

Tensors

“ML generalization of vectors and matrices to any number of dimensions”

scalar

x

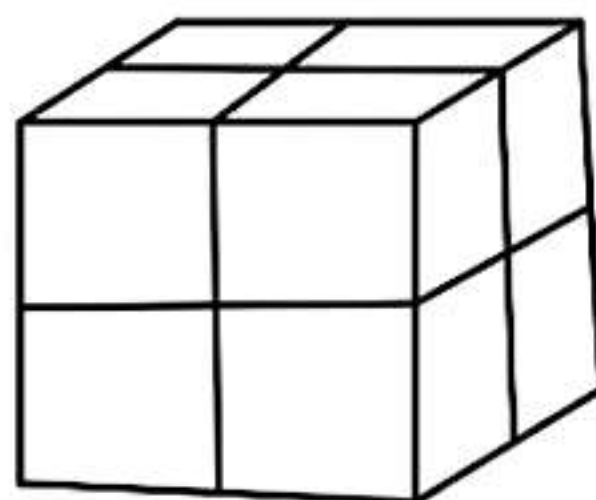
vector

$[x_1 \ x_2 \ x_3]$

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

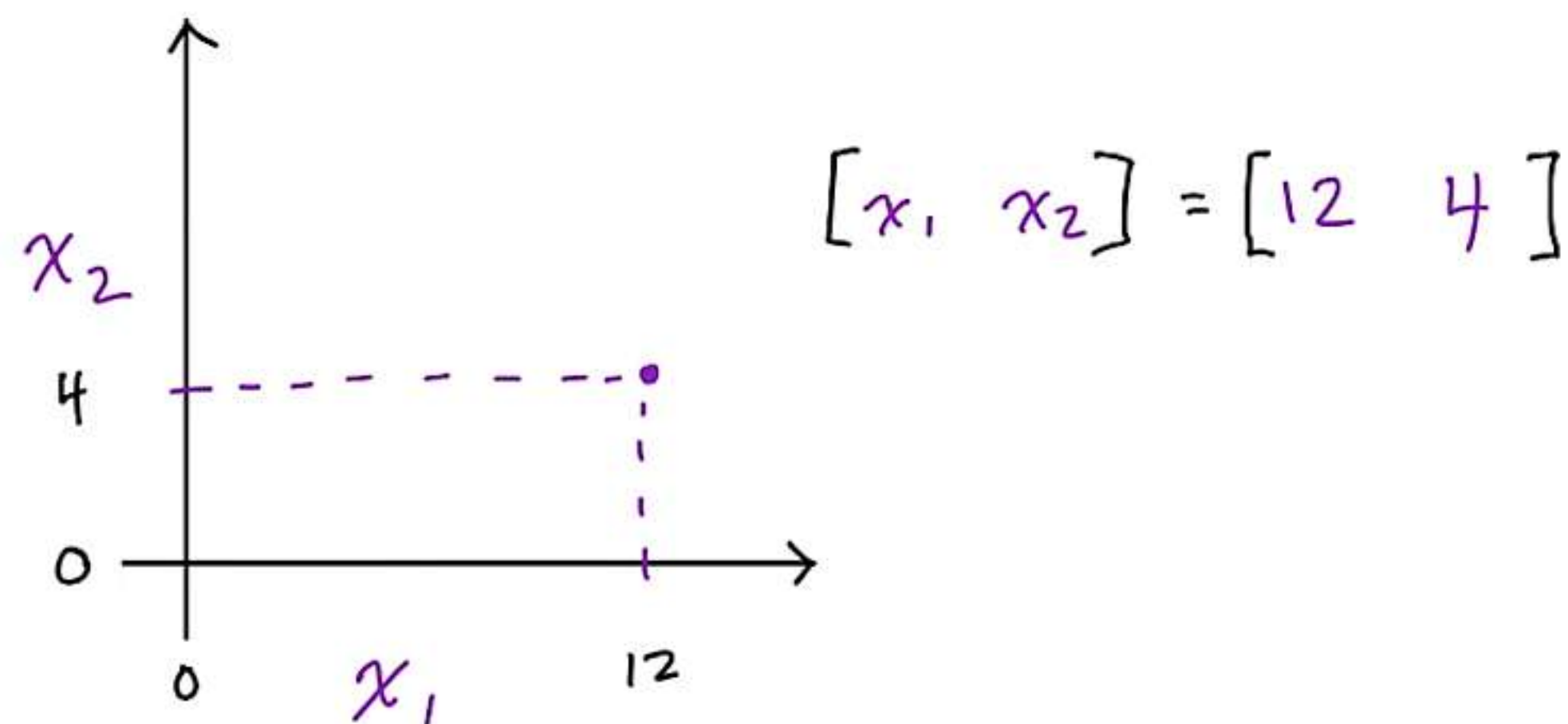
higher dimensional

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.: \mathbf{x}
- Arranged in an order, so element can be accessed by its index
 - Elements are scalars so *not* bold, e.g., second element of \mathbf{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



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

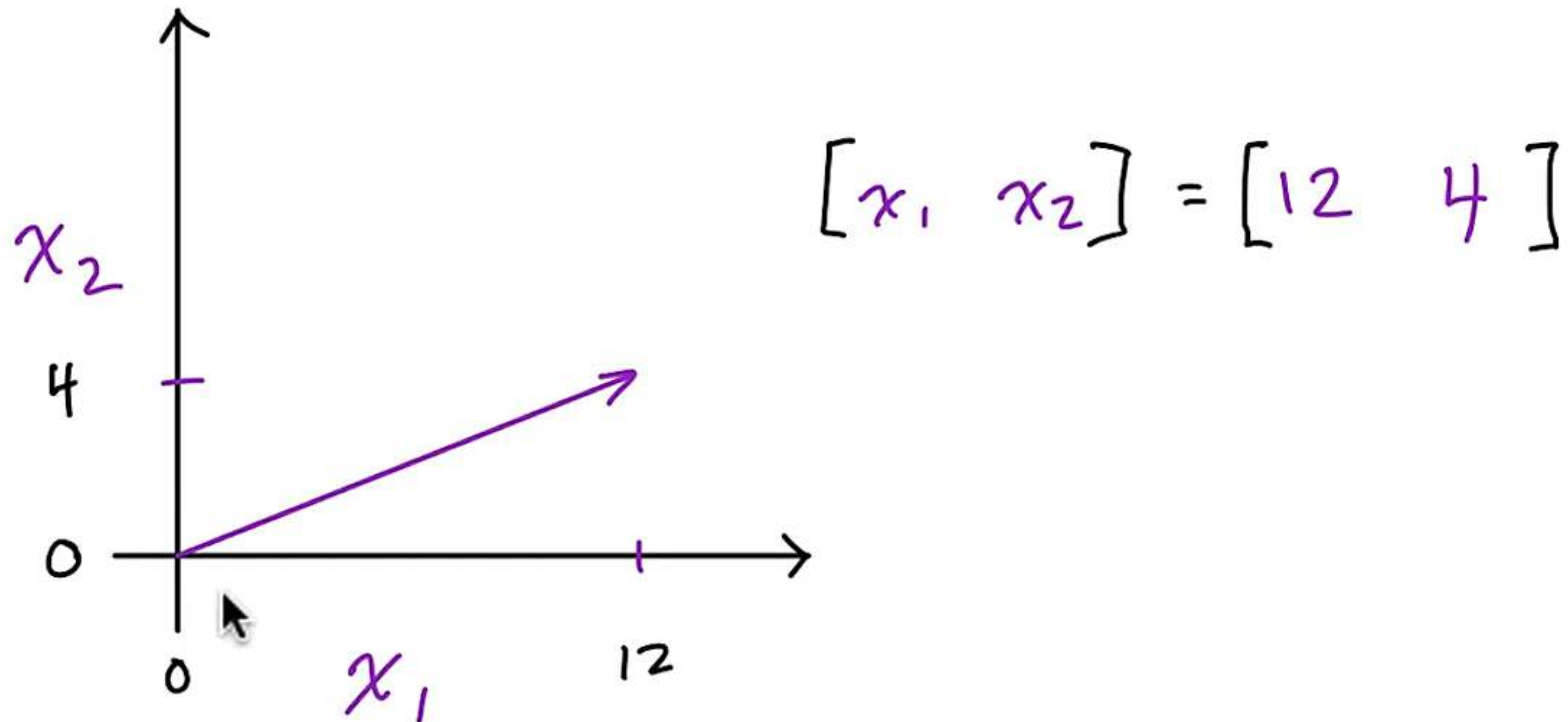
shape is (1, 3)

column
vector

(3, 1)

Norms

Vectors represent a magnitude and direction from origin:



Norms are functions that quantify vector magnitude.

L^2 Norm

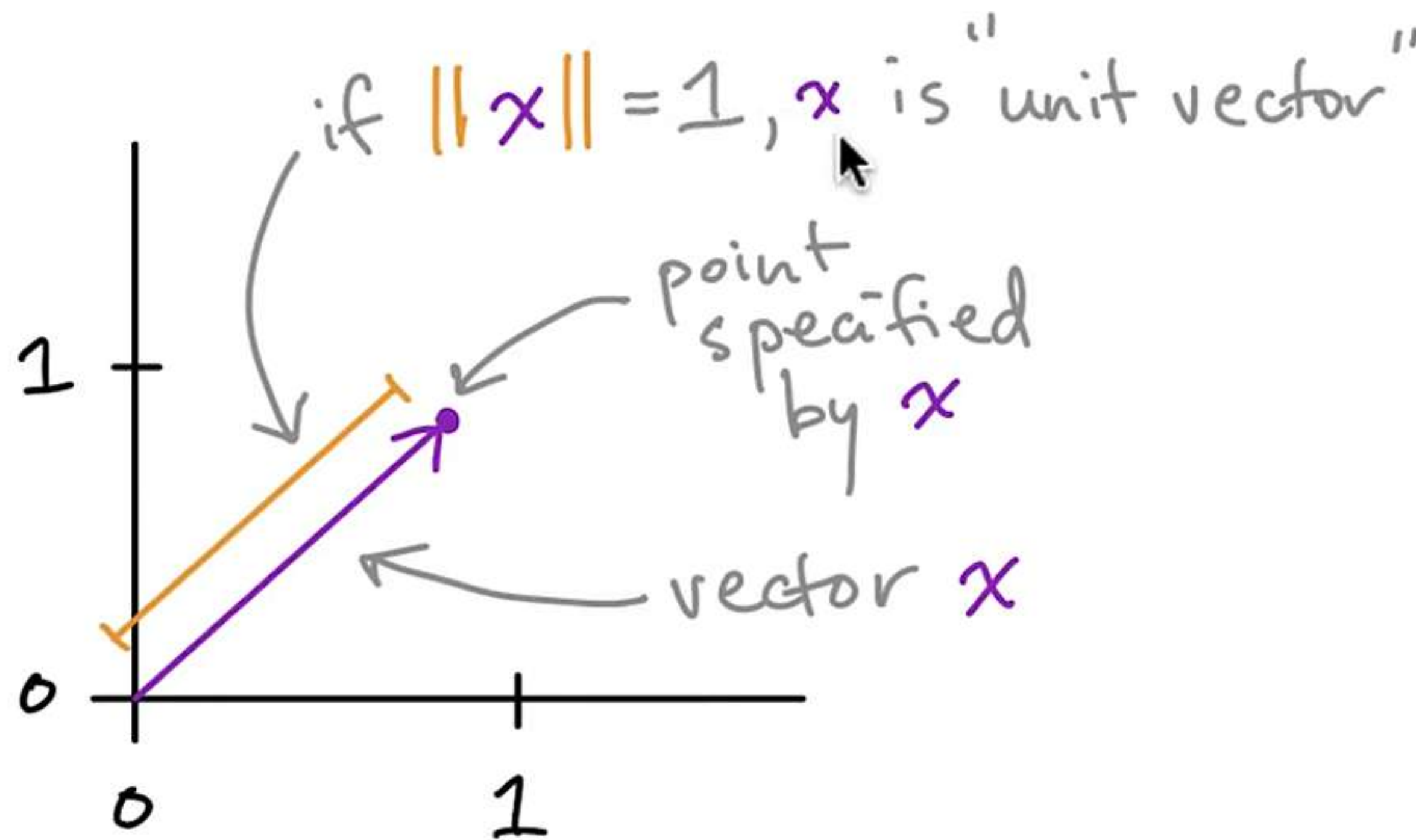
- Described by:

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

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

Unit Vectors

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



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: $\|\mathbf{x}\|_2^2 = \sum_i x_i^2$
- Computationally cheaper to use than L^2 norm because:
 - Squared L^2 norm equals simply $\mathbf{x}^T \mathbf{x}$
 - Derivative (used to train many ML algorithms) of element x requires that element alone, whereas L^2 norm requires \mathbf{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:

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

- 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