Statistical Learning and Data Mining
CS 363D/ SSC 358

Lecture: Linear Algebra Foundations

Prof. Pradeep Ravikumar
pradeepr@cs.utexas.edu

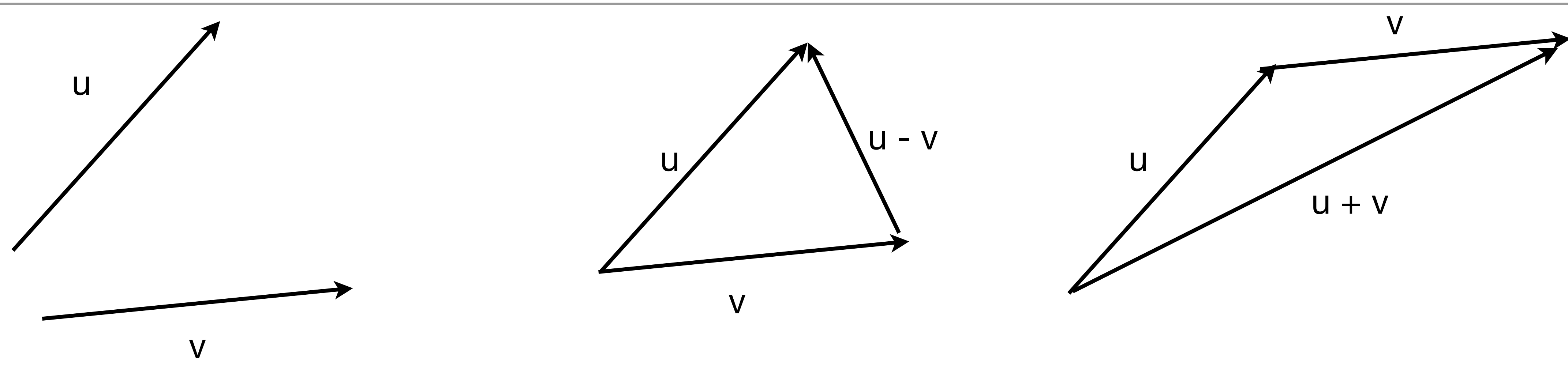
Outline

- Vectors (Norms, Distances, Inner Products, Orthogonality, Linear Combinations, Linear Independence, Linear Subspace, Basis, Orthogonal Basis)

Vectors

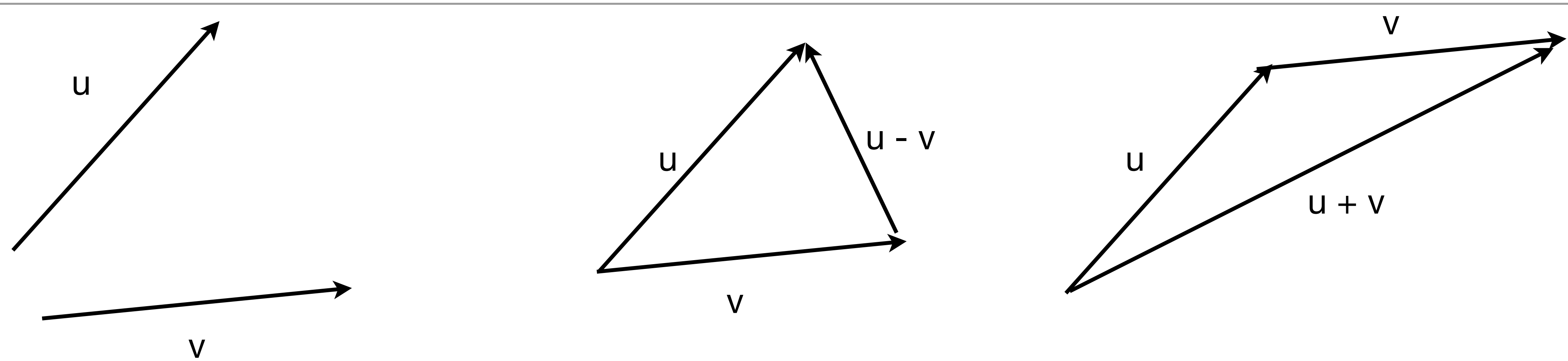
- Think of a vector as an abstract mathematical representation of an object
- Can be imbue such “vectors” with properties possessed by real numbers (also called scalars)?

Vectors



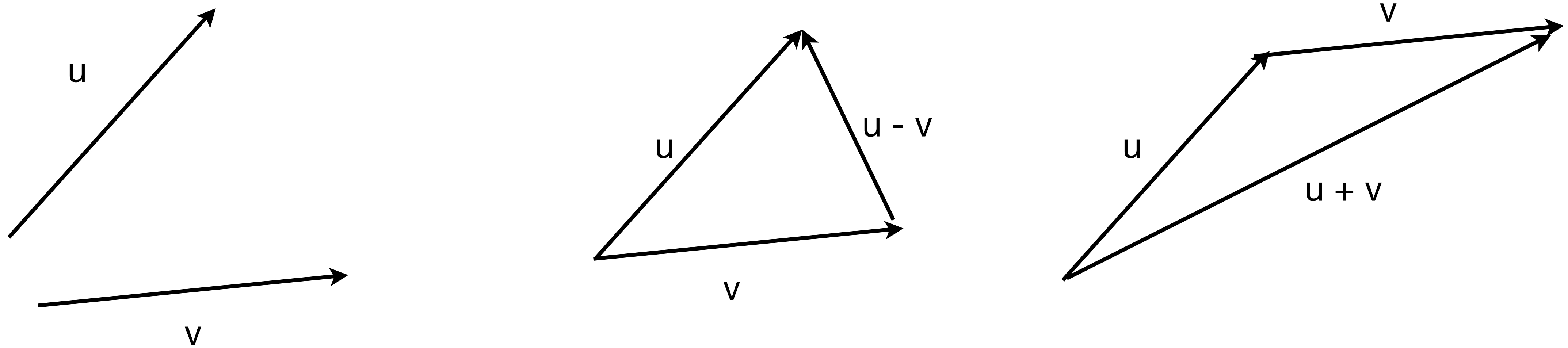
- Can add and subtract vectors

Vectors



- Can add and subtract vectors
- Commutative: u + v = v + u

Vectors

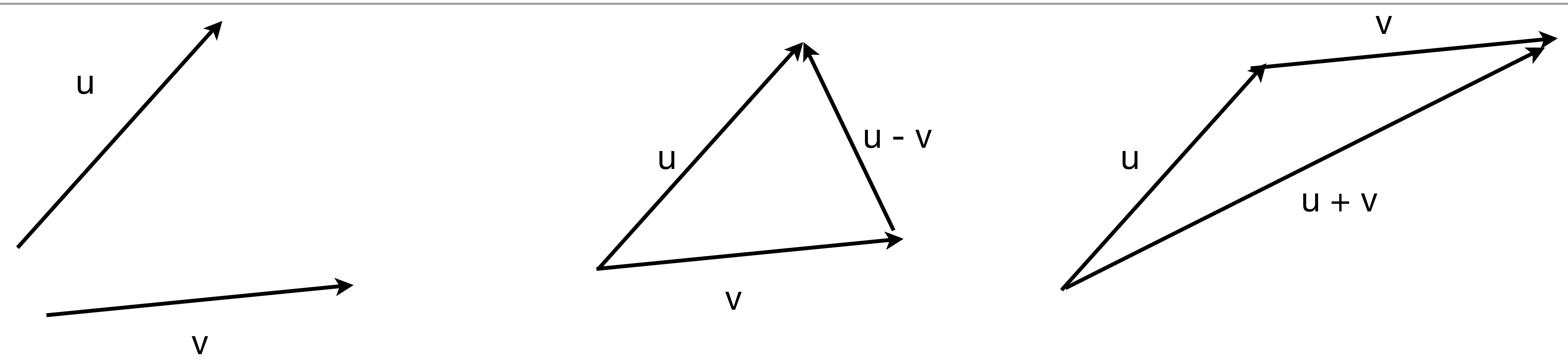


- Can add and subtract vectors

- Commutative: u + v = v + u

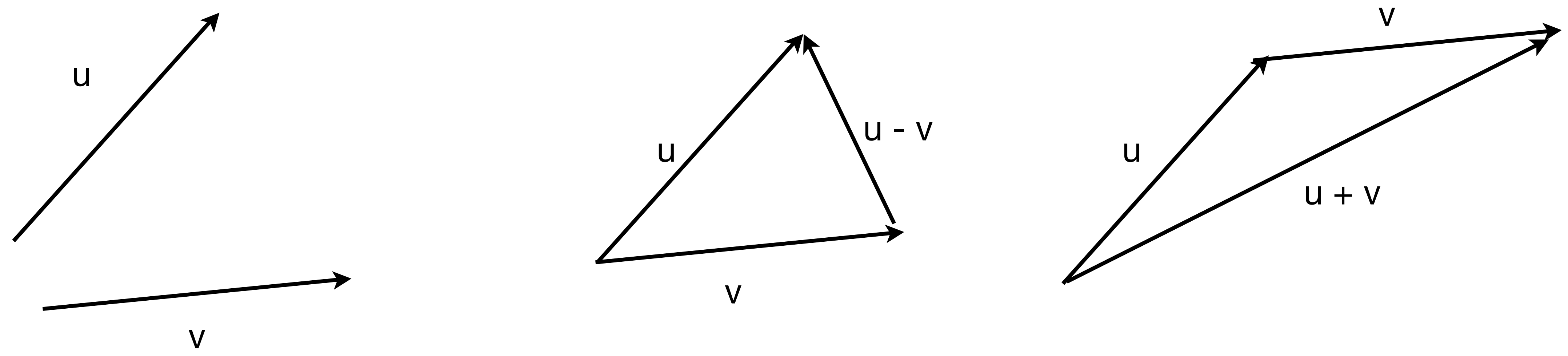
- Associative: u + (v + w) = (u + v) + w

Vectors



- Can add and subtract vectors
- Commutative: u + v = v + u
- Associative: u + (v + w) = (u + v) + w
- Zero: There exists a vector 0, such that u + 0 = u

Vectors



- Can add and subtract vectors
- Commutative: u + v = v + u
- Associative: u + (v + w) = (u + v) + w
- Zero: There exists a vector 0, such that u + 0 = u
- Inverse: For every u, there is a vector - u, such that u + (-u) = 0

Vectors

- Can multiply vectors with scalars

Vectors

- Can multiply vectors with scalars
- Associative: a (b u) = (ab) u

Vectors

- Can multiply vectors with scalars
- Associative: a (b u) = (ab) u
- Distributive I: (a + b) u = a u + b u

Vectors

- Can multiply vectors with scalars
- Associative: a (b u) = (ab) u
- Distributive I: (a + b) u = a u + b u
- Distributive II: a (u + v) = a u + a v

Vectors

- Can multiply vectors with scalars
- Associative: a (b u) = (ab) u
- Distributive I: (a + b) u = a u + b u
- Distributive II: a (u + v) = a u + a v
- Identity: 1 u = u

Vector Space

- A vector space is a set of vectors, along with associated scalars (typically: real numbers), that satisfy properties in previous two slides, and that are **closed** under vector addition and scalar multiplication

- An abstraction for many “sets of objects”
  - not just in data mining/machine learning but in many applications across science and engineering

- And from the previous two slides, we can “treat” them like ordinary numbers for the most part

Vector Space: Linear Independence

- Suppose we have three vectors x_1, x_2, and x_3, and that $x_1 = \alpha_2 x_2 + \alpha_3 x_3$. Then x_1 is **linearly dependent** on x_2 and x_3.

Vector Space: Linear Independence

- Suppose we have three vectors x_1, x_2, and x_3, and that $x_1 = \alpha_2 x_2 + \alpha_3 x_3$. Then x_1 is **linearly dependent** on x_2 and x_3.
- When are $x_1, x_2, \ldots, x_n$ linearly *independent*?

Vector Space: Linear Independence

- Suppose we have three vectors x_1, x_2, and x_3, and that $x_1 = \alpha_2 x_2 + \alpha_3 x_3$. Then x_1 is **linearly dependent** on x_2 and x_3.

- When are $x_1, x_2, \ldots, x_n$ linearly *independent*?

- $x_1, x_2, \ldots, x_n$ are **linearly independent** $\equiv$
  If $\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n = 0$, then $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$.

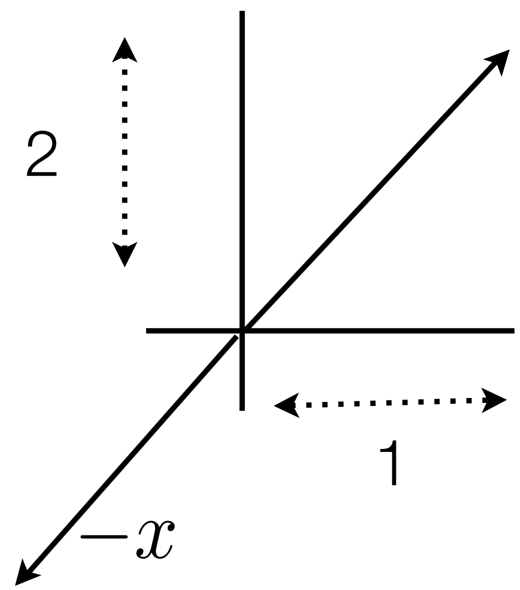
Vector Space: Subspace

- A **linear subspace** is a set of vectors that is closed under vector addition and scalar multiplication: if x_1 and x_2 belong to the subspace, then so do $\alpha_1 x_1 + \alpha_2 x_2$.

Vector Space: Subspace

- A **linear subspace** is a set of vectors that is closed under vector addition and scalar multiplication: if x_1 and x_2 belong to the subspace, then so do $\alpha_1 x_1 + \alpha_2 x_2$.
- A **basis** of the subspace is the maximal set of vectors in the subspace that are linearly independent of each other.

Vectors



$$x = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad = 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- A vector space is thus the set of vectors obtained as linear combinations of its “basis” vectors

- Can thus represent a vector as an array of numbers: where the numbers are the coefficients of the basis vectors in the linear combination

Vectors

- A vector can be thought of as having a **direction** and a **magnitude**

Vectors

- A vector can be thought of as having a **direction** and a **magnitude**
  - This “magnitude” is called a vector norm

Vectors

- A vector can be thought of as having a **direction** and a **magnitude**
  - This “magnitude” is called a vector norm
- Properties satisfied by vector norms || . ||

Vectors

- A vector can be thought of as having a **direction** and a **magnitude**
  - This “magnitude” is called a vector norm
- Properties satisfied by vector norms $|| \, . \, ||$
  - $|| x || >= 0$ and $|| x || = 0$ if and only if $x = 0$

Vectors

- A vector can be thought of as having a **direction** and a **magnitude**
  - This “magnitude” is called a vector norm
- Properties satisfied by vector norms || . ||
  - $\| x \| >= 0$ and $\| x \| = 0$ if and only if $x = 0$
  - $\| a x \| = | a | \; \| x \|$ (Homegeneity)

Vectors

- A vector can be thought of as having a **direction** and a **magnitude**
  - This “magnitude” is called a vector norm
- Properties satisfied by vector norms || . ||
  - || x || >= 0 and || x || = 0 if and only if x = 0
  - || a x || = | a | || x || (Homegeneity)
  - || x + y || <= || x || + || y || (Triangle Inequality)

Examples: Vector Norms

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

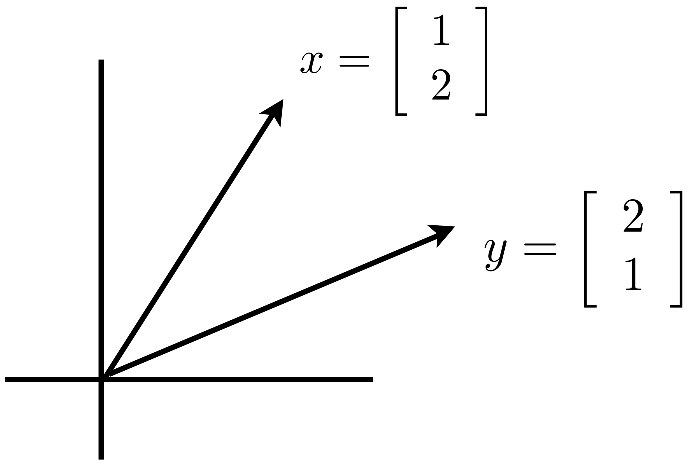
$$\|x\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2} \quad : \text{ 2-norm; ``Euclidean'' norm}$$

$$\|x\|_1 = |x_1| + |x_2| + \ldots + |x_n| \quad : \text{ 1-norm}$$

$$\|x\|_p = \sqrt[p]{|x_1|^p + |x_2|^p + \ldots + |x_n|^p} \quad : \text{ p-norm}$$

$$\|x\|_\infty = \max_{i=1}^{n} |x_i| \quad : \infty\text{-norm}$$

Distances



- How do we measure the "distance" between two vectors?
- We looked at a few distance measures in the previous class; which could be looked at as distances between vectors
- One could also use vector norms to compute distances:

$$\|x - y\|_2 = \left\| \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\|_2 = \sqrt{(1-2)^2 + (2-1)^2} = \sqrt{2}$$

$$\|x - y\|_1 = 2$$

$$\|x - y\|_\infty = 1$$

Metrics

A distance $d(x, y)$ is a metric iff

- $d(x, y) \geq 0$, and $d(x, y) = 0$ iff $x = y$
- $d(x, y) = d(y, x)$ (Symmetry)
- $d(x, z) \leq d(x, y) + d(y, z)$ (Triangle Inequality)

Metrics

A distance $d(x, y)$ is a metric iff

- $d(x, y) \geq 0$, and $d(x, y) = 0$ iff $x = y$
- $d(x, y) = d(y, x) \quad$ (Symmetry)
- $d(x, z) \leq d(x, y) + d(y, z) \quad$ (Triangle Inequality)

Candidate metric: $d(x, y) = \|x - y\|$. Is this a valid metric?

Metrics

A distance $d(x, y)$ is a metric iff

- $d(x, y) \geq 0$, and $d(x, y) = 0$ iff $x = y$
- $d(x, y) = d(y, x)$ (Symmetry)
- $d(x, z) \leq d(x, y) + d(y, z)$ (Triangle Inequality)

Candidate metric: $d(x, y) = \|x - y\|$. Is this a valid metric?

$$d(x, z) = \|x - z\| = \|(x - y) + (y - z)\| \leq \|x - y\| + \|y - z\| = d(x, y) + d(y, z).$$

Metrics

A distance $d(x,y)$ is a metric iff

- $d(x,y) \geq 0$, and $d(x,y) = 0$ iff $x = y$
- $d(x,y) = d(y,x)$ (Symmetry)
- $d(x,z) \leq d(x,y) + d(y,z)$ (Triangle Inequality)

Candidate metric: $d(x,y) = \|x - y\|$. Is this a valid metric?

$$d(x,z) = \|x - z\| = \|(x-y) + (y-z)\| \leq \|x - y\| + \|y - z\| = d(x,y) + d(y,z).$$

✓ $d(x,y) = \|x - y\|$ is a valid metric.

Inner Products (Also: Dot Products)

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Inner Product: $x^T y = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = \sum_{i=1}^{n} x_i y_i$

Can be viewed as: $\begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

Inner Products (Also: Dot Products)

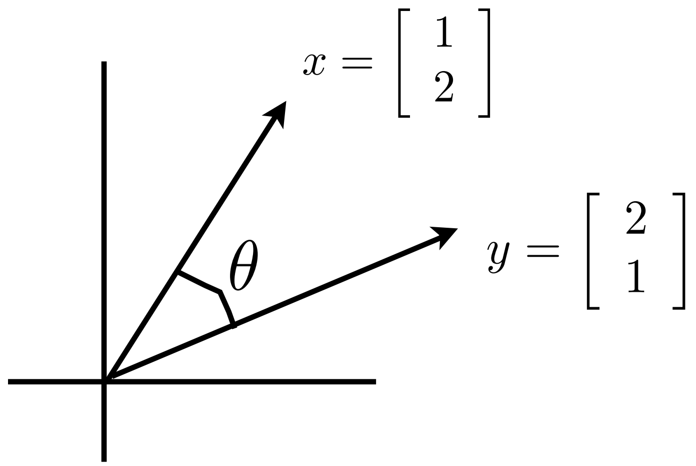
$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Inner Product: $x^T y = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = \sum_{i=1}^{n} x_i y_i$

Can be viewed as: $\begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

Examples: $x^T x = \|x\|_2^2$, $(x-y)^T(x-y) = \|x-y\|_2^2$

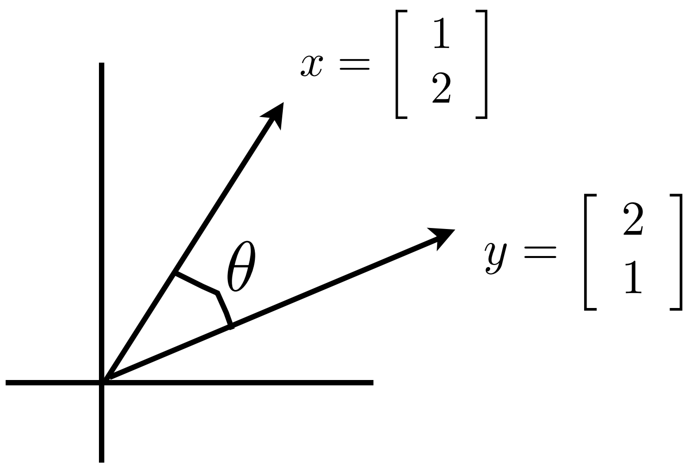
Projections



$$x^T y = \|x\|_2 \|y\|_2 \cos\theta$$

$$\cos\theta = \frac{x^T y}{\|x\|_2 \|y\|_2}$$

Projections



Projection of x onto y:

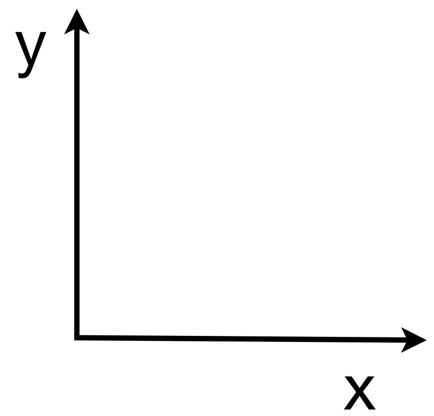
Magnitude: $\|x\|_2 \cos\theta = x^T \left(\frac{y}{\|y\|_2}\right) = x^T \underbrace{\widehat{y}}_{\text{Unit norm}}$

$$x^T y = \|x\|_2 \|y\|_2 \cos\theta$$

$$\cos\theta = \frac{x^T y}{\|x\|_2 \|y\|_2}$$

Vector: $(\|x\|_2 \cos\theta)\,\widehat{y} = (x^T \widehat{y})\,\widehat{y}$

Orthogonal



$x \perp y \iff x^T y = 0$: x and y are said to be orthogonal to each other

Vector Space: Subspace

- A **linear subspace** is a set of vectors that is closed under vector addition and scalar multiplication: if x_1 and x_2 belong to the subspace, then so do $\alpha_1 x_1 + \alpha_2 x_2$.
- A **basis** of the subspace is the maximal set of vectors in the subspace that are linearly independent of each other.
- An **orthogonal basis** is a basis where all basis vectors are *orthogonal* to each other.