girafe
ai

Probability and Calculus Recap

Lecture 2

Last Time

- Vectors
  - Vector spaces
  - Inner products
  - Lengths
  - Distances
  - Angles
- Analytic Geometry
  - Projections
  - Hyperplanes
  - Normal vector

Today

- More on matrices
    - matrix operations;
    - determinant.
- Calculus
- Probability Theory

Matrices: a small review

A Matrix

- $A \in \mathbb{R}^{m \times n}$ - a **matrix** with m rows and n columns:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- *Examples:*

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{2 \times 2}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{R}^{2 \times 3}$$

Special Matrices

- Diagonal matrix: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ $\left(a_{ii} \neq 0,\ a_{ij} = 0\ \forall i \neq j\right)$
- Identity matrix: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $\left(a_{ii} = 1,\ a_{ij} = 0\ \forall i \neq j\right)$
- Symmetric matrix: $\begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 6 \\ 5 & 6 & 3 \end{bmatrix}$ $\left(a_{ij} = a_{ji}\right)$
- Triangular matrix: $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{bmatrix}$ $\left(a_{ij} = 0\ \forall i > j\ or\ \ \forall i < j\right)$

Basic Operations with Matrices

- Addition:

$$A = \{a_{ij}\}_{i=1,\ldots,m,j=1,\ldots,n}, \qquad B = \{b_{ij}\}_{i=1,\ldots,m,j=1,\ldots,n}, \qquad A + B = \{a_{ij} + b_{ij}\}_{i=1,\ldots,m,j=1,\ldots,n}$$

Basic Operations with Matrices

- Addition:

$$A = \{a_{ij}\}_{i=1,\dots,m,j=1,\dots,n}, \qquad B = \{b_{ij}\}_{i=1,\dots,m,j=1,\dots,n}, \qquad A + B = \{a_{ij} + b_{ij}\}_{i=1,\dots,m,j=1,\dots,n}$$

- Multiplication by a scalar:

$$A = \{a_{ij}\}_{i=1,\dots,m,j=1,\dots,n}, \qquad \lambda \in \mathbb{R}, \qquad \lambda A = \{\lambda a_{ij}\}_{i=1,\dots,m,j=1,\dots,n}$$

Matrix Multiplication

- Matrix multiplication:

$$A = \{a_{ij}\}_{i=1,\ldots,m,j=1,\ldots,n}, \qquad B = \{b_{ij}\}_{i=1,\ldots,n,j=1,\ldots,k}$$

$$A \cdot B = \{(A_i, B^j)\}_{i=1,\ldots,m,j=1,\ldots,k} = \left\{\sum_{l=1,\ldots,n} a_{il} \cdot b_{lj}\right\}_{i=1,\ldots,m,j=1,\ldots,k}$$

Matrix Multiplication

- Matrix multiplication:

$$A = \{a_{ij}\}_{i=1,\dots,m,j=1,\dots,n}, \qquad B = \{b_{ij}\}_{i=1,\dots,n,j=1,\dots,k}$$

$$A \cdot B = \{(A_i, B^j)\}_{i=1,\dots,m,j=1,\dots,k} = \left\{\sum_{l=1,\dots,n} a_{il} \cdot b_{lj}\right\}_{i=1,\dots,m,j=1,\dots,k}$$

- Example $\mathbb{R}^{2\times 2}$:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

Matrix Multiplication

- For numbers: $2 \times 3 = 3 \times 2 = 6$.

Matrix Multiplication

- For numbers: $2 \times 3 = 3 \times 2 = 6$.

- Matrix multiplication is *(in general)* not commutative:

$$\boldsymbol{AB} \neq \boldsymbol{BA}$$

Matrix Multiplication

- For numbers: $2 \times 3 = 3 \times 2 = 6$.
- Matrix multiplication is *(in general)* not commutative:

$$\boldsymbol{AB \neq BA}$$

- Example:

$$A = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 6 & 2 \\ 3 & 2 \end{bmatrix}, \quad AB = \begin{bmatrix} 30 & 14 \\ 12 & 6 \end{bmatrix}, \quad BA = \begin{bmatrix} 20 & 28 \\ 11 & 16 \end{bmatrix}$$

Matrix Multiplication

- Multiplication by identity matrix E:

$$AE = EA = A$$

Matrix Multiplication

- Multiplication by identity matrix E:

$$AE = EA = A$$

- Multiplication by zero matrix O:

$$AO = OA = O$$

Transposing a Matrix

- The transpose of a matrix results from "flipping" the rows and columns:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \qquad A^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}.$$

girafe ai

Transposing a Matrix

- The following properties of transposes are easily verified:
  - A – symmetric matrix $\Longrightarrow A^T = A$
  - $(A^T)^T = A$
  - $(A + B)^T = A^T + B^T$
  - $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$

Linear Transforms

A more interesting way of looking at matrices.

Linear Transformation

Linear Transformation

Linear Transformation

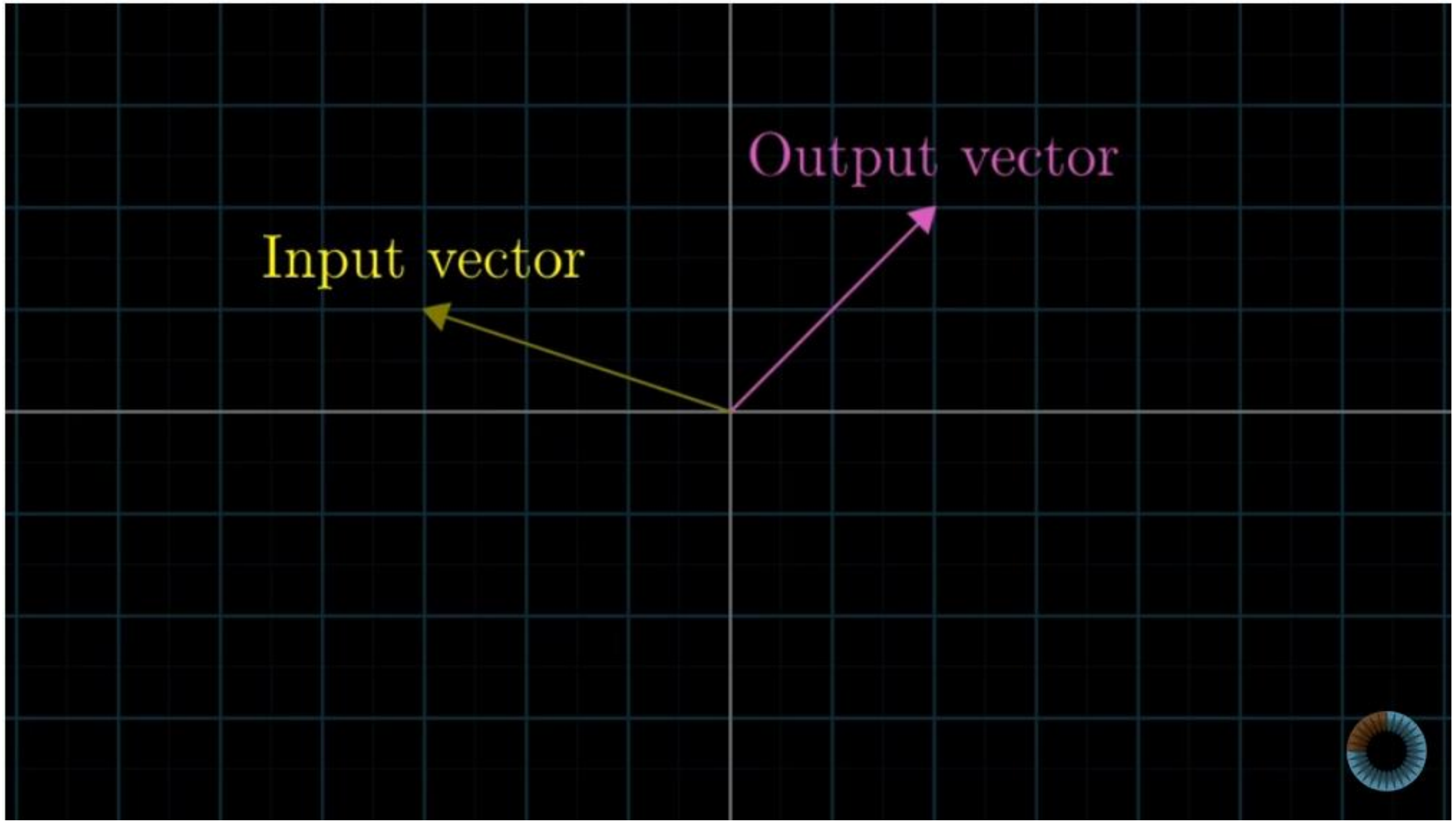
Linear Transformation

$$x_{input} \rightarrow A \rightarrow x_{output}$$

A – transformation

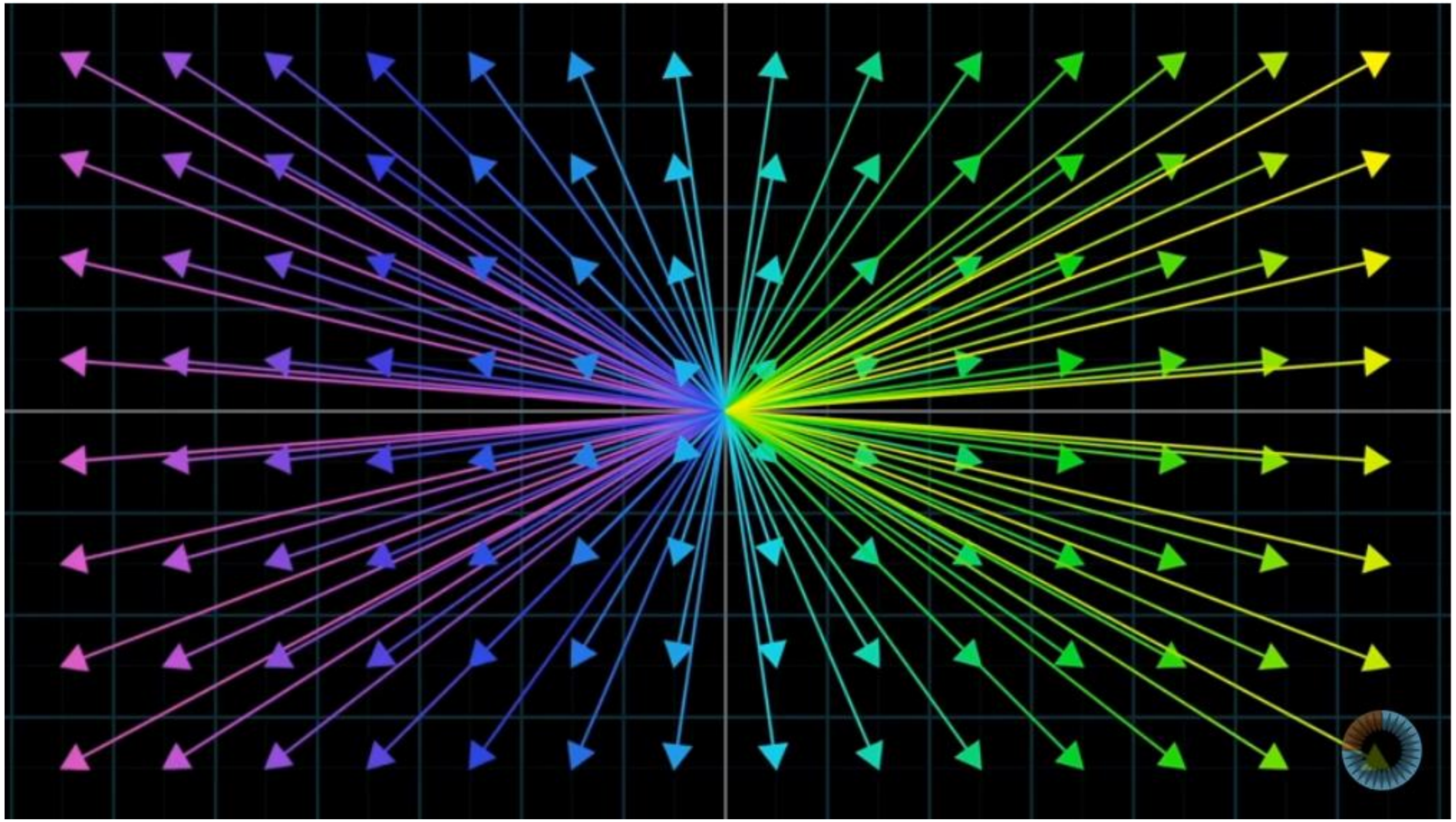
x_{input}, x_{output} – vectors

Transformation



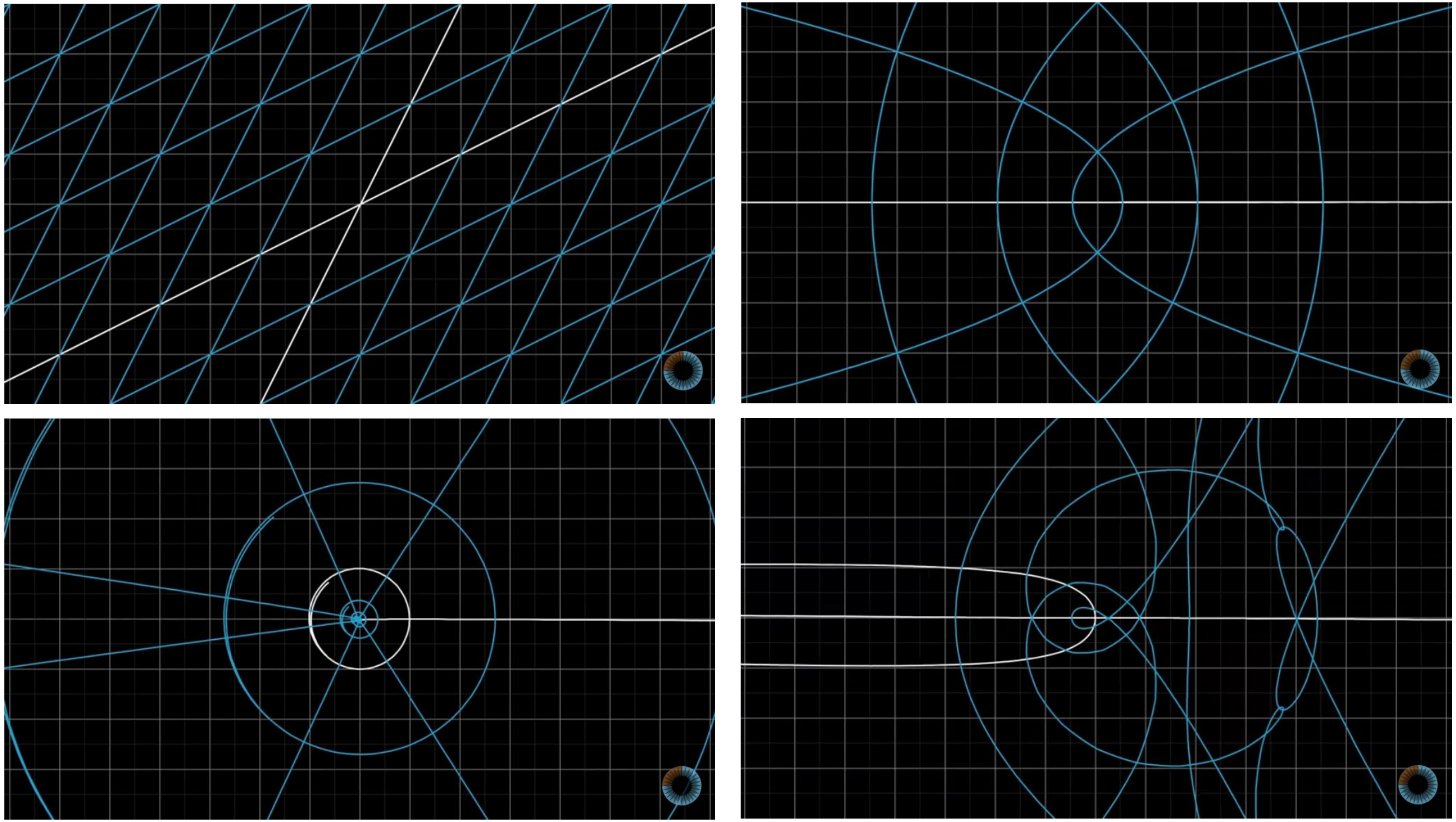
Source: Linear Transformations and Matrices

Transformation



Source: Linear Transformations and Matrices

Transformation: Examples



Source: [Linear Transformations and Matrices]()

Linear Transformation

Linear Transformation

$$x_{input} \rightarrow A \rightarrow x_{output}$$

A – transformation

x_{input}, x_{output} – vectors

Linear Transformation

<u>Linear</u> Transformation

A transformation that satisfies two properties:

1. $A(x+y) = A(x) + A(y)$
2. $A(\lambda x) = \lambda A x$

$$x_{input} \rightarrow A \rightarrow x_{output}$$

A – transformation

x_{input}, x_{output} – vectors

Linear Transformation

- How to describe a linear transformation numerically?
-

Linear Transformation

- How to describe a linear transformation numerically?
- With matrices! How?

Linear Transformation

- How to describe a linear transformation numerically?

- With matrices! How?

$$x_{input} = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n,\ \ e_1, \dots e_n - \text{basis},\ \ x_1, \dots, x_n - \text{coordnates}$$

Linear Transformation

- How to describe a linear transformation numerically?
- With matrices! How?

$$x_{input} = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n,\ \ e_1, \dots e_n - \text{basis},\ \ x_1, \dots, x_n - \text{coordnates}$$

$$x_{output} = A\left(x_{input}\right) = A(x_1 e_1 + x_2 e_2 + \cdots + x_n e_n) =$$

girafe ai

Linear Transformation

- How to describe a linear transformation numerically?
- With matrices! How?

$$x_{input} = x_1 e_1 + x_2 e_2 + \dots + x_n e_n,\ \ e_1, \dots e_n - \text{basis},\ \ x_1, \dots, x_n - \text{coordnates}$$

$$x_{output} = A\left(x_{input}\right) = A(x_1 e_1 + x_2 e_2 + \dots + x_n e_n) = x_1 A(e_1) + x_2 A(e_2) + \dots + x_n A(e_n)$$

Linear Transformation

- How to describe a linear transformation numerically?
- With matrices! How?

$$x_{input} = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n,\ \ e_1, \dots e_n - \text{basis},\ \ x_1, \dots, x_n - \text{coordnates}$$

$$x_{output} = A\left(x_{input}\right) = A(x_1 e_1 + x_2 e_2 + \cdots + x_n e_n) = x_1 A(e_1) + x_2 A(e_2) + \cdots + x_n A(e_n)$$

$$A \coloneqq [A(e_1) \mid A(e_2) \mid \dots \mid A(e_n)]$$

Linear Transformation

- How to describe a linear transformation numerically?
- With matrices! How?

$$x_{input} = x_1 e_1 + x_2 e_2 + \dots + x_n e_n,\ \ e_1, \dots e_n - \text{basis},\ \ x_1, \dots, x_n - \text{coordnates}$$

$$x_{output} = A(x_{input}) = A(x_1 e_1 + x_2 e_2 + \dots + x_n e_n) = x_1 A(e_1) + x_2 A(e_2) + \dots + x_n A(e_n)$$

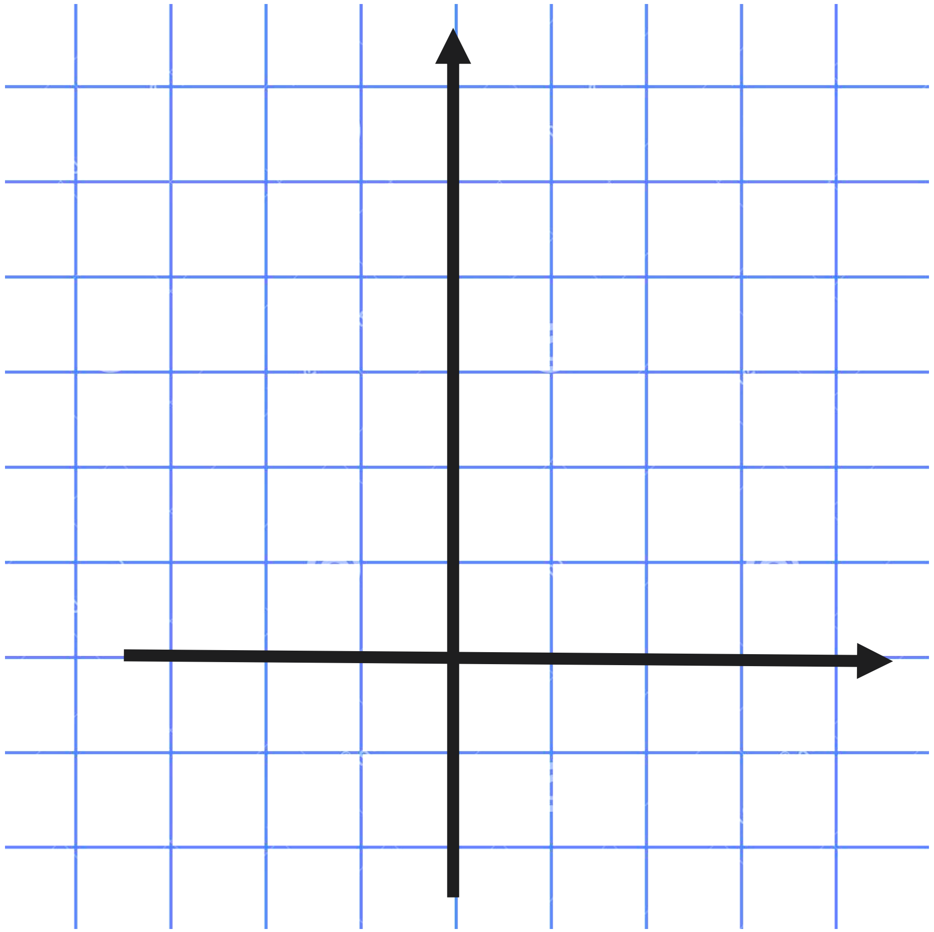
$$A \coloneqq [A(e_1) \mid A(e_2) \mid \dots \mid A(e_n)]$$

$$\Longrightarrow x_{output} = A(x_{input}) = A \cdot x_{input}$$

girafe ai

Example: Rotation

- Imagine that we want to rotate vectors in $\mathbb{R}^2$ 90° anti-clockwise.

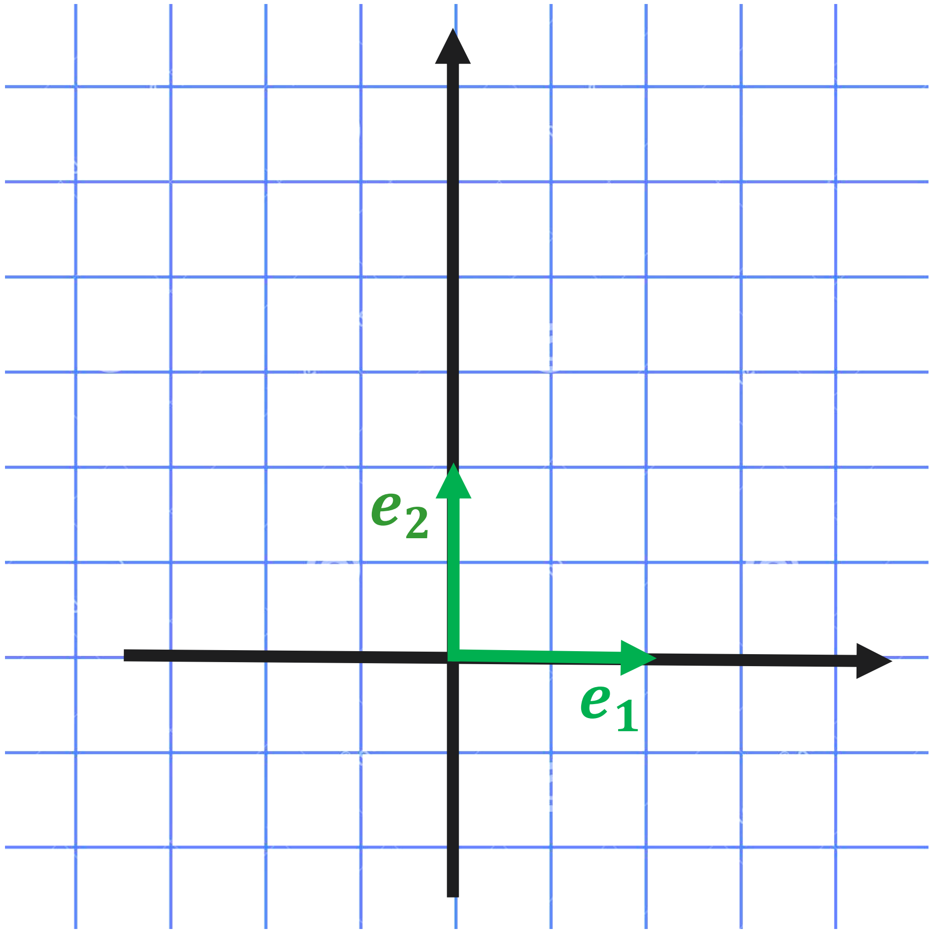


Example: Rotation

- Imagine that we want to rotate vectors in $\mathbb{R}^2$ 90° anti-clockwise.

- What happens to the basis vectors?

  $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \rightarrow \quad , \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \rightarrow$

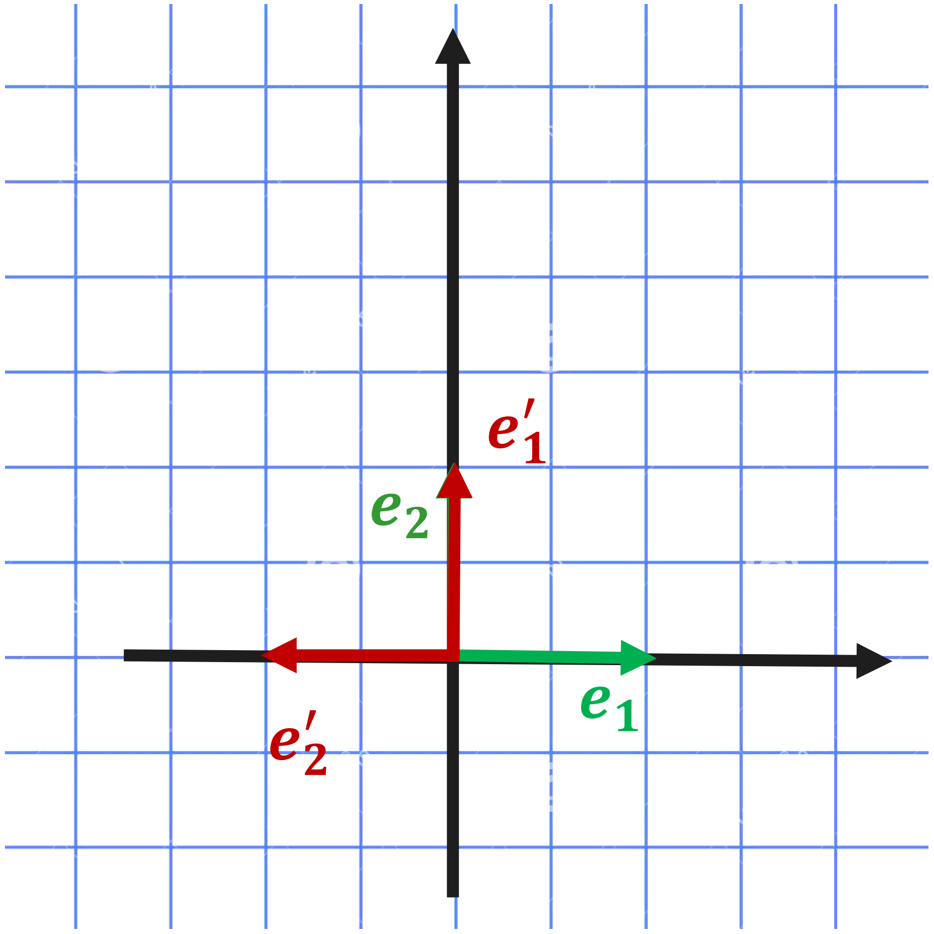


Example: Rotation

- Imagine that we want to rotate vectors in $\mathbb{R}^2$ 90° anti-clockwise.

- What happens to the basis vectors?

$$e_1 = \begin{bmatrix}1\\0\end{bmatrix} \rightarrow \qquad , \qquad e_2 = \begin{bmatrix}0\\1\end{bmatrix} \rightarrow$$

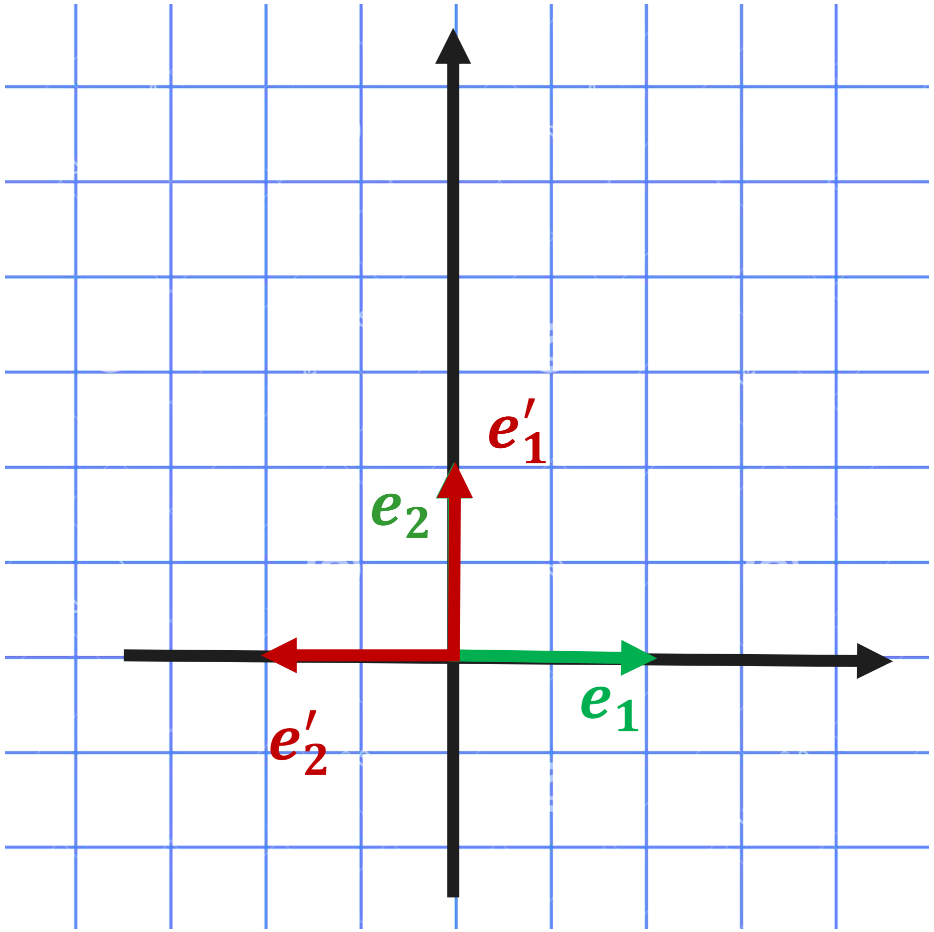


Example: Rotation

- Imagine that we want to rotate vectors in $\mathbb{R}^2$ 90° anti-clockwise.

- What happens to the basis vectors?

$$e_1 = \begin{bmatrix}1\\0\end{bmatrix} \to \begin{bmatrix}0\\1\end{bmatrix}, \qquad e_2 = \begin{bmatrix}0\\1\end{bmatrix} \to \begin{bmatrix}-1\\0\end{bmatrix}$$



Example: Rotation

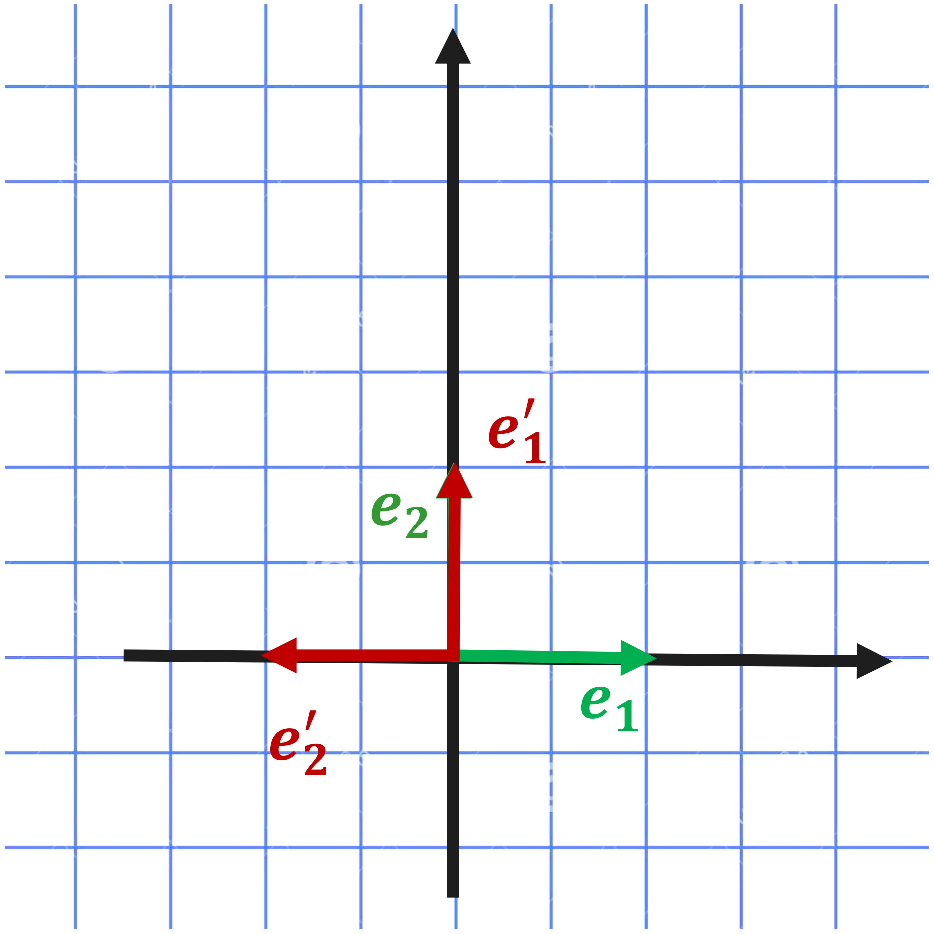
- Imagine that we want to rotate vectors in $\mathbb{R}^2$ 90° anti-clockwise.

- What happens to the basis vectors?

$$e_1 = \begin{bmatrix}1\\0\end{bmatrix} \rightarrow \begin{bmatrix}0\\1\end{bmatrix}, \qquad e_2 = \begin{bmatrix}0\\1\end{bmatrix} \rightarrow \begin{bmatrix}-1\\0\end{bmatrix}$$

- Therefore,

$$R = \begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}$$



Example: Rotation

- Imagine that we want to rotate vectors in $\mathbb{R}^2$ 90° anti-clockwise.

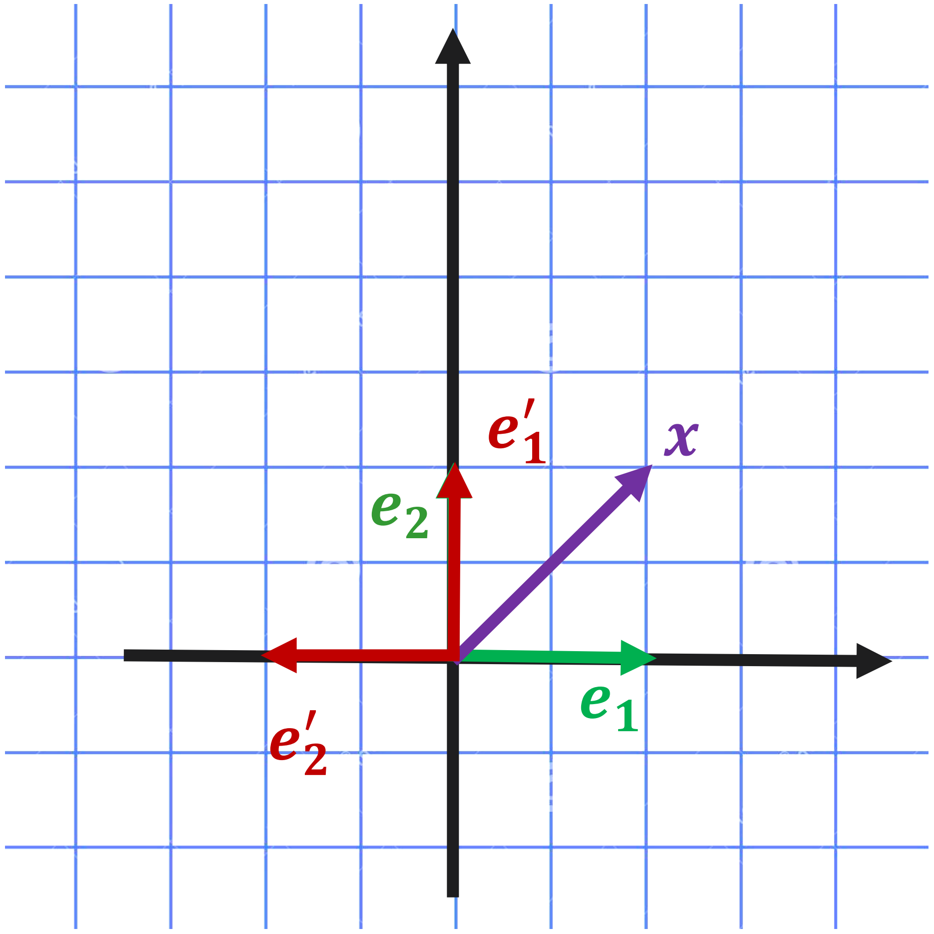
- What happens to the basis vectors?

$$e_1 = \begin{bmatrix}1\\0\end{bmatrix} \to \begin{bmatrix}0\\1\end{bmatrix}, \qquad e_2 = \begin{bmatrix}0\\1\end{bmatrix} \to \begin{bmatrix}-1\\0\end{bmatrix}$$

- Therefore,

$$R = \begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}$$

- Consider $x = [1, 1]^T$.



Example: Rotation

- Imagine that we want to rotate vectors in $\mathbb{R}^2$ 90° anti-clockwise.

- What happens to the basis vectors?

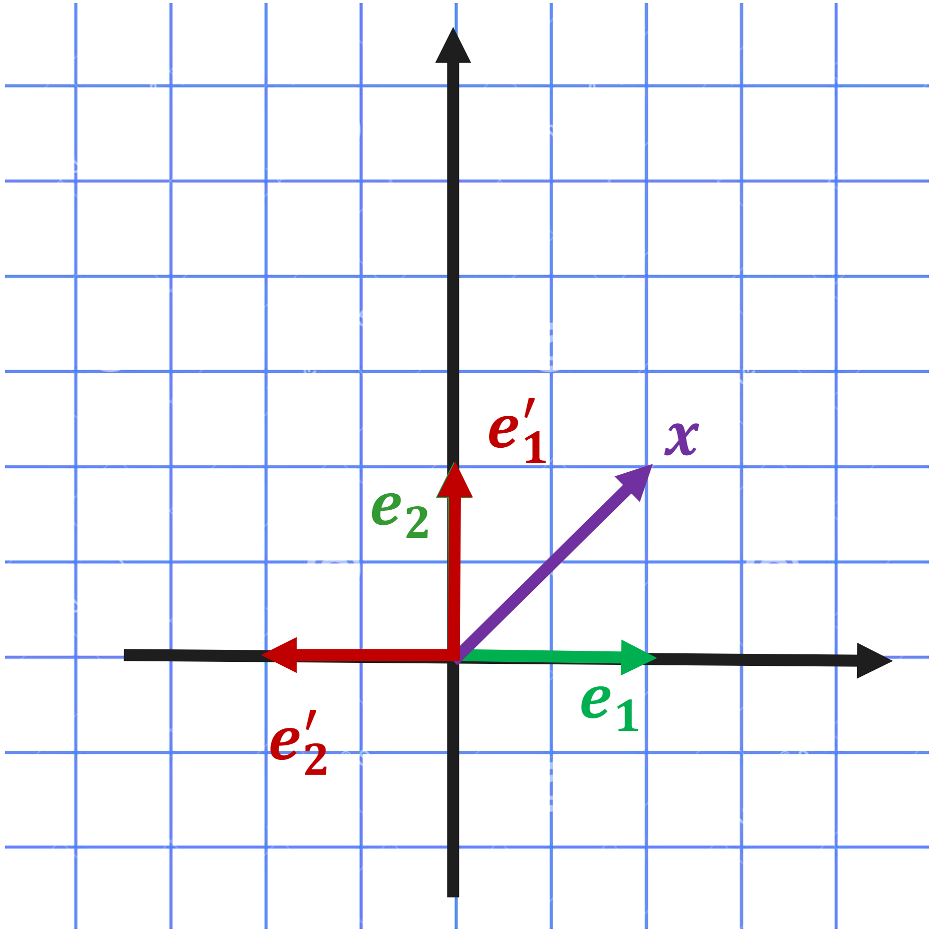
$$e_1 = \begin{bmatrix}1\\0\end{bmatrix} \rightarrow \begin{bmatrix}0\\1\end{bmatrix}, \qquad e_2 = \begin{bmatrix}0\\1\end{bmatrix} \rightarrow \begin{bmatrix}-1\\0\end{bmatrix}$$

- Therefore,

$$R = \begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}$$

- Consider $x = [1, 1]^T$. After rotation:

$$x_{rotated} =$$



Example: Rotation

- Imagine that we want to rotate vectors in $\mathbb{R}^2$ 90° anti-clockwise.

- What happens to the basis vectors?

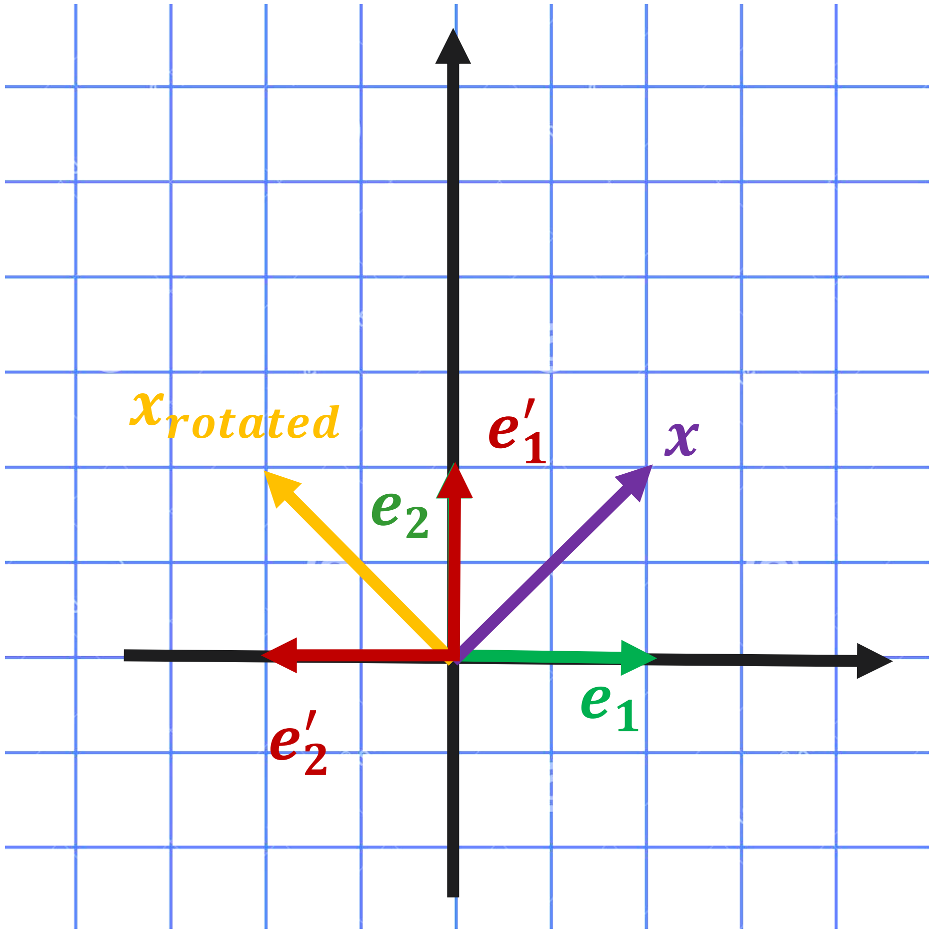
  $e_1 = \begin{bmatrix}1\\0\end{bmatrix} \to \begin{bmatrix}0\\1\end{bmatrix}, \qquad e_2 = \begin{bmatrix}0\\1\end{bmatrix} \to \begin{bmatrix}-1\\0\end{bmatrix}$

- Therefore,

$$R = \begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}$$

- Consider $x = [1, 1]^T$. After rotation:

$x_{rotated} =$



Example: Rotation

- Imagine that we want to rotate vectors in $\mathbb{R}^2$ 90° anti-clockwise.

- What happens to the basis vectors?

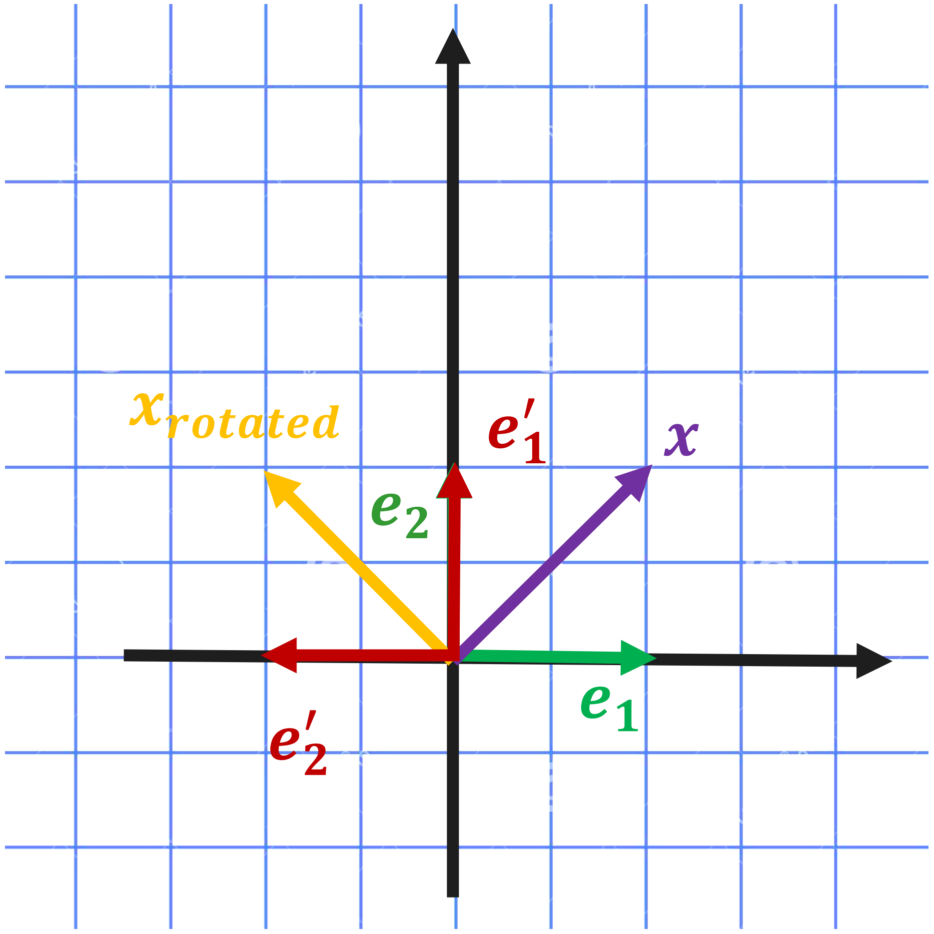
$$e_1 = \begin{bmatrix}1\\0\end{bmatrix} \to \begin{bmatrix}0\\1\end{bmatrix}, \qquad e_2 = \begin{bmatrix}0\\1\end{bmatrix} \to \begin{bmatrix}-1\\0\end{bmatrix}$$

- Therefore,

$$R = \begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}$$

- Consider $x = [1, 1]^T$. After rotation:

$$x_{rotated} = Rx = \begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}-1\\1\end{bmatrix}.$$



Linear Transformation

- Every linear transformation can be defined by its matrix.

  Columns = how this transformation changes the vectors in the selected basis.

Linear Transformation

- Every linear transformation can be defined by its matrix.

  Columns = how this transformation changes the vectors in the selected basis.

- Vice versa: every square matrix defines some linear transformation.

Common Transforms

Identity Transformation

- Doesn't change anything.

- Transformation matrix E:

$$Ex = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

girafe ai

Identity Transformation

- Doesn't change anything.

- Transformation matrix E:

$$Ex = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- Example:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

girafe ai

Stretching / Squeezing

- Enlarge (compress) all distances in a particular direction by a constant factor.

- Transformation matrix:

$$Kx = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

girafe ai

Stretching / Squeezing

- Enlarge (compress) all distances in a particular direction by a constant factor.

- Transformation matrix:

$$Kx = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- Example: stretch x-axis (x3) and squeeze y-axis (x 0.5):

$$\begin{bmatrix} 3 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Projection on an Axis

- Consider $\mathbb{R}^3$. Project on the XY −plane.

- Transformation matrix:

$$Px = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Projection on an Axis

- Consider $\mathbb{R}^3$. Project on the XY −plane.

- Transformation matrix:

$$Px = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

- Example:

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

Rotation

- Rotating points anticlockwise by θ.

- Rotation matrix R_θ:

$$R_\theta\, x = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Rotation

- Rotating points anticlockwise by $\boldsymbol{\theta}$.

- Rotation matrix R_θ:

$$R_\theta\, x = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- Example: rotate by $\mathbf{45}^{\circ}$ anticlockwise:

$$\begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix}$$

Combining Transforms

Composition

- Let A and B be two linear transforms.
  What if we first apply A and then B?

Composition

- Let A and B be two linear transforms.
  What if we first apply A and then B?

- Example: first rotate by 90°, then squeeze.

Composition

- Let A and B be two linear transforms.
  What if we first apply A and then B?

- Example: first rotate by 90°, then squeeze.

$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ – rotation, $B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$ – squeezing

Composition

- Let A and B be two linear transforms.
  What if we first apply A and then B?

- Example: first rotate by 90°, then squeeze.

$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ – rotation, $B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$ – squeezing

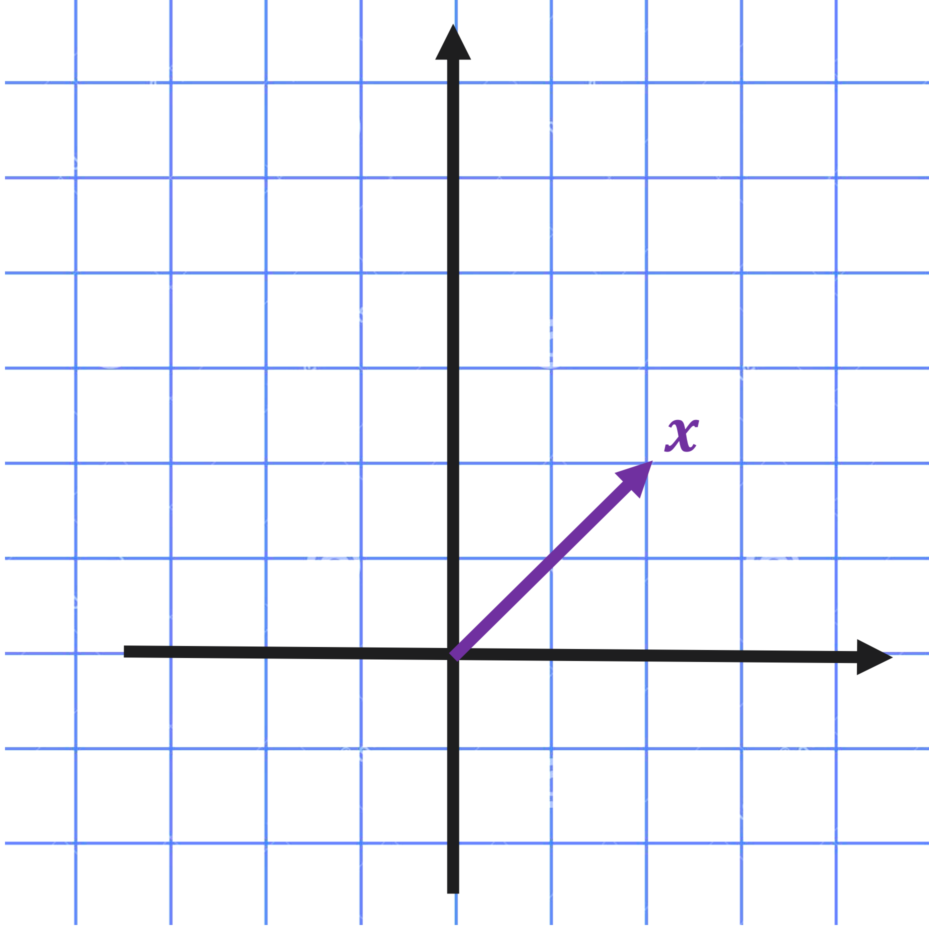
$$B(Ax) =$$

girafe ai

Composition

- Let A and B be two linear transforms.
  What if we first apply A and then B?

- Example: first rotate by 90°, then squeeze.



$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ – rotation, $B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$ – squeezing

$$B(Ax) =$$

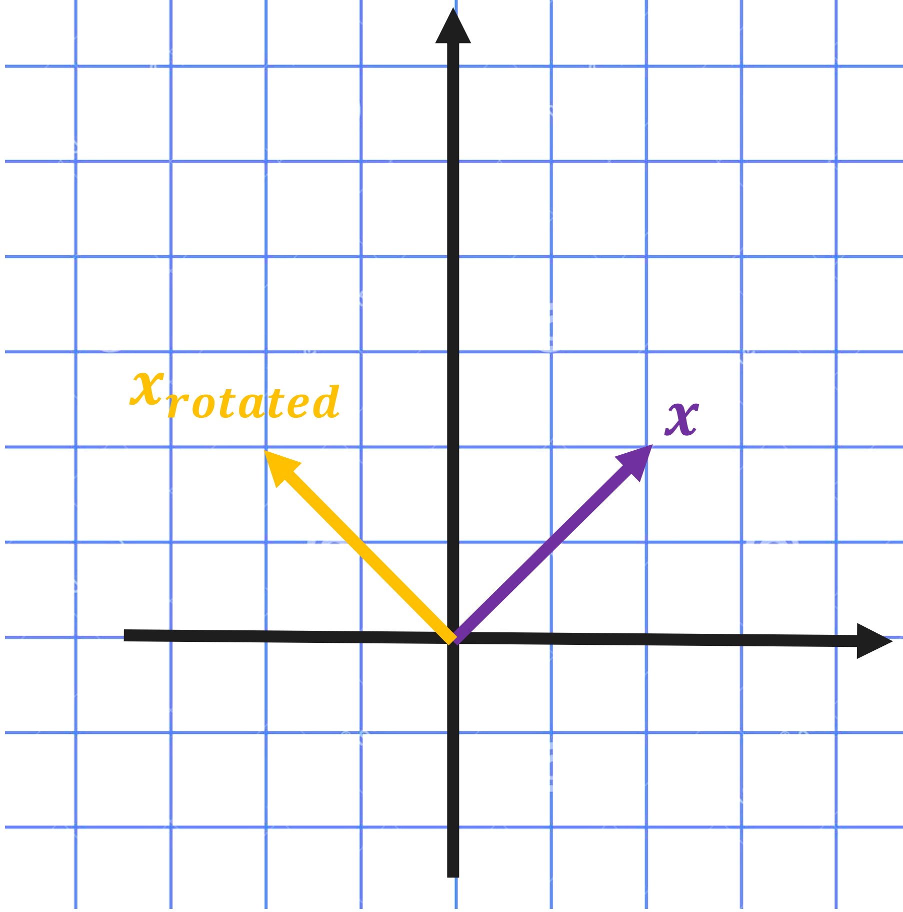
Composition

- Let A and B be two linear transforms.
  What if we first apply A and then B?

- Example: first rotate by 90°, then squeeze.

$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ – rotation, $B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$ – squeezing

$$B(Ax) =$$

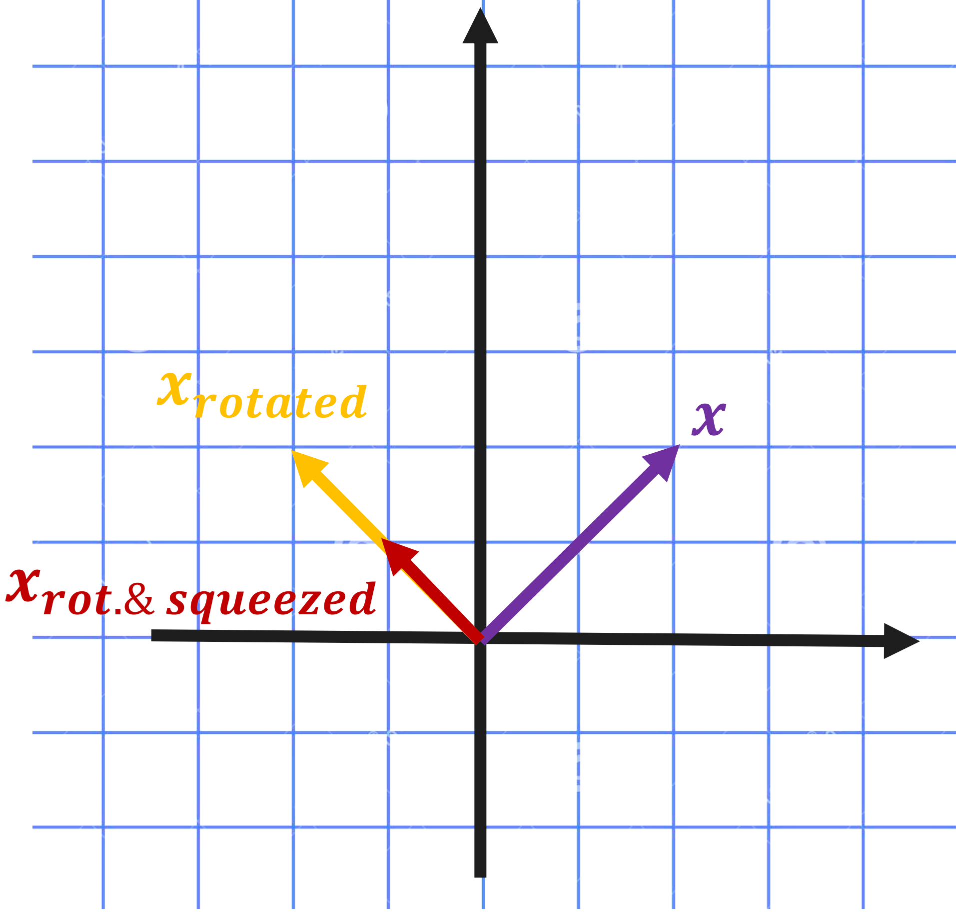


Composition

- Let A and B be two linear transforms.
  What if we first apply A and then B?

- Example: first rotate by 90°, then squeeze.

$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ – rotation, $B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$ – squeezing



$B(Ax) =$

girafe ai

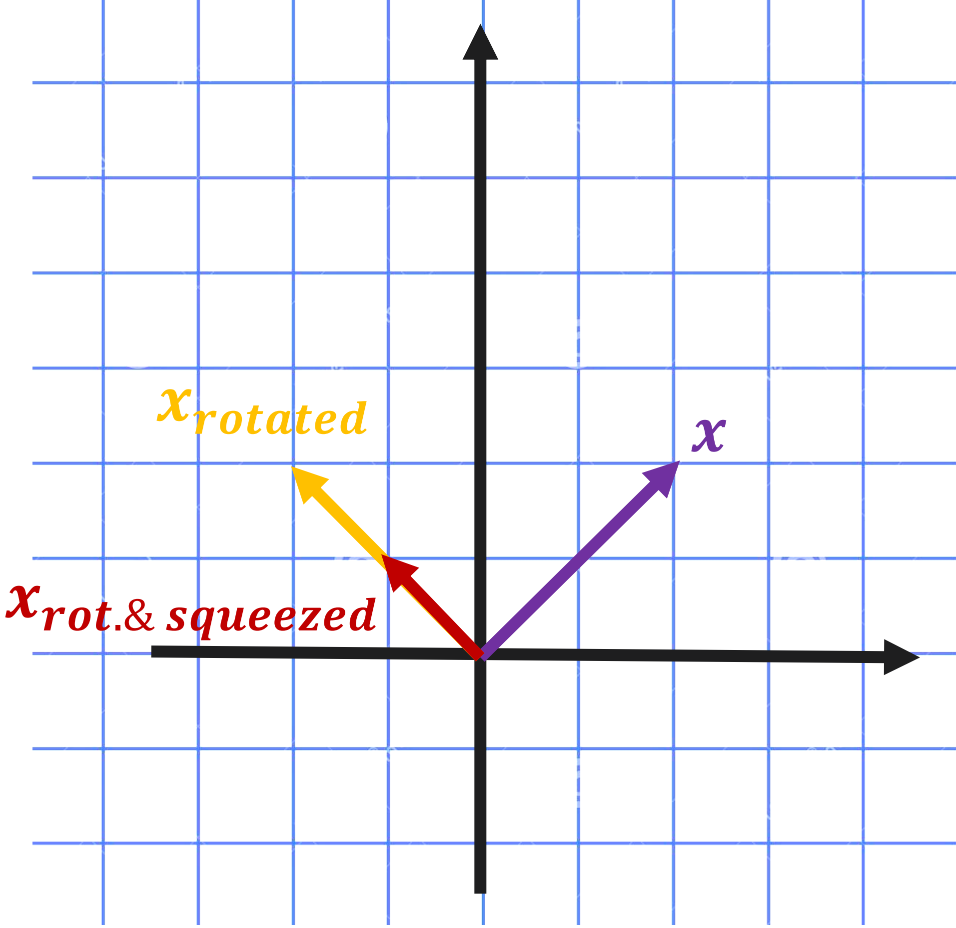
Composition

- Let A and B be two linear transforms.
  What if we first apply A and then B?

- Example: first rotate by 90°, then squeeze.

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \text{ – rotation,} \quad B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \text{ – squeezing}$$

$$B(Ax) = (BA)x = Cx$$



Composition

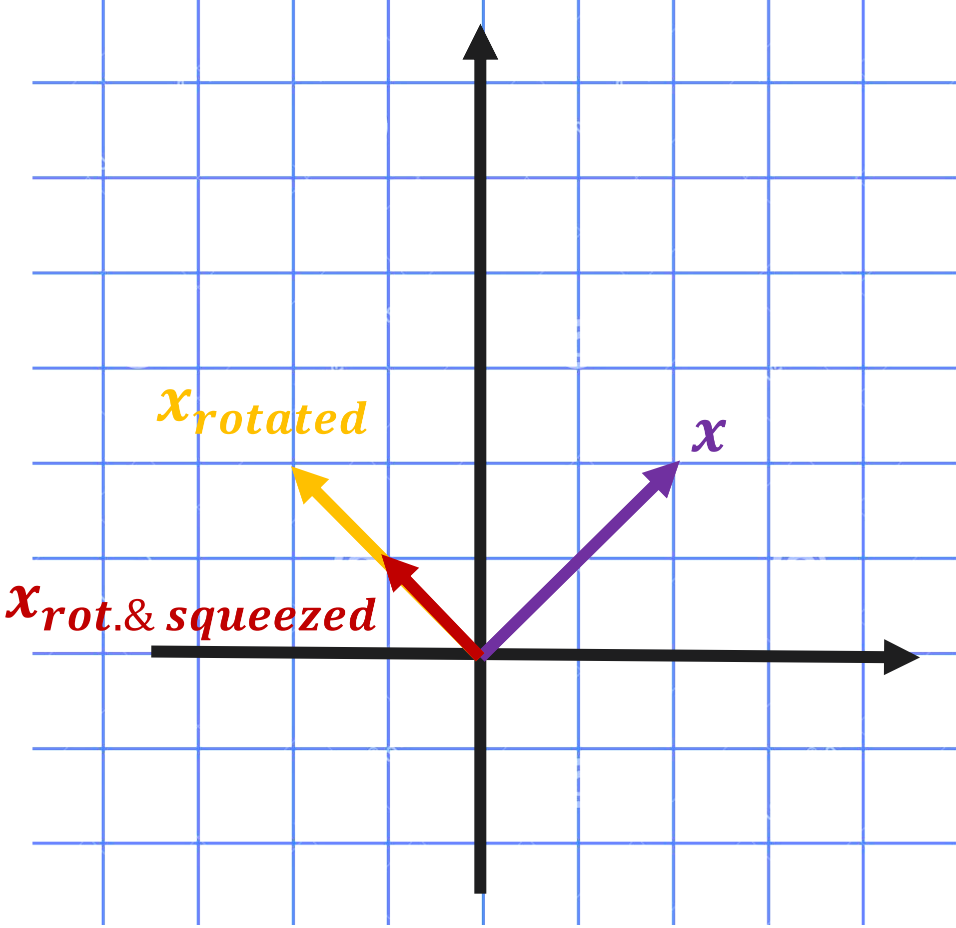
- Let A and B be two linear transforms.
  What if we first apply A and then B?

- Example: first rotate by 90°, then squeeze.

$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ – rotation, $B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$ – squeezing

$$B(Ax) = (BA)x = Cx$$

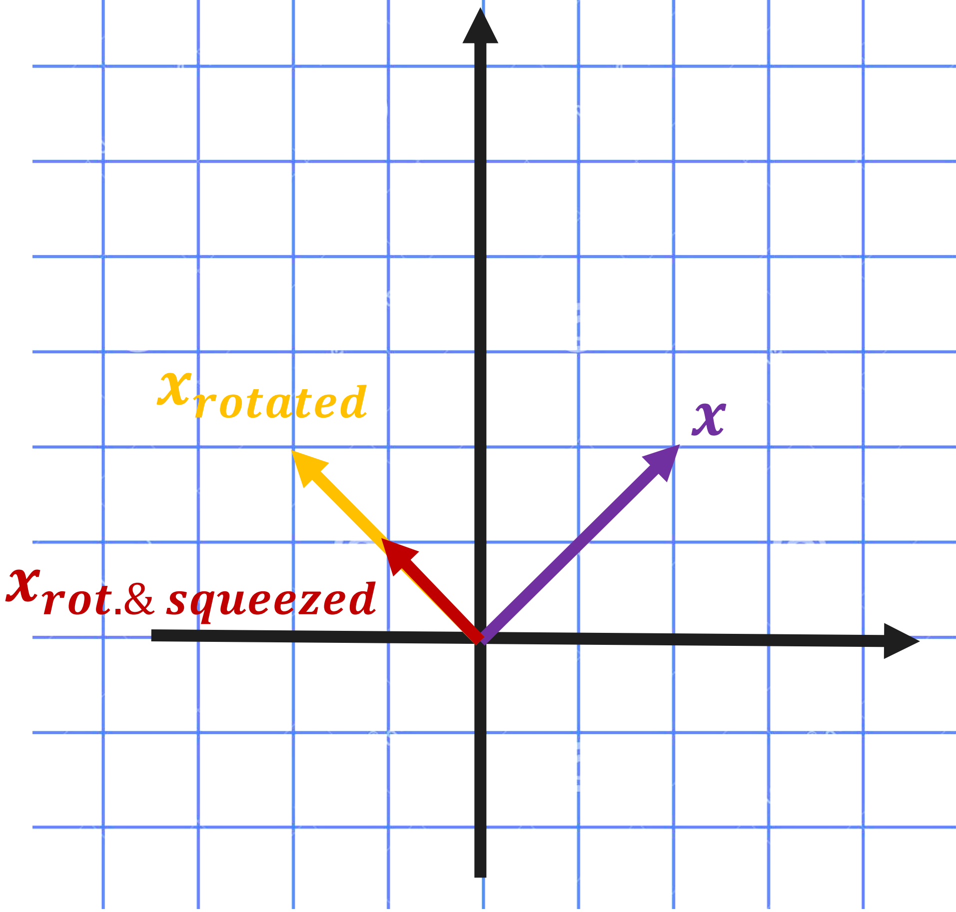
$$C = BA = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -0.5 \\ 0.5 & 0 \end{bmatrix}$$



Composition

- Let A and B be two linear transforms.
  What if we first apply A and then B?

- Example: first rotate by 90°, then squeeze.



$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ – rotation, $B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$ – squeezing

$$B(Ax) = (BA)x = Cx$$

$$C = BA = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -0.5 \\ 0.5 & 0 \end{bmatrix}$$

= “rotate by 90° and squeeze”

Sum of Transforms

- Let A and B be two linear transforms.

Sum of Transforms

- Let A and B be two linear transforms.

- $C = A + B$ – also a linear transform:

$$C = Ax + Bx$$

Sum of Transforms

- Let A and B be two linear transforms.
- $C = A + B$ – also a linear transform:

$$C = Ax + Bx$$

- Example:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} - \text{rotation}, \quad B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} - \text{squeezing}$$

Sum of Transforms

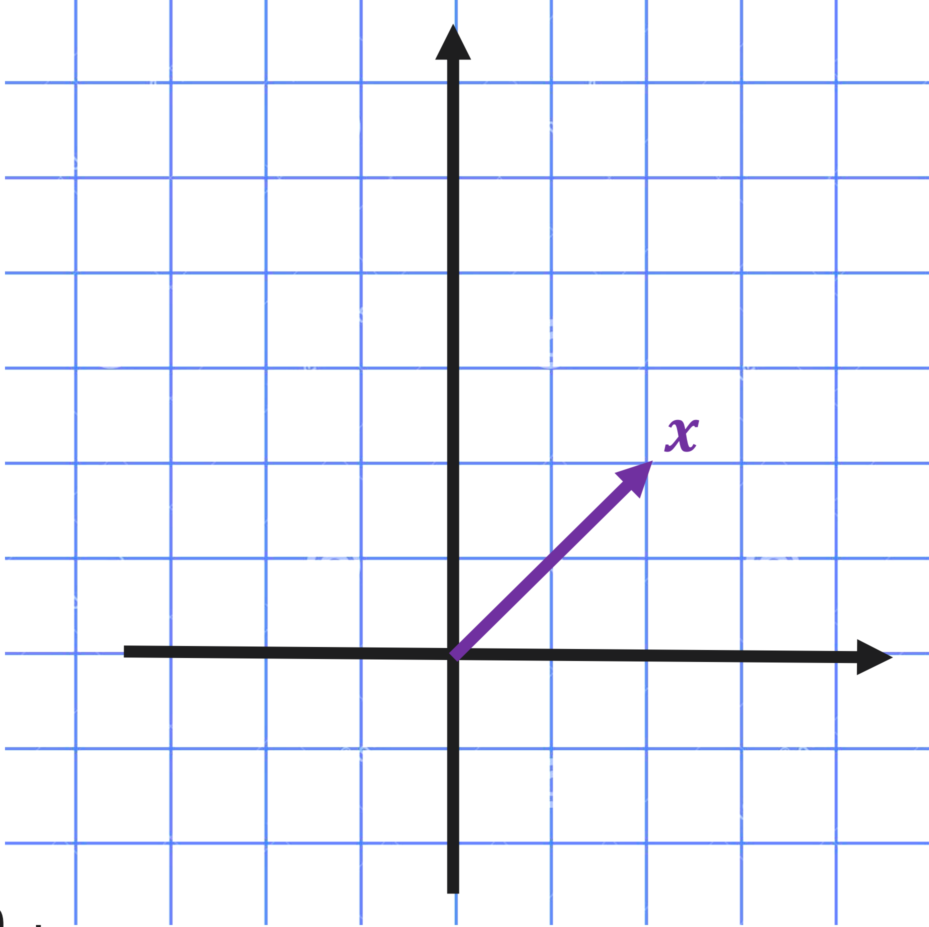
- Let A and B be two linear transforms.
- $C = A + B$ – also a linear transform:

$$C = Ax + Bx$$

- Example:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \text{ – rotation,} \quad B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \text{ – squeezing}$$

$$\mathrm{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$



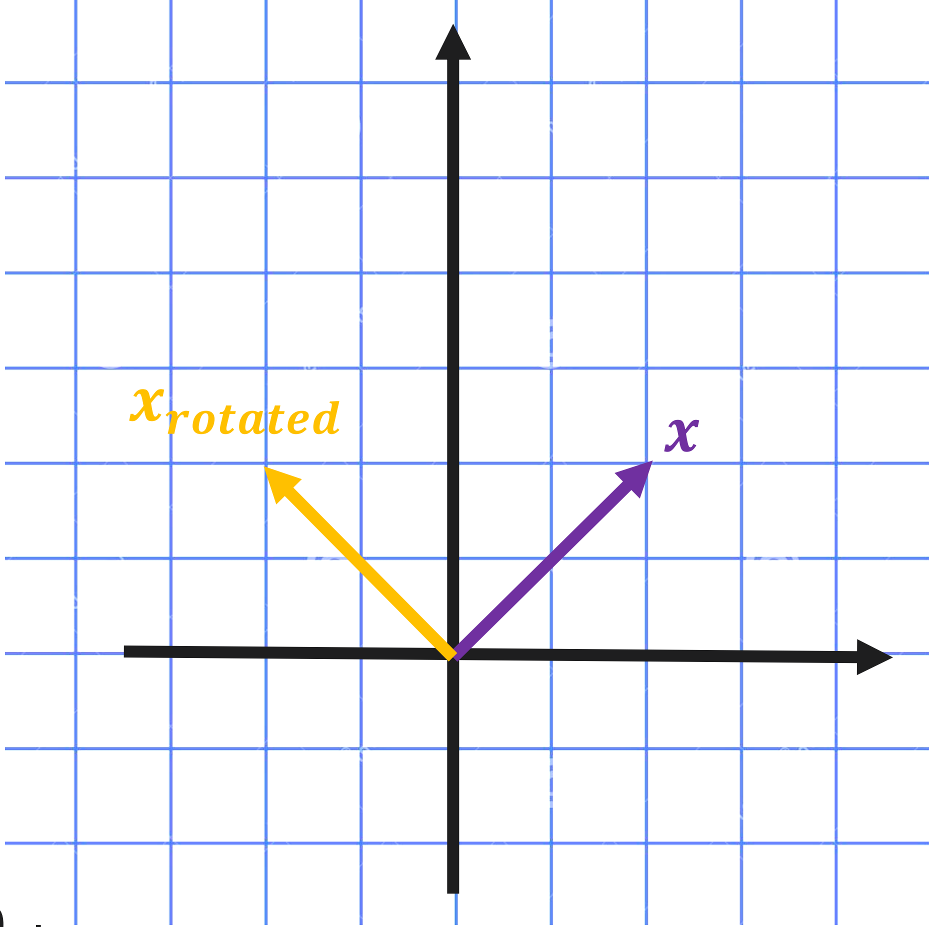
girafe ai

Sum of Transforms

- Let A and B be two linear transforms.
- $C = A + B$ – also a linear transform:

$$C = Ax + Bx$$

- Example:



$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \text{ – rotation,} \quad B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \text{ – squeezing}$$

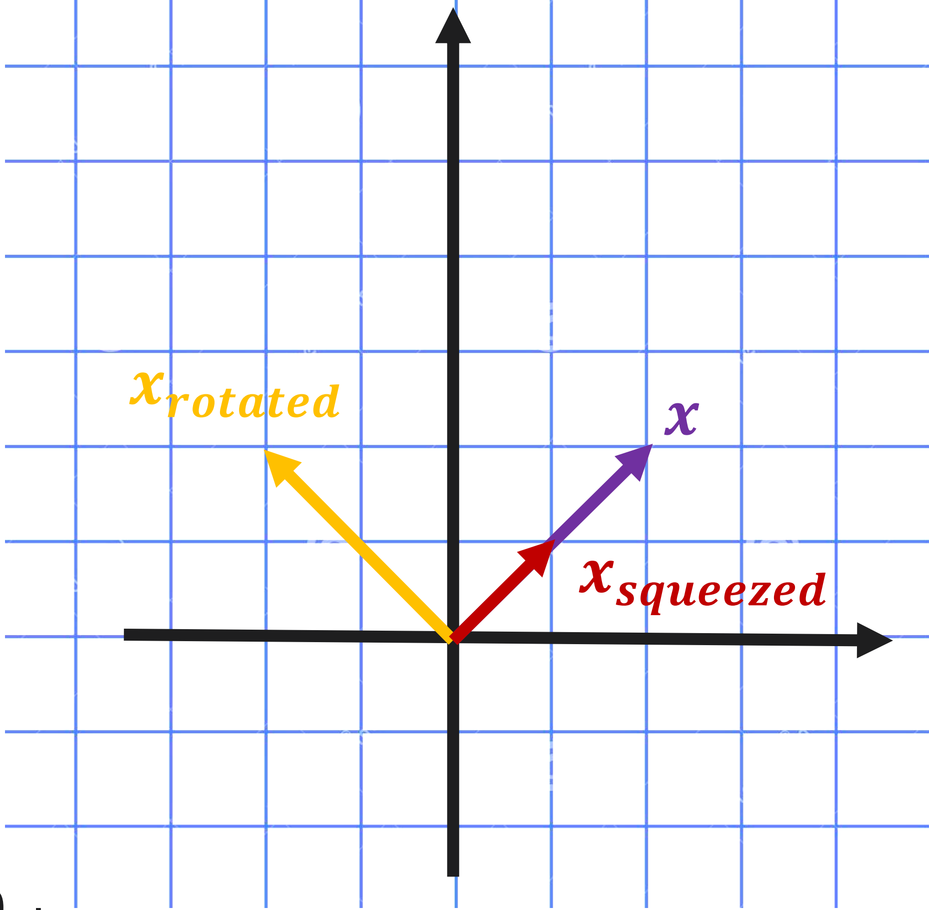
$$\mathrm{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Sum of Transforms

- Let A and B be two linear transforms.
- $C = A + B$ – also a linear transform:

$$C = Ax + Bx$$

- Example:



$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \text{ – rotation,} \quad B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \text{ – squeezing}$$

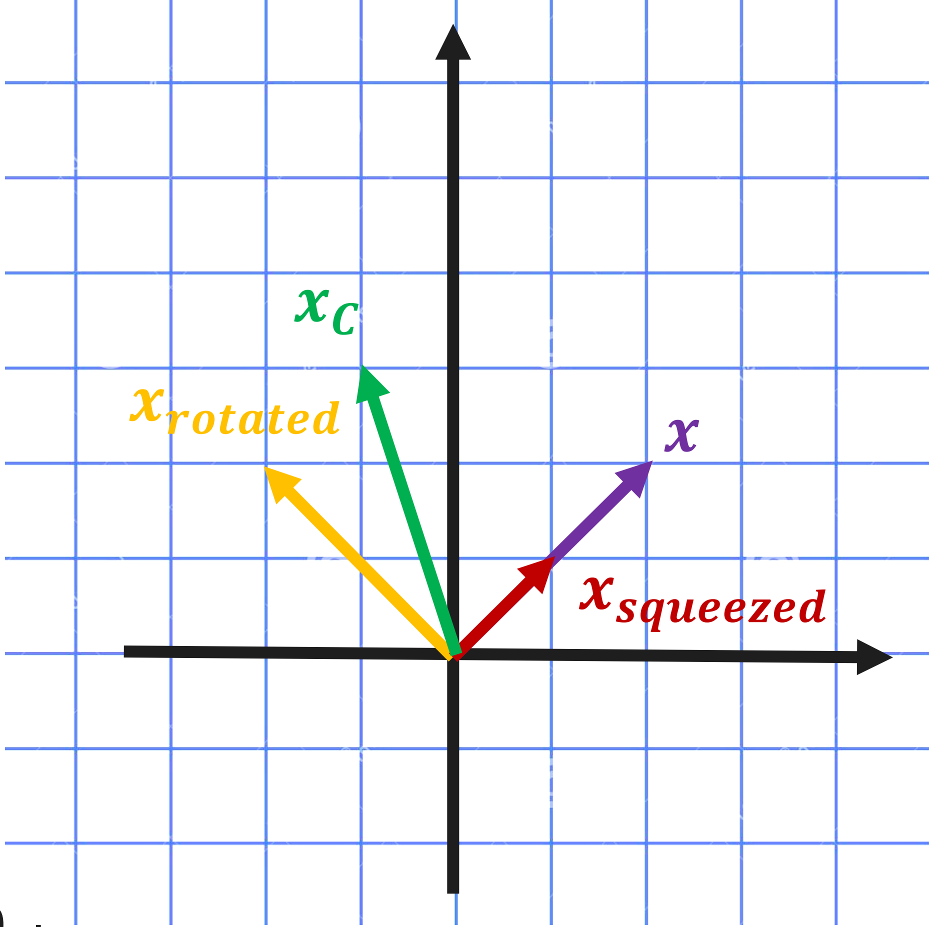
$$\mathrm{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Sum of Transforms

- Let A and B be two linear transforms.
- $C = A + B$ – also a linear transform:

$$C = Ax + Bx$$

- Example:



$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$ – rotation, $$B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$ – squeezing

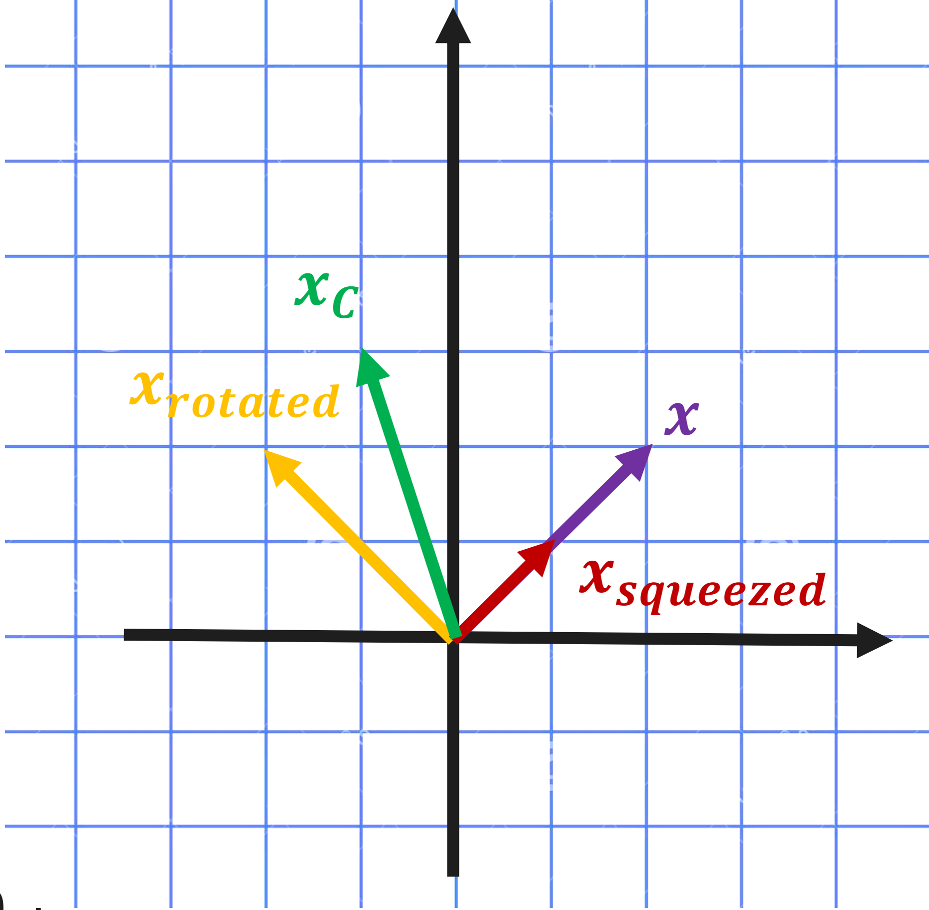
$$\mathrm{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Sum of Transforms

- Let A and B be two linear transforms.
- $C = A + B$ – also a linear transform:

$$C = Ax + Bx$$

- Example:



$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \text{ – rotation,} \quad B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \text{ – squeezing}$$

$$\mathrm{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad Ax + Bx = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 1.5 \end{bmatrix}$$

Inverse Transform

Inverse transform

Inverse transform

- Consider a transform A.

Inverse transform

- Consider a transform A.
- Inverse transform A^{-1} “maps everything back to where it was”:

Inverse transform

- Consider a transform A.
- Inverse transform A^{-1} “maps everything back to where it was”:

$$A^{-1}(Ax) = x \iff A^{-1}A = E - \text{identity transformation.}$$

Inverse transform

- Consider a transform A.
- Inverse transform A^{-1} "maps everything back to where it was":

$$A^{-1}(Ax) = x \iff A^{-1}A = E - \text{identity transformation.}$$

- Not every transform has an inverse!

Inverse transform

- Consider a transform A.
- Inverse transform A^{-1} “maps everything back to where it was”:

$$A^{-1}(Ax) = x \iff A^{-1}A = E$$ – identity transformation.

- Not every transform has an inverse!
    - Rotation: yes (rotate it back)

Inverse transform

- Consider a transform A.
- Inverse transform A^{-1} "maps everything back to where it was":

$$A^{-1}(Ax) = x \iff A^{-1}A = E \text{ – identity transformation.}$$

- Not every transform has an inverse!
  - Rotation: yes (rotate it back)
  - Projection: no

Inverse of a Matrix

- An n x n matrix A has an inverse if there exists A^{-1} such that

$$AA^{-1} = A^{-1}A = E$$

Inverse of a Matrix

- An n x n matrix A has an inverse if there exists A^{-1} such that

$$AA^{-1} = A^{-1}A = E$$

- A matrix that doesn't have an inverse is called singular or degenerate.

Inverse of a Matrix

- An n x n matrix A has an inverse if there exists A^{-1} such that

$$AA^{-1} = A^{-1}A = E$$

- A matrix that doesn’t have an inverse is called singular or degenerate.

- Which matrices have an inverse?

Determinant

Determinant

- A numerical way to characterize a linear transformation (and its matrix):
    - absolute value = how much area changes;
    - sign = change of orientation.

- More info on the interpretation: see video.

Determinant

- A – linear transform.

Determinant

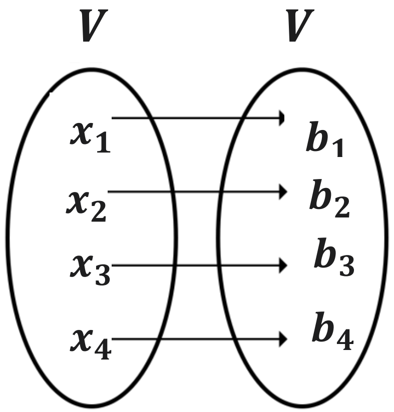
- A – linear transform.

- The determinant is nonzero $\Leftrightarrow A$ is invertible and the linear map represented by matrix A is an isomorphism:

Determinant

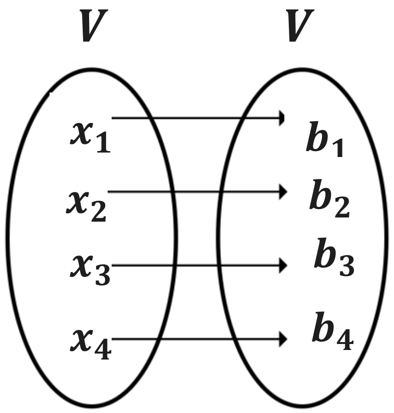
- A – linear transform.

- The determinant is nonzero $\Leftrightarrow$ A is invertible and the linear map represented by matrix A is an isomorphism:
  - $\det A \neq 0$:
    - A:

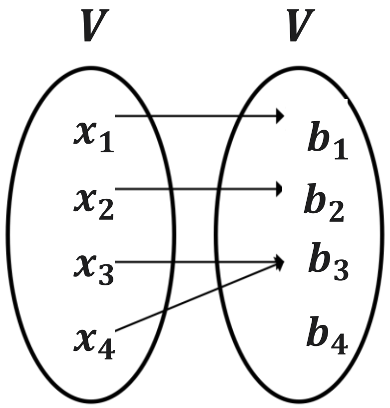


Determinant

- A – linear transform.
- The determinant is nonzero $\Leftrightarrow$ A is invertible and the linear map represented by matrix A is an **isomorphism**:
  - $\det A \neq 0$:
    - A:



  - $\det A = 0$:
    - Several vectors are mapped onto the same vector $\Leftrightarrow$ A maps original vector space onto a lower-dimensional space.



Computing Determinant

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Computing Determinant

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

- Example:

$$R_{90} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = 0 - (-1) = 1 \iff$$

"there is a transform inverse to rotation by $\mathbf{90}^{\circ}$ anticlockwise".

Computing Determinant

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

girafe ai

Computing Determinant

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

- Example:

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{vmatrix} = 1 \cdot \begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix} - 0 + 0 = 0 \Leftrightarrow$$

“there is no transpose inverse to projection onto XY-plane”

girafe ai

Computing Determinant

- $A = \{a_{ij}\} - n \times n$ matrix.

Computing Determinant

- $A = \{a_{ij}\} - n \times n$ matrix.

- M_{ij} – its minor $\Leftrightarrow M_{ij}$ is an $(n-1) \times (n-1)$ matrix resulting from removing i-th row and j-th column from A.

Computing Determinant

- $A = \{a_{ij}\} - n \times n$ matrix.

- M_{ij} – its minor $\Leftrightarrow M_{ij}$ is an $(n-1) \times (n-1)$ matrix resulting from removing i-th row and j-th column from A.

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det M_{ij} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det M_{ij}$$

Computing Determinant

- $A = \{a_{ij}\} - n \times n$ matrix.

- M_{ij} – its minor $\Leftrightarrow M_{ij}$ is an $(n-1) \times (n-1)$ matrix resulting from removing i-th row and j-th column from A.

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det M_{ij} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det M_{ij}$$

- Laplace extension.

Some Properties of the Determinant

- $\det A^T = \det A$
- $\det AB = \det A \cdot \det B$
- $\det A^{-1} = \dfrac{1}{\det A}$

girafe ai

Rank

Column Space

- Consider a square matrix A.

- Its columns $A^1, \ldots, A^n$ can be seen as vectors.

Column Space

- Consider a square matrix A.

- Its columns $A^1, \dots, A^n$ can be seen as vectors.

- $U = span\{A^1, \dots, A^n\}$ – column space of A.
    - All vectors that can be obtain by linearly combining columns of A.
    - $\Leftrightarrow$ image of linear transformation A (= all the vectors we can get by applying A).

Rank

- Column space $U = span\{A^1, \dots, A^n\}$ is the image of linear transformation A.

- Rank of a matrix is the number of dimensions in its column space.

Rank

- Column space $U = span\{A^1, \dots, A^n\}$ is the image of linear transformation A.
- Rank of a matrix is the number of dimensions in its column space.
    - Full rank matrix: n columns, all linearly independent.
    - Lower-rank matrices: linearly dependent columns present.

Rank: Examples

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

- $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$

- $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Rank: Examples

- $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ rank(A) = 1$

- $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$

- $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Rank: Examples

- $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ rank(A) = 1$

- $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}, \ rank(A) = 2$

- $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Rank: Examples

- $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ rank(A) = 1$

- $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}, \ rank(A) = 2$

- $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ rank(A) = 3$

Column vs Row Rank

- Column space of A = span of A's columns.
  Its dimensionality = (column) rank.

Column vs Row Rank

- Column space of A = span of A's columns.
  Its dimensionality = (column) rank.

- We can also define the row space of A as a span of A's rows.
  Its dimensionality = row rank.

Column vs Row Rank

- Column space of A = span of A's columns.
  Its dimensionality = (column) rank.

- We can also define the row space of A as a span of A's rows.
  Its dimensionality = row rank.

- Column rank vs. row rank?

Column vs Row Rank

- Column space of A = span of A's columns.
  Its dimensionality = (column) rank.

- We can also define the row space of A as a span of A's rows.
  Its dimensionality = row rank.

- Column rank vs. row rank?

- Fundamental result: the column rank and the row rank are always equal.
  See proofs.

Rank

- This is why there cannot be more than n linearly independent vectors in $\mathbb{R}^n$!

Rank

- This is why there cannot be more than n linearly independent vectors in $\mathbb{R}^n$!

- Imagine $x_1, x_2, \dots, x_m \in \mathbb{R}^n$:

Rank

- This is why there cannot be more than n linearly independent vectors in $\mathbb{R}^n$!

- Imagine $x_1, x_2, \dots, x_m \in \mathbb{R}^n$:

$$X = [x_1 \mid x_2 \mid \dots \mid x_n] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

Rank

- This is why there cannot be more than n linearly independent vectors in $\mathbb{R}^n$!

- Imagine $x_1, x_2, \dots, x_m \in \mathbb{R}^n$:

$$X = [x_1 \mid x_2 \mid \dots \mid x_n] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

$$rank(X) \leq \min\{n, m\}$$

girafe ai

Rank: Examples

- $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},\ rank(A) = 1$

- $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix},\ rank(A) = 2$

- $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ rank(A) = 3$

Rank: Examples

- $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $rank(A) = 1$ $\Leftrightarrow$ $\mathbb{R}^3$ is mapped onto a line
- $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$, $rank(A) = 2$
- $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $rank(A) = 3$

Rank: Examples

- $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $rank(A) = 1 \Leftrightarrow \mathbb{R}^3$ is mapped onto a line

- $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$, $rank(A) = 2 \Leftrightarrow \mathbb{R}^3$is mapped onto a plane

- $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $rank(A) = 3$

Rank: Examples

- $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $rank(A) = 1 \iff \mathbb{R}^3$ is mapped onto a line

- $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$, $rank(A) = 2 \iff \mathbb{R}^3$is mapped onto a plane

- $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $rank(A) = 3 \iff \mathbb{R}^3$is mapped on itself (isomorphism)

girafe
ai

Rank: Examples

- $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $rank(A) = 1 \iff \mathbb{R}^3$ is mapped onto a line

  Infinitely many vectors are mapped into a zero vector.

- $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$, $rank(A) = 2 \iff \mathbb{R}^3$is mapped onto a plane

  Infinitely many vectors are mapped into a zero vector.

- $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $rank(A) = 3 \iff \mathbb{R}^3$is mapped on itself (isomorphism)

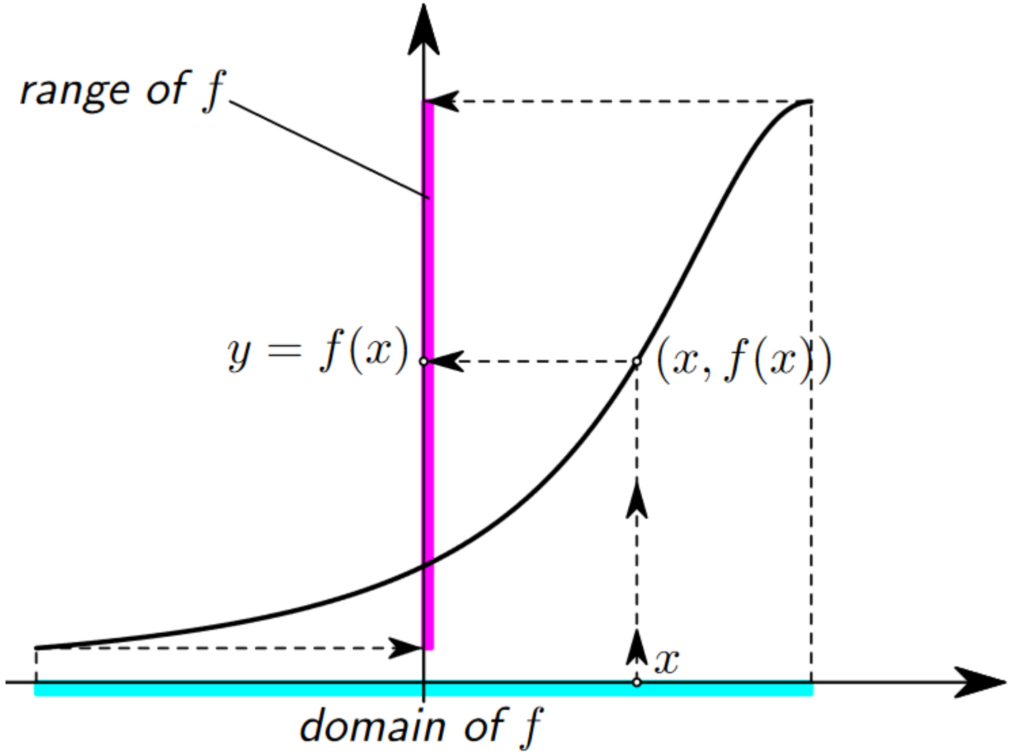
  Only a zero vector is mapped into a zero vector.

girafe ai

Functions

What is a Function?

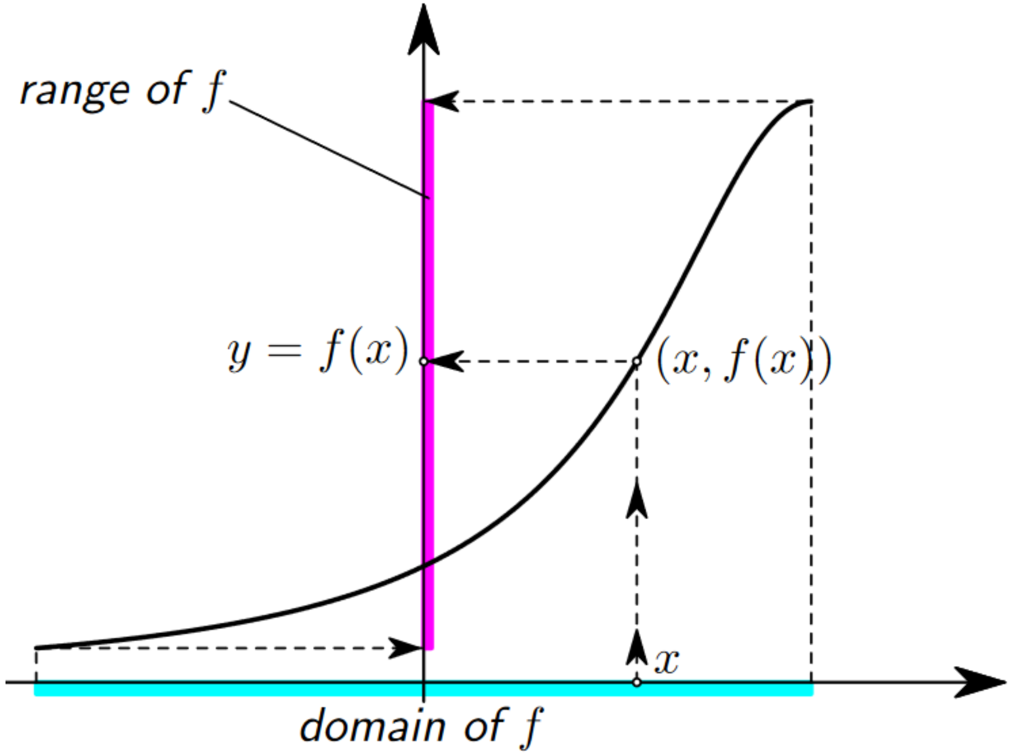
- **Function** describes the relationship between x and y



What is a Function?

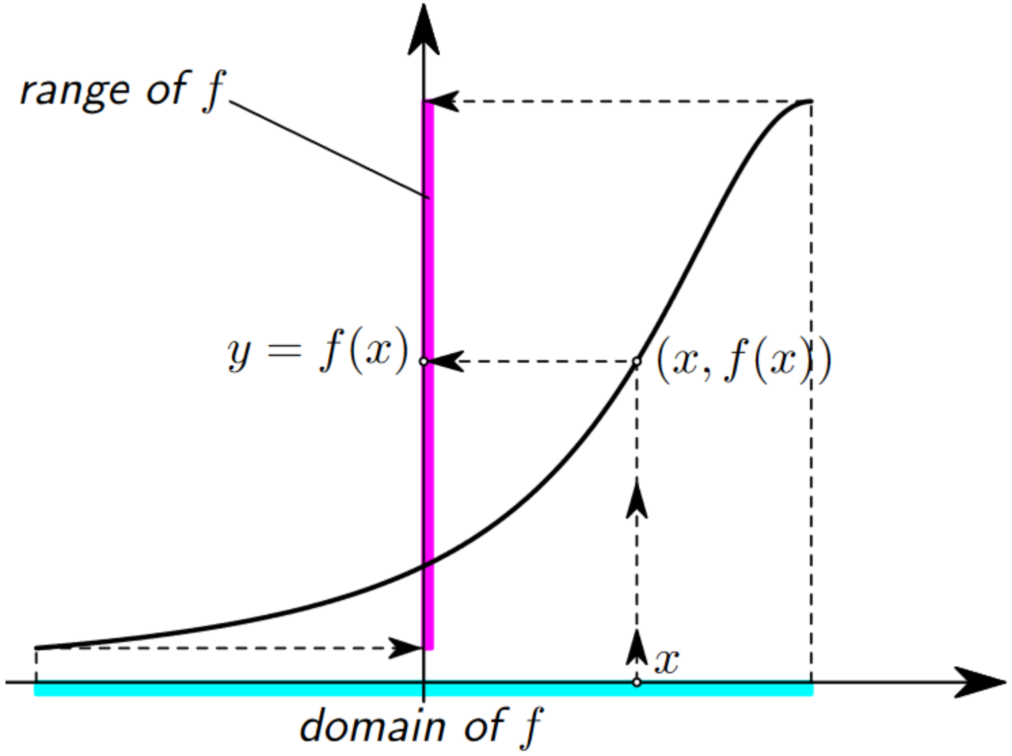
- Function describes the relationship between x and y

- *Domain:* is the set of numbers for which the function is defined



What is a Function?

- **Function** describes the relationship between x and y
- *Domain:* is the set of numbers for which the function is defined
- *Range:* the set of all possible numbers $f(x)$ as x runs over its domain



Some Univariate Functions

- A linear function:

$$f(x) = 2x + 1, \qquad f: \mathbb{R} \rightarrow \mathbb{R}$$

Some Univariate Functions

- A linear function:

intercept

$$f(x) = 2x + 1, \qquad f: \mathbb{R} \to \mathbb{R}$$

slope

girafe ai

Some Univariate Functions

- A linear function:

$$f(x) = 2x + 1, \qquad f: \mathbb{R} \to \mathbb{R}$$

- A polynomial function:

$$f(x) = x^2 - 2x + 1, \qquad f: \mathbb{R} \to \mathbb{R}$$

Some Univariate Functions

- A linear function:

$$f(x) = 2x + 1, \qquad f: \mathbb{R} \to \mathbb{R}$$

- A polynomial function:

$$f(x) = x^2 - 2x + 1, \qquad f: \mathbb{R} \to \mathbb{R}$$

- An exponential function:

$$f(x) = 10^x, \qquad f: \mathbb{R} \to \mathbb{R}^+$$

girafe ai

Some Univariate Functions

- A linear function:

$$f(x) = 2x + 1, \qquad f: \mathbb{R} \to \mathbb{R}$$

- A polynomial function:

$$f(x) = x^2 - 2x + 1, \qquad f: \mathbb{R} \to \mathbb{R}$$

- An exponential function:

$$f(x) = 10^x, \qquad f: \mathbb{R} \to \mathbb{R}^+$$

- A trigonometric function:

$$f(x) = \sin\ x\,, \qquad f: \mathbb{R} \to [0,1]$$

Limit of a Function

Limit

$$\lim_{x \to a} f(x) = L$$

- “The limit of $f(x)$ as x approaches a is L”

- Informally: for x close to a, f(x) is close to L.
  The closer x gets to a, the closer f(x) gets to L.

Limit

$$\lim_{x \to a} f(x) = L$$

- “The limit of $f(x)$ as x approaches a is L”
- Informally: for x close to a, f(x) is close to L.
  The closer x gets to a, the closer f(x) gets to L.
- Formally:

$$\forall \varepsilon > 0\ \exists \delta > 0: |x - a| < \delta \rightarrow |f(x) - L| < \varepsilon$$

Limit - Examples

$$\lim_{x \to +\infty} \frac{1}{x} = 0$$

Limit - Examples

$$\lim_{x \to +\infty} \frac{1}{x} = 0$$

$$\lim_{x \to 0} \frac{1}{x} = +\infty$$

Limit - Examples

$$\lim_{x \to +\infty} \frac{1}{x} = 0$$

$$\lim_{x \to 0} \frac{1}{x} = +\infty$$

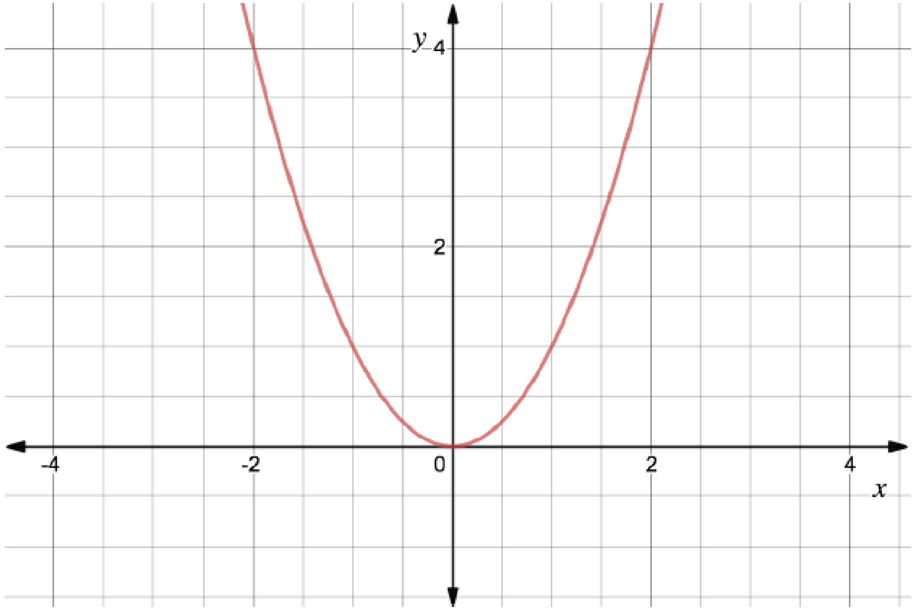
$$\lim_{x \to 2} \frac{x^2 - 2x}{x^2 - 4} = \lim_{x \to 2} \frac{x(x-2)}{(x-2)(x+2)} = \lim_{x \to 2} \frac{x}{x+2} = 0.5$$

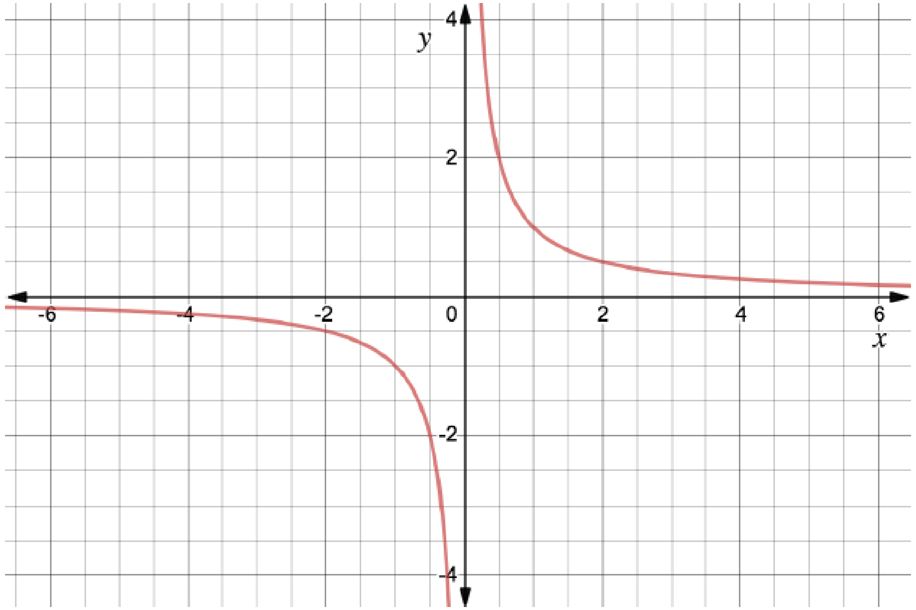
girafe ai

Properties of Functions

Continuity - Informally

- Very basic definition: a **continuous function** is one that can be drawn in one continuous stroke.

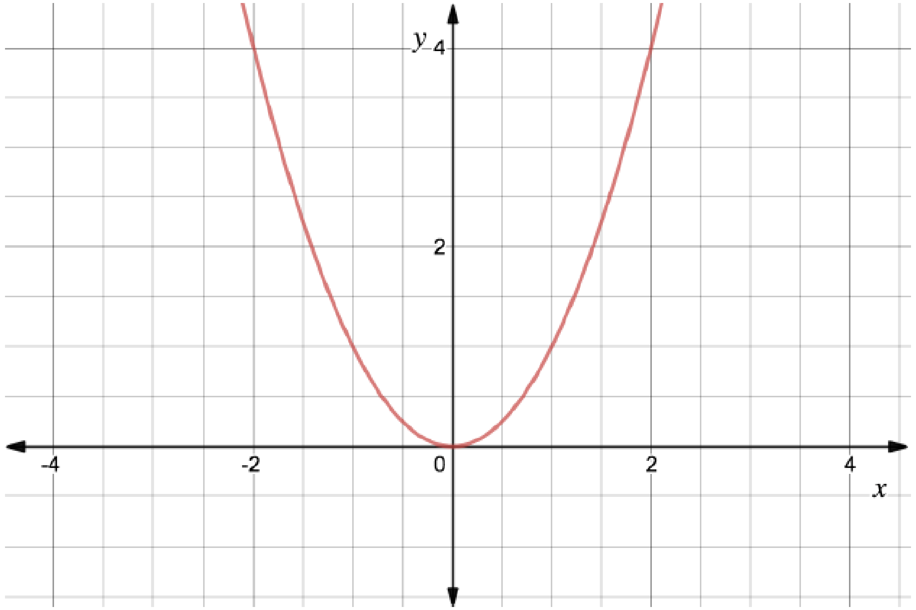


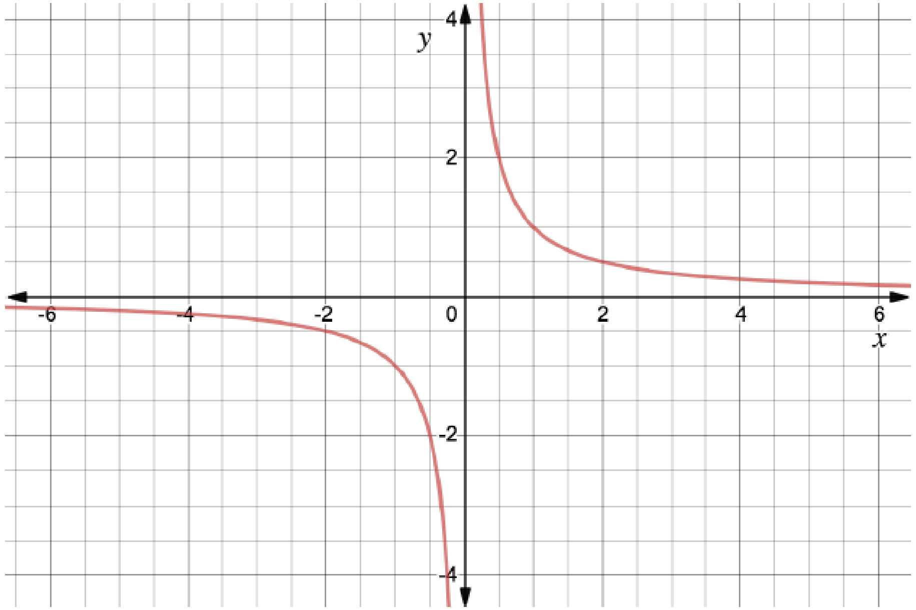


Source: https://www.expii.com/t/precalculus-concepts-continuous-vs-discontinuous-4767

Continuity - Informally

- Very basic definition: a **continuous function** is one that can be drawn in one continuous stroke.





- **Intermediate value property:** if a continuous function takes on two values, it must also take on all values in between.

Continuity - Formally

- A function $f(x)$ is continuous if for every x_0 in its domain

$$\lim_{x \to x_0} f(x) = f(x_0)$$

Continuity - Formally

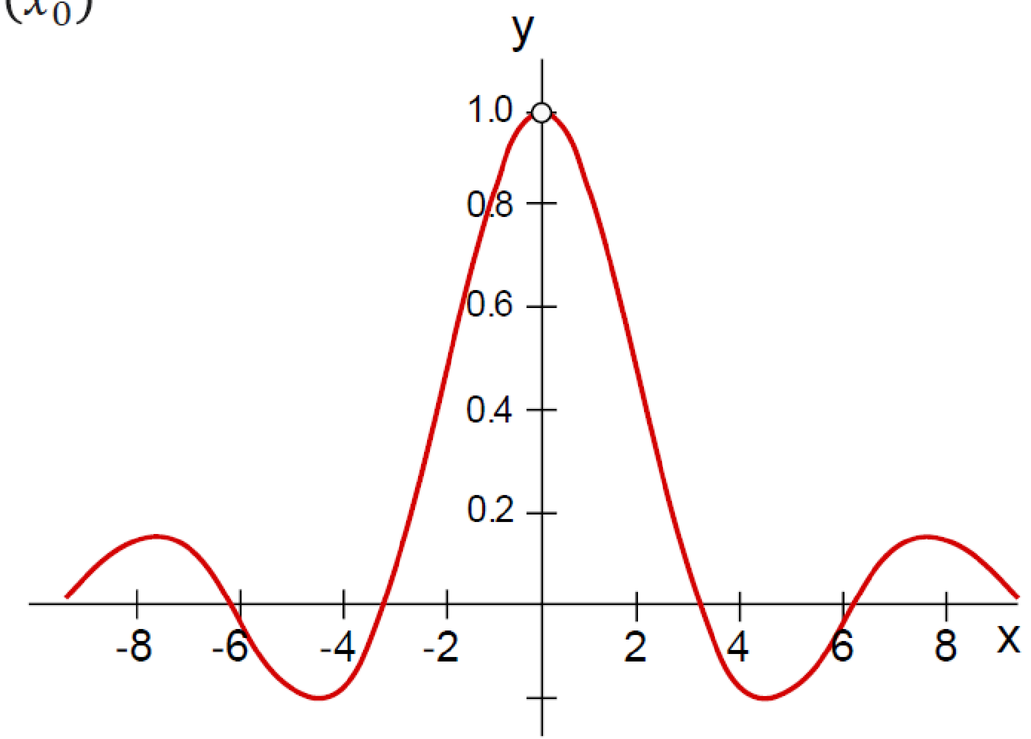
- A function $f(x)$ is continuous if for every x_0 in its domain

$$\lim_{x \to x_0} f(x) = f(x_0)$$

- Example:

$$f(x) = \frac{\sin x}{x}$$

Not defined at $x_0 = 0$, but $\lim_{x \to 0} \frac{\sin x}{x} = 1$.



Continuity - Formally

- A function $f(x)$ is continuous if for every x_0 in its domain

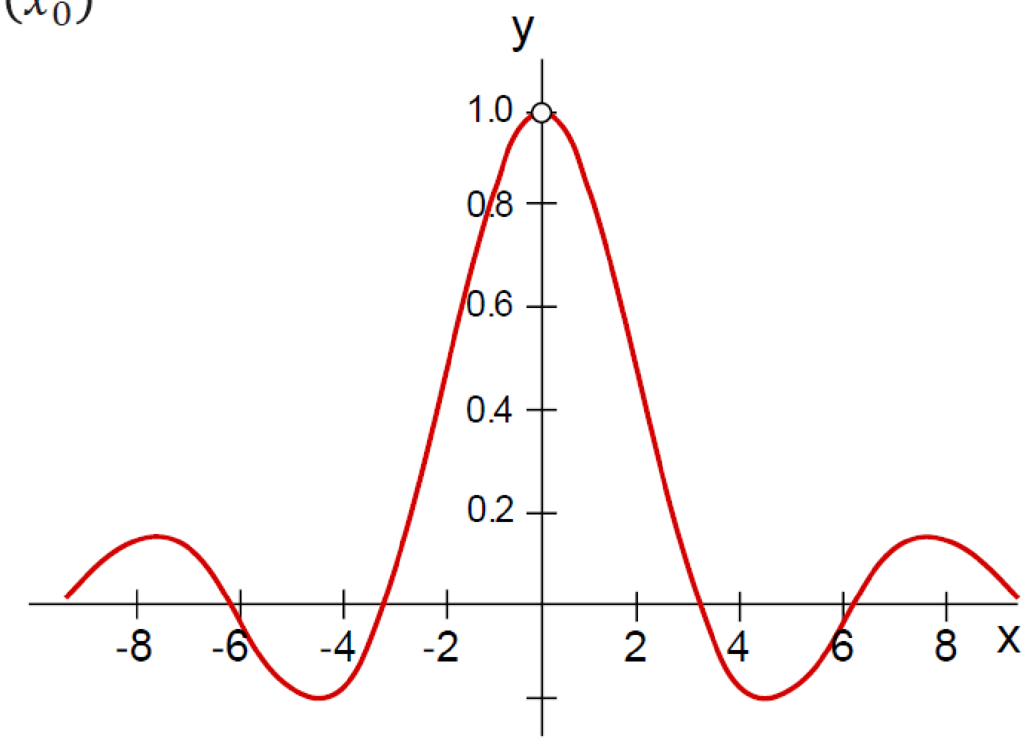
$$\lim_{x \to x_0} f(x) = f(x_0)$$

- Example:

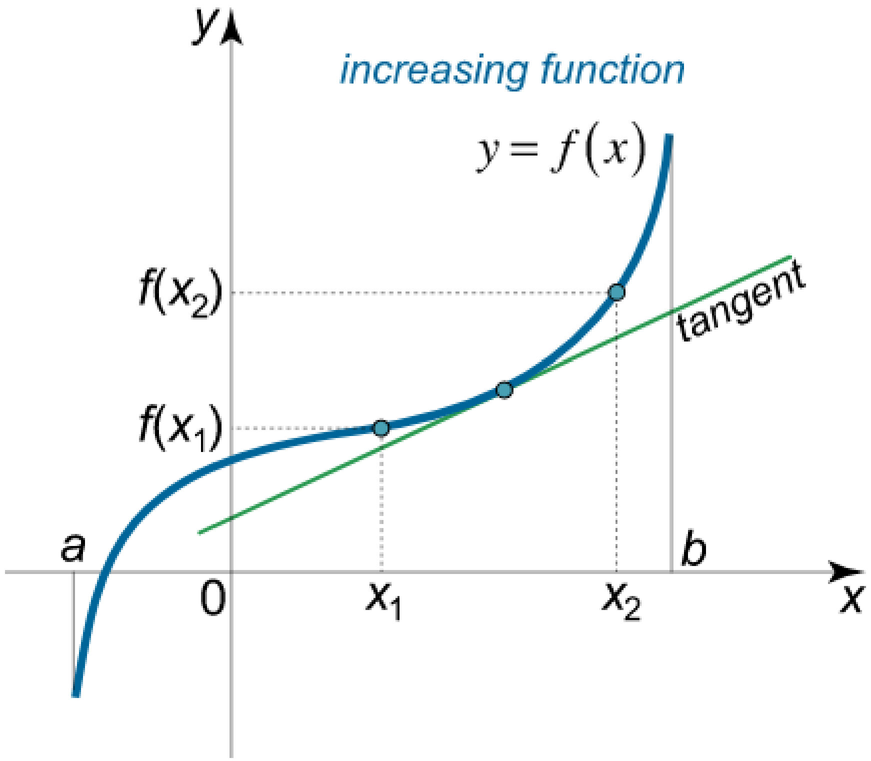
$$f(x) = \frac{\sin x}{x}$$

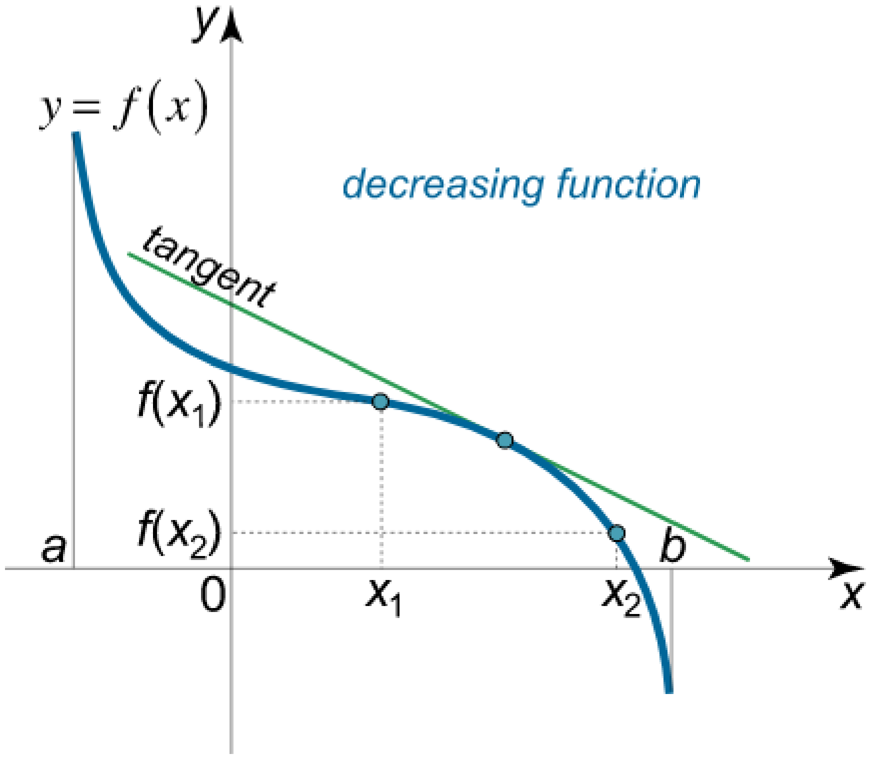
Not defined at $x_0 = 0$, but $\lim_{x \to 0} \frac{\sin x}{x} = 1$.

$f(x) = \begin{cases} \frac{\sin x}{x}, & x \neq 0 \\ 1, & x = 0 \end{cases}$ is a continuous function!



Increasing / Decreasing

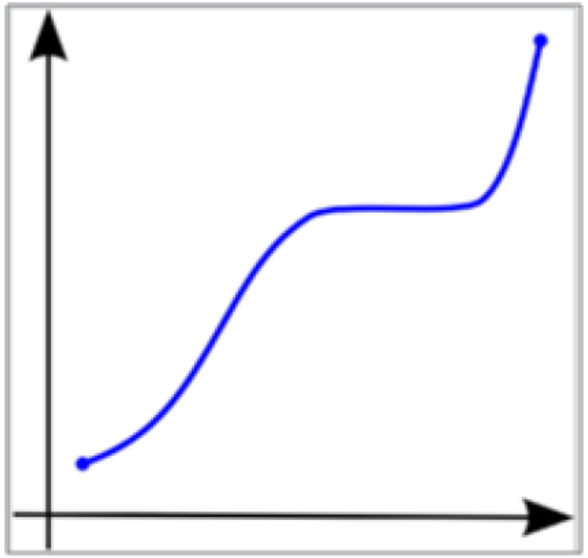




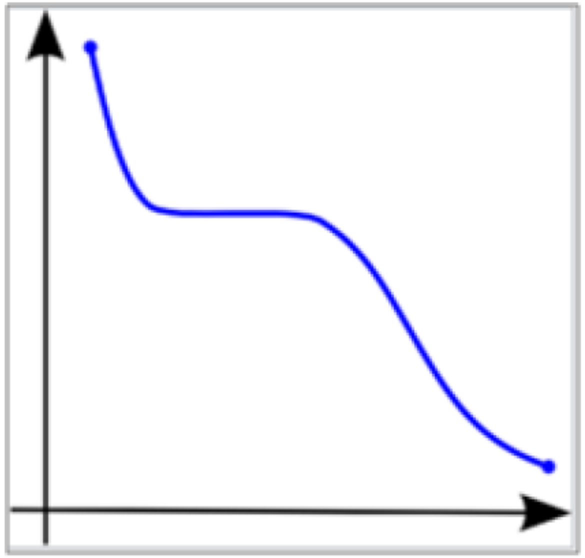
Source: https://math24.net/increasing-decreasing-functions.html

Monotonicity

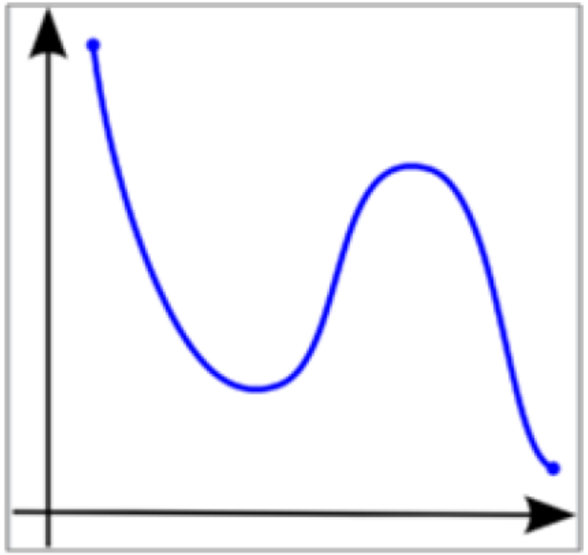
- A **monotonic function** = a (non-) increasing / decreasing function over the whole domain.



A monotonically non-decreasing function.



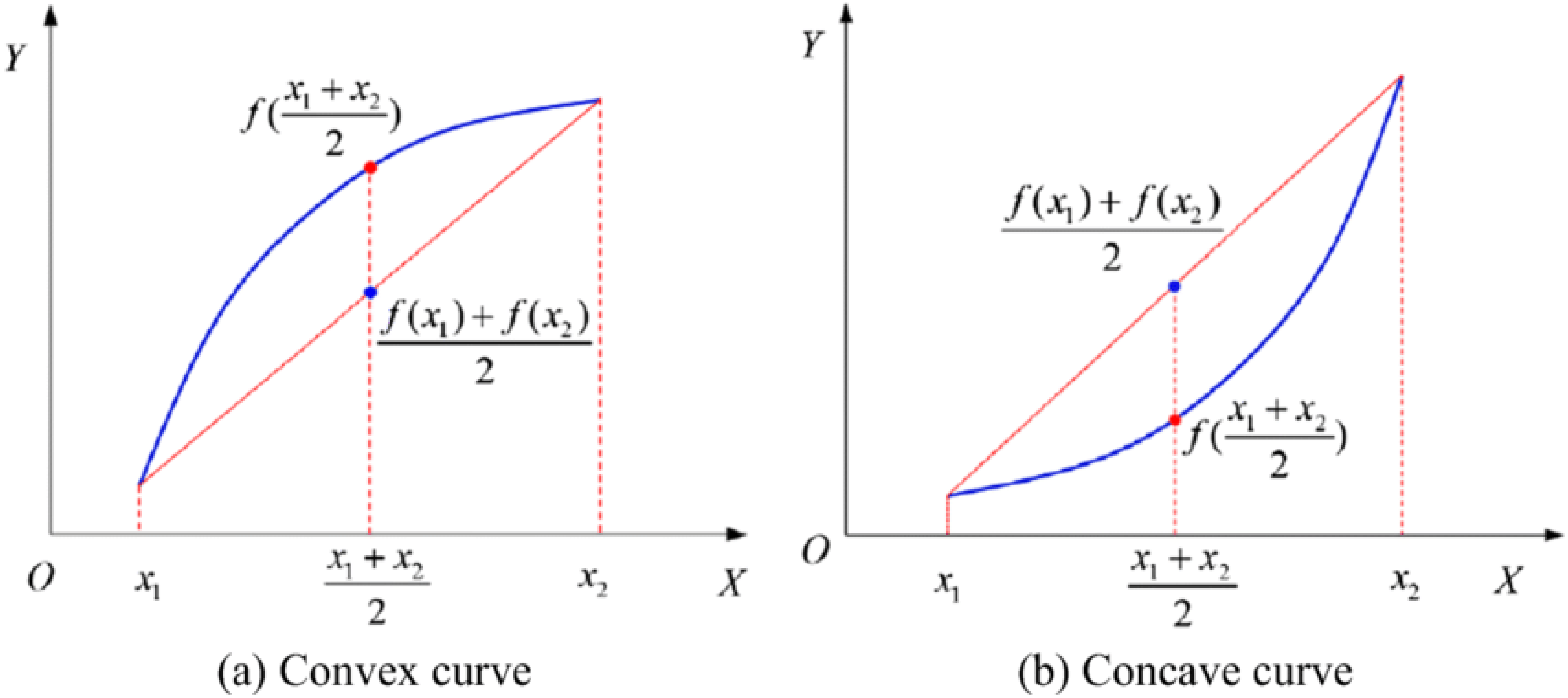
A monotonically non-increasing function.



A non-monotonic function.

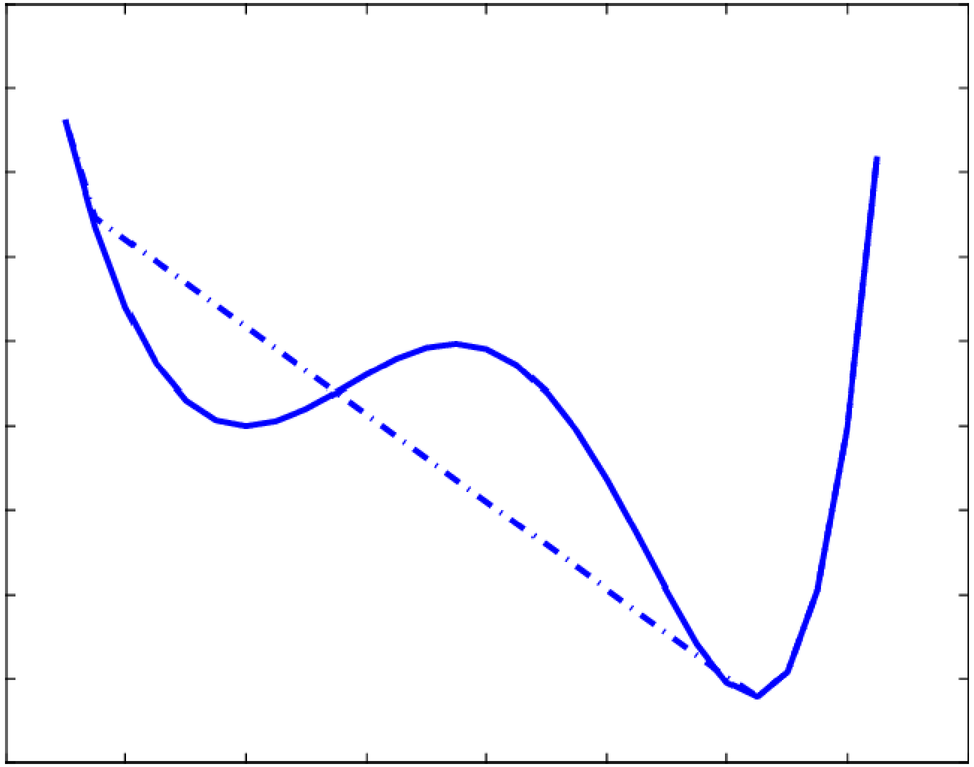
Source: https://en.wikipedia.org/wiki/Monotonic_function

Convexity



Source: https://www.researchgate.net/figure/Judgment-of-concavity-and-convexity-a-Convex-curve-b-Concave-curve_fig3_339939083

Convexity



A non-convex curve

Derivatives

Derivative

- A way to measure change:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

Derivative

- A way to measure change:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

- Derivative of the function f at the point x tells us how much the function f changes as the input x changes by a small amount Δx:

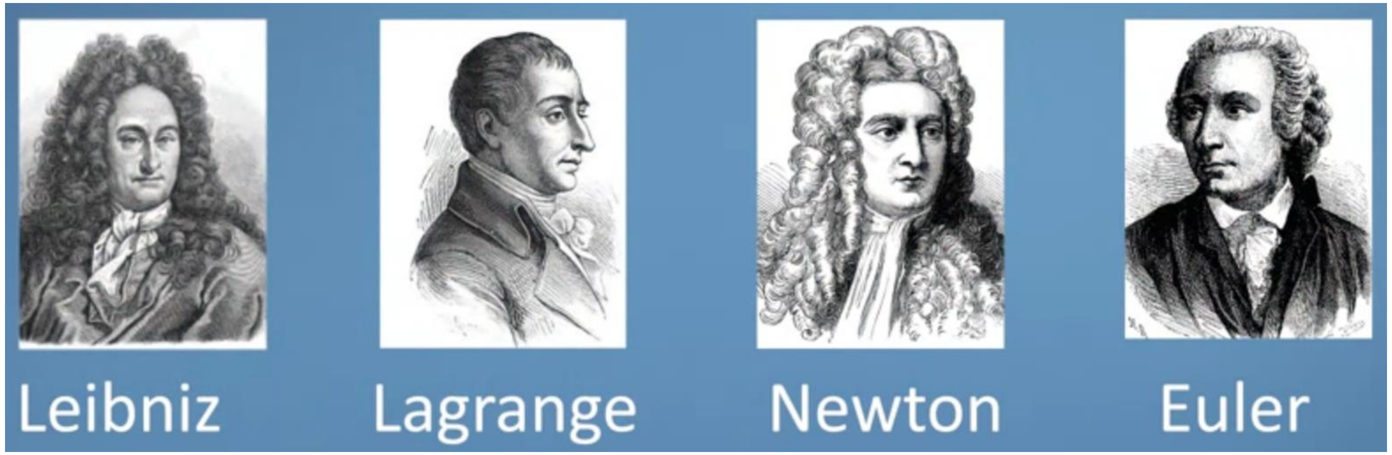
$$f(x + \Delta x) \approx f(x) + \Delta x \cdot f'(x)$$

Derivatives - Example

$$\left(\frac{1}{x}\right)' = \lim_{\Delta x \to 0} \frac{\frac{1}{x + \Delta x} - \frac{1}{x}}{\Delta x} = \lim_{\Delta x \to 0} \frac{-\Delta x}{\Delta x \cdot x(x + \Delta x)} = \lim_{\Delta x \to 0} \frac{-1}{x^2 + x\Delta x} = -\frac{1}{x^2}.$$

girafe ai

Derivatives – Other Notation



$$f'(x) = f'_x(x) = \frac{d}{dx} f(x) = \frac{\partial}{\partial x} f(x)$$

Derivatives

$$(c)' = 0 \quad (c = \text{const}), \qquad (x^\alpha)' = \alpha x^{\alpha-1},$$

$$(e^x)' = e^x, \qquad (a^x)' = a^x \ln a,$$

$$(\ln x)' = \frac{1}{x}, \qquad (\log_a x)' = \frac{1}{x \ln a},$$

$$(\sin x)' = \cos x, \qquad (\cos x)' = -\sin x,$$

$$(\operatorname{tg} x)' = \frac{1}{\cos^2 x}, \qquad (\operatorname{ctg} x)' = -\frac{1}{\sin^2 x},$$

$$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}, \qquad (\arccos x)' = -\frac{1}{\sqrt{1-x^2}},$$

$$(\operatorname{arctg} x)' = \frac{1}{1+x^2}, \qquad (\operatorname{arcctg} x)' = -\frac{1}{1+x^2}.$$

Sum Rule

$$[u(x) + v(x)]' = u'(x) + v'(x)$$

- Example:

$$(x^2 + x^3)' = 2x + 3x^2$$

Product Rule

$$[u(x) \cdot v(x)]' = u'(x) \cdot v(x) + u(x) \cdot v'(x)$$

- Example:

$$(xe^x)' = 1 \cdot e^x + x \cdot e^x$$

$$\left(\frac{1-x}{x}\right)' = (1-x) \cdot \frac{1}{x} = -\frac{1}{x} - \frac{1-x}{x^2}$$

Chain Rule

- Tells us how to compute the derivative of the composition of functions:

$$f\left(g(x)\right)' = f'\left(g(x)\right) \cdot g'(x)$$

- Other notation:

$$\frac{df}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx}$$

Chain Rule - Example

$$\left(\frac{1}{1-x}\right)' = -\frac{1}{(1-x)^2} \cdot (1-x)' = \frac{1}{(1-x)^2}$$

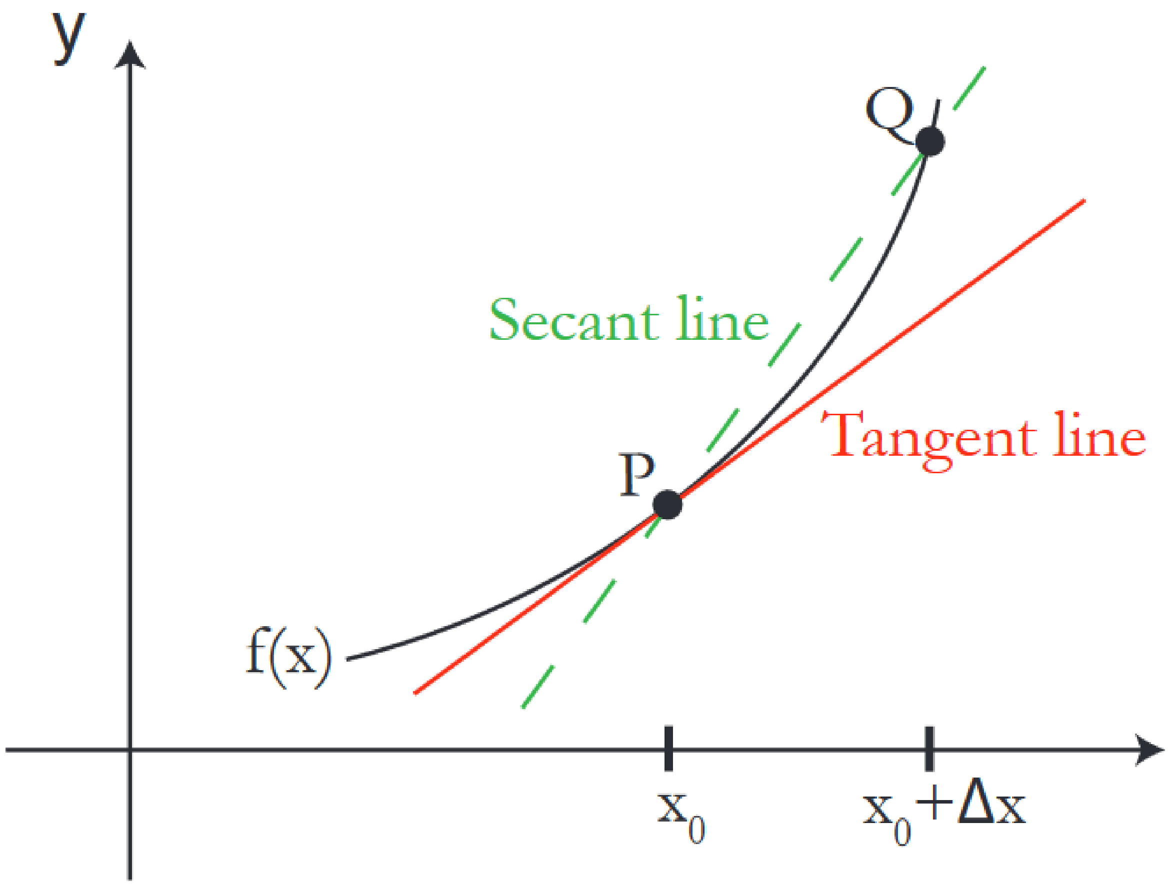
$$\left(e^{x^2}\right)' = e^{x^2} \cdot (x^2)' = e^{x^2} \cdot 2x$$

Quotient Rule

$$\frac{u(x)}{v(x)} = \left[u(x) \cdot \frac{1}{v(x)} = u'(x) \cdot \frac{1}{v(x)} - u(x) \cdot \frac{1}{\left(v(x)\right)^2} \cdot v'(x)\right] =$$

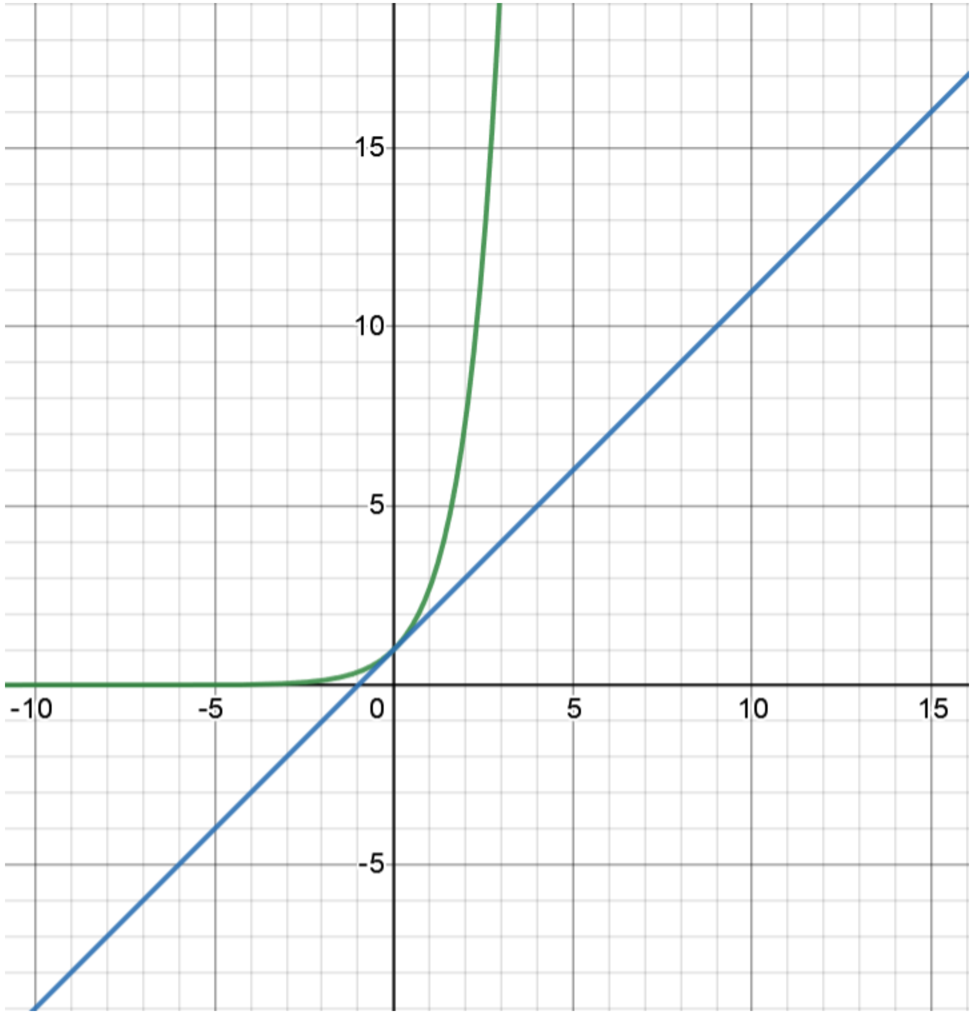
$$= \frac{u'(x)v(x) - u(x)v'(x)}{\left(v(x)\right)^2}$$

Geometric Meaning of a Derivative



Tangent Line - Example

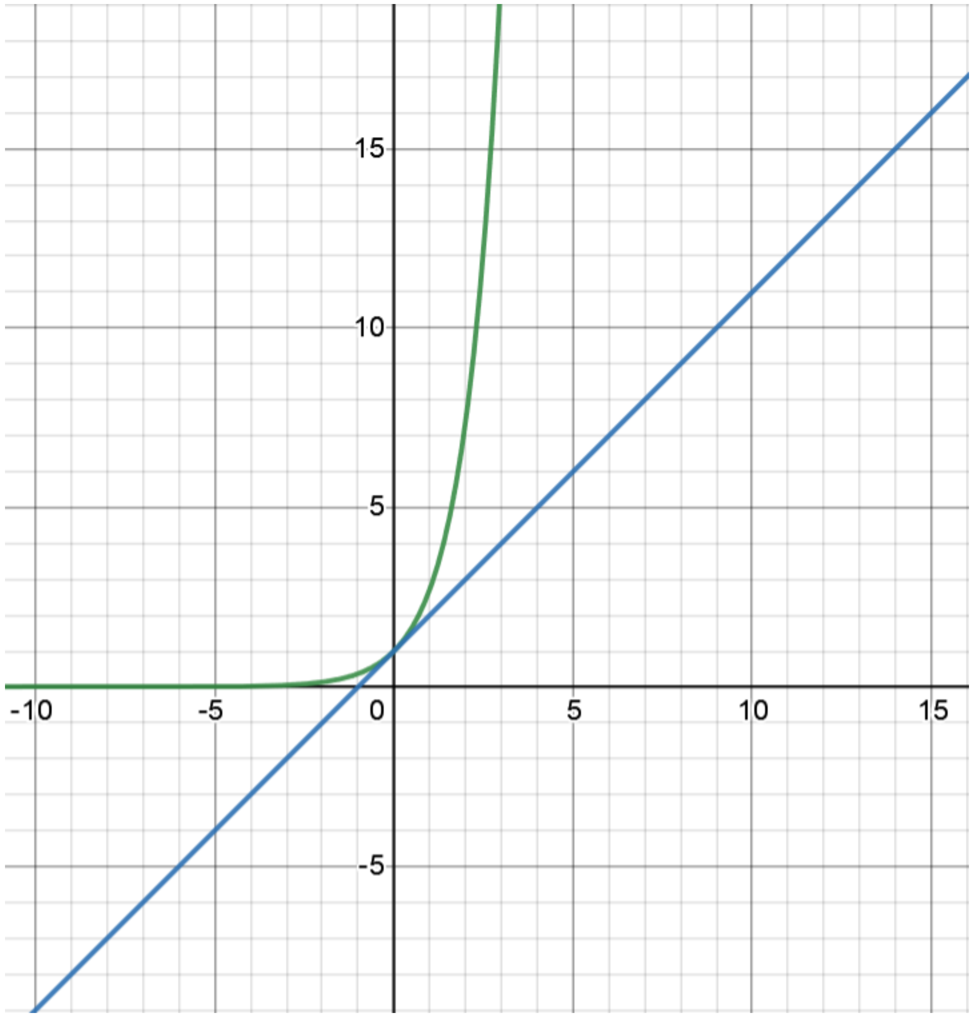
- Find a tangent line to $y = e^x$ at $x_0 = 0$.



Tangent Line - Example

- Find a tangent line to $y = e^x$ at $x_0 = 0$.
- Solution:

Tangent line: $y = kx + b$

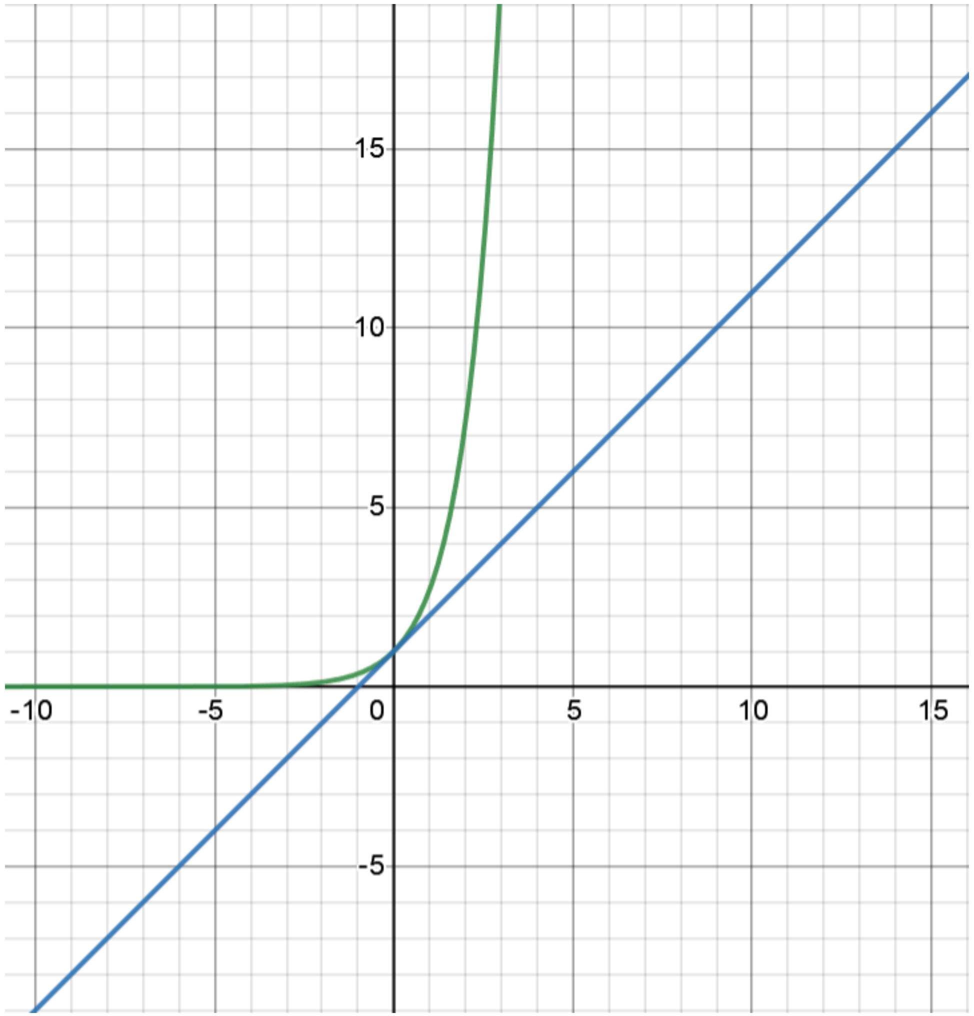


Tangent Line - Example

- Find a tangent line to $y = e^x$ at $x_0 = 0$.
- Solution:

Tangent line: $y = kx + b$

$$f'(x) = e^x, \qquad k = f'(x_0) = f'(0) = 1$$



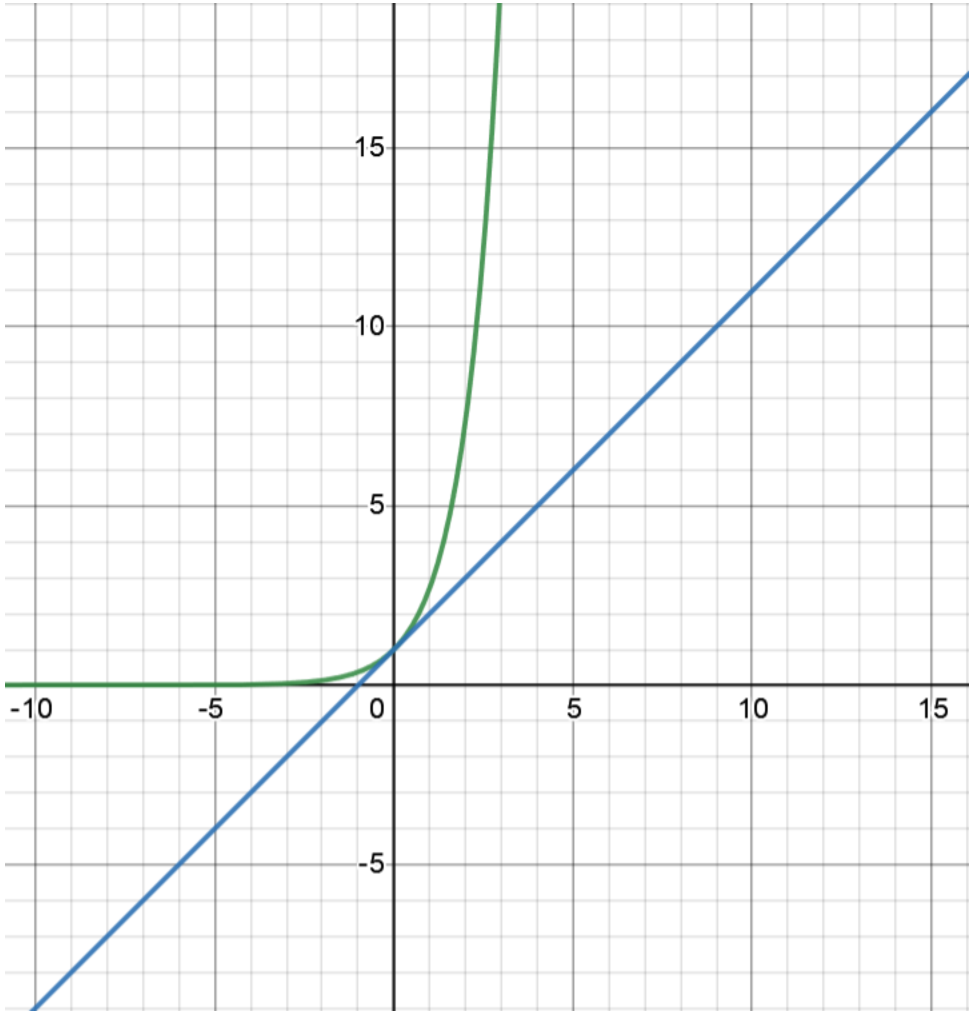
Tangent Line - Example

- Find a tangent line to $y = e^x$ at $x_0 = 0$.
- Solution:

Tangent line: $y = kx + b$

$$f'(x) = e^x, \qquad k = f'(x_0) = f'(0) = 1$$

Tangent line touches the graph at $x_0 = 0$:



Tangent Line - Example

- Find a tangent line to $y = e^x$ at $x_0 = 0$.
- Solution:

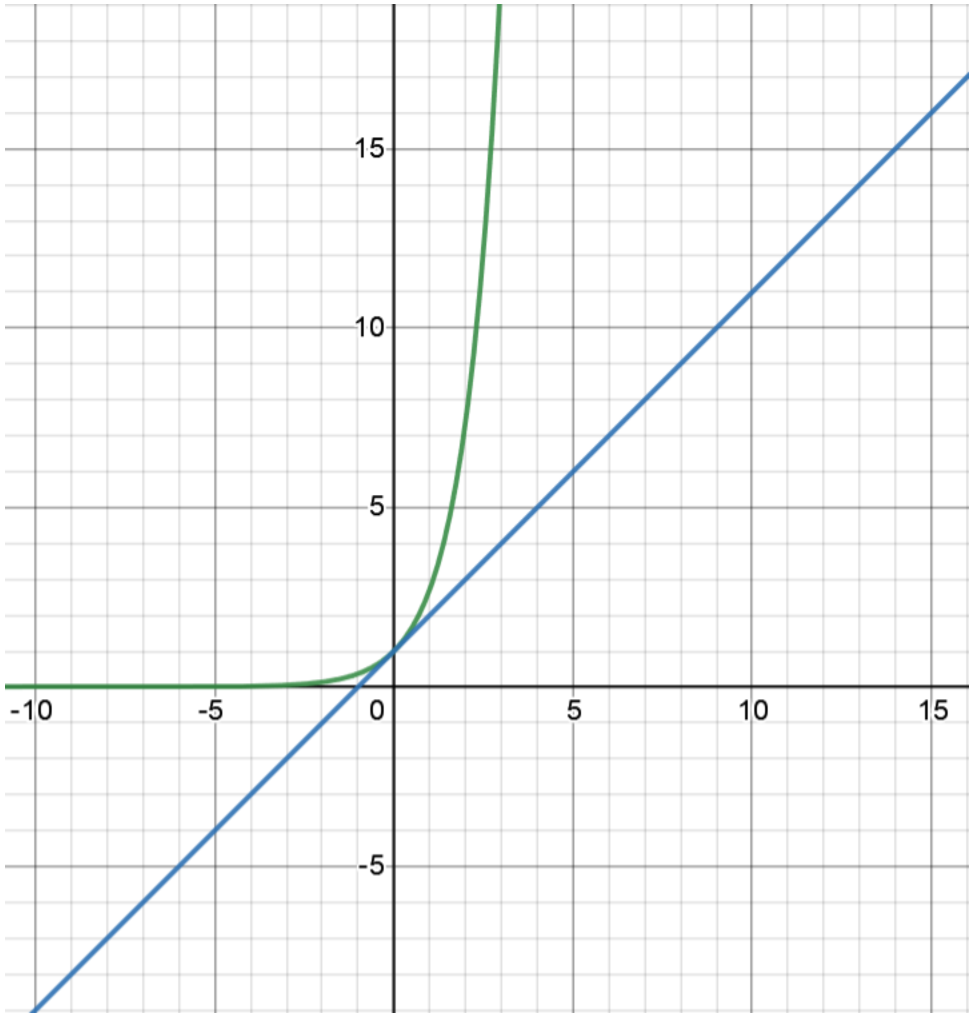
Tangent line: $y = kx + b$

$$f'(x) = e^x, \qquad k = f'(x_0) = f'(0) = 1$$

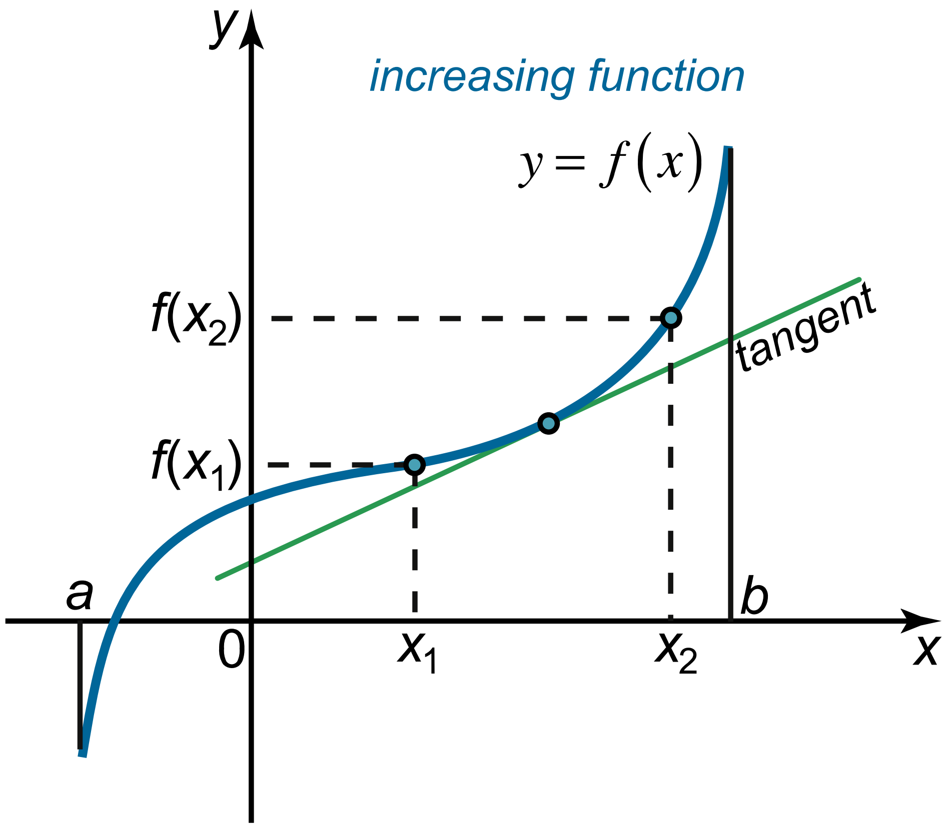
Tangent line touches the graph at $x_0 = 0$:

$$1 \cdot 0 + b = e^0 = 1, \qquad b = 1$$

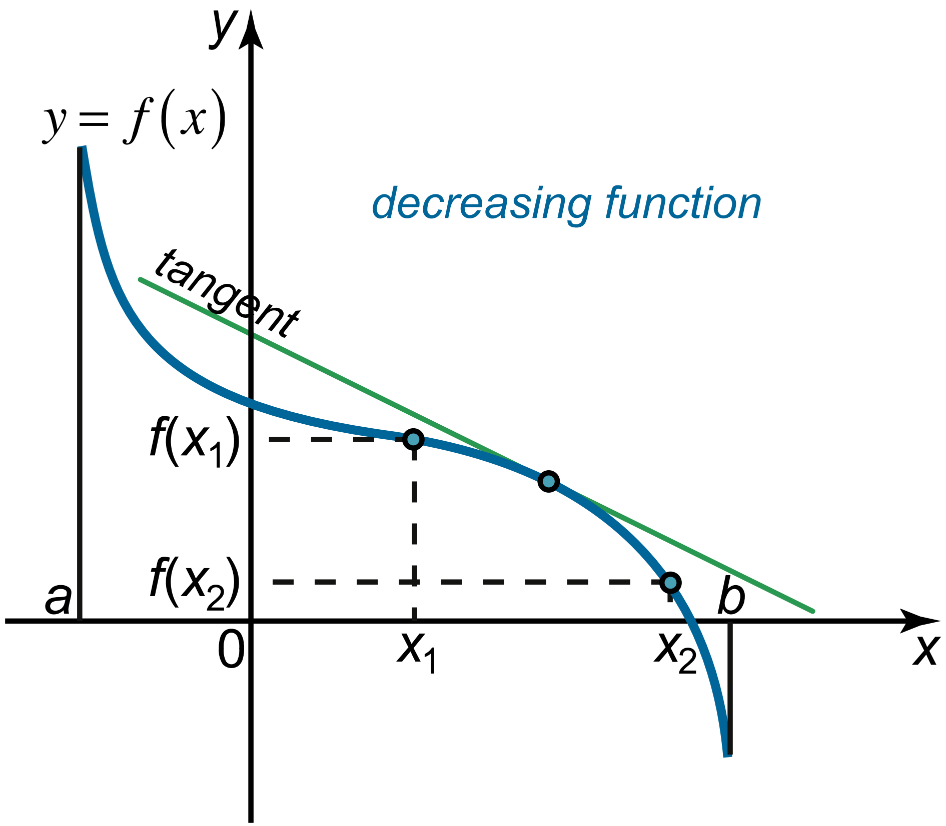
Tangent line: $y = x + 1$



Increasing / Decreasing



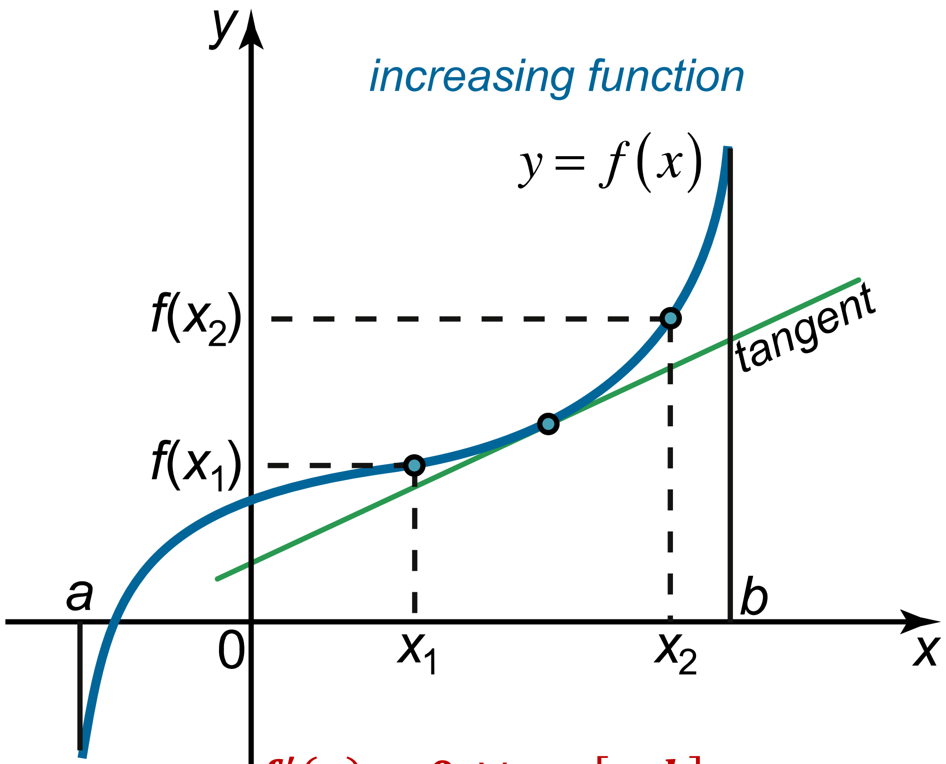
$f(x_2) > f(x_1) \quad \forall x_2 > x_1 \in [a; b]$



$f(x_2) < f(x_1) \quad \forall x_2 > x_1 \in [a; b]$

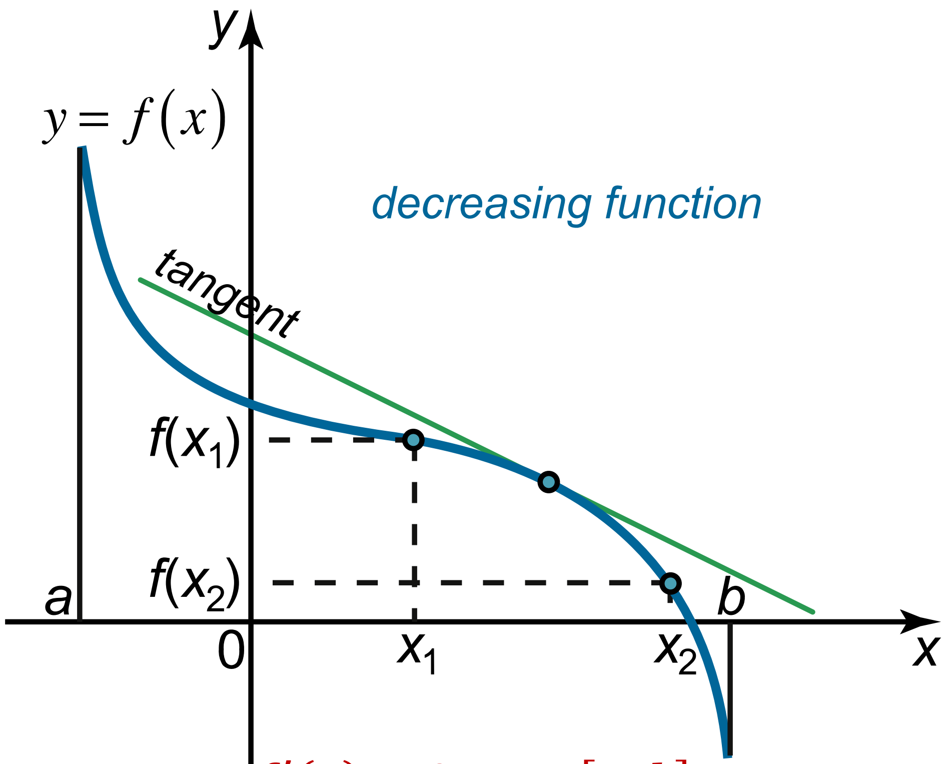
Source: https://math24.net/increasing-decreasing-functions.html

Increasing / Decreasing



$f'(x) > 0 \;\; \forall x \in [a; b]$

$f(x_2) > f(x_1) \quad \forall x_2 > x_1 \in [a; b]$



$f'(x) < 0 \;\; \forall x \in [a; b]$

$f(x_2) < f(x_1) \quad \forall x_2 > x_1 \in [a; b]$

Source: https://math24.net/increasing-decreasing-functions.html

Exploring a Function with Its Derivative

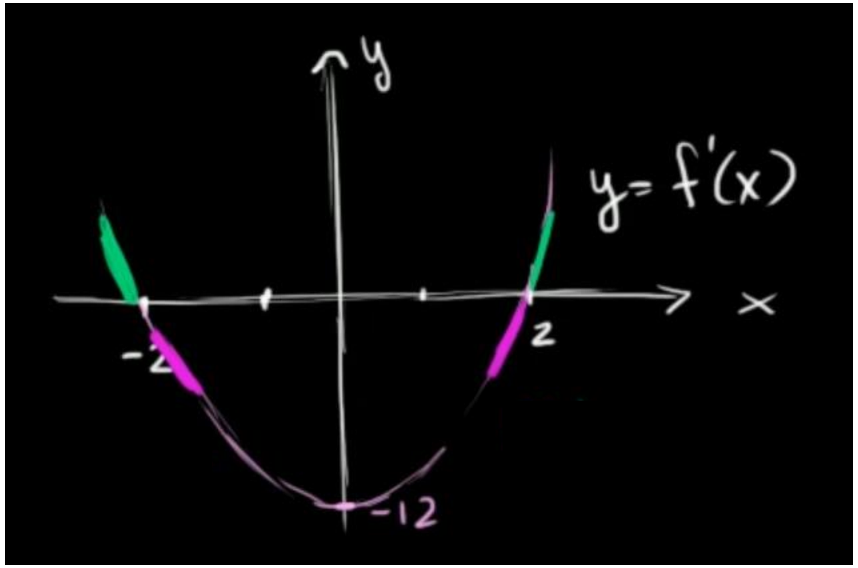
- Consider a function $f(x) = x^3 - 12x + 2$.

Source: https://www.khanacademy.org/math/ap-calculus-ab/ab-diff-analytical-applications-new/ab-5-8/v/analyzing-a-function-with-its-derivative

Exploring a Function with Its Derivative

- Consider a function $f(x) = x^3 - 12x + 2$.

Derivative: $f'(x) = 3x^2 - 12 = 0$



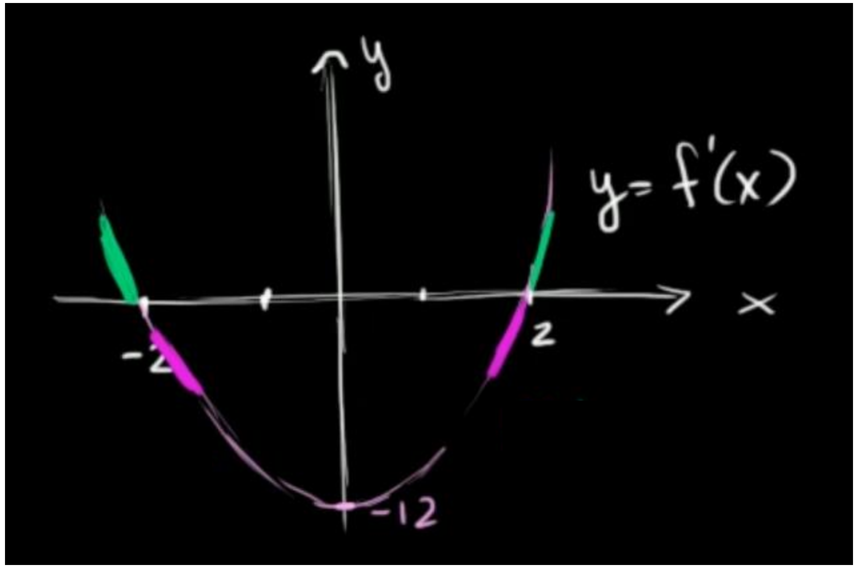
Source: https://www.khanacademy.org/math/ap-calculus-ab/ab-diff-analytical-applications-new/ab-5-8/v/analyzing-a-function-with-its-derivative

Exploring a Function with Its Derivative

- Consider a function $f(x) = x^3 - 12x + 2$.

$$\text{Derivative: } f'(x) = 3x^2 - 12 = 0$$

$$f'(x) \Leftrightarrow x = \pm 2$$



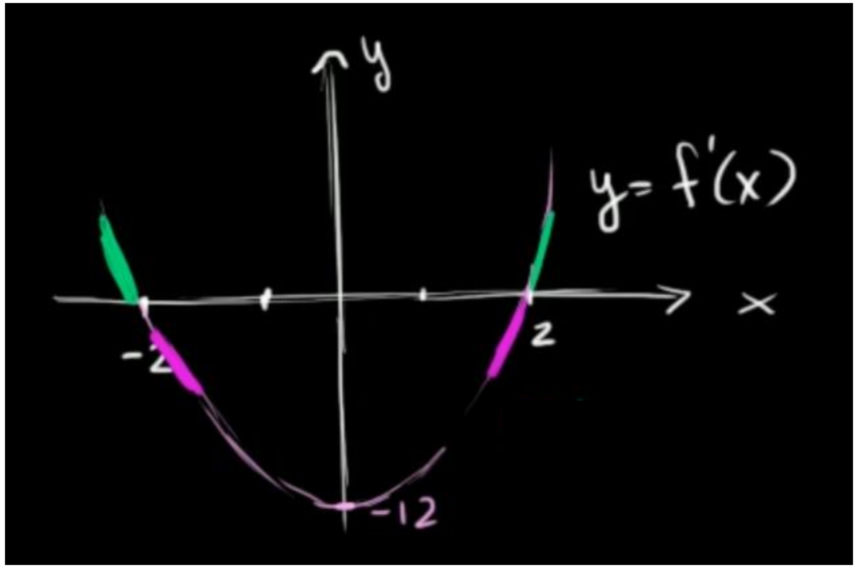
Source: https://www.khanacademy.org/math/ap-calculus-ab/ab-diff-analytical-applications-new/ab-5-8/v/analyzing-a-function-with-its-derivative

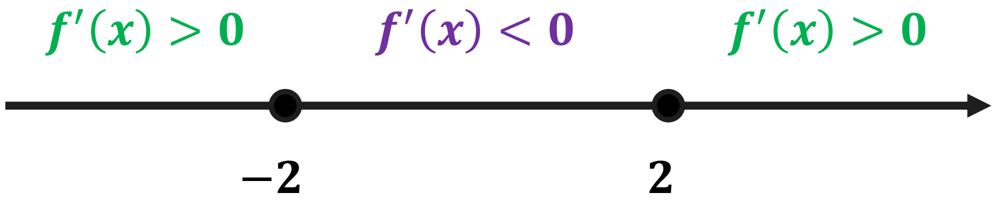
Exploring a Function with Its Derivative

- Consider a function $f(x) = x^3 - 12x + 2$.

Derivative: $f'(x) = 3x^2 - 12 = 0$

$f'(x) \Leftrightarrow x = \pm 2$





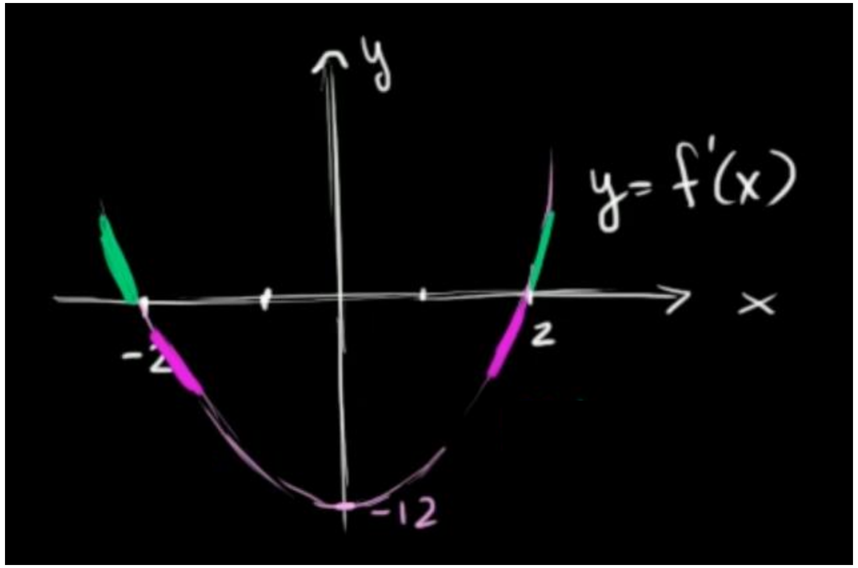
Source: https://www.khanacademy.org/math/ap-calculus-ab/ab-diff-analytical-applications-new/ab-5-8/v/analyzing-a-function-with-its-derivative

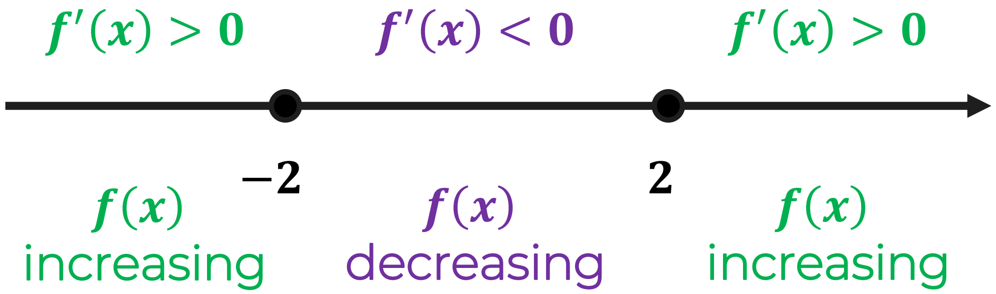
Exploring a Function with Its Derivative

- Consider a function $f(x) = x^3 - 12x + 2$.

Derivative: $f'(x) = 3x^2 - 12 = 0$

$$f'(x) \Leftrightarrow x = \pm 2$$





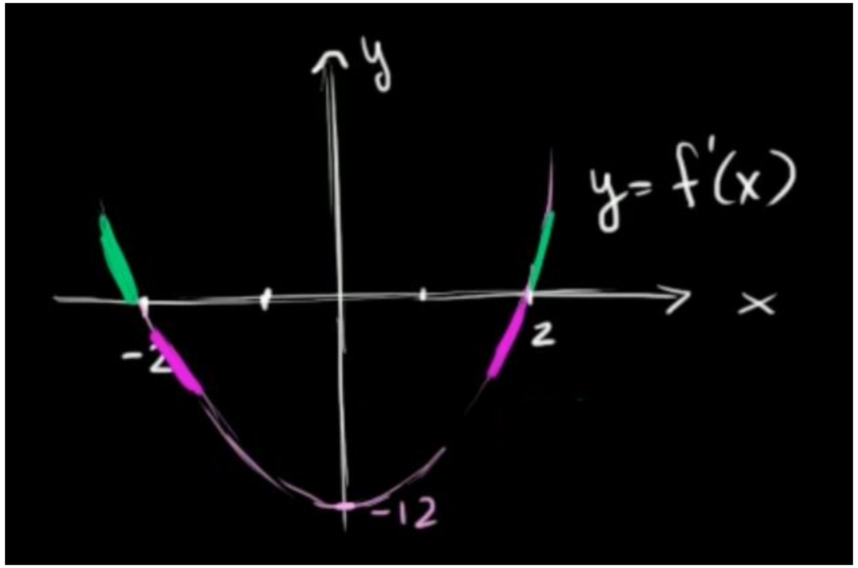
Source: https://www.khanacademy.org/math/ap-calculus-ab/ab-diff-analytical-applications-new/ab-5-8/v/analyzing-a-function-with-its-derivative

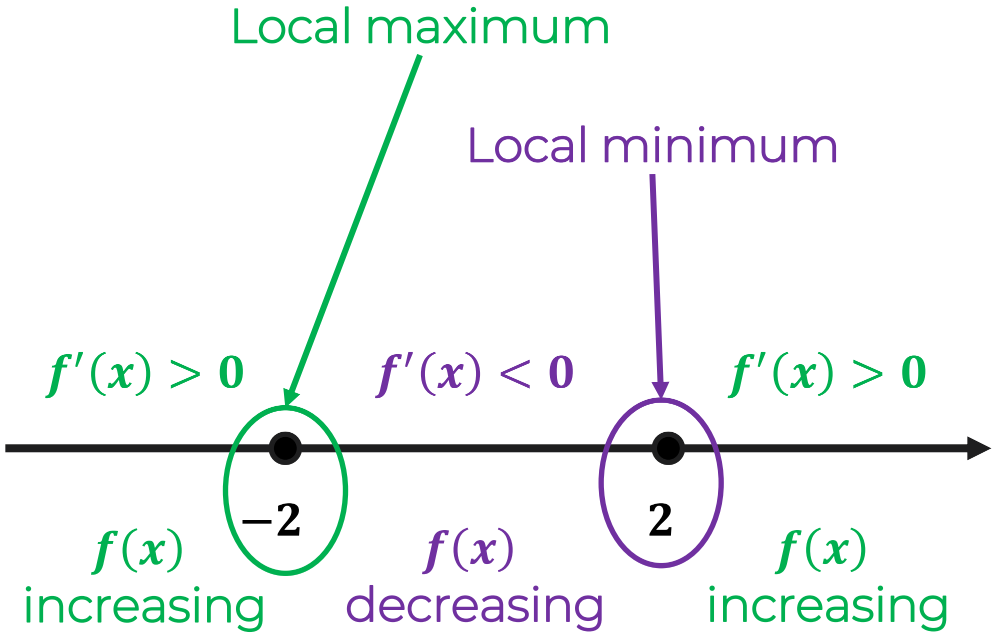
Exploring a Function with Its Derivative

- Consider a function $f(x) = x^3 - 12x + 2$.

Derivative: $f'(x) = 3x^2 - 12 = 0$

$$f'(x) \Leftrightarrow x = \pm 2$$



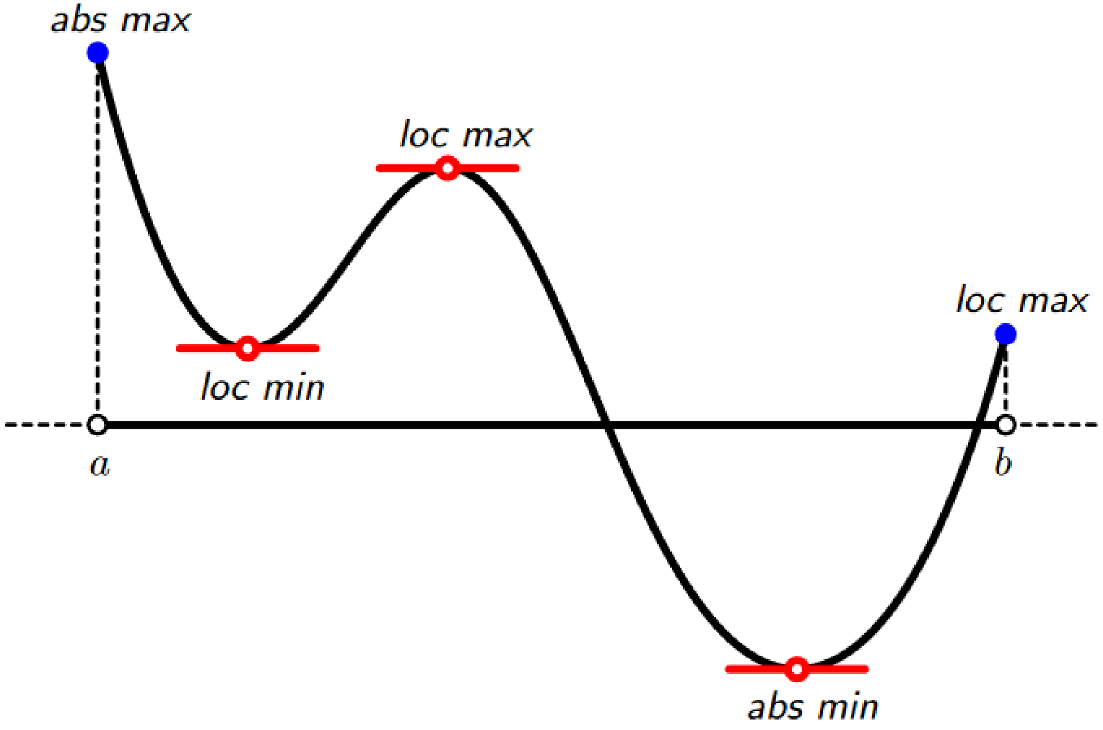


Source: https://www.khanacademy.org/math/ap-calculus-ab/ab-diff-analytical-applications-new/ab-5-8/v/analyzing-a-function-with-its-derivative

Extrema

Extrema of a Function

- $f(x)$ reaches its local minima *(maxima)* at x_0 if $f(x_0)$ is the smallest *(highest)* value of $f(x)$ around x_0 .



- $f(x)$ reaches its global minima *(maxima)* at x_0 if $f(x_0)$ is the smallest *(highest)* value of $f(x)$ on the interval of interest.

Critical Point

- A stationary point of $f(x)$ is a point x_0 such that $f'(x_0) = 0$

Source:http://www.conejousd.org/Portals/49/Departments/Math/Sansing/Math%20AnalyCP/Notes%203-6%20Critical%20Points%20and%20Extrema.pdf

Critical Point

- A **stationary point** of $f(x)$ is a point x_0 such that $f'(x_0) = 0$

- A **critical point** of $f(x)$ is a point x_0 such that
    - $f(x_0) = 0$ (x_0 is a stationary point) or
    - $f'(x_0)$ doesn't exit.

Source:http://www.conejousd.org/Portals/49/Departments/Math/Sansing/Math%20AnalyCP/Notes%203-6%20Critical%20Points%20and%20Extrema.pdf

Critical Point

- A **stationary point** of $f(x)$ is a point x_0 such that $f'(x_0) = 0$

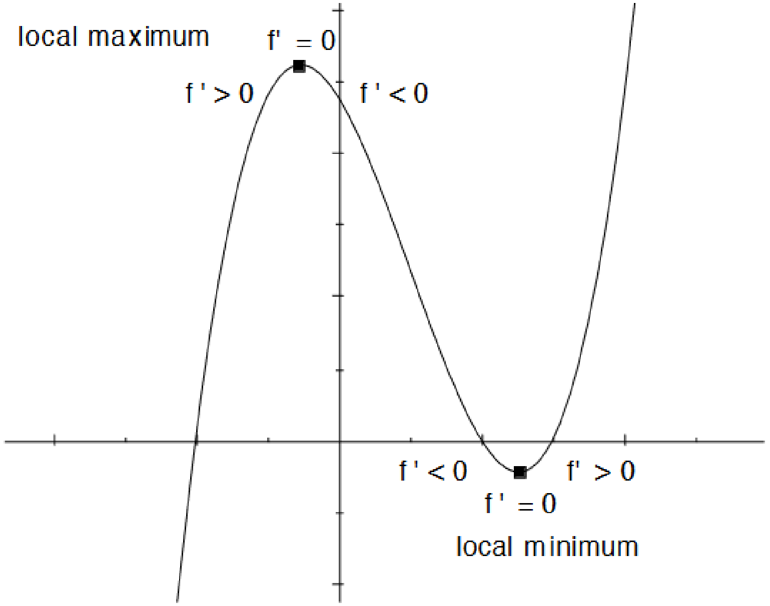
- A **critical point** of $f(x)$ is a point x_0 such that
    - $f(x_0) = 0$ (x_0 is a stationary point) or
    - $f'(x_0)$ doesn't exit.

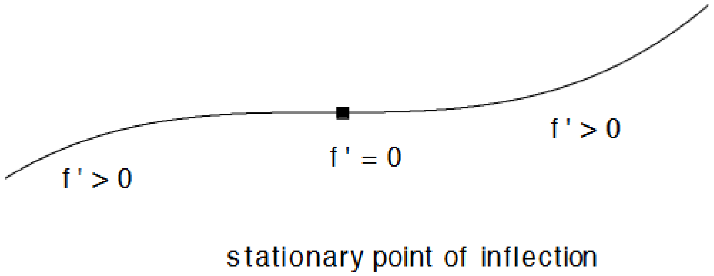
- Critical points: those points on a graph at which a line drawn tangent to the curve is horizontal or vertical.

Source:http://www.conejousd.org/Portals/49/Departments/Math/Sansing/Math%20AnalyCP/Notes%203-6%20Critical%20Points%20and%20Extrema.pdf

First Derivative Test

- Let x_0 be a critical point of $f(x)$.
- If $f'(x) < 0$ for $x < x_0$ and $f'(x) > 0$ for $x > x_0$ then x_0 is a point of a local minimum.



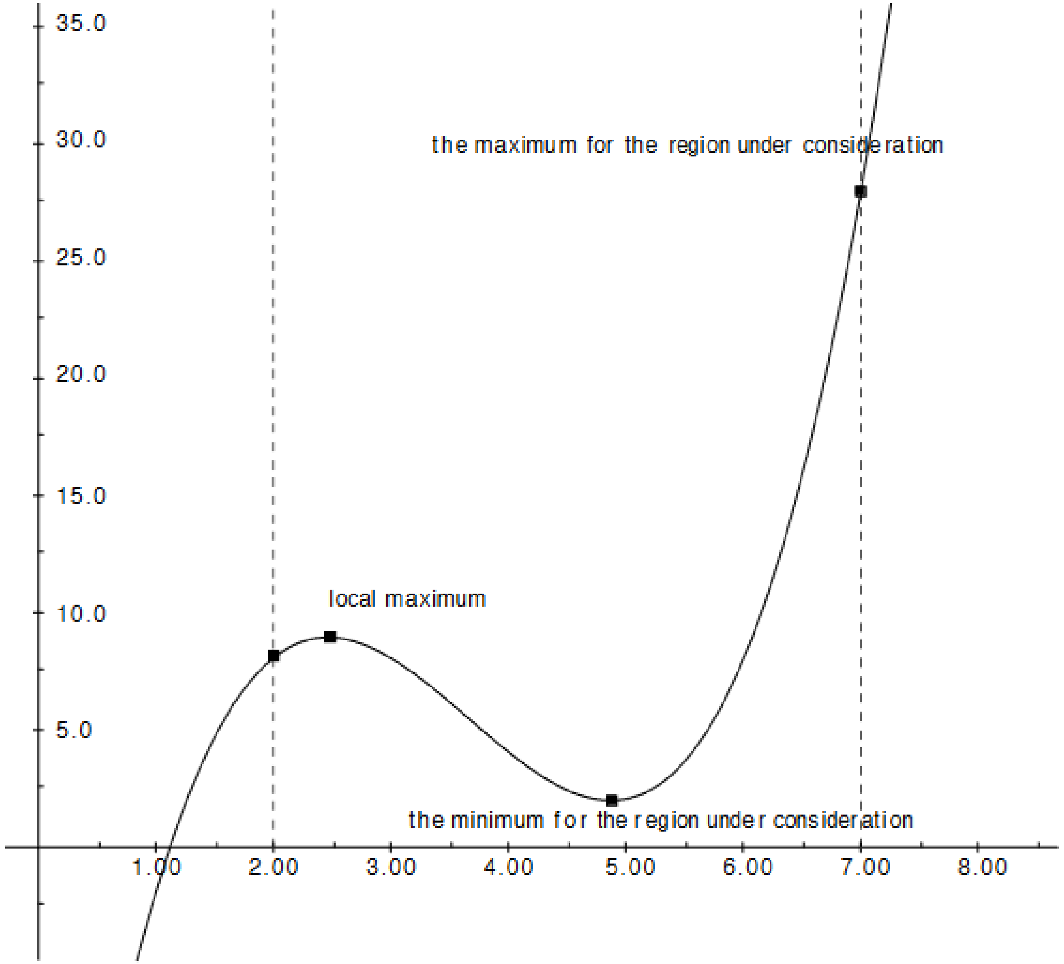


- If $f'(x) > 0$ for $x < x_0$ and $f'(x) < 0$ for $x > x_0$ then x_0 is a point of a local maximum.

girafe ai

Source: https://www.sydney.edu.au/content/dam/students/documents/mathematics-learning-centre/introduction-to-differential-calculus.pdf

Don’t Forget the Endpoints!



Source: https://www.sydney.edu.au/content/dam/students/documents/mathematics-learning-centre/introduction-to-differential-calculus.pdf

Algorithm for Finding Global Extrema

- Suppose you need to find global maxima (minima) of $f(x)$ on $[a; b]$.

- Here is s recipe:

    1. Find all critical points of $f(x)$ on $[a; b]$;
    2. Determine which of them are the local maxima (minima);
    3. Compute $f(x)$ at the endpoints: $f(a)$ and $f(b)$.
    4. Pick the point from (2) – (3) corresponding to the largest (smallest) function value.

Finding Extrema - Example

- Find the global minimum of $f(x) = x^2 e^x$ on $[-4, 1]$.

Finding Extrema - Example

- Find the global minimum of $f(x) = x^2 e^x$ on $[-4, 1]$.

Derivative: $f'(x) = 2xe^x + x^2 e^x = xe^x(x+2)$

Finding Extrema - Example

- Find the global minimum of $f(x) = x^2e^x$ on $[-4, 1]$.

Derivative: $f'(x) = 2xe^x + x^2e^x = xe^x(x+2)$

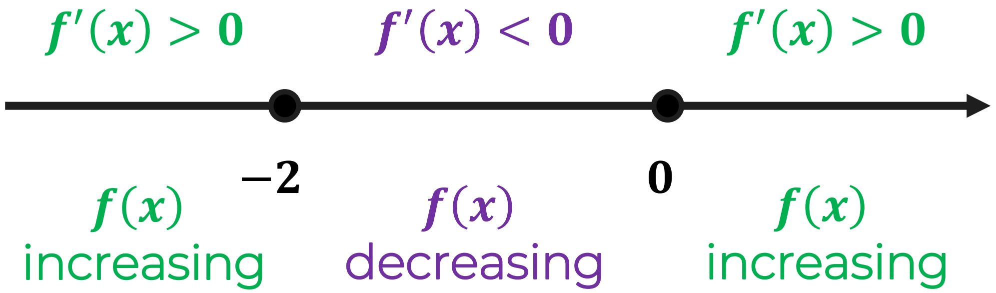
Stationary points: $f'(x) = 0 \Leftrightarrow x = 0,\ x = -2$

Finding Extrema - Example

- Find the global minimum of $f(x) = x^2 e^x$ on $[-4, 1]$.

Derivative: $f'(x) = 2xe^x + x^2 e^x = xe^x(x+2)$

Stationary points: $f'(x) = 0 \Leftrightarrow x = 0,\ x = -2$



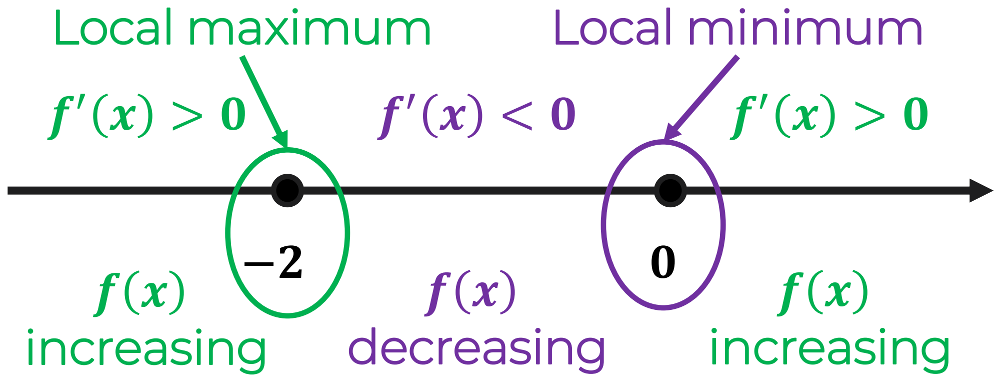
Finding Extrema - Example

- Find the global minimum of $f(x) = x^2e^x$ on $[-4, 1]$.

Derivative: $f'(x) = 2xe^x + x^2e^x = xe^x(x+2)$

Stationary points: $f'(x) = 0 \Leftrightarrow x = 0,\ x = -2$

$$f(-2) = 4e^{-2} \approx 0.54, \qquad f(0) = 0$$



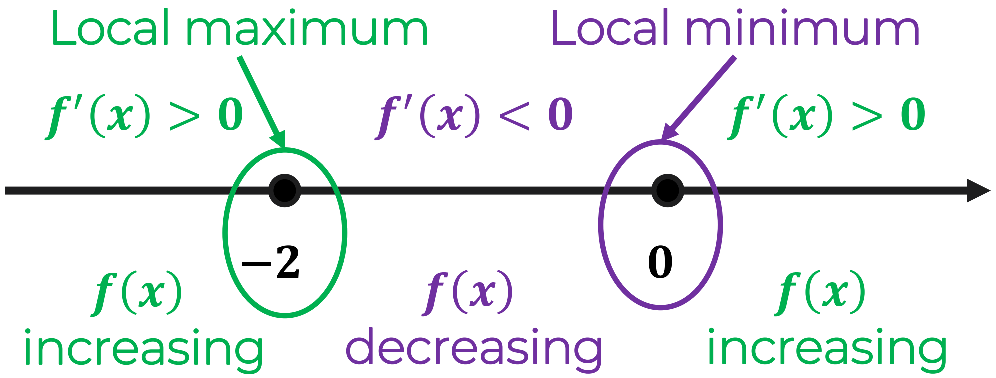
Finding Extrema - Example

- Find the global minimum of $f(x) = x^2 e^x$ on $[-4, 1]$.

Derivative: $f'(x) = 2xe^x + x^2e^x = xe^x(x+2)$

Stationary points: $f'(x) = 0 \Leftrightarrow x = 0,\ x = -2$

$$f(-2) = 4e^{-2} \approx 0.54, \qquad f(0) = 0$$



Endpoints: $f(-4) = 16e^{-4} \approx 0.29, \qquad f(1) = e \approx 2.7$

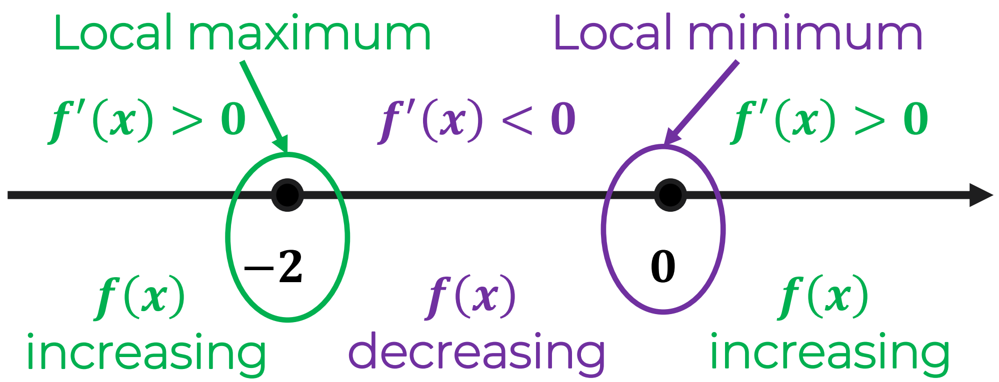
Finding Extrema - Example

- Find the global minimum of $f(x) = x^2 e^x$ on $[-4, 1]$.

Derivative: $f'(x) = 2xe^x + x^2 e^x = xe^x(x+2)$

Stationary points: $f'(x) = 0 \Leftrightarrow x = 0,\ x = -2$

$f(-2) = 4e^{-2} \approx 0.54,\qquad \boldsymbol{f(0) = 0}$



Endpoints: $f(-4) = 16e^{-4} \approx 0.29,\qquad \boldsymbol{f(1) = e \approx 2.7}$

Higher Derivatives

Higher Derivatives

- Derivatives of the derivatives:

$$f''(x) = \big(f'(x)\big)', \qquad f'''(x) = \big(f''(x)\big)', \qquad \ldots$$

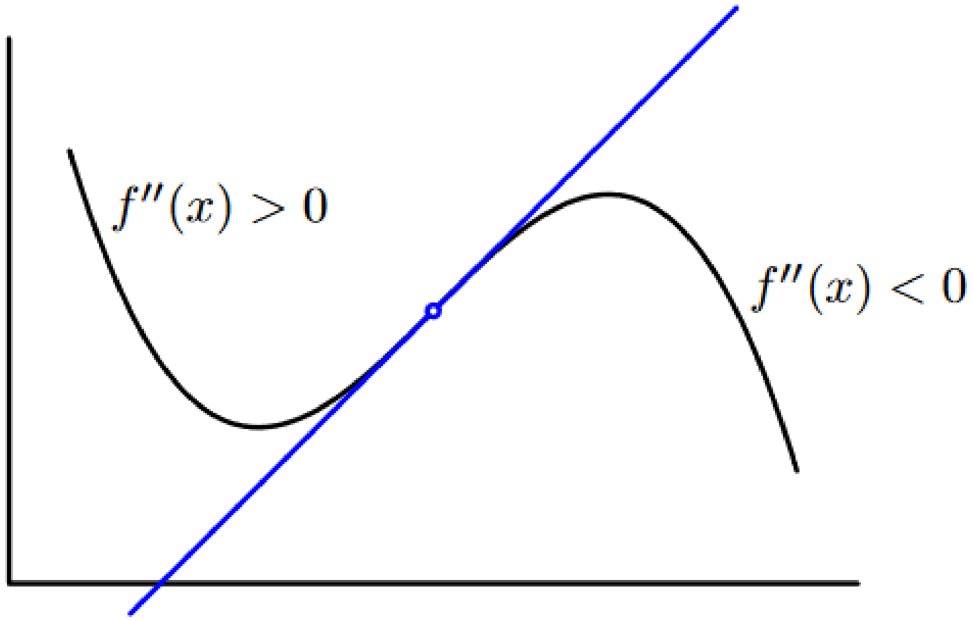
- Pretty straightforward!

- Example:

$$(3x^3 + 2x^2 + x)'' = (9x^2 + 4x + 1)' = 18x + 4$$

Second Derivative and Convexity

- A function is convex on some interval $[a; b]$ if and only if $\boldsymbol{f''(x) > 0}$ for all $x \in [a; b]$.



Source: https://people.math.wisc.edu/~angenent/Free-Lecture-Notes/free221.pdf

Second Derivative Test

- Consider a differentiable function $f(x)$.
- Let x_0 be its stationary point: $f'(x_0) = 0$.
- If $\boldsymbol{f''(x_0) < 0}$ then $f(x)$ has a local maximum at x_0, and
  if $\boldsymbol{f''(x_0) > 0}$ then $f(x)$ has a local minimum at x_0.

girafe ai

Second Derivative Test

- Consider a differentiable function $f(x)$.

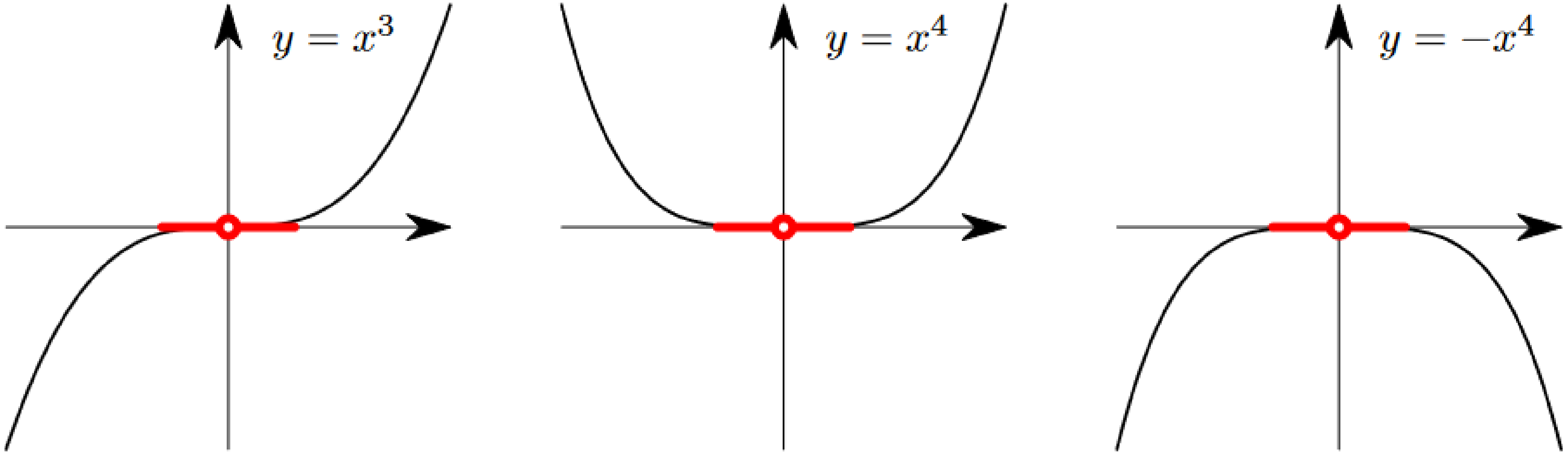
- Let x_0 be its stationary point: $f'(x_0) = 0$.

- If $\boldsymbol{f''(x_0) < 0}$ then $f(x)$ has a local maximum at x_0, and
if $\boldsymbol{f''(x_0) > 0}$ then $f(x)$ has a local minimum at x_0.

- We don't know what happens when $f''(x_0) = 0$: need to check manually.

Second Derivative Test

- We don't know what happens when $f''(x_0) = 0$: need to check manually.



Source: https://people.math.wisc.edu/~angenent/Free-Lecture-Notes/free221.pdf

Probability Theory

Probability Theory - Introduction

- Statistics and probability theory constitute a branch of mathematics for dealing with uncertainty. The probability theory provides a basis for the science of statistical inference from data

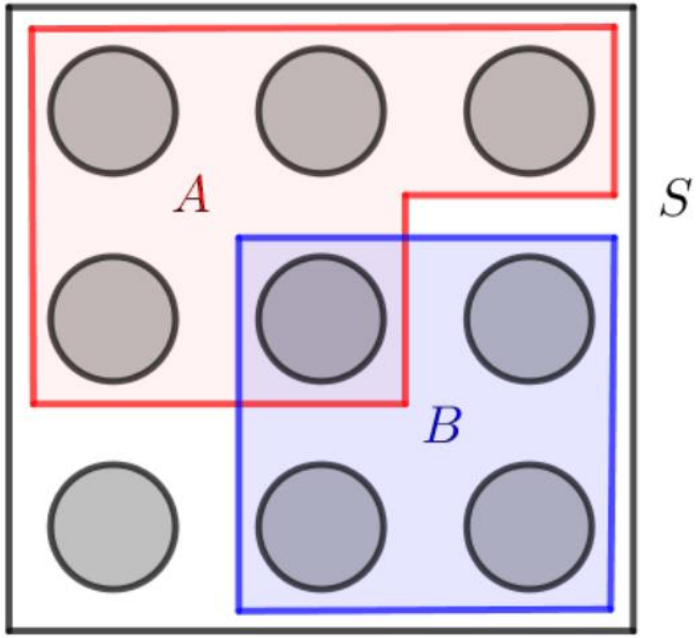
- Sample: (of size N) obtained from a mother population assumed to be represented by a probability

- Descriptive statistics: description of the sample

- Inferential statistics: making a decision or an inference from a sample of our problem

Sample Space

- Experiment: any process or procedure for which more than one outcome is possible
- Sample space: set of all the possible experimental outcomes
- Elementary outcome – the simplest of all possible outcomes – denoted by grey circle
- Possible experimental outcome: A, B

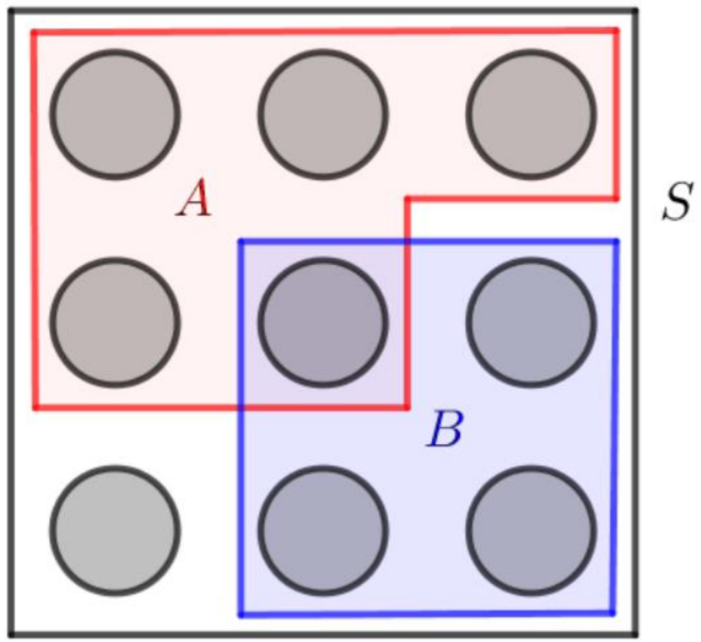


girafe ai

Sample Space

- The **union** $A \cup B$ is the event that occurs **iff** (if and only if) **at least one of A, B occurs.**

- The intersection $A \cap B$ is the event that occurs **iff** **both A and B occur.**



Sample Space

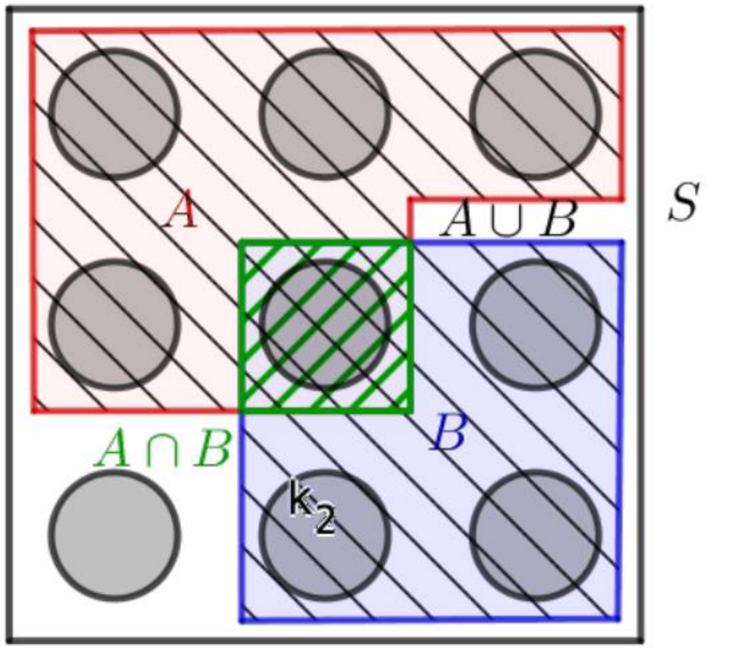
- The **union** $A \cup B$ is the event that occurs **iff** (if and only if) **at least one of A, B occurs.**

- The intersection $A \cap B$ is the event that occurs **iff both A and B occur.**

- The complement A^c occurs iff A does **not occur**

- De Morgan's laws:

$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$



Sample Space - Example

- A coin if flipped 5 times: heads is H and tails is T
- Possible outcome: HHTHT
- Sample space: set of all possible strings of length 5 with H and T symbols

Sample Space - Example

- Let A_j=event that the j-th flip is Heads

  $A_j = \{(H, s_2, \dots, s_5): s_j \in \{H, T\} for\ 2 < j < 5\}$

- Let B=event that **at least** one flip was Heads → $B = \bigcup_{j=1}^{5} A_j$

- Let C = event that all one flips were Heads → $\mathrm{C} = \bigcap_{j=1}^{5} A_j$

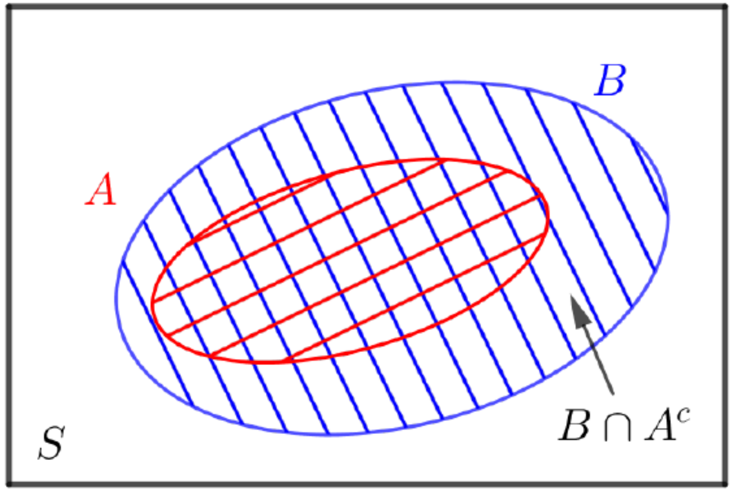
- Let D = event that there were at least two consecutive Heads →

$$\mathrm{D} = \bigcup_{j=1}^{4} (A_j \cap A_{j+1})$$

girafe ai

Sample Space

- Some other relationships between events:
- A implies B = $A \subset B$ (A is a subset of B)
- A and B are mutually exclusive = $A \cap B = \emptyset$



girafe ai

Probability: Naïve Definition

- Let A be an event for an experiment with a finite sample space S. Naive probability of A is

$$P_{naive}(A) = \frac{|A|}{|S|} = \frac{number\ of\ outcomes\ favourable\ to\ A}{total\ number\ of\ outcomes\ in\ S}$$

- In general the following is correct:

$$P_{naive}(A^c) = \frac{|A^c|}{|S|} = \frac{|S| - |A|}{|S|} = 1 - \frac{|A|}{|S|} = 1 - P_{naive}(A)$$

Probability: Non-naïve Definition

- A **probability space** consists of a **sample space** S and a **probability function** P which takes an event $A \subset S$ as input and returns $\mathrm{P}(A)$, a real number between 0 and 1, as output.

- P satisfies these axioms:

$$P(\emptyset) = 0, P(S) = 1$$

if $A_1, A_2, \ldots$ are disjoint events (=mutually exclusive, $A_i \cap B_j = \emptyset, i \neq j$), then

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$$

girafe ai

Probability: Non-naïve Definition

- A **probability space** consists of a **sample space** S and a **probability function** P which takes an event $A \subset S$ as input and returns $\mathrm{P}(A)$, a real number between 0 and 1, as output.

- **Frequentist** view: probability = long-run frequency over a large number of repetitions of an experiment

- **Bayesian** view: probability = degree of belief about the event in question

girafe ai

Probability: Non-naïve Definition

- The following properties can be derived from the axioms:

$$P(A^c) = 1 - P(A)$$

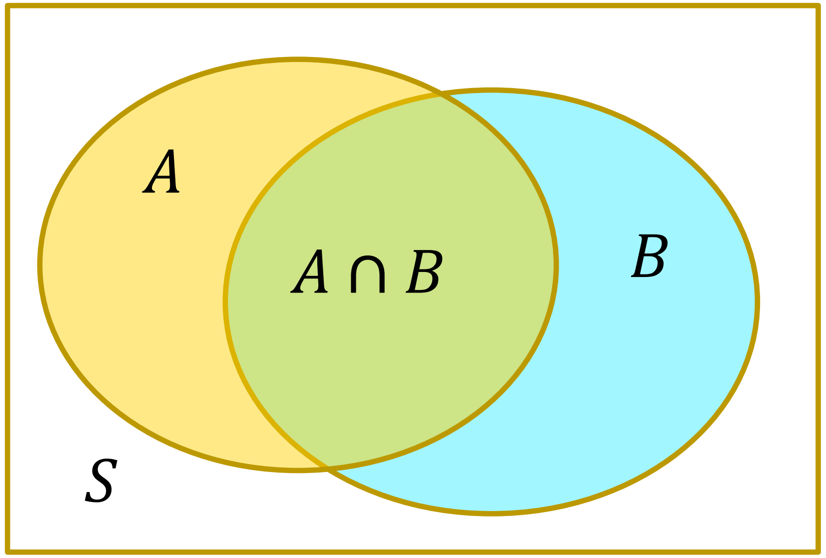
$$A \subset B \;\rightarrow\; P(A) \leq P(B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Conditional Probability

- Conditional Probability: of an event A conditional on event B:

$$\mathrm{P}(A|B) = \frac{P(A \cap B)}{P(B)}, \text{for } \mathrm{P(B)} > 0$$



Conditional Probability

- General Multiplication Law:

$$\mathrm{P}(A|B) = \frac{P(A \cap B)}{P(B)} \leftrightarrow P(A \cap B) = \mathrm{P}(A|B)P(B)$$

$$\mathrm{P}(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} \leftrightarrow P(A \cap B \cap C) = P(B \cap C)\mathrm{P}(A|B \cap C)$$

....

$$\mathrm{P}(A_1 A_2 \dots A_n) = P(A_1)P(A_2|A_1) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$$

Conditional Probability

- Two events A and B are independent if:

$$\mathrm{P}(A|B) = P(A)$$

or

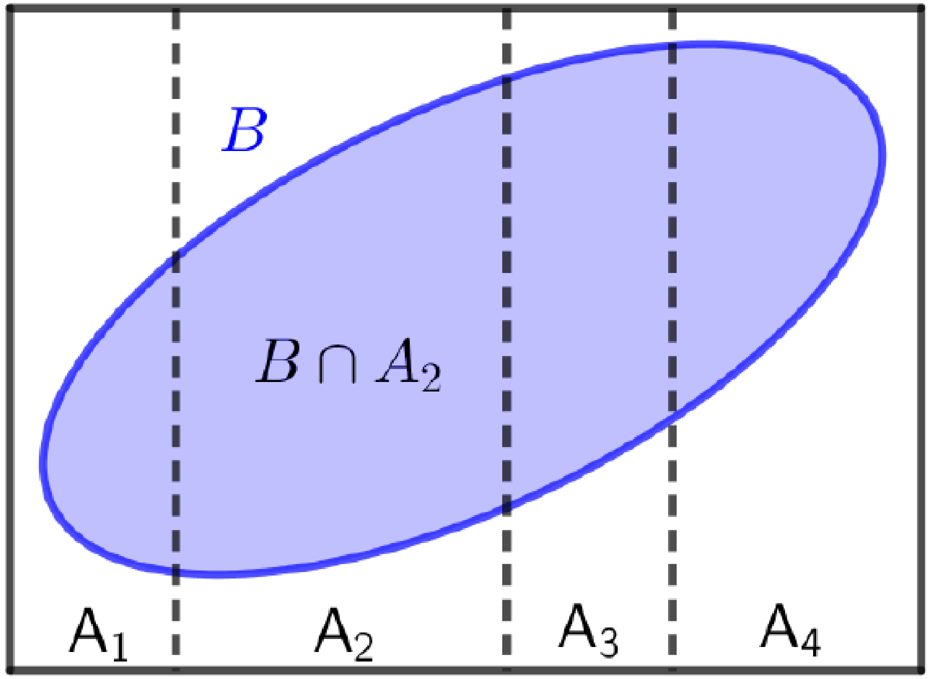
$$\mathrm{P}(B|A) = P(B)$$

or

$$P(A \cap B) = P(A)P(B)$$

Law of Total Probability

- Given a partition $\{A_1, A_2, \ldots, A_n\}$ of the sample space S, the probability of an event B can be expressed as:



$$\mathrm{P}(B) = \sum_{j=1}^{\infty} P(A_j) P(B|A_j)$$

Bayes' Theorem

- Given a partition $\{A_1, A_2, \ldots, A_n\}$ of the sample space S
- posterior probabilities of the event A_j conditional on event B
- can be obtained from $\mathrm{P}(A_j)$ and $\mathrm{P}(A_j|B)$:

$$\mathrm{P}(A_j|B) = \frac{P(A_j)P(B|A_j)}{P(B)} = \frac{P(A_j)P(B|A_j)}{\sum_{j=1}^{\infty} P(A_j)P(B|A_j)}$$

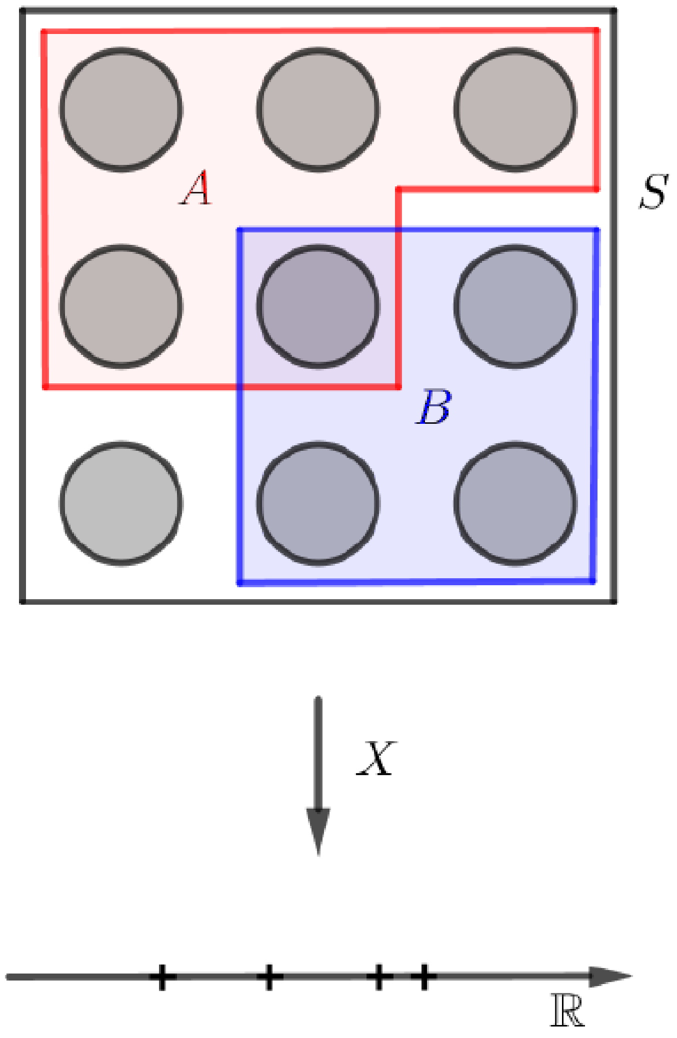
girafe ai

Random Variables

Discrete Case

- Random variable (r.v.) X : mapping from sample space S to real line

- In other words: numerical value $\mathrm{X(w)}$ mapped for each outcome w of particular experiment



Discrete Case - Example

Consider double coin tosses → Sample space: $\mathrm{S} = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$

Possible random variables:

- $\mathrm{X} = \{\ \#\ \text{of Heads}\ \}$ → $\mathrm{X(HH)} = 2, \mathrm{X(HT)} = \mathrm{X(TH)} = 1, \mathrm{X(TT)} = 0$

- $\mathrm{Y} = \{\ \#\ \text{of Tails}\}$ → $\mathrm{Y} = 2 - \mathrm{X}$

- $\mathrm{I} = \begin{cases} 1, if\ 1st\ toss = Heads \\ 0, otherwise \end{cases}$ - indicator random variable

Probability Mass Function

- Probability Mass Function is a set of probability values $\mathbf{p_i}$, assigned to each value x_i taken by the discrete random variable $\mathbf{X}$:

$$\mathrm{P(X = x_i) = p_i} = P_X(x_i)$$

- $0 \leq \mathrm{p_i} \leq 1,\ \sum_{i=1}^{\infty} p_i = 1$

- $\{\mathrm{X = x_i}\}$ actually denotes an **event**: $\{A \in S\colon \mathrm{X(A) = x_i}\}$

girafe ai

Probability Mass Function – Example

Consider double coin tosses:

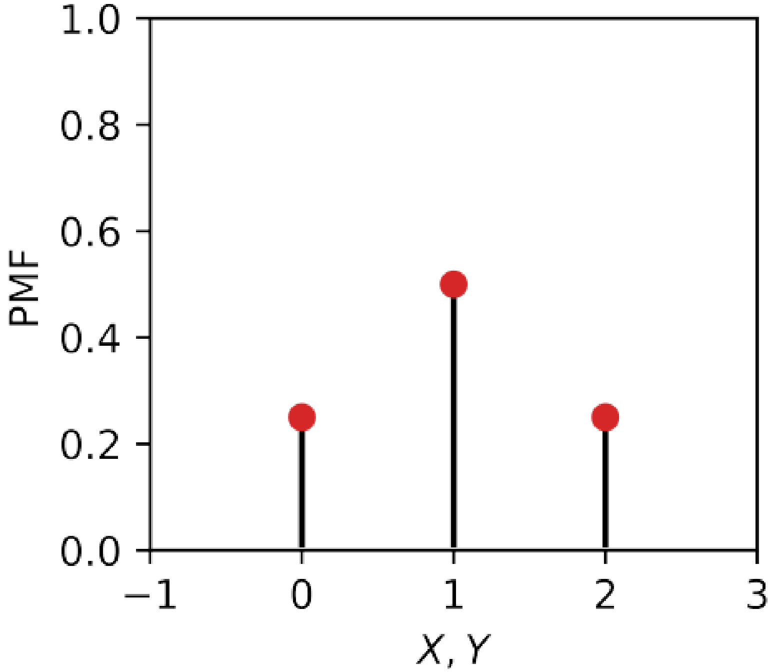
- $X = \{\# \text{ of Heads}\}$

$$P_X(0) = P(X = 0) = \frac{1}{4} = P_X(2),$$

$$P_X(1) = \frac{1}{4}$$

- $Y = \{\# \text{ of Tails}\}$

$$Y = 2 - X \rightarrow P_Y = P_X$$



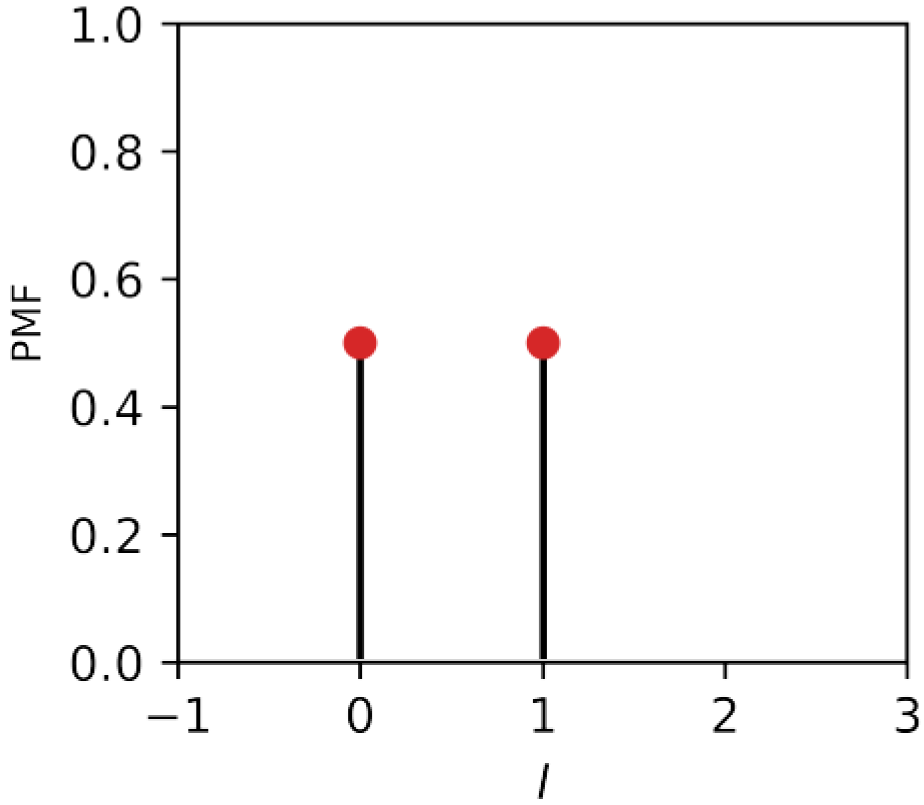
Probability Mass Function – Example

Consider double coin tosses:

- $I = \begin{cases} 1, if\ 1st\ toss = Heads \\ \quad 0, otherwise \end{cases}$

$$P_I(1) = P(I = 0) = \frac{1}{2}$$

$$P_I(0) = P(I = 1) = \frac{1}{2}$$



girafe ai

Continuous Case

- Consider non-discrete sample space S and random variable X

- Describe value occurrence with Probability Density Function (p.d.f):

$$f_X(x) \geq 0$$

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

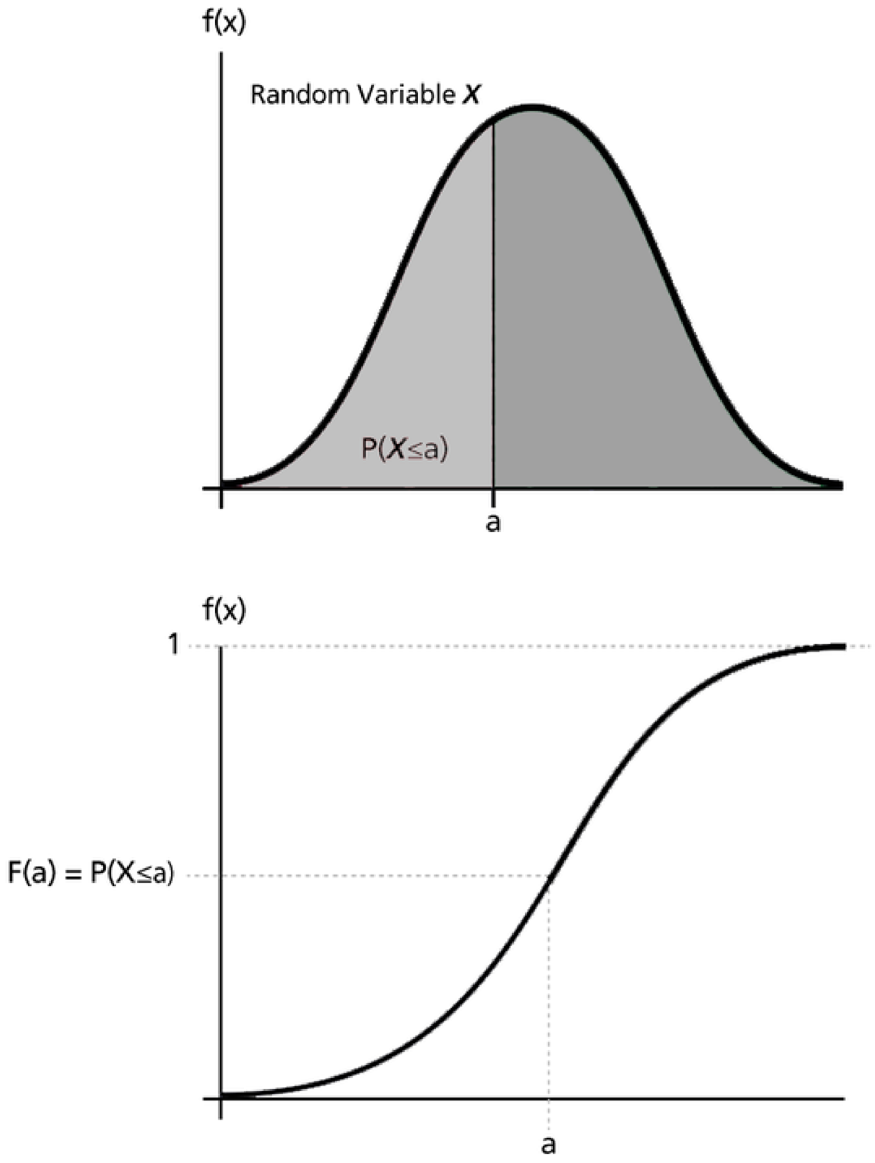


Image Source

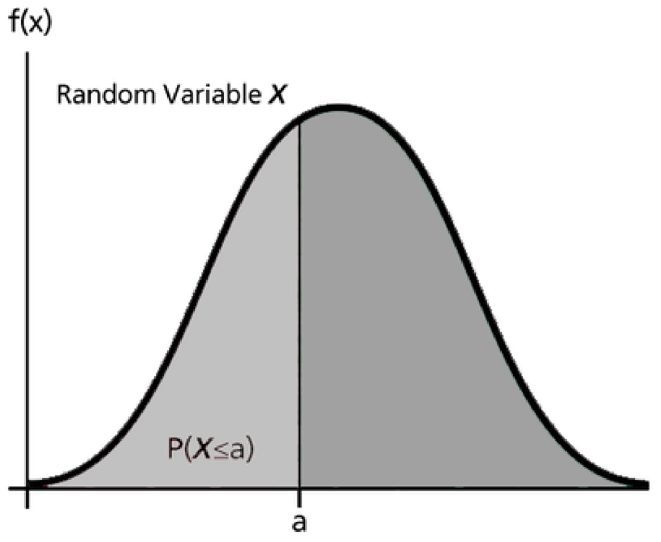
Continuous Case

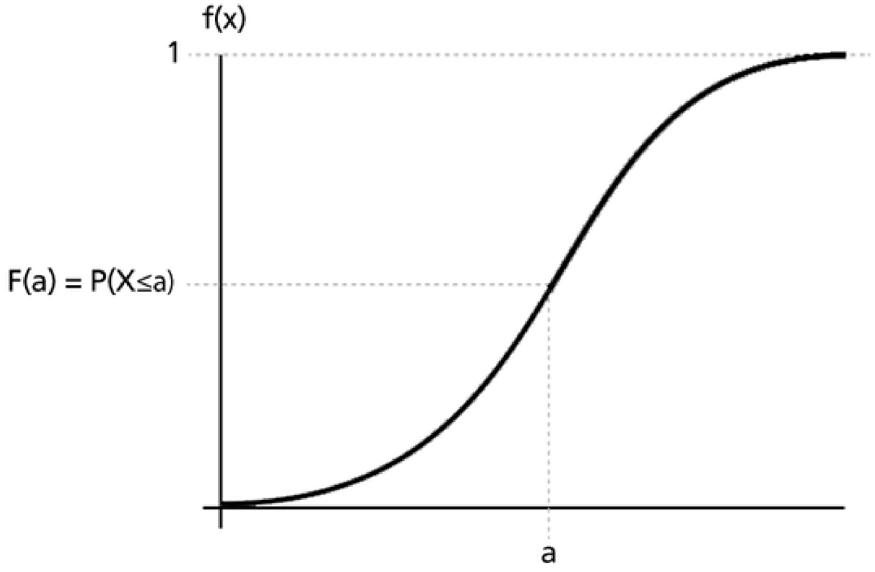
- Cumulative Distribution Function:

$$F_X(a) = \int_{-\infty}^{a} f_X(y)dy = P(X < a)$$

- Estimate probability of having value from range:

$$P(a < X < b) = \int_{b}^{a} f_X(y)dy = F_X(b) - F_X(a)$$





Expected Value

- Estimate of the most frequently occurring value for random variable X

- Discrete case:

$$E(X) = \sum_i x_i p_i = \sum_i x_i P(X = x_i)$$

- Continuous case:

$$E(X) = \int_{-\infty}^{+\infty} y \, f_X(y) dy$$

Variance

- Positive quantity measuring the spread of the distribution about its mean value

- General formula:

$$Var(X) = E[\left(X - E(X)\right)^2]$$

- Standard deviation:

$$Std(X) = \sqrt{Var(X)}$$

Independence and Covariance

- Two r.v. X and Y are independent if the following is correct for their p.d.f.'s for all possible x, y:

$$f(x,y) = f_X(x)f_y(y)$$

- For any r.v. X and Y *covariance* may be estimated:

$$Cov(X,Y) = E[\big(X - E(X)\big)\big(Y - E(Y)\big)]$$

- *Correlation* is estimated using covariance:

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

girafe ai

Independence and Covariance

- Two r.v. X and Y are independent if the following is correct for their p.d.f.'s for all possible x, y:

$$f(x,y) = f_X(x)f_y(y)$$

- Independent r.v. X and Y have covariance and correlation equal to 0

- The contrary is not always true

Functions of Random Variables

- Given r.v. X with expectation $\mathrm{E(X) = E_X}$ and variance $\mathrm{Var(X) = V_X}$
- We may construct a new r.v. $\mathrm{Y} = \alpha\,\mathrm{X} + \beta$ and estimate:
- Expectation:

$$\mathrm{E(Y)} = E(\alpha X + \beta) = \alpha E_X + \beta$$

- Variance:

$$Var(Y) = Var(\alpha X + \beta) = \alpha^2 V_X$$

Functions of Random Variables

- Given r.v. X with expectation $\mathrm{E(X)} = \mathrm{E_X}$ and variance $\mathrm{Var(X)} = \mathrm{V_X}$

- We may construct a new r.v. $\mathrm{Y} = \frac{(\mathrm{X}-\mathrm{E_X})}{\sqrt{V_X}} = \frac{1}{\sqrt{V_X}}X - \frac{E_X}{\sqrt{V_X}}$ and estimate:

- Expectation:

$$\mathrm{E(Y)} = 0$$

- Variance:

$$Var(Y) = 1$$

girafe ai

Functions of Random Variables

- Given any r.v. X_1 and X_2, for which expected value may be estimated

- Resulting expectation:

$$\mathrm{E}(\mathrm{X}_1 + \mathrm{X}_2) = E(X_1) + E(X_2)$$

- Resulting variance:

$$\mathrm{E}(\mathrm{X}_1 + \mathrm{X}_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$$

girafe ai

Functions of Random Variables

- Given any r.v. X_1 and X_2, for which expected value may be estimated

- Resulting expectation:

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

- Resulting variance:

$$E(X_1 + X_2) = Var(X_1) + Var(X_2) + \cancel{2Cov(X_1, X_2)}$$

For independent r.v.

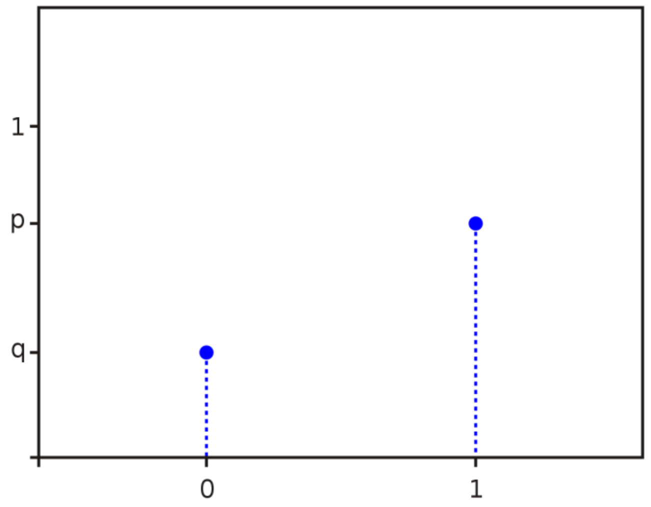
Bernoulli Distribution

- Discrete distribution with p.m.f.:

$$X \sim Bernoulli(p)$$

$$X = \begin{cases} 1, P_X = p \\ 0, P_X = 1 - p \end{cases}$$

- Used to represent random guessing with predefined probability



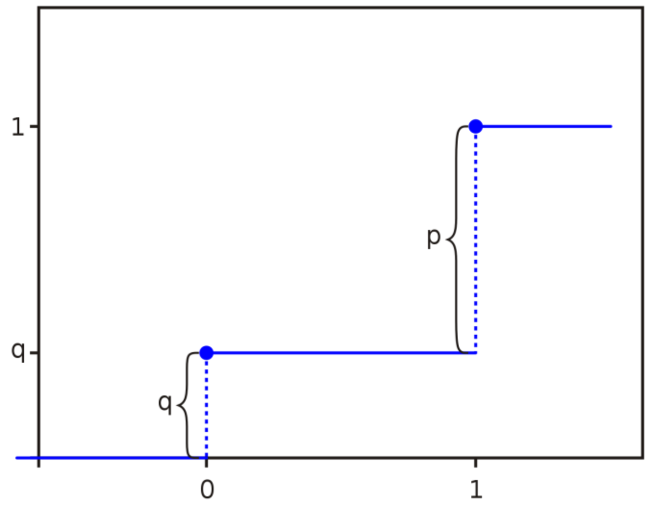


Image Source

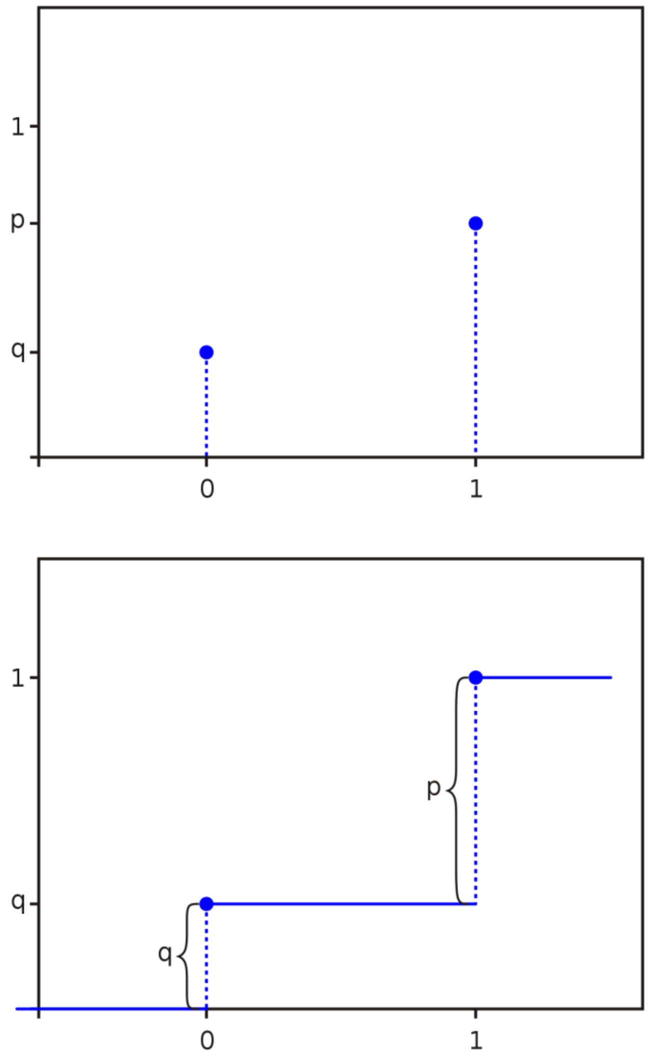
Bernoulli Distribution

- Discrete distribution with p.m.f.:

$$X \sim Bernoulli(p)$$

$$X = \begin{cases} 1, P_X = p \\ 0, P_X = 1 - p \end{cases}$$

- $E(X) = p, \mathrm{Var(X)} = \mathrm{p(1-p)}$



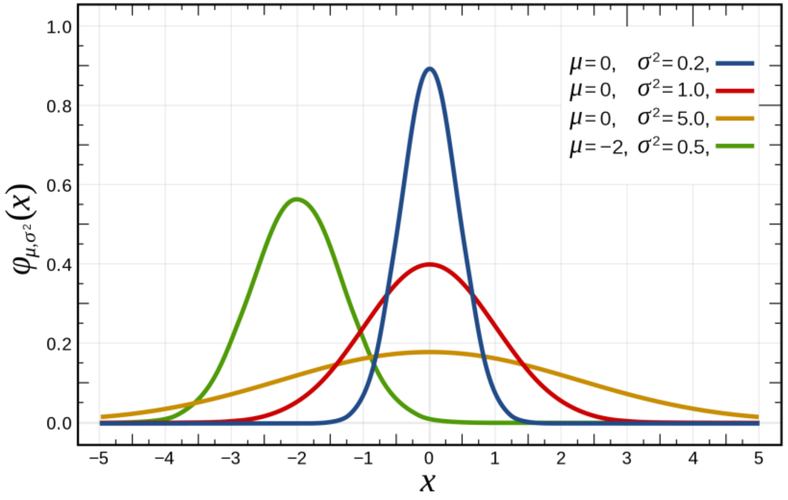
Normal Distribution

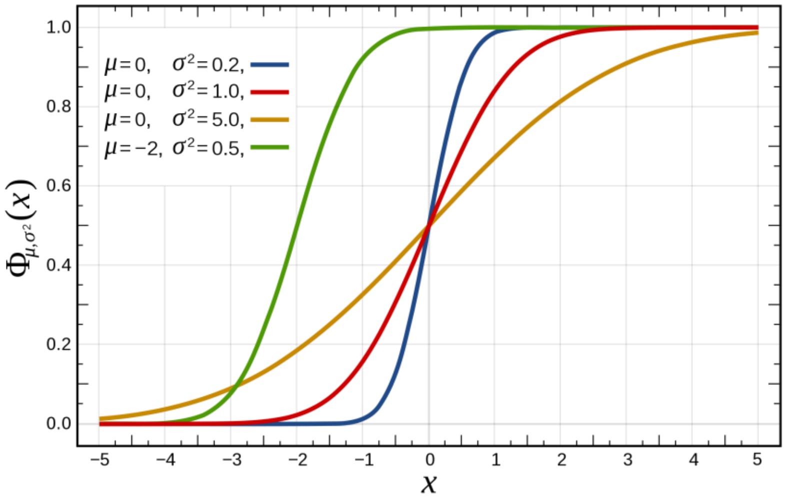
- Continuous distribution with p.d.f.:

$$X \sim N(\mu, \sigma^2)$$

$$f_X(y; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- Used to represent random variables which distributions are not known





Central Limit Theorem

- Let $X_1, \dots, X_n$ - r.v. which are *i.i.d. (independent and identically distributed) and having mean* μ and variance σ^2

- Then a new random variable :

$$\bar{X} = \frac{\sum_{i=0}^{n} X_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

girafe ai

Image Source

Law of Large Numbers

- Let $X_1, \dots, X_n$ - r.v. which are *i.i.d. (independent and identically distributed) and having same mean* μ

- Let's construct a *random sequence* by averaging the r.v.'s:

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{\sum_{i=0}^{n} X_i}{n}$$

- Then $\bar{X}_n \to \mu$ as n $\to \infty$

- Resulting variance: $\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2} Var(X_1 + \cdots + X_n) =$

girafe ai

Image Source

Law of Large Numbers

- Let $X_1, \dots, X_n$ - r.v. which are *i.i.d. (independent and identically distributed) and having same mean μ*

- Let's construct a *random sequence* by averaging the r.v.'s:

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{\sum_{i=0}^{n} X_i}{n}$$

- If all r.v.'s have the same variance σ^2, then resulting variance:

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2} Var(X_1 + \cdots + X_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

To sum up

- Matrices as linear transforms
- Examples of common transforms
- Inverse
- Determinant
- Rank
- Calculus
- Probability theory basics
- Random Variables
- Main theoretical concepts

Next Time

- Statistics
- Maximum Likelihood Estimation (MLE)
- ML Introduction & Supervised Learning
- K-Nearest Neighbors Algorithm (KNN)