Advanced Machine Learning and Deep Learning

Dr. Mohammad Pourhomayoun

Assistant Professor

Computer Science Department

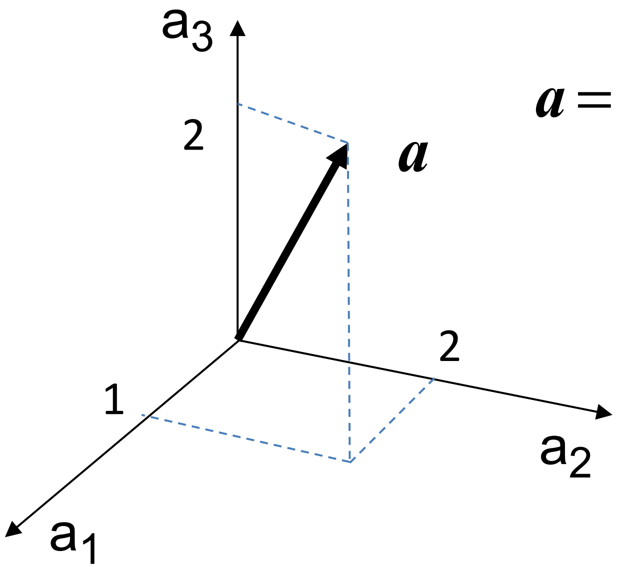
California State University, Los Angeles

Continue:
Matrices and Vectors Review

Vector Norm

- **Vector Norm2 (or simply called Vector Norm or Vector Length)**:

$$\boxed{\|\boldsymbol{v}\| = \sqrt{\sum_{i=1}^{N} v_i^2}} \qquad \|\boldsymbol{v}\|^2 = \sum_{i=1}^{N} v_i^2$$



$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

Vector Length (norm2) $= \sqrt{1^2 + 2^2 + 2^2} = 3$

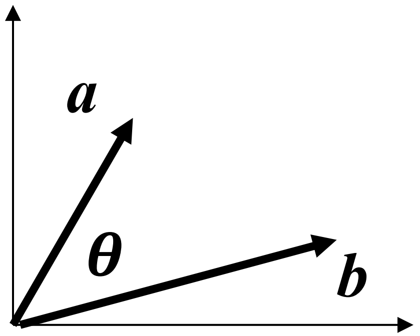
Inner Product (Dot Product)

- **Inner Product Between Vectors :** Define the inner product between two N-dimensional vectors by:

$$\boldsymbol{a}.\boldsymbol{b} = \sum_{i=1}^{N} a_i\, b_i$$

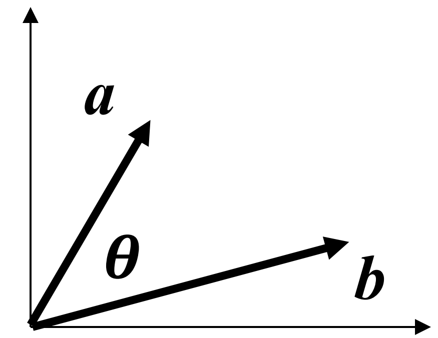
$$= \left\| \boldsymbol{a} \right\| \times \left\| \boldsymbol{b} \right\| \times \mathrm{Cos}\,\theta$$

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix}$$



Inner Product (Dot Product)

$$\boldsymbol{a}.\boldsymbol{b} = \sum_{i=1}^{N} a_i\, b_i \;= \|\boldsymbol{a}\| \times \|\boldsymbol{b}\| \times \mathrm{Cos}\,\theta$$



$$\mathrm{Cos}\,\theta = \frac{\boldsymbol{a}.\boldsymbol{b}}{\|\boldsymbol{a}\| \times \|\boldsymbol{b}\|} = \frac{\sum_{i=1}^{N} a_i\, b_i}{\|\boldsymbol{a}\| \times \|\boldsymbol{b}\|}$$

- (i) This lies between -1 and 1.
- (ii) It measures directional <u>SIMILARITY</u> of ***a*** and ***b***

  = +1 when ***a*** and ***b*** point to the same direction.

  = 0 when ***a*** and ***b*** are a "right angle".

  = -1 when ***a*** and ***b*** point to opposite directions.

Inner Product (Dot Product)

- **Example:**

$$\boldsymbol{a} = \begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix} \Rightarrow$$

$$\|\boldsymbol{a}\| = \sqrt{3^2 + (-1)^2 + 4^2} = \sqrt{26}$$

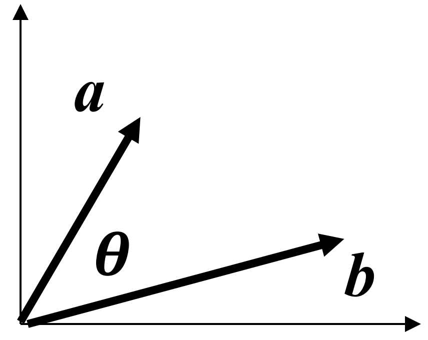
$$\|\boldsymbol{b}\| = \sqrt{1^2 + 5^2 + 2^2} = \sqrt{30}$$

$$\text{Cos}\,\theta = \frac{\boldsymbol{a}.\boldsymbol{b}}{\|\boldsymbol{a}\| \times \|\boldsymbol{b}\|} = \frac{\sum_{i=1}^{N} a_i b_i}{\|\boldsymbol{a}\| \times \|\boldsymbol{b}\|} = \frac{3\times 1 + (-1)\times 5 + 4\times 2}{\sqrt{26}\times\sqrt{30}} \approx 0.21$$

$$\Rightarrow \quad \theta = Cos^{-1}(0.21) \approx 78^o$$

Orthogonal and Orthonormal

$$\boldsymbol{a}.\boldsymbol{b} = \sum_{i=1}^{N} a_i\, b_i = \|\boldsymbol{a}\| \times \|\boldsymbol{b}\| \times \mathrm{Cos}\,\theta$$

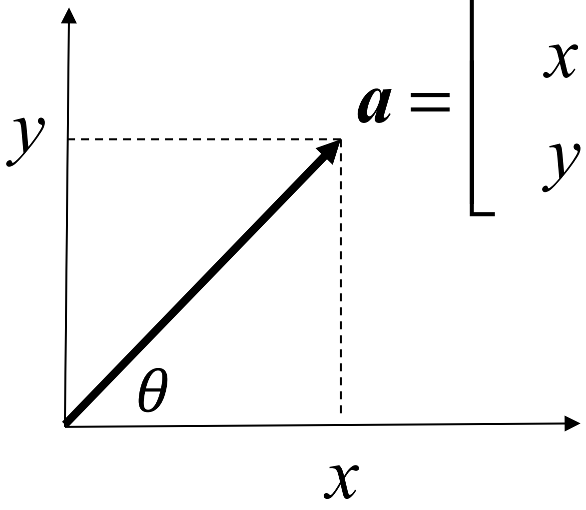


$$\Rightarrow \quad \mathrm{Cos}\,\theta = \frac{\boldsymbol{a}.\boldsymbol{b}}{\|\boldsymbol{a}\| \times \|\boldsymbol{b}\|} = \frac{\sum_{i=1}^{N} a_i\, b_i}{\|\boldsymbol{a}\| \times \|\boldsymbol{b}\|}$$

- Two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are called **orthogonal** when $\boldsymbol{a}.\boldsymbol{b} = 0$
  - The angle between them is 90° (perpendicular).
- Two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are called **orthonormal** when $\boldsymbol{a}.\boldsymbol{b} = 0$ and each has unit length (Norm =1).

Vector Projection on Coordinate Axis

- Vector Projection on Coordinate Axis:



$$\boldsymbol{a} = \begin{bmatrix} x \\ y \end{bmatrix}$$

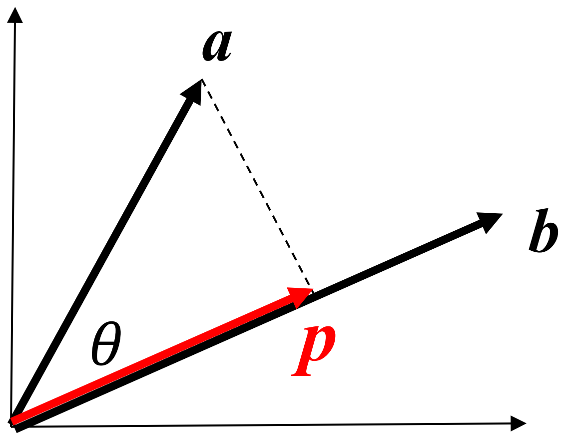
$$x = \|\boldsymbol{a}\| \operatorname{Cos}\theta$$

$$y = \|\boldsymbol{a}\| \operatorname{Sin}\theta$$

$$\|\boldsymbol{a}\| = \sqrt{x^2 + y^2}$$

Vector Projection on Another Vector

- A vector can be completely determined by its ***length*** and ***direction***
- Thus, to find $\boldsymbol{p}$, we have to find its length and direction.



Length: $\|\boldsymbol{p}\| = \|\boldsymbol{a}\| \,\mathrm{Cos}\,\theta$

Direction: $\dfrac{\boldsymbol{b}}{\|\boldsymbol{b}\|}$

$$\boldsymbol{p} = \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}\|\boldsymbol{a}\|\,\mathrm{Cos}\,\theta = \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|^2}\|\boldsymbol{a}\|\,\|\boldsymbol{b}\|\,\mathrm{Cos}\,\theta$$

$$\boxed{\boldsymbol{p} = \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|^2}(\boldsymbol{a}.\boldsymbol{b})}$$

Building Vectors From Other Vectors

- Can we find a set of "prototype" vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_N\}$ from which we can build ***all other vectors*** in some given vector space by using linear combinations of the $\boldsymbol{v}_i$?
- e.g: for any vector $\boldsymbol{a}$ and any vector $\boldsymbol{b}$, there are some scalar numbers α_i's and β_i's such that we can write the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ as linear combinations of vector set $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_N\}$:

$$\boldsymbol{a} = \sum_{i=1}^{N} \alpha_i \boldsymbol{v}_i \qquad \boldsymbol{b} = \sum_{i=1}^{N} \beta_i \boldsymbol{v}_i$$

- Same "***Ingredients***", just different amounts of them!!!

Building Vectors From Other Vectors

- **Span of a Set of Vectors:** A set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_\mathrm{N}\}$ is said to span the vector space V if it is possible to write every vector $\boldsymbol{a}$ in V as a linear combination of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_\mathrm{N}\}$:
- Thus, for any arbitrary vector $\boldsymbol{a}$, there are some numbers α_i's that:

$$\boldsymbol{a} = \sum_{i=1}^{N} \alpha_i \boldsymbol{v}_i$$

- **Note**: We would also prefer to have the smallest prototype set. In other word, we are looking for the smallest set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_\mathrm{N}\}$ (smallest N) that can build every other vector in the space.

Building Vectors From Other Vectors

- **Example**: it is possible to build any 2-D vector using the following two vectors. Thus, $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ span the 2-D space:

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Example:

$$\boldsymbol{a} = \begin{bmatrix} 5 \\ 6 \end{bmatrix} = (-1) \times \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 6 \times \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\Rightarrow \quad \boldsymbol{a} = -1\boldsymbol{v}_1 + 6\boldsymbol{v}_2$$

Linear Independence

- **Linear Independence:** A set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N\}$ is said to be **linearly independent** if none of them can be made as a linear combination of the others.
- If a set of vectors is NOT linearly independent, then we have "***redundancy***" in the set (the set has more vectors than needed to be a prototype set!)
- For example, if we have a set of four vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ and we know that we can build $\boldsymbol{v}_2$ from $\boldsymbol{v}_1$ and $\boldsymbol{v}_3$. Then, every vector that we can build from $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ can also be built from only $\{\boldsymbol{v}_1, \boldsymbol{v}_3, \boldsymbol{v}_4\}$.
- **Note: Any set of mutually orthogonal vectors is linear independent.**

Building Vectors From Other Vectors

- **Example**: it is possible to build any 2-D vector in terms of the following three vectors:

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \boldsymbol{v}_3 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

- However, $\boldsymbol{v}_3 = 2\,\boldsymbol{v}_1 + \boldsymbol{v}_2$. Thus, they are not ***linear independent***. In other word, we have **redundancy** and one of them is unnecessary!
- Thus, rather than building a vector in terms of three of them, we can use only two of them to build any arbitrary 2-D vector!

Basis of a Vector Space

- **Basis of a Vector Space:** A basis of a vector space is a set of **linear independent** vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N\}$ that **span** the space.
  - "Span" says there are enough vectors to build everything.
  - "Linear Independence" says there are not more than needed.
- **NOTE:** After writing a vector $\boldsymbol{a}$ in terms of basis vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N\}$, then from now on, we can represent vector $\boldsymbol{a}$ only using numbers $(\alpha_1, \alpha_2, \ldots, \alpha_N)$. This is called **"Decomposition"**:

$$\boldsymbol{a} = \sum_{i=1}^{N} \alpha_i \boldsymbol{v}_i$$

It is like a ***recipe*** for making a cake! To make a cake, I do not need to take the ingredients everywhere with myself. The ingredients are standard and I can find them everywhere. I just need to keep the recipe!

Orthonormal (ON) Basis

- **Orthonormal (ON) Basis:** If a basis of a "vector space" contains vectors that are **orthonormal** to each other (all pairs of basis vectors are orthogonal and each basis vector has unit norm), we call it Orthonormal Basis.

- **NOTE:** Often, we prefer to have **Orthonormal Basis** because:
  - **It guarantees that they are linear independent, so NO REDUNDANCY!**
  - **Easy computation via inner products.**
  - **It is possible to show that ON basis preserves the geometry (closeness, size, orientation, etc.)**

Orthonormal (ON) Basis

- **A very simple example of Orthonormal (ON) Basis in R^3 (3-D):**

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad \boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- **NOTE:** In many cases, the key to solving a ***data science*** problem or ***big data processing*** problem lies in finding the correct basis. this is equivalent to finding the right data transform.

Symmetric Matrix

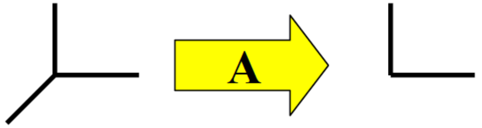
- Matrix $\boldsymbol{A}$ is ***symmetric*** if $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$.
- In other word, a matrix is symmetric if for every element $\boldsymbol{a_{ij}}$, we have $\boldsymbol{a_{ij}} = \boldsymbol{a_{ji}}$.
- *Example:*

$$\boldsymbol{A} = \begin{bmatrix} 1 & 7 & -5 \\ 7 & 3 & 21 \\ -5 & 21 & 45 \end{bmatrix}$$

Matrix to Vector Multiplication

- **<u>Matrix as Transform:</u>** Our main view of matrices will be as "***operators***" that transform one vector into another vector.
- Consider a 2x3 matrix below. We can use that matrix to ***transform*** the ***3-dimensional*** vector $\boldsymbol{x}$ into a ***2-dimensional*** vector $\boldsymbol{y}$:

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 3 & -2 & 5 \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$



$$\boldsymbol{y} = A\boldsymbol{x} = \begin{bmatrix} 1 & 2 & -1 \\ 3 & -2 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix} = \begin{bmatrix} -2 \\ 33 \end{bmatrix}$$

Matrix to Vector Multiplication

- **Matrix as Transform:** Our main view of matrices will be as "***operators***" that transform one vector into another vector.
- In general:

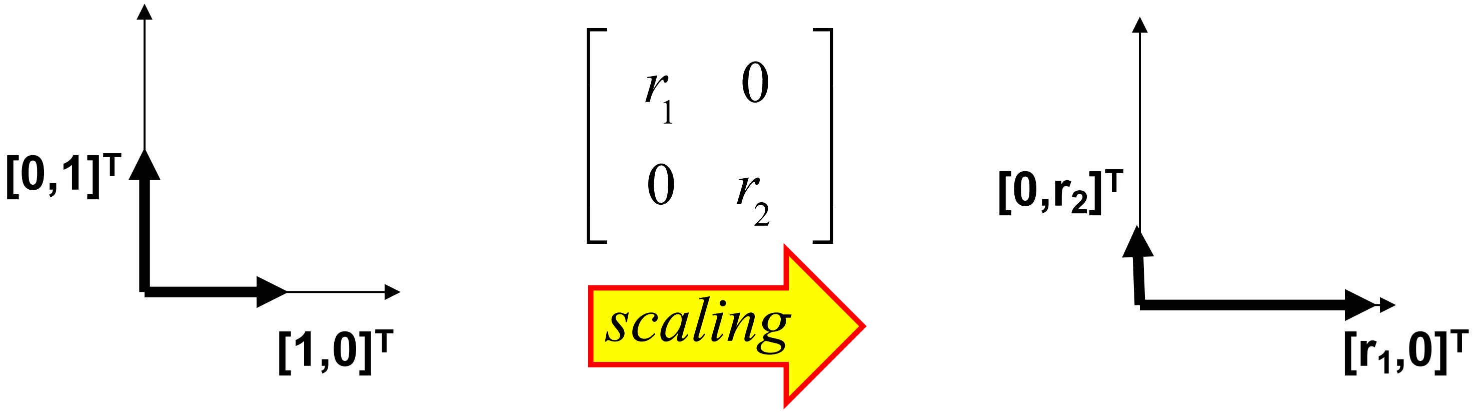
$$\underbrace{\boldsymbol{y}}_{m\times 1} = \underbrace{\boldsymbol{A}}_{m\times n}\underbrace{\boldsymbol{x}}_{n\times 1}$$

- **Special Case:** If the mapping matrix ***A*** is ***square,*** then the space that vectors get mapped to has the same dimension.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \Rightarrow \quad \begin{aligned} y_1 &= ax_1 + bx_2 \\ y_2 &= cx_1 + dx_2 \end{aligned}$$

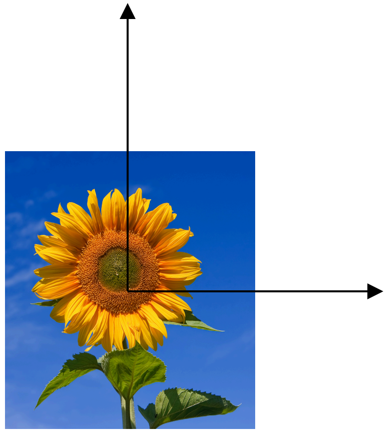
Transform Example: Scaling

- ***Scaling Transform*** is a **diagonal** matrix with scaling ratios on the main diagonal and zero everywhere else.
- note: x-axis and y-axis could be scaled with different ratios.



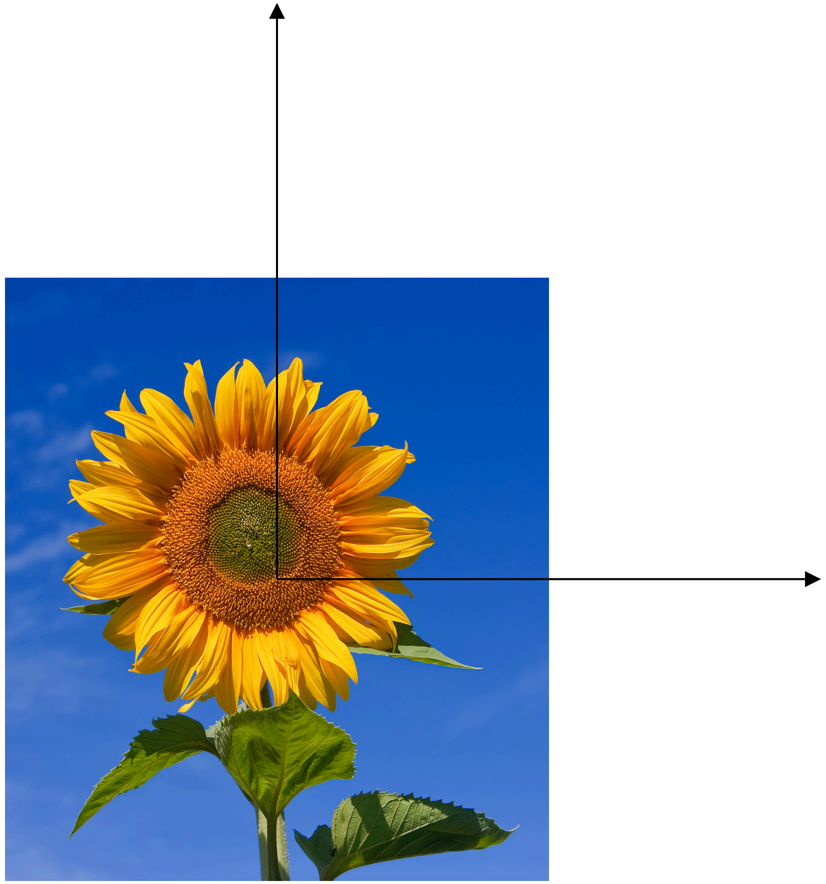
Transform Example: Scaling

- r > 1: dilation
- r < 1: contraction



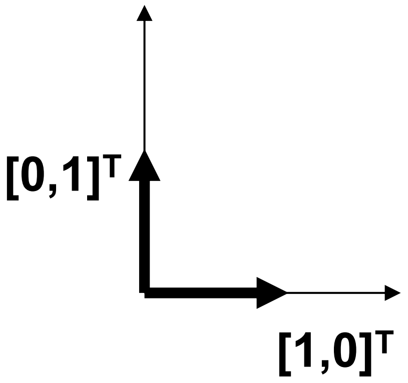
$$\begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix}$$

scaling

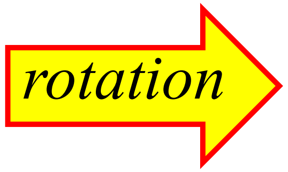


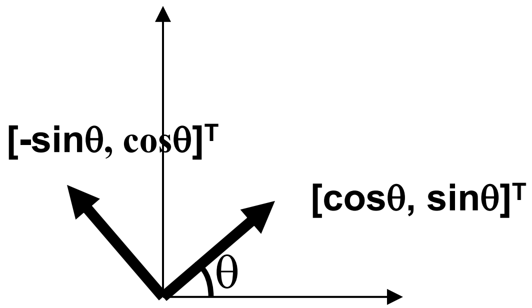
Transform Example: Rotation

- ***Rotation Transform*** is an orthogonal matrix with the following format:

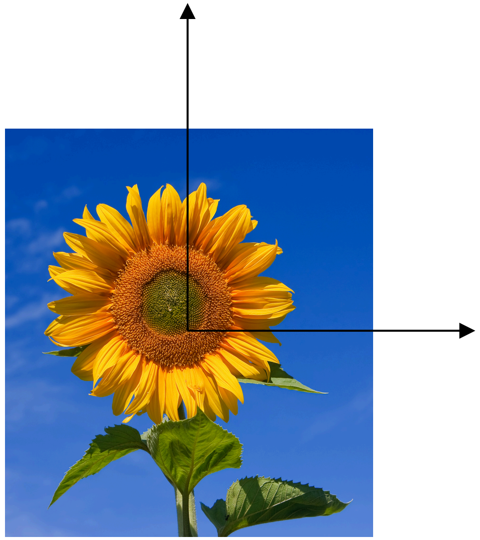


$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$



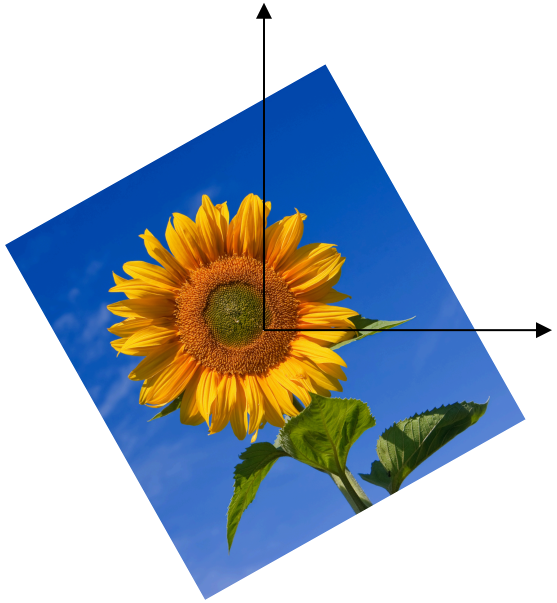


Transform Example: Rotation



$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

rotation



Eigenvalues and Eigenvectors

- **Question:** For a given nxn matrix A, which vectors get mapped into being almost themselves???
- More precisely, which vectors get mapped to a scalar multiple of themselves???
- Even more precisely, which vectors v satisfy the following:

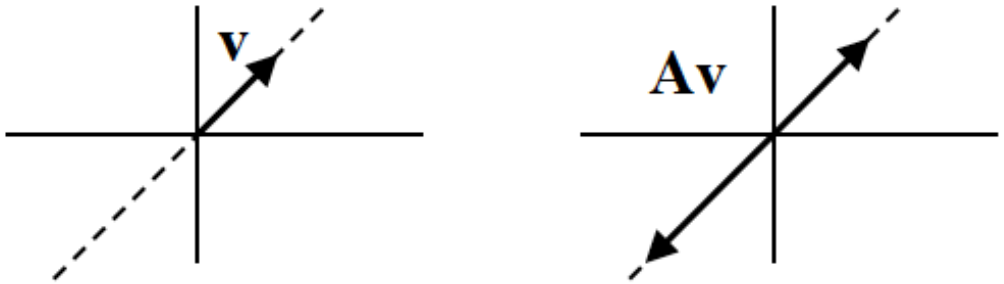
$$\boxed{Av = \lambda v}$$

- These vectors v are "special" and are called the ***eigen-vectors*** of A.
- The scalars λ is the corresponding ***eigen-values***.

Eigenvalues and Eigenvectors

$$\boxed{Av = \lambda v}$$

- These vectors $\boldsymbol{v}$ are called the ***eigenvectors*** of ***A***.
- The scalars λ is the corresponding ***eigenvalues***.
- **In other word, Transformation using matrix *A* on its eigenvectors acts just like Multiplying these vectors by a Scalar!**



Eigenvalues and Eigenvectors

- **Special Case:** If square matrix A is ***symmetric***, then
  - eigenvectors corresponding to distinct eigenvalues are orthonormal.
  - eigenvalues are real valued.
  - We can decompose A as:

$$\boxed{\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T}$$

  - where $\mathbf{V}$ is a matrix with $\boldsymbol{A}$'s ***eigenvectors as columns***, and $\boldsymbol{\Lambda}$ is a diagonal matrix with ***eigenvalues on main diagonal*** and 0 everywhere else.

- **We will use eigenvalues/eigenvectors later in "Dimensionality Reduction".**

Thank You!

Questions?