x) = mx + b$
  - m slope
  - b y-intercept
- different types of lines based on different values of m and b



Frequent Classes of functions: Polynomials

- $y = f(x) = P(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0$
  - n nonnegative ineteger
  - $a_n, a_{n-1}, \ldots, a_2, a_1, a_0$ coefficients
  - if $a_n \neq 0$, the degree of polynomial is n
  - domain is $(-\infty, +\infty) = \mathbb{R}$
- Example: $P(x) = 2x^5 - x^3 + \frac{3}{7}x^2 + \sqrt{3}$
- Linear function: polynomial of degree 1, $y = ax + b$
- Quadratic function: polynomial of degree 2. $a \neq 0$

$$P(x) = ax^2 + bx + c, \quad a \neq 0$$

- Cubic function $P(x) = ax^3 + bx^2 + cx + d$ given that $a \neq 0$

Frequent Classes of functions: Power and Rational Functions

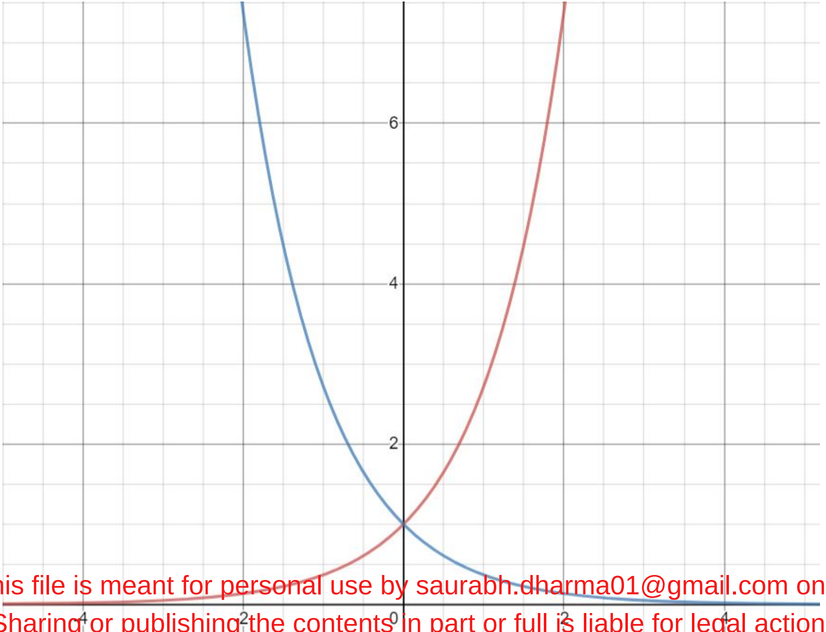
- Power functions
  - $f(x) = x^a$, where a is a constant
    - $f(x) = x^n$, n is a positive integer
    - $f(x) = x^{\frac{1}{n}} = \sqrt[n]{x}$, n is a positive integer, root function
    - $f(x) = x^{-1} = \frac{1}{x}$ reciprocal function
- Rational functions
  - $f(x) = \frac{P(x)}{Q(x)}$ where P and Q are polynomials

- $y = \sin(x)$, $-1 \leq \sin(x) \leq 1$, $\sin(n\pi) = 0$
- $y = \cos(x)$, $-1 \leq \cos(x) \leq 1$
- $y = \tan(x) = \frac{\sin(x)}{\cos(x)}$
- $y = \cot(x) = \frac{\cos(x)}{\sin(x)}$
- $y = \sec(x) = \frac{1}{\cos(x)}$
- $y = \csc(x) = \frac{1}{\sin(x)}$

Frequent Classes of functions: Exponential

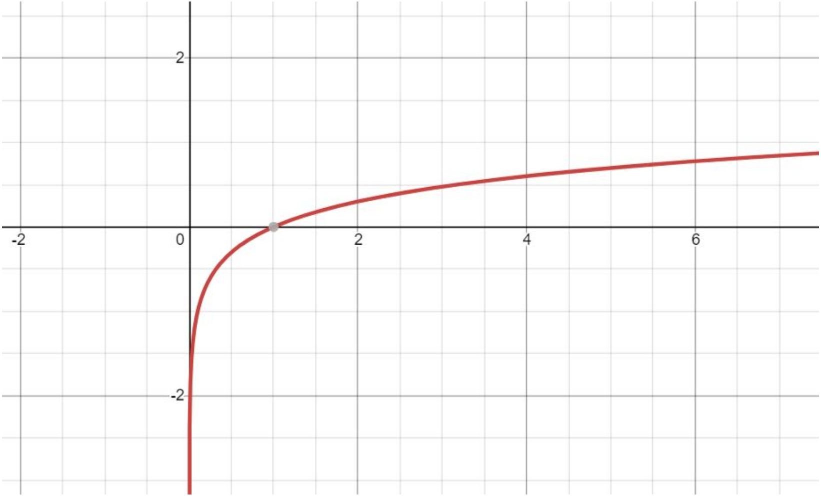
- $y = f(x) = b^x$ where the base b is a positive constant
  - $b > 1$ growth function
  - $b < 1$ decay function
- Index laws
  - $(a^n)(a^m) = a^{n+m}$
  - $\frac{a^n}{a^m} = a^{n-m}$
  - $(ab)^n = (a^n)(b^n)$
  - $(\frac{a}{b})^n = \frac{a^n}{b^n}$
  - $(a^n)^m = a^{mn}$
  - for $a \neq 0$, $a^0 = 1$ and $a^1 = a$
  - $a^{-n} = \frac{1}{a^n}$
- $y = e^x$ where $e \approx 2.718$ is called natural exponential function

Growth and Decay



Frequent Classes of functions: Logarithmic

- $y = f(x) = \log_b(x)$ where the base b is a positive constant
- $y = \log_b(x)$ same as $x = b^y$
- Logrtaithimic properties
  - $\log_b(xy) = \log_b(x) + \log_b(y)$
  - $\log_b(\frac{x}{y}) = \log_b(x) - \log_b(y)$
  - $\log_b(x^r) = r\log_b(x)$
- if $b = e$, $\log_e(x) = \ln(x)$
  - $\ln(x) = y \iff e^y = x$
  - $\ln(e^x) = x$ for all $x \in \mathbb{R}$
  - $e^{\ln(x)} = x$ for all $x > 0$
  - $\ln(e) = 1$
- $\log(x_1x_2x_3) = \log(x_1) + \log(x_2) + \log(x_3)$
- $\log\left(\prod_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} \log(x_i)$ and $\log\left(\prod_{i=1}^{n} f(x_i)\right) = \sum_{i=1}^{n} \log(f(x_i))$

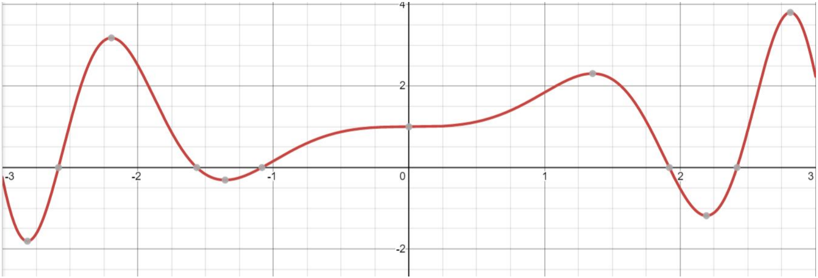

2
-2
-2
0
2
4
6

Composition of functions

- $(f \circ g)(x) = f(g(x))$
- Example
  - $f(x) = x^2,$ and $g(x) = x - 2$, then $(f \circ g)(x) = f(g(x)) = f(x-2) = (x-2)^2$
  - $(g \circ f)(x) = g(f(x)) = g(x^2) = x^2 - 2$
- In general $(f \circ g)(x) \neq (g \circ f)(x)$
- $(f \circ g \circ h)(x) = f(g(h(x)))$

Derivatives

Why do we need derivative?



Derivative: the rate of change

- slope of $y = f(x)$ at $x = a$

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

- sign and magnitude of slope
  - Slope is positive at $x = a$: increasing
  - Slope is negative at $x = a$: decreasing
  - $f'(a) \leq f'(b)$
  - Slope is zero at $x = a$: stationary point
- for $y = f(x)$, $f'(x) = y' = \frac{dy}{dx} = \frac{df}{dx}$ are the same

Derivative Table

- $f(x) = c$, then $f'(x) = 0$
- $y = x$, then $y' = 1$
- $y = x^n$, then, $y' = nx^{n-1}$
- $(cf(x))' = cf'(x)$
- $y = af_1(x) + bf_2(x)$, then $y' = af_1'(x) + bf_2'(x)$
- $y = f_1(x)f_2(x)$, then $y' = f_1'(x)f_2(x) + f_1(x)f_2'(x)$
- $y = \frac{f(x)}{g(x)}$, then

$$y' = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$$

Derivative of Trigonometric functions

- $y = \sin(x)$, then $y' = \cos(x)$
- $y = \cos(x)$, then $y' = -\sin(x)$
- find the derivative of $y = \tan(x)$. (ans. $\sec^2(x)$)
- find the derivative of $y = \cot(x)$. (ans: $-\csc^2(x)$)
- find the derivative of $y = \sec(x)$. (ans: $\sec(x)\tan(x)$)
- find the derivative of $y = \csc(x)$. (ans: $-\csc(x)\cot(x)$)

More derivatives

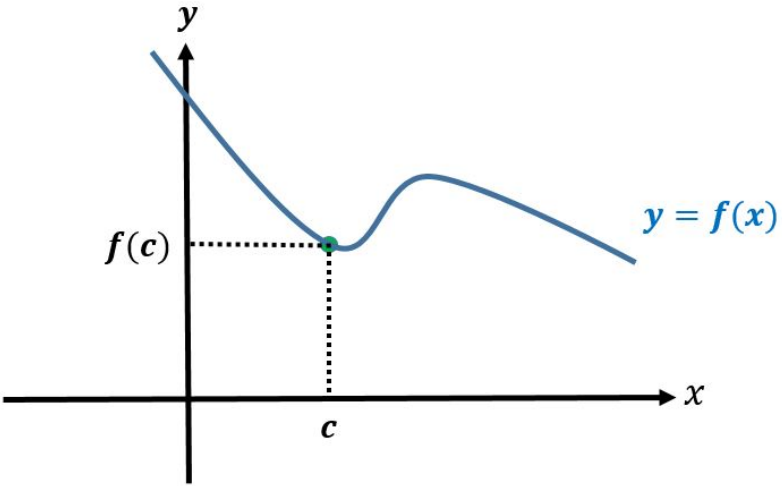
- $\frac{d}{dx}(\ln(x)) = \frac{1}{x}$
- $\frac{d}{dx}[\ln(g(x))] = \frac{g'(x)}{g(x)}$
- $\frac{d}{dx}(\sin(kx)) = k\cos(kx)$
- $\frac{d}{dx}(\cos(kx)) = -k\sin(kx)$
- $\frac{d}{dx}(e^{kx}) = ke^{kx}$

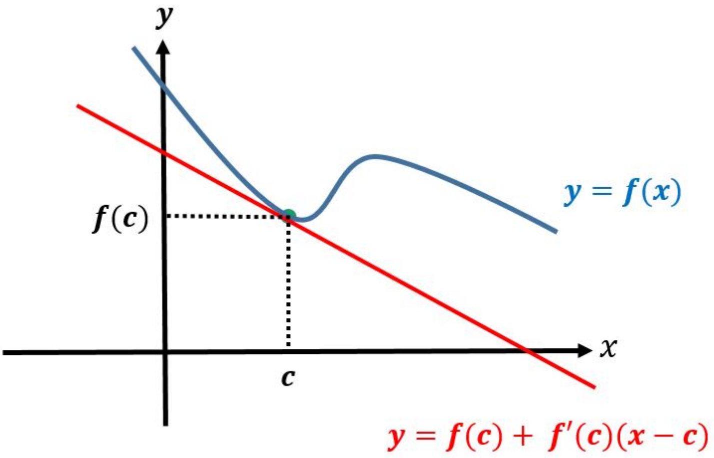
The Chain Rule

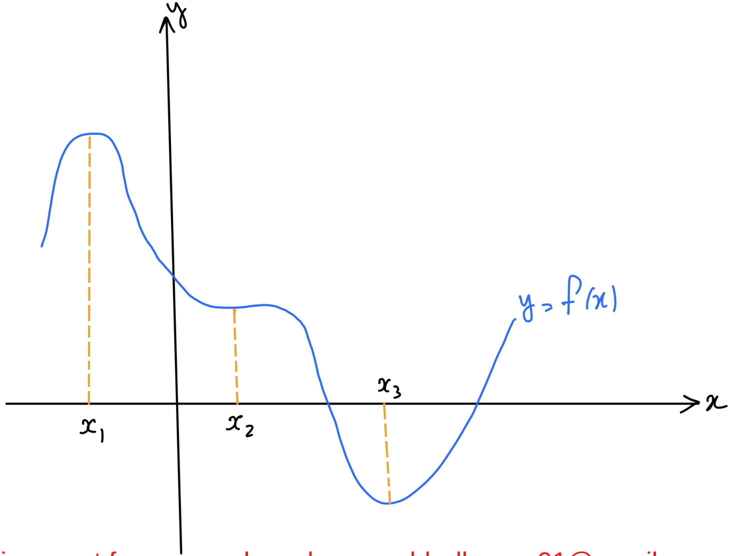
- When $y = f \circ g(x)$ or $y = f(g(x))$
- $y' = g'(x)f'(g(x))$
- Another interpretation:
  - if $y = f(u)$, we can talk about the rate of change of y when independent variable u is changing, which is $\frac{dy}{du}$.
  - If $u = g(x)$, then we can talk about the rate of change of u, when independent variable x is changing, which is $\frac{du}{dx}$.
  - $y = f(u)$ and $u = g(x)$, then $y = f(g(x))$, and y is a function of x. The rate of change of y with respect to x is

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

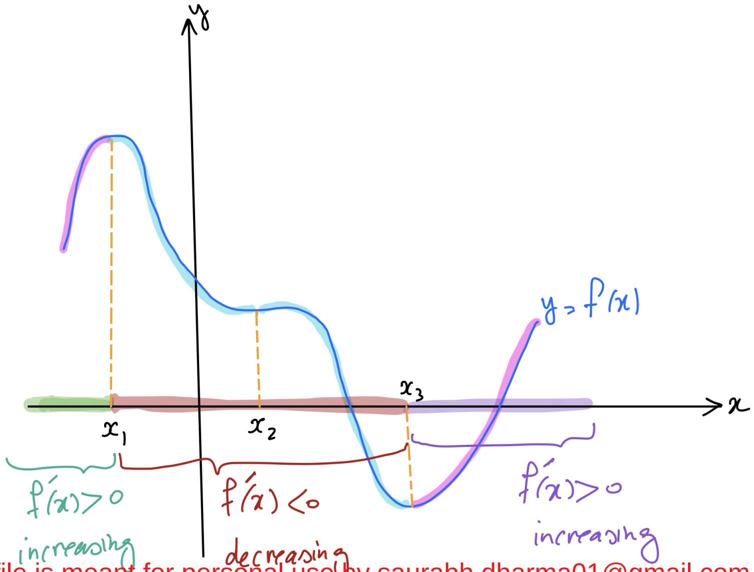
Linear Approximation



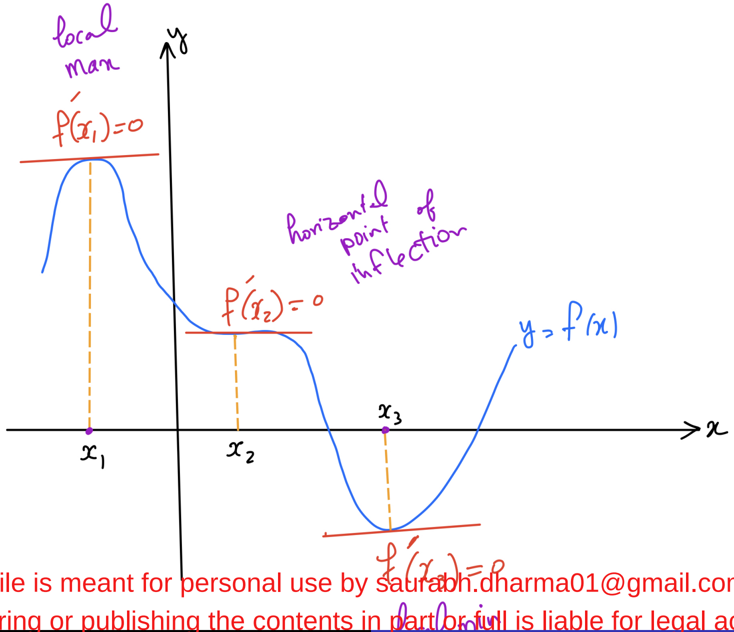

y
x
$f(c)$
c
$y = f(x)$
$y = f(c) + f'(c)(x - c)$


y
x
x_1
x_2
x_3
$y = f(x)$

Finding maxima and minima

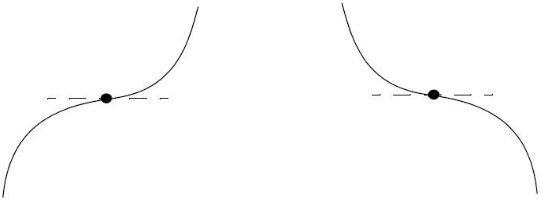


Finding maxima and minima



Finding Maxima and minima

- The points on a curve $y = f(x)$ at which the slope (gradient) is 0 are called **stationary points** (or **critical points**).
- For $y = f(x)$ if $f'(c) = 0$, then c is a stationary point
- A stationary point could be
  - a local maximum,
  - a local minimum, or
  - a horizontal point of inflection



Finding Maxima and minima of $y = f(x)$

- solve $f'(c) = 0$ and find c values
  - The stationary point $x = c$ is a local maximum point if

$$\begin{cases} x < c & f'(x) > 0 \\ x > c & f'(x) < 0 \end{cases}$$

  - The stationary point $x = c$ is a local minimum point if

$$\begin{cases} x < c & f'(x) < 0 \\ x > c & f'(x) > 0 \end{cases}$$

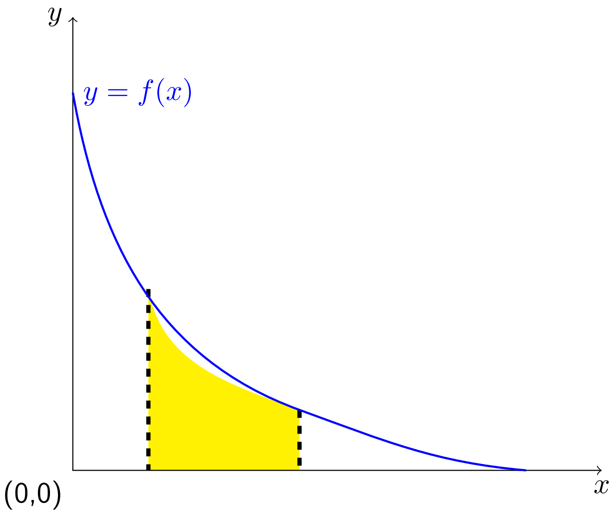
  - a horizontal point of inflection otherwise.

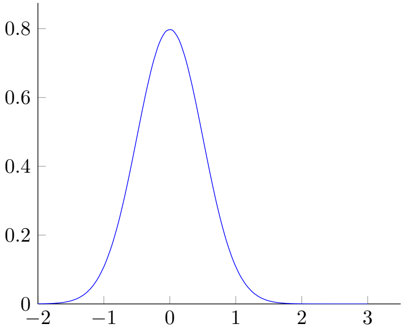
Maxima and minima of $f(x)$ using second derivative

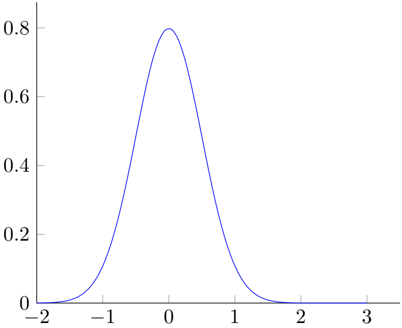
- Stationary (critical) points: $f'(c) = 0$
  - if $f''(c) > 0$, $x = c$ is a local minima
  - if $f''(c) < 0$, $x = c$ is a local maxima
  - if $f''(c) = 0$, $x = c$ no conclusion
- $f''(x) > 0$ at a point, the curve is concave up (convex) at that point
- $f''(x) < 0$ at a point, the curve is concave down (concave) at that point

Integrals

Integrals







$$P(a \leq X \leq b)$$

Integrals

- Indefinite Integral: If $F'(x) = f(x)$, then

$$\int f(x)dx = F(x) + c.$$

  $F(x)$ is called an antiderivetaive of f.

- Definite Integral: $\int_a^b f(x)dx = F(b) - F(a)$

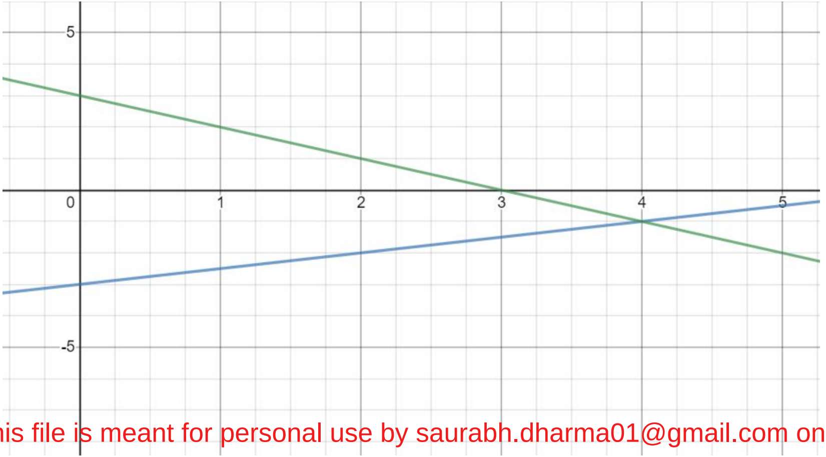
- $\int f(x)dx = F(x)$ means $F'(x) = f(x)$
- $\int cf(x)dx = c\int f(x)dx$
- $\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$
- $\int kdx = kx + C$
- $\int x^n dx = \frac{x^{n+1}}{n+1} + C, \quad (n \neq -1)$
- $\int \sin(x)dx = -\cos(x) + C$
- $\int \cos(x)dx = \sin(x) + C$

Integrals

- $\int \sec^2(x)dx = \tan(x) + C$
- $\int \csc^2(x)dx = -\cot(x) + C$
- $\int \sec(x)\tan(x)dx = \sec(x) + C$
- $\int \csc(x)\cot(x)dx = -\csc(x) + C$
- $\int e^{kx}dx = \frac{1}{k}e^{kx} + C$
- $\int \frac{1}{x}dx = \ln(x) + C$

System of Simultaneous Equations

$$\begin{cases} x - 2y = 6, & \text{line 1} \\ x + y = 3, & \text{line 2} \end{cases}$$



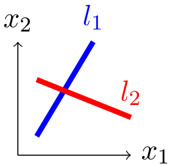
System of simultaneous equations

- Substitution
  1. Make one variable the subject of one of the equations
  2. Substitute for this variable in the other equation
  3. Solve this equation to find the solution for one variable
  4. Substitute the answer found in 3 into the equation obtained in 1 to find the solution for the remaining variable
- solve

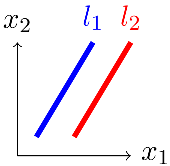
$$\begin{cases} x - 2y = 6, & \text{line 1} \\ x + y = 3, & \text{line 2} \end{cases}$$

Linear System of two equations

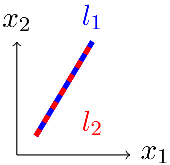
$$\begin{cases} a_1x_1 + b_1x_2 = c_1 \dashrightarrow (l_1) \\ a_2x_1 + b_2x_2 = c_2 \dashrightarrow (l_2) \end{cases}$$



Definite system
single solution



Inconsistent system
no solution



Indefinite system
many solutions

- Solve by substitution

$$\begin{cases} x + y + z = 6, & \text{plane 1} \\ x + 2y = 5, & \text{plane 2} \\ 2x + z = 5, & \text{plane 3} \end{cases}$$

Which system is easier to solve?

- System 1

$$\begin{cases} x + y + z = 6, & \text{plane 1} \\ x + 2y + 0z = 5, & \text{plane 2} \\ 2x + 0y + z = 5, & \text{plane 3} \end{cases}$$

- System 2

$$\begin{cases} x + 0y + 2z = 7, & \text{plane 1} \\ 0x + y - z = -1, & \text{plane 2} \\ 0x + 0y + z = 3, & \text{plane 3} \end{cases}$$

- System 3

$$\begin{cases} x + 0y + 0z = 1, & \text{plane 1} \\ 0x + y + 0z = 2, & \text{plane 2} \\ 0x + 0y + z = 3, & \text{plane 3} \end{cases}$$

Useful but harmless operations

$$\begin{cases} x - 2y = 6, \\ x + y = 3, \end{cases}$$

- Interchange two equations
- Multiply each element in an equation by a non-zero number
- Multiply an equation by a non-zero number and add the result to another equation.