CS 316: Introduction to Deep Learning

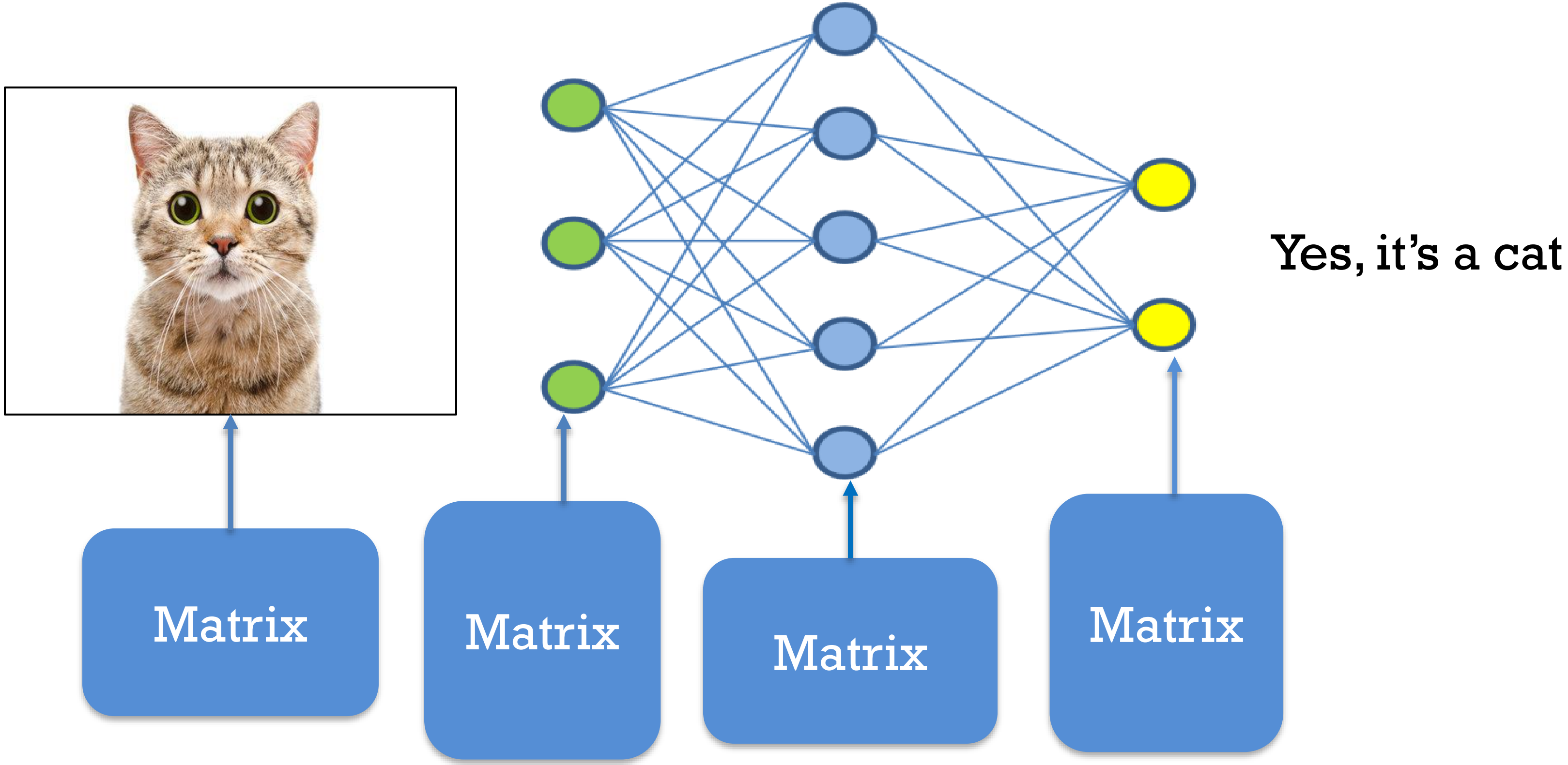
Linear Algebra Recap & Overview
Week 2

Dr Abdul Samad

Lecture Outline

- Motivation
- Notation
- Scalars
- Vectors
- Matrices
- Vector and matrix calculations
- Eigenvectors and eigenvalues
- Additional Topics

Neural Networks – Matrix Operations

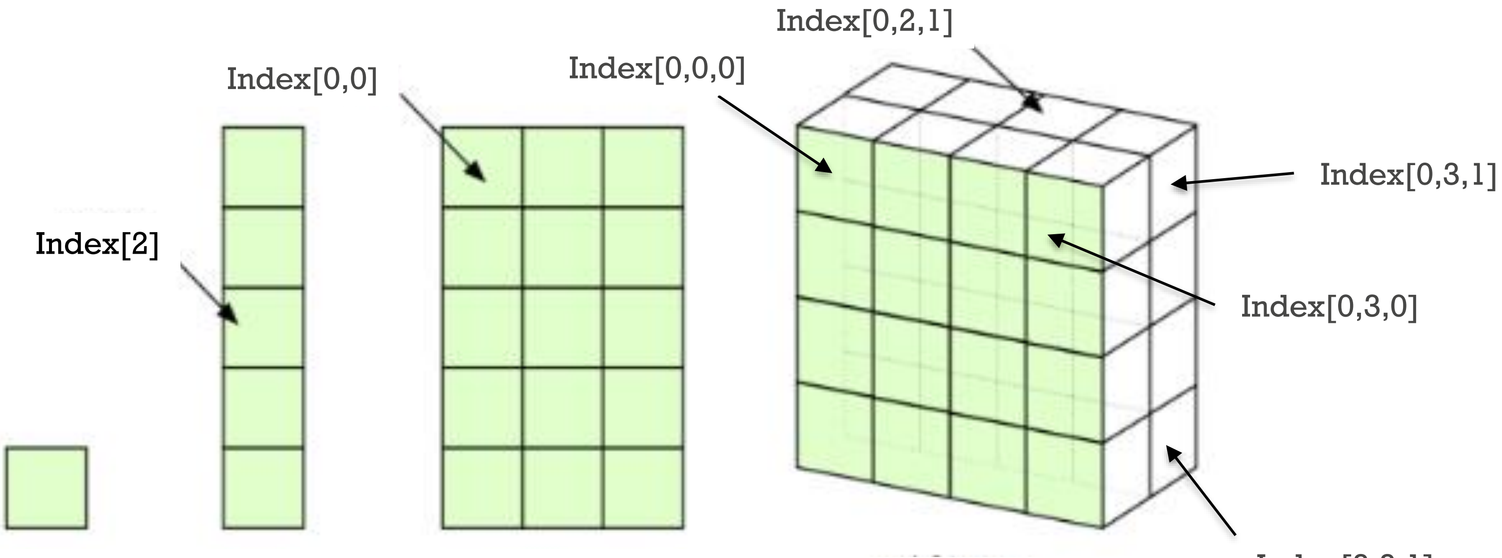


Notation

- Scalar – represented by lower case letter (a)
- Vector – represented by bold lower case letter (**a**)
- Matrix – represented by bold upper case letter (**A**)

This notation will be used throughout the course.

Scalars vs Vectors vs Matrices vs Tensors



rank 0 tensor
dimensions []
Scalar
e.g. Number

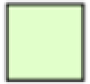
rank 1 tensor
dimensions [5]
Vector
e.g. Data point

rank 2 tensor
dimensions [5,3]
2D - Matrix
e.g. Grey Scale Image

rank 3 tensor
dimensions [4,4,2]
3D - Matrix
e.g. RGB Image

Tensors

Scalar



rank 0 tensor
dimensions []
Scalar
e.g. Number

- Variable described by a single number
- 0^{th} order tensor
- Simple Operations
  - Addition
  - Multiplication
  - Function
- Length

Scalar Operations

$$c = a + b$$

$$c = a \cdot b$$

$$c = \sin a$$

Length

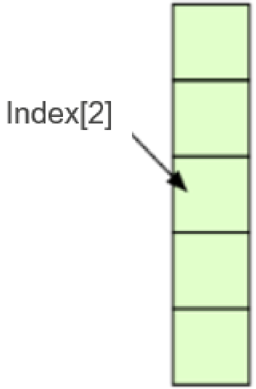
$$|a| = \begin{cases} a & \text{if} a > 0 \\ -a & \text{otherwise} \end{cases}$$

$$|a + b| \leq |a| + |b|$$

$$|a \cdot b| = |a| \cdot |b|$$

Vector

- Variable described by a list of scalars
- 1^{st} order tensor
- Simple Operations
  - Addition
  - Multiplication
  - Function
- Length



rank 1 tensor
dimensions [5]
Vector
e.g. Data point

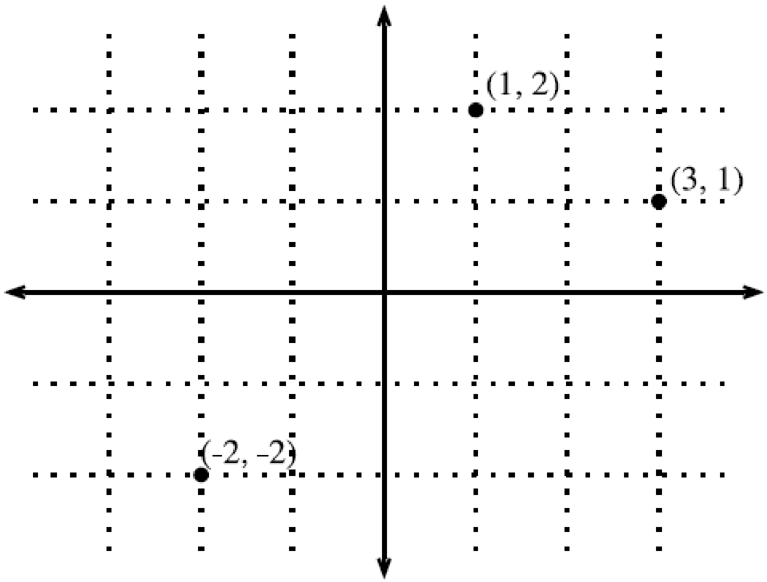
Vector orientation

- Vector can be either written as a row vector or a column vector.

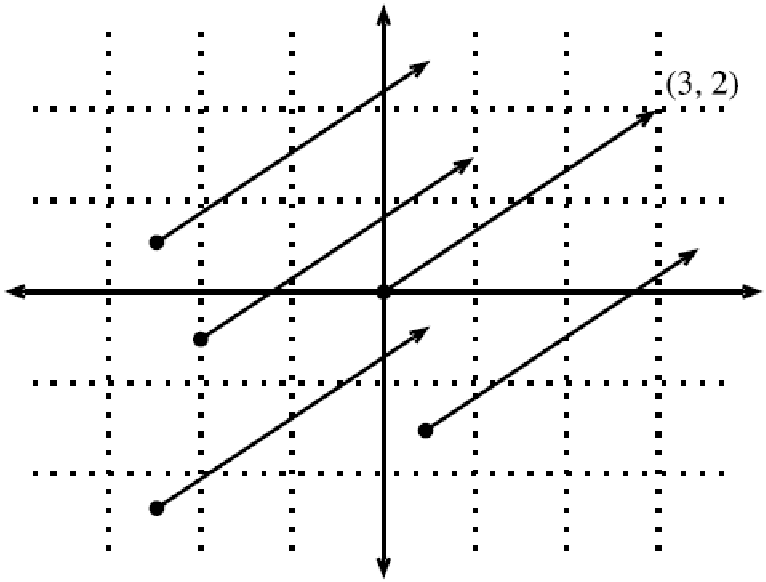
$$\mathbf{x} = \begin{bmatrix} 1 \\ 7 \\ 0 \\ 1 \end{bmatrix},$$

$$\mathbf{x}^\top = \begin{bmatrix} 1 & 7 & 0 & 1 \end{bmatrix}.$$

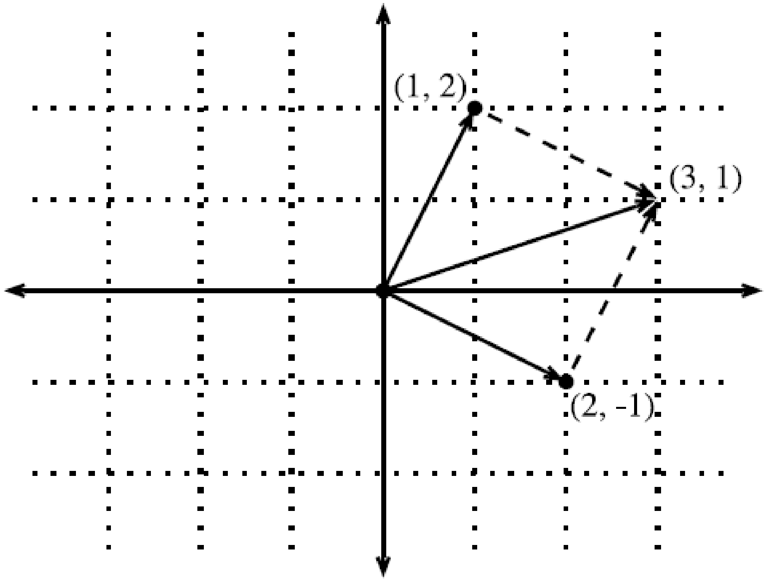
Geometric view – Point in Space



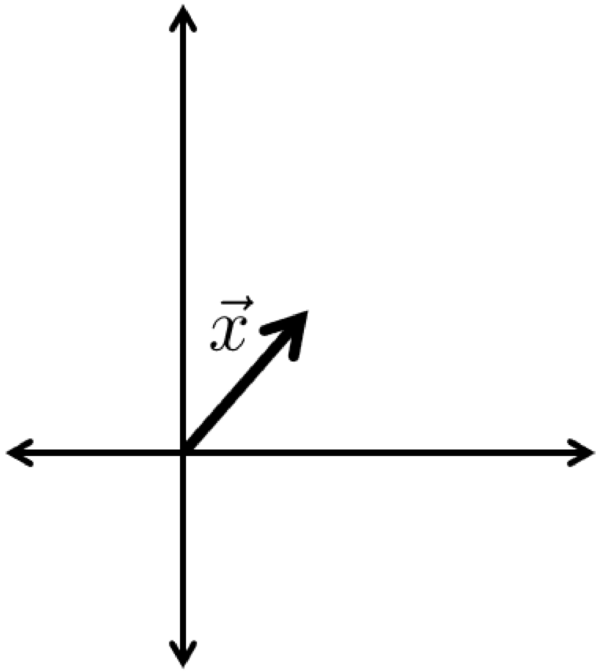
Geometric view – Direction in Space

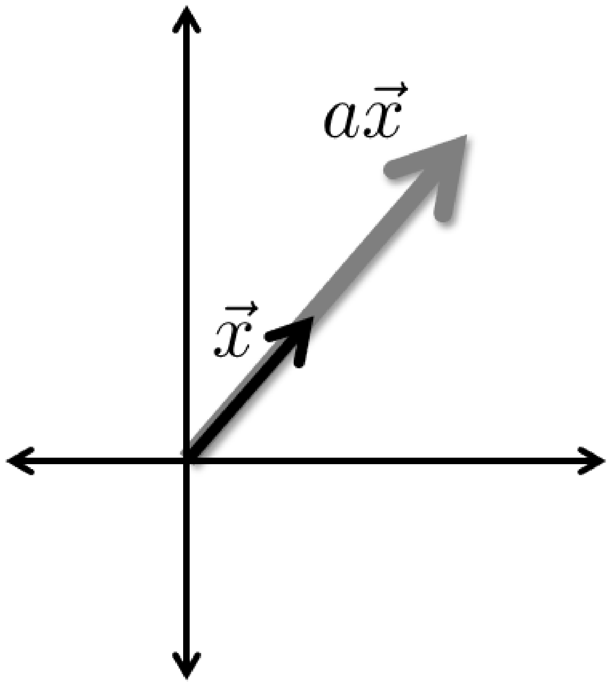


Geometric view - Vector Addition



Geometric view – Scaling a Vector





Vector Operations

$$c = a + b \quad \text{where } c_i = a_i + b_i$$

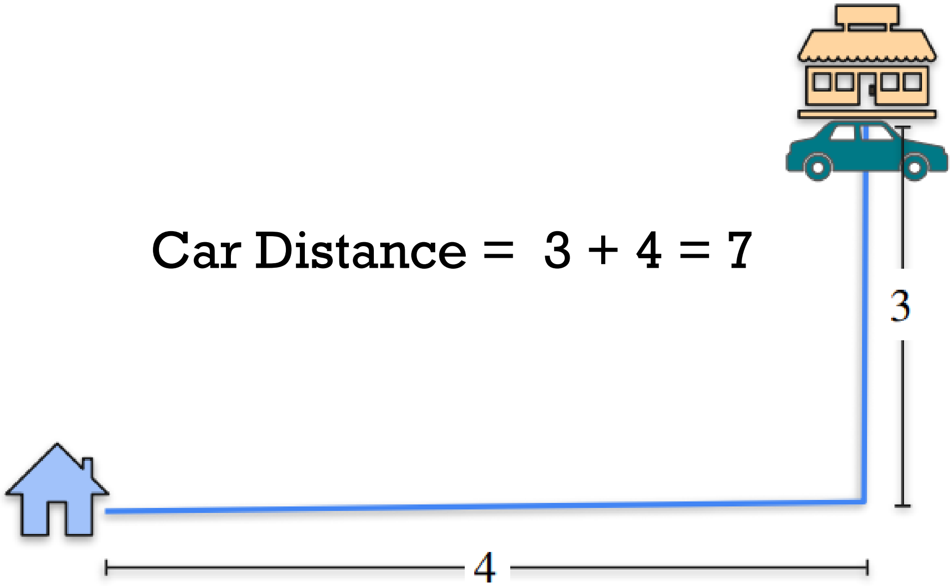
$$c = \alpha \cdot b \quad \text{where } c_i = \alpha b_i$$

$$c = \sin a \quad \text{where } c_i = \sin a_i$$

How to get from point A to point B ?

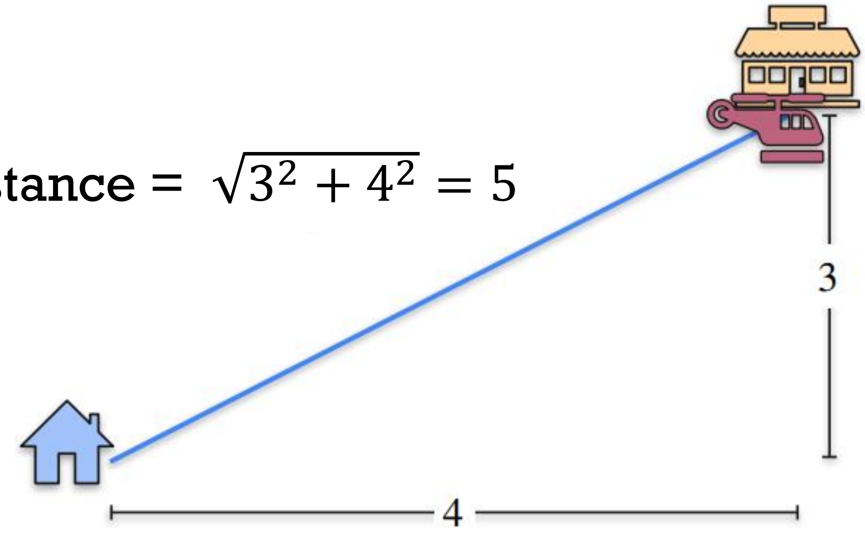
How to get from point A to point B ?

Car Distance = 3 + 4 = 7

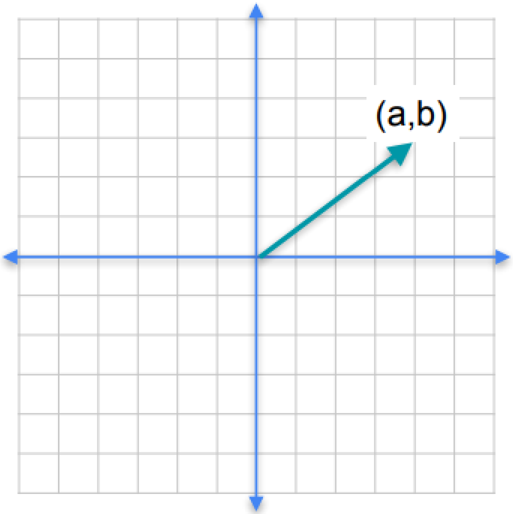


How to get from point A to point B ?

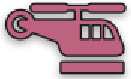
Helicopter Distance = $\sqrt{3^2 + 4^2} = 5$



Norms



L1-norm $= |(a, b)|_1 = |a| + |b|$



L2-norm $= |(a, b)|_2 = \sqrt{a^2 + b^2}$

Vector Norms

$$||a||_2 = \left[\sum_{i=1}^{m} a_i^2\right]^{\frac{1}{2}} \qquad ||\mathbf{a}||_1 = \sum_{i=1}^{m} |a_i|$$

$$||a|| \geq 0 \text{ for all } a$$

$$||a+b|| \leq ||a|| + ||b||$$

$$||a \cdot b|| = |a| \cdot ||b||$$

Vector – Vector Multiplication

- Element-wise multiplication
- Dot Product (Inner Product)

Element-wise multiplication

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} .* \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_2 b_2 \end{pmatrix}$$

Dot Product

$$\mathbf{u}^\top \mathbf{v} = \sum u_i \cdot v_i.$$

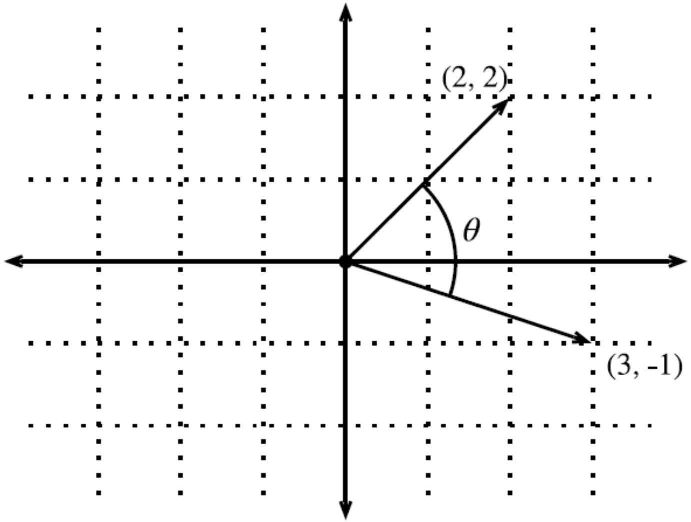
$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \mathbf{v} = \mathbf{v}^\top \mathbf{u},$$

Dot Product – Geometric View

$$\mathbf{u} \cdot \mathbf{v} = \|u\| \, \|v\| \, cos(\theta)$$

$$\theta = arccos(\frac{u \cdot v}{\|u\|\|v\|})$$

Linear Algebra



Geometric View

Example

Compute the dot product for the vectors $u = [-1, 1], v = [1, 1]$

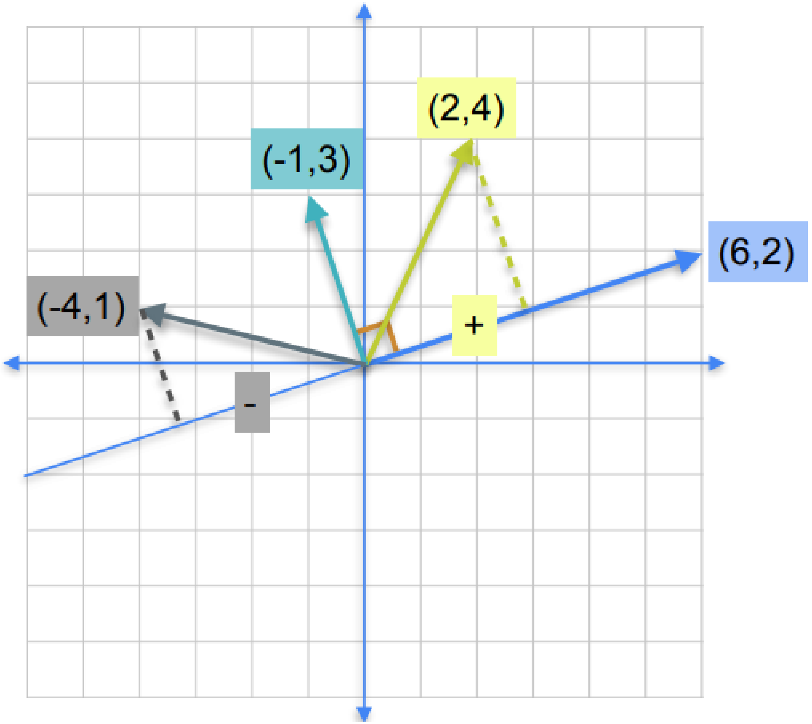
Example

Compute the dot product for the vectors $u = [-1, 1], v = [1, 1]$

$$
\begin{aligned}
u \cdot v &= -1 + 1 \\
u \cdot v &= 0
\end{aligned}
$$

Consequently, the vectors u and v are orthogonal

Visualizing Dot Product

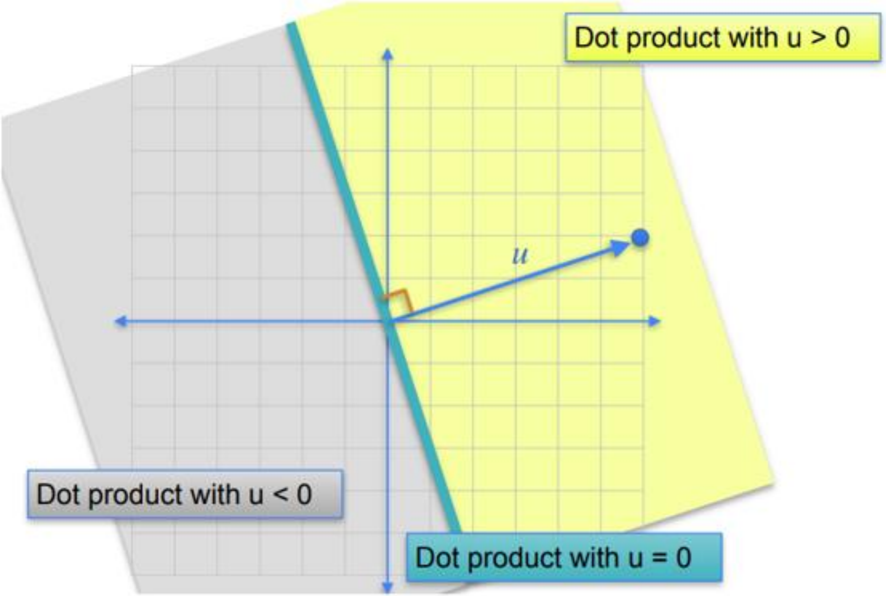


$\begin{bmatrix}6 & 2\end{bmatrix}\begin{bmatrix}2\\4\end{bmatrix} = 20$ Positive

$\begin{bmatrix}6 & 2\end{bmatrix}\begin{bmatrix}-1\\3\end{bmatrix} = 0$

$\begin{bmatrix}6 & 2\end{bmatrix}\begin{bmatrix}-4\\1\end{bmatrix} = -22$ Negative

Visualizing Dot Product



$\langle u, v \rangle > 0$

$\langle u, v \rangle = 0$

$\langle u, v \rangle < 0$

Cosine similarity

Cosine similarity is used to determine the closeness of two vectors.

$$\cos(\theta) = \frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}||\,||\mathbf{w}||}.$$

Example

Compute cosine similarity for the following:

(a) $u = [-1, -1], v = [1, 1]$

(b) $x = [1, 1], y = [2, 2]$

(c) $a = [1, 1], b = [-1, 1]$

Example

Compute cosine similarity for the following:

(a) $u = [-1, -1], v = [1, 1]$

(b) $x = [1, 1], y = [2, 2]$

(c) $a = [1, 1], b = [-1, 1]$

$$cos(\theta) = \frac{u \cdot v}{||u||\,||v||} = -1$$

$$cos(\theta) = \frac{x \cdot y}{||x||\,||y||} = 1$$

$$cos(\theta) = \frac{a \cdot b}{||a||\,||b||} = 0$$

Cosine Similarity - NLP

- Suppose are give two sentences.
  - Sentence 1: Julie loves me more than Linda loves me
  - Sentence 2: Jane likes me more than Julie loves me
- We want to know to how similar these texts are, purely in terms of frequency (and ignoring word order).
- We begin by making a list of the words from both texts:
  - Me Julie loves Linda than more likes Jane.

Word	S1	S2
Me	2	2
Julie	0	1
loves	1	1
Linda	1	0
Than	0	1
More	2	1
Likes	1	1
Jane	1	1

Cosine Similarity - NLP

- Consequently, we have two vectors:

$$\mathbf{a} = [2, 1, 0, 2, 0, 1, 1, 1]$$

$$\mathbf{b} = [2, 1, 1, 1, 1, 0, 1, 1]$$

- Now we can compute the cosine similarity between the two vectors.

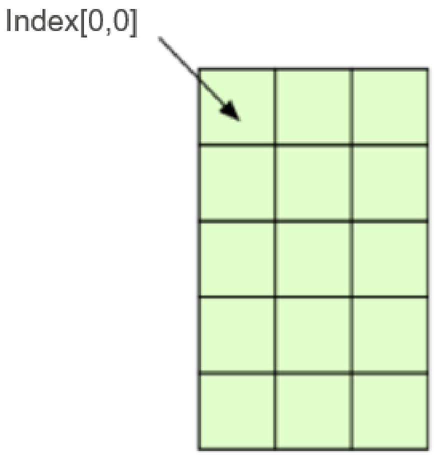
$$\cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{||a||||b||}$$

$$\cos(\theta) = \frac{4 + 1 + 0 + 2 + 0 + 0 + 1 + 1}{\sqrt{12} \times \sqrt{9}}$$

$$\cos(\theta) = \frac{9}{\sqrt{108}} = 0.866$$

Matrix

- Variable described by a list of vectors.
- 2^{nd} order tensor
- Simple Operations
  - Addition
  - Multiplication
  - Function
- Length



rank 2 tensor
dimensions [5,3]
Matrix
e.g. Grey Scale Image

Matrix Operations

$$C = A + B \qquad \text{where } C_{ij} = A_{ij} + B_{ij}$$

$$C = \alpha \cdot B \qquad \text{where } C_{ij} = \alpha B_{ij}$$

$$C = \sin A \qquad \text{where } C_{ij} = \sin A_{ij}$$

Matrix Norm

$$\|A\|_{\mathrm{Frob}} = \left[\sum_{ij} A_{ij}^2\right]^{\frac{1}{2}}$$

Matrix – Vector Multiplication

$$\overrightarrow{y} = \overleftrightarrow{W}\overrightarrow{x}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} W_{11} & W_{12} & \cdots & W_{1N} \\ W_{21} & W_{22} & \cdots & W_{2N} \\ \vdots & \vdots & & \vdots \\ W_{M1} & W_{M2} & \cdots & W_{MN} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$$

M X 1   M X N   N X 1

The i^{th} element of **y** is the dot product of the i^{th} row of W with **x**

Linear transformation

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad v = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{aligned} \mathbf{Av} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} \\ &= x \begin{bmatrix} a \\ c \end{bmatrix} + y \begin{bmatrix} b \\ d \end{bmatrix} \\ &= x \left\{ \mathbf{A} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\} + y \left\{ \mathbf{A} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}. \end{aligned}$$

Example

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

Based on the equation derived, on the previous slide compute Av

Example

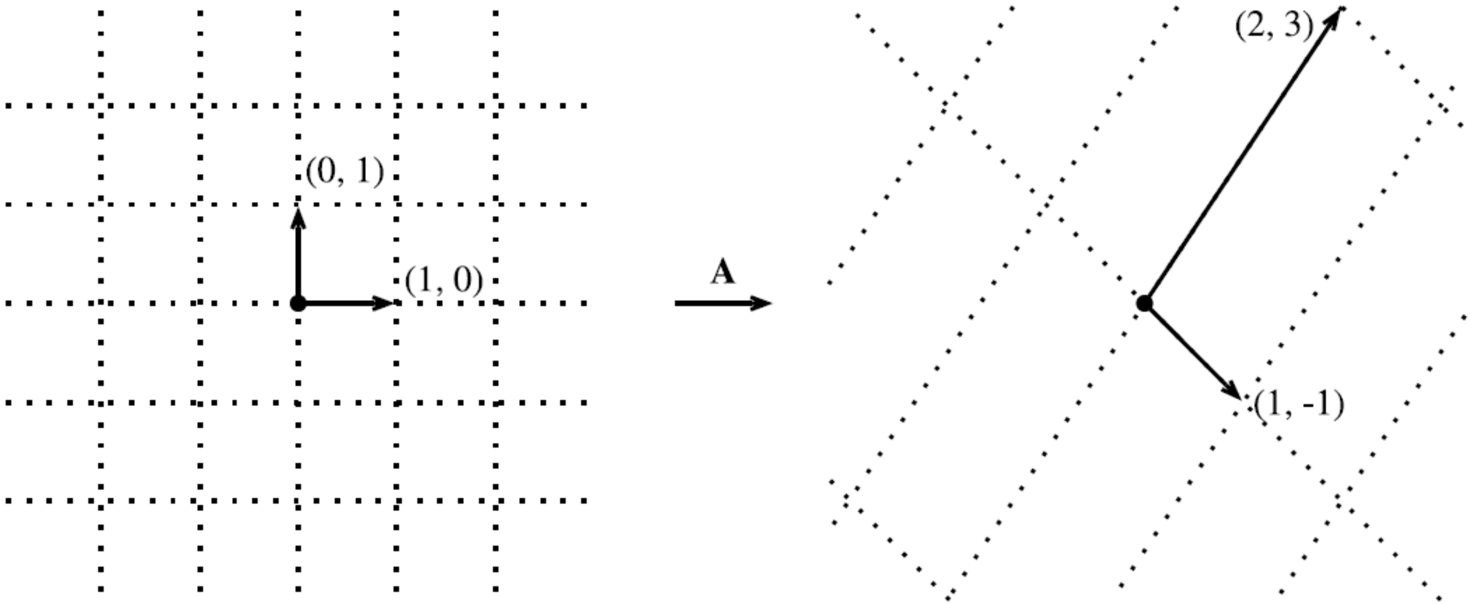
$$A = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} \qquad v = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

Based on the equation derived, on the previous slide compute Av

$$A([0,1]^T) = [1,-1]^T$$
$$A([1,0]^T) = [2,3]^T$$
$$Av = 2 \cdot A([1,0]^T) - 1 \cdot A([0,1]^T)$$
$$Av = 2 \cdot [1,-1]^T - 1 \cdot [2,3]^T = [0,5]^T$$

Example – Geometric view

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



Matrix – Matrix Multiplication

$$C_{ij} = \sum_{k=1}^{P} A_{ik} B_{kj}$$

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1P} \\ A_{21} & A_{22} & \cdots & A_{2P} \\ \vdots & \vdots & & \vdots \\ A_{i1} & A_{i2} & \cdots & A_{iP} \\ \vdots & \vdots & & \vdots \\ A_{N1} & A_{N2} & \cdots & A_{NP} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1j} & \cdots & B_{1M} \\ B_{21} & B_{22} & \cdots & B_{2j} & \cdots & B_{2M} \\ \vdots & \vdots & & \vdots & & \vdots \\ B_{P1} & B_{P2} & \cdots & B_{Pj} & \cdots & B_{PM} \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1M} \\ C_{21} & C_{22} & \cdots & C_{2M} \\ \vdots & \vdots & C_{ij} & \vdots \\ C_{N1} & C_{N2} & \cdots & C_{NM} \end{pmatrix}$$

N X P **P X M** **N X M**

C_{ij} is the inner product of the i^{th} row of **A** with the j^{th} column of **B**

Special Matrices - Identity

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

For all A, $AI = IA = A$

Special Matrices - Diagonal

$$\overleftrightarrow{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix} \qquad \overleftrightarrow{D}\,\overrightarrow{x} = \begin{pmatrix} d_1 x_1 \\ d_2 x_2 \\ \vdots \\ d_n x_n \end{pmatrix}$$

Compressed Notation

- $\mathbf{v} \cdot \mathbf{w} = \sum_i v_i w_i$
- $\|\mathbf{v}\|_2^2 = \sum_i v_i v_i$
- $(\mathbf{Av})_i = \sum_j a_{ij} v_j$
- $(\mathbf{AB})_{ik} = \sum_j a_{ij} b_{jk}$
- $\mathrm{tr}(\mathbf{A}) = \sum_i a_{ii}$

Example

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

Compute Av

Example

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

Compute Av

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

v is a special vector because multiplying by A is like a scalar. v is an **eigenvector** of A, and the scaling factor is called the **eigenvalue** associated with the eigen vector.

Eigenvector & eigenvalues

In general, if we can find a number λ and a vector v such that
$Av = \lambda v$
We say that v is an eigenvector for A and λ is an eigenvalue

Additional Topics

Outer product

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \begin{pmatrix} y_1 & y_2 & \cdots & y_M \end{pmatrix} = \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_M \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_M \\ \vdots & \vdots & \ddots & \cdots \\ x_N y_1 & x_N y_2 & \cdots & x_N y_M \end{pmatrix}$$

N X 1   1 X M   N X M

Determinant

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

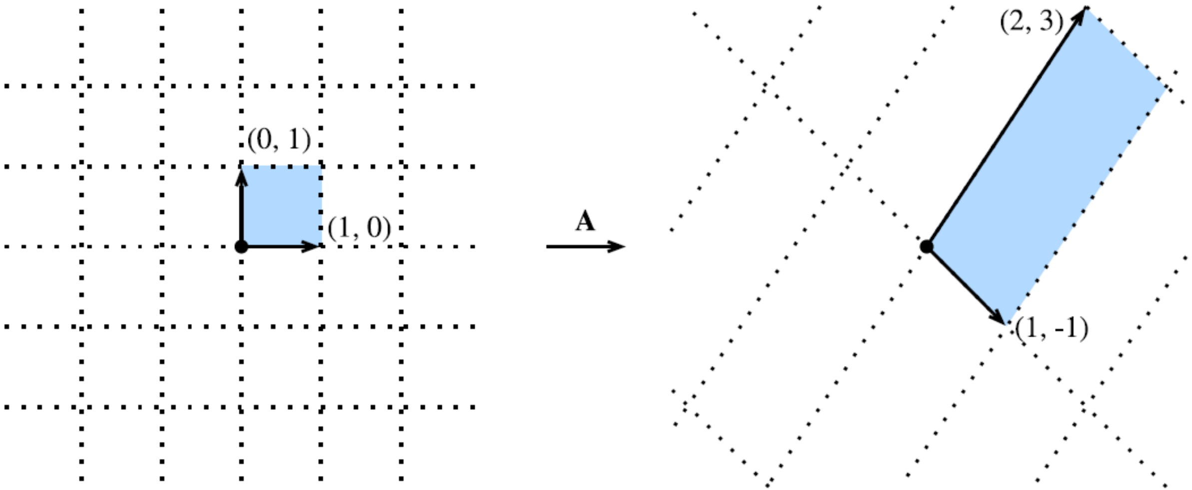
$$det\mathbf{A} = ad - bc$$

Geometric view - determinant

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

$$\text{Determinant of A} = \text{Area of parallelogram}$$

$$det\mathbf{A} = ad - bc = 5$$



Special Matrices - Inverse

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}.$$

If A is a general 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$,

then the inverse is $\dfrac{1}{ab - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$