Tensors

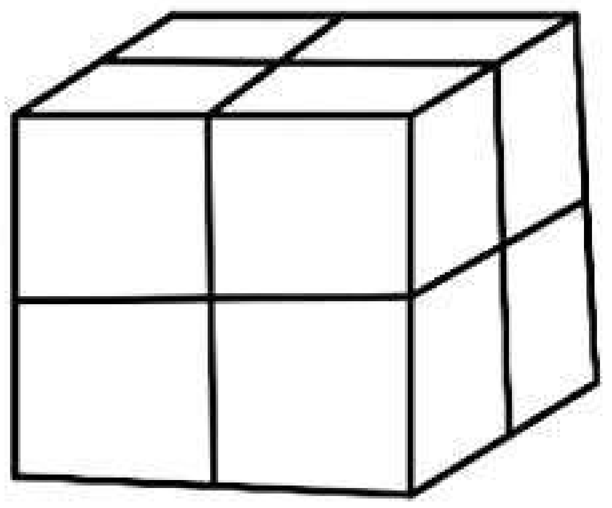
"ML generalization of vectors and matrices to any number of dimensions"

scalar x

vector $\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$

matrix $\begin{bmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{bmatrix}$

3-tensor



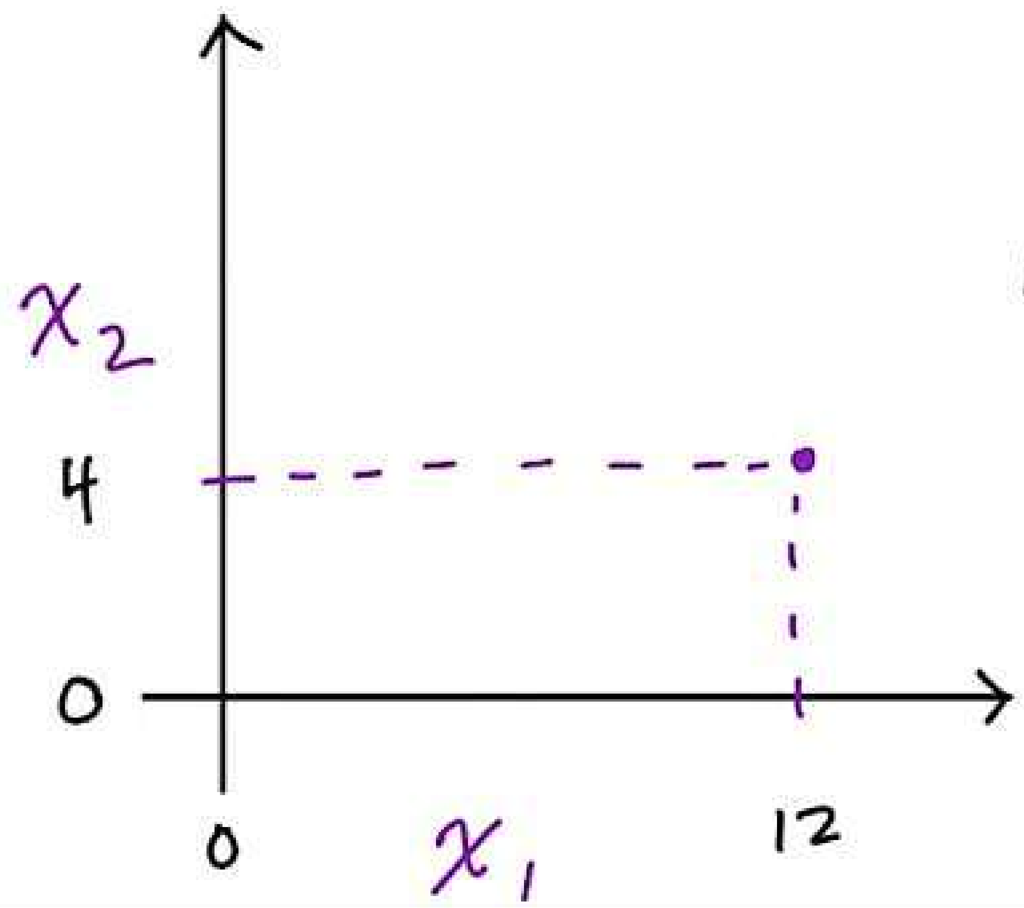
Dimensions	Mathematical Name	Description
0	scalar	magnitude only
1	vector	array
2	matrix	flat table, e.g., square
3	3-tensor	3D table, e.g., cube
n	*n*-tensor	higher dimensional

Scalars

- No dimensions
- Single number
- Denoted in lowercase, italics, e.g.: x
- Should be *typed*, like all other tensors: e.g., int, float32

Vectors

- One-dimensional array of numbers
- Denoted in lowercase, italics, bold, e.g.: $\boldsymbol{x}$
- Arranged in an order, so element can be accessed by its index
    - Elements are scalars so *not* bold, e.g., second element of $\boldsymbol{x}$ is x_2
- Representing a point in space:
    - Vector of length two represents location in 2D matrix (shown)
    - Length of three represents location in 3D cube
    - Length of *n* represents location in *n*-dimensional tensor



$[x_1 \;\; x_2] = [12 \;\; 4]$

Vector Transposition

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

row vector

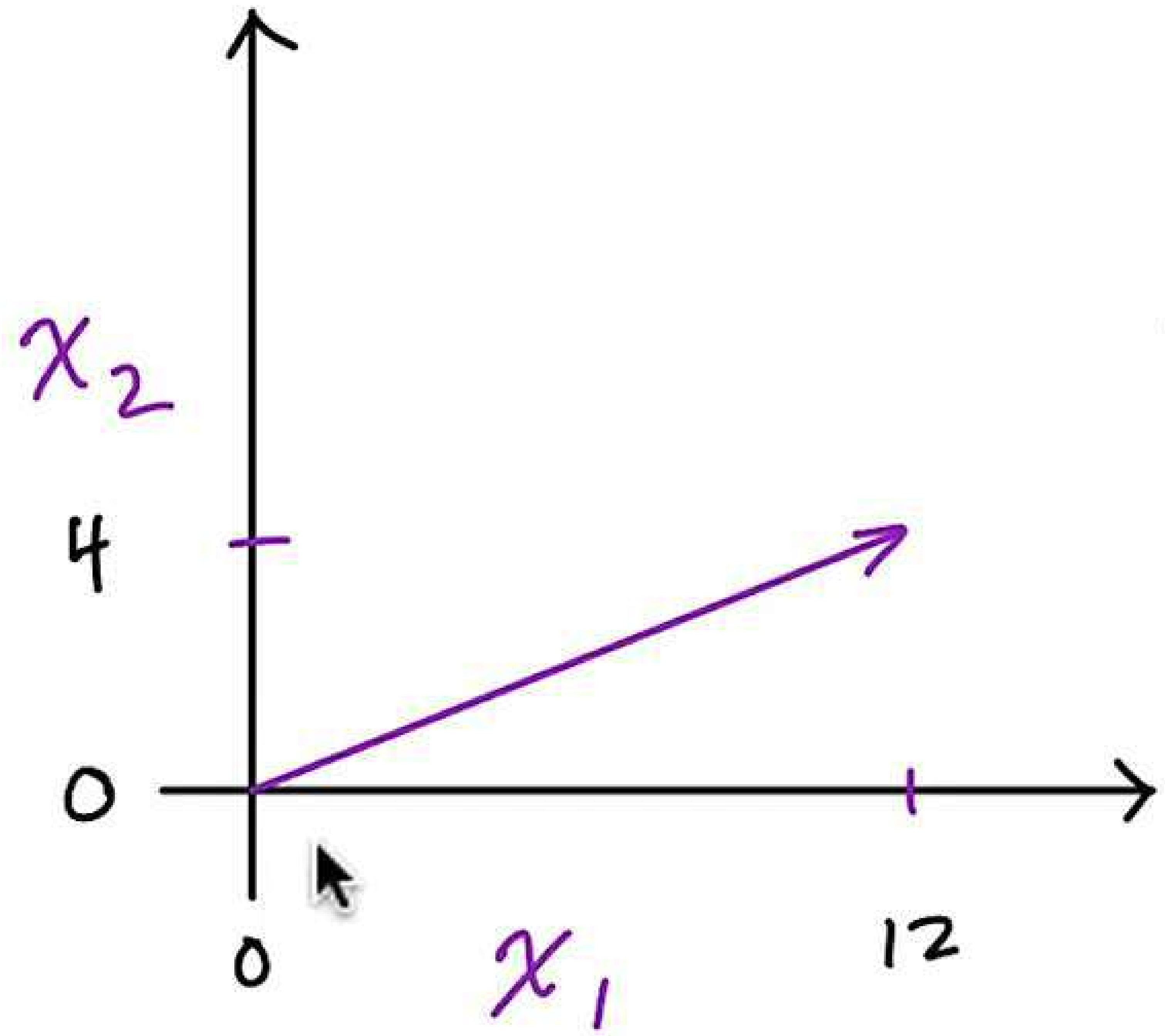
column vector

shape is (1, 3)

(3, 1)

Norms

Vectors represent a magnitude and direction from origin:



$[x_1 \quad x_2] = [12 \quad 4]$

Norms are functions that quantify vector magnitude.

L^2 Norm

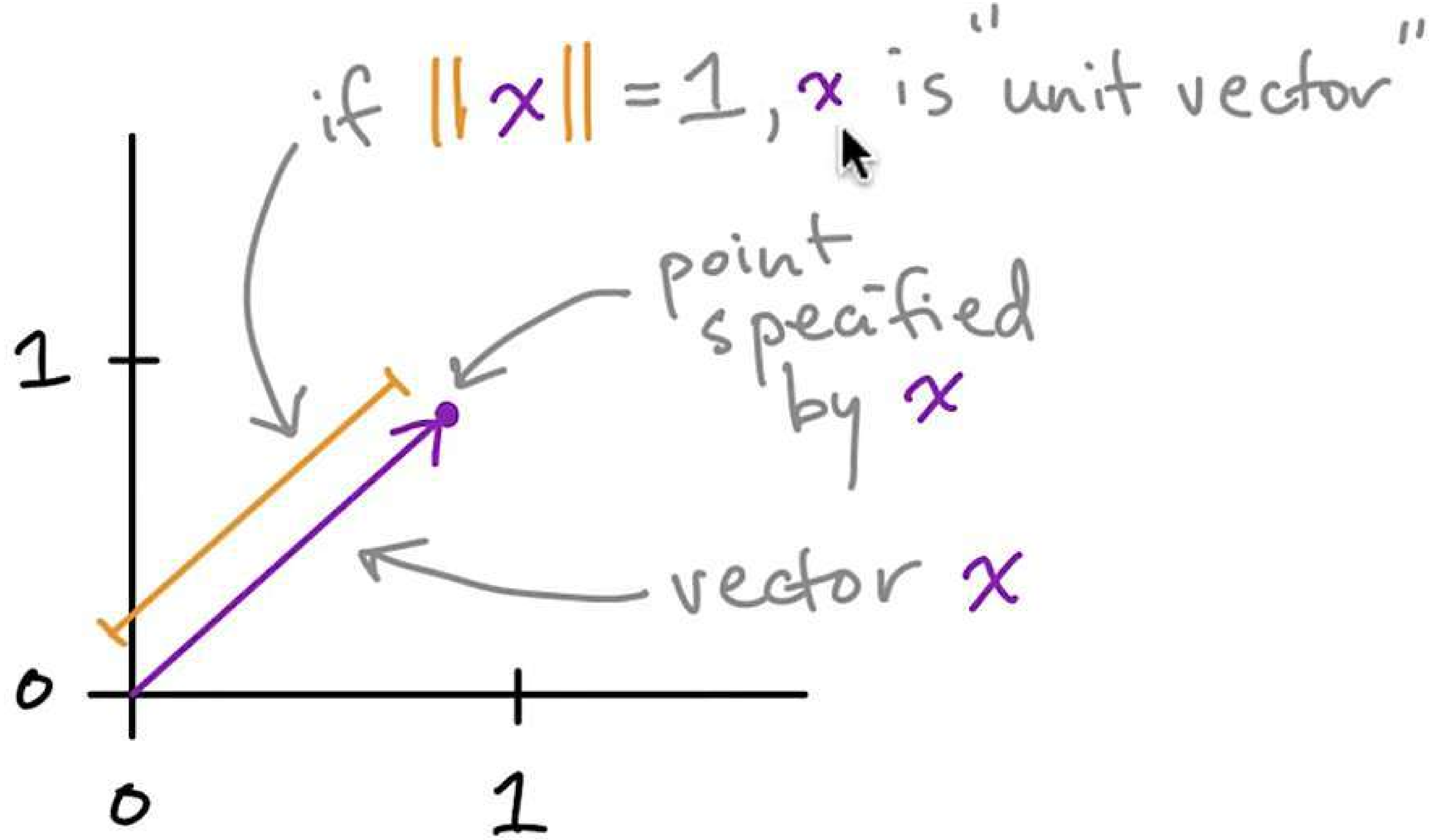
- Described by:

$$||x||_2 = \sqrt{\sum_i x_i^2}$$

- Measures simple (Euclidean) distance from origin
- Most common norm in machine learning
  - Instead of $||\boldsymbol{x}||_2$, it can be denoted as $||\boldsymbol{x}||$

Unit Vectors

- Special case of vector where its length is equal to one
- Technically, $\boldsymbol{x}$ is a unit vector with "unit norm", i.e.: $||\boldsymbol{x}|| = 1$



L^1 Norm

- Described by:

$$\|x\|_1 = \sum_i |x_i|$$

- Another common norm in ML
- Varies linearly at all locations whether near or far from origin
- Used whenever difference between zero and non-zero is key

Squared L^2 Norm

- Described by: $$\|x\|_2^2 = \sum_i x_i^2$$
- Computationally cheaper to use than L^2 norm because:
  - Squared L^2 norm equals simply $\boldsymbol{x}^\mathrm{T}\boldsymbol{x}$
  - Derivative (used to train many ML algorithms) of element x requires that element alone, whereas L^2 norm requires $\boldsymbol{x}$ vector
- Downside is it grows slowly near origin so can't be used if distinguishing between zero and near-zero is important

Max Norm (or L^∞ Norm)

- Described by:

$$\|x\|_\infty = \max_i |x_i|$$

- Final norm we'll discuss; also occurs frequently in ML
- Returns the absolute value of the largest-magnitude element

Generalized L^p Norm

- Described by:

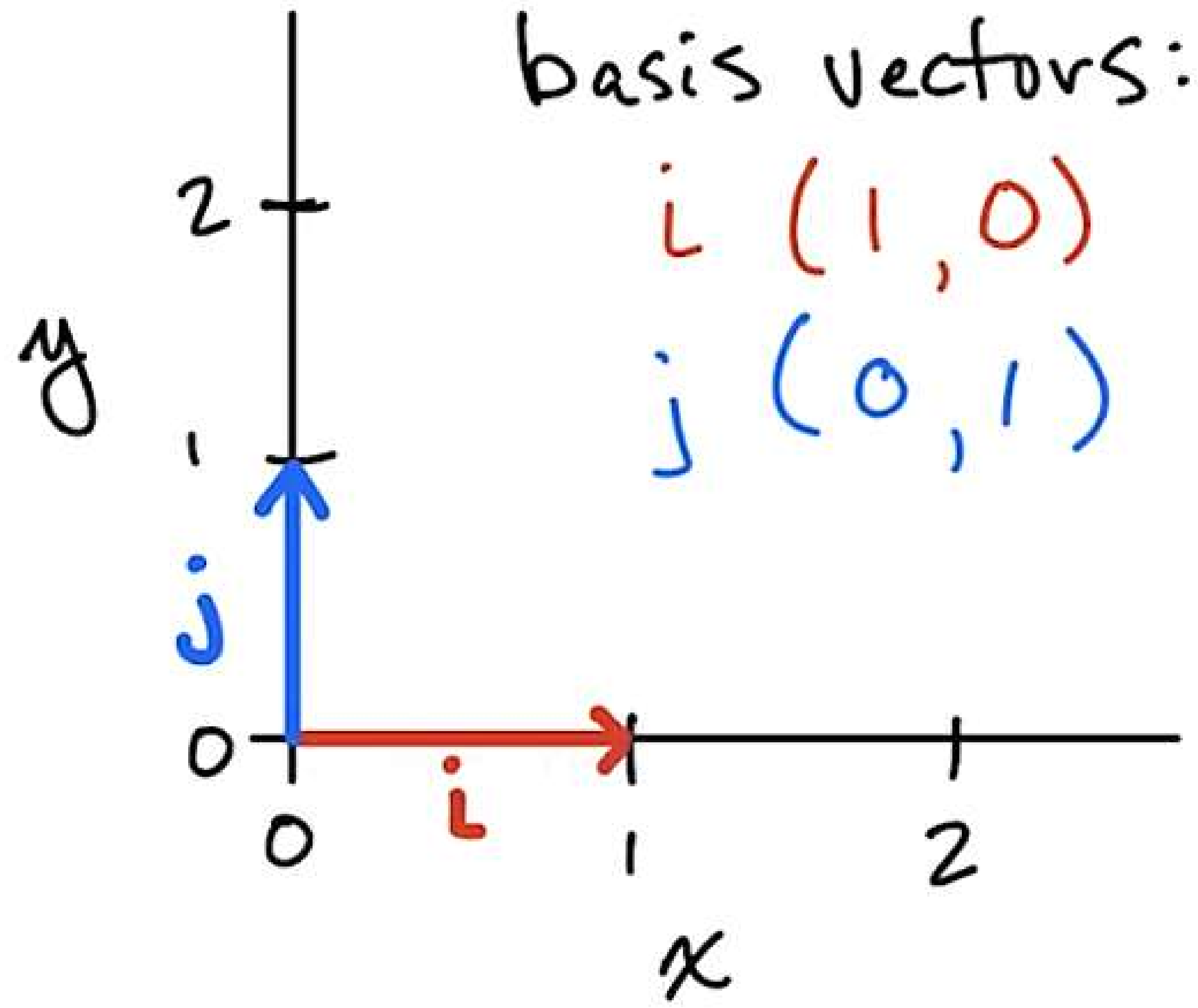
$$\|x\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$$

- p must be:
  - A real number
  - Greater than or equal to one
- Can derive L^1, L^2, and L^∞ norm formulae by substituting for p
- Norms, particularly L^1 and L^2, used to regularize objective functions

Orthogonal Vectors

- $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal vectors if $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0$
- Are at 90° angle to each other (assuming non-zero norms)
- n-dimensional space has max n mutually orthogonal vectors (again, assuming non-zero norms)
- **Orthonormal** vectors are orthogonal *and* all have unit norm
    - Basis vectors are an example

Hands-on code demo



Basis Vectors

- Can be scaled to represent *any* vector in a given vector space
- Typically use unit vectors along axes of vector space (shown)