Artificial Intelligence Algorithms and Mathematics

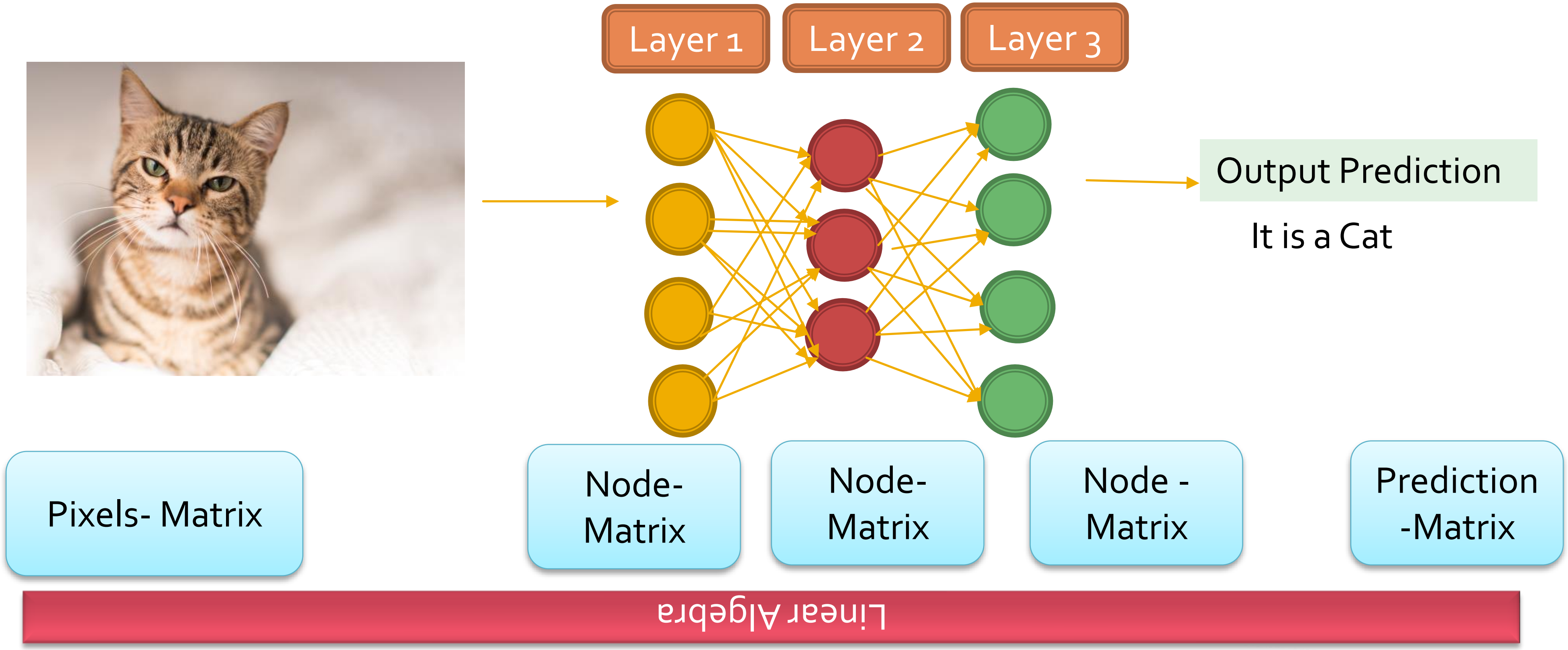
CSCN 8000

Linear Algebra

- Vectors
- Matrix
  - Inverse
  - Rank
- Linear Transformation
- Eigenvalues
- Eigenvectors
- These are used in machine learning to store and compute on data.

Motivation to Machine Learning

- Linear Algebra most useful in machine learning
- Most popular application –Neural Networks –image recognition –using matrix operations.



Recap: Singular vs Non-Singular

- Linear dependence between rows – Second row is multiple of the first row. Here they are linearly dependent
  - For Example : x+y = 0

    2x+2y=0

1	1	First Row
2	2	Second Row/multiple of the first row

- Otherwise, they are referred to be linearly independent
  - For Example : x+y = 0

    x+2y=0

1	1	First Row
1	2	Second Row/not a multiple of the first row

Rank of Matrix

- The rank of a matrix is the maximum number of linearly independent rows (or columns) in the matrix.
- A matrix is considered to be of **full rank** if none of its columns (or rows) can be expressed as a linear combination of the others. This implies linear independence. **[Non-Singular]**
- A matrix is **non-full rank** if its columns (or rows) are linearly dependent. In this case, at least one column (or row) can be expressed as a linear combination of the others. **[Singular]**

Solving System of Linear Equations

- To solve the linear equations, we could get to a solved system
  - Multiplying by a constant [Scaling]
  - Adding two equations [Elimination]
  - Subtracting the equations [Elimination]
  - Substitution with values
  - For example :
    - $Eq\ 1:\ 4x + y = \ 16$
    - $Eq\ 2: x\ - \frac{1}{2}y = 4$
    - Multiply Eq 2 by 2 → $Eq\ 2: 2x\ - y = 8$
    - Add Eq 1 & Eq2 → $6x\ = 24\ \ \rightarrow x = 4$
    - Substitute x in Eq1 → $(4 * 4) + y = \ 16 \rightarrow y = 0$

Matrix Row Reduction –Gaussian Elimination

- Apply the operations to convert the matrix into a much *simpler* representation and solve the system.

System (Actual)

- 5x+y =17
- 4x-3y=6

5	1	17
4	-3	6

Scaling

$R1 = \frac{1}{5} R1$

- $1x + \frac{1}{5}y = \frac{17}{5}$
- 4x-3y=6

1	1/5	17/5
4	-3	6

Addition

$R2 = R2 - 4R1$

- $1x + \frac{1}{5}y = \frac{17}{5}$
- 0x-3.8y=-7.6

1	1/5	17/5
0	-3.8	-7.6

Scaling

$R2 = \frac{1}{-3.8} R2$

- $1x + \frac{1}{5}y = \frac{17}{5}$
- 0x+1y=2

1	1/5	17/5
0	1	2

Row Echelon Form

Substitution

$R2 \rightarrow R1$

- $1x + \frac{1}{5}y = \frac{17}{5}$
- 0x+1y=2

1	1/5	17/5
0	1	2

Example

- Actual System
  - $5x+y=11$
  - $10x+2y=22$

Manipulated system

$x+0.2y=2.2$

$0x+0y=0$

- Matrix form

5	1
10	2

Upper Diagonal Matrix

1	0.2
0	0

Row Echelon Matrix

Row Echelon Form

- The Row Echelon Form (REF), also known as row-reduced form, is a specific arrangement of a matrix that simplifies solving systems of linear equations. A matrix is in row echelon form if it satisfies the following conditions:
  - Zero Rows: Any row containing only zeros is at the bottom.
  - Leading Entry: In each nonzero row, the leftmost nonzero entry is 1, and the element directly below this leading 1 is zero.
  - Leading 1s: The leading 1 in the second row (if it exists) is to the right of the leading 1 in the first row. The leading 1 in the third row (if it exists) is to the right of the leading 1 in the second row, and so on.

A

1	2	5	2
0	1	3	3
0	0	0	1
0	0	0	0

B

1	2	5	2
0	1	3	3
0	0	1	0
0	0	1	1

Rank of Matrix

- Number of pivot elements –Rank of matrix
- The rank of the matrix is the number of non-zero rows in the row echelon form. To find the rank, we need to perform the following steps:
  - Find the row-echelon form of the given matrix
  - Count the number of non-zero rows.

Trace of a Matrix

- The trace of a matrix, typically denoted as $tr(A)$ for a matrix A, is the sum of the elements on the main diagonal of the matrix.
- $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$
- $tr(A) = 1 + 4 = 5$
- Applications:
  - Often used in optimization problems
  - Understanding the properties of certain matrices like covariance matrices

Which Matrices have inverse?

- An invertible matrix is one that is:
  - *Full Rank*
  - *Has Determinant not equal to zero*
  - *All rows and columns are linearly independent*
  - *Non-Singular*
  - *Square Matrix*
- For example, for the following 2x2 matrix:
  - $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$
  - $A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$
  - $A\,A^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$

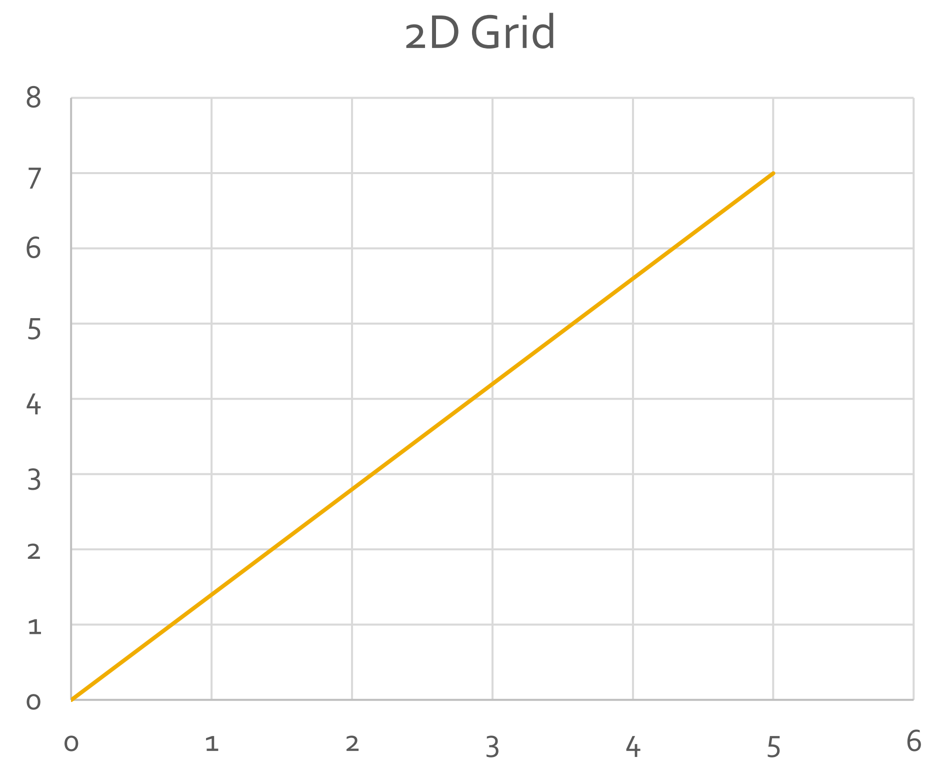
Recall: Vector

- A vector is a quantity defined by a magnitude and a direction. For example, a rocket's velocity is a 3-dimensional vector: its magnitude is the rocket's speed, and its direction is (hopefully) up.
- A vector can be represented by an array of numbers called *scalars*.
- Each scalar corresponds to the magnitude of the vector with regard to each dimension.
- For example, say the rocket is going up at a slight angle: it has a vertical speed of 5,000 m/s, and also a slight speed towards the East at 10 m/s, and a slight speed towards the North at 50 m/s.

$$\textbf{velocity} = \begin{pmatrix} 10 \\ 50 \\ 5000 \end{pmatrix}$$

Norms

- L1 –norm:
  - Formulation: $||[a, b]|| = |a| + |b|$
  - Geometric Interpretation: the distance you would travel if you could only move along the grid lines.
- L2 –norm:
  - Formulation: $||[a, b]|||| = \sqrt{a^2 + b^2}$
  - Geometric Interpretation: measures the straight-line distance, which is the Euclidean distance between two points.



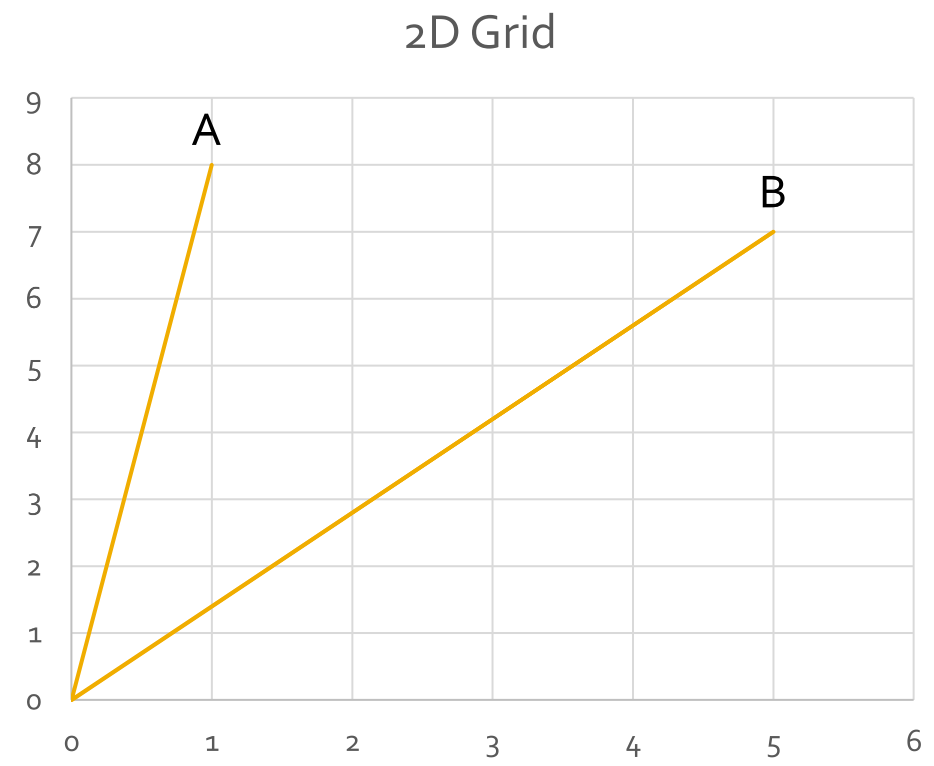
Sum and Difference of Vectors

- Sum and Difference of Vectors : $z = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, y = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$

  - Add the coordinates of vectors , $z + y = \begin{bmatrix} 2+1 \\ 1+4 \end{bmatrix}$
  - Subtract the coordinates of vectors , $z - y = \begin{bmatrix} 2-1 \\ 1-4 \end{bmatrix}$

Distance between the Vectors

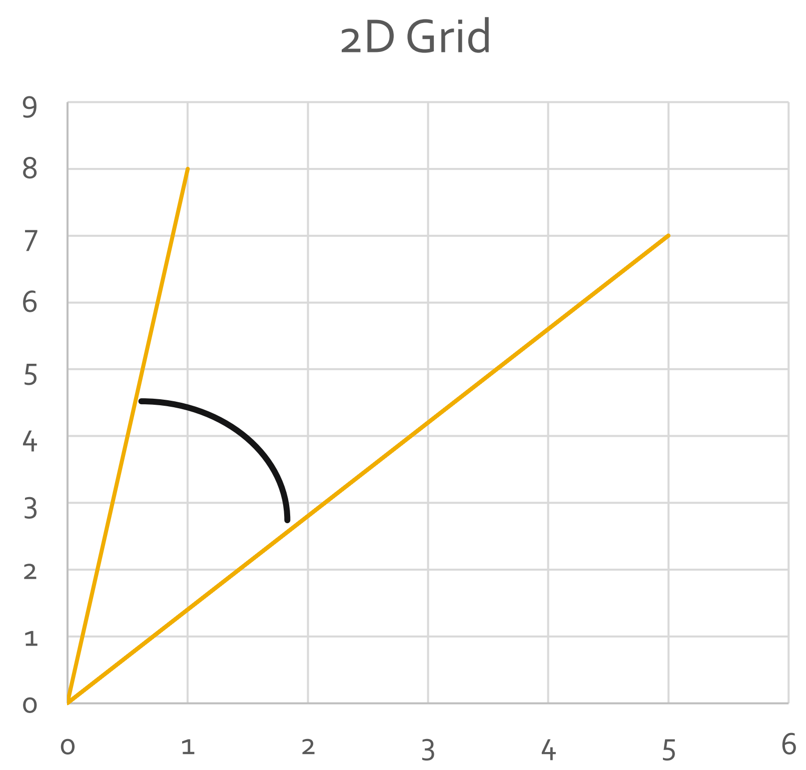
- How different are two vectors $A = \begin{bmatrix}1\\8\end{bmatrix}$ and B $= \begin{bmatrix}5\\7\end{bmatrix}$ from each other?
  - L1-distance $|A - B|$:
    - $|A - B| = |1 - 5| + |8 - 7| = 5$
  - L2-distance $||A - B||$:
    - $||A - B|| = \sqrt{(1-5)^2 + (8-7)^2} = 4.12$



Geometric Definition of the Dot Product

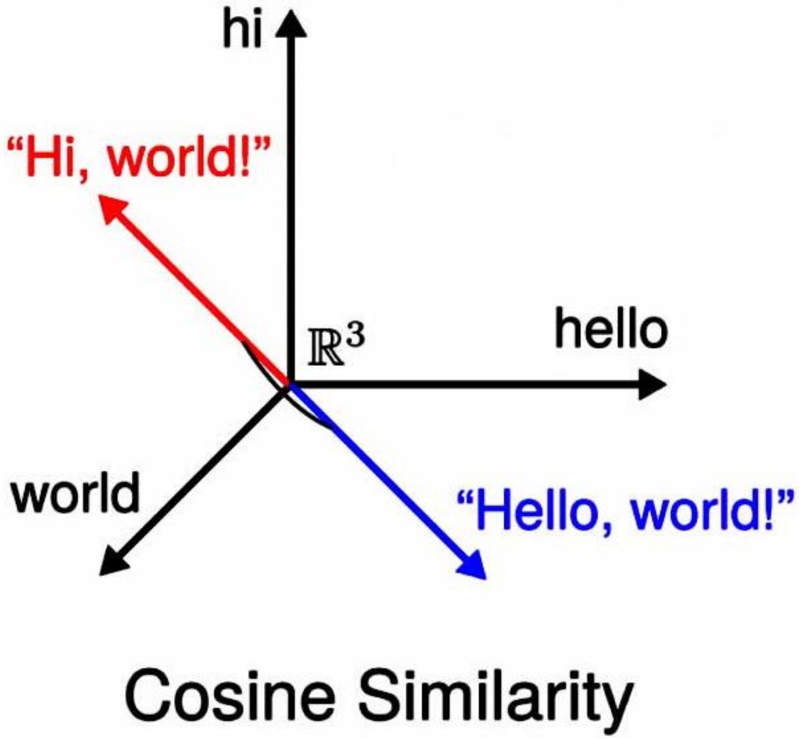
- In Euclidean space, a Euclidean vector has both magnitude and direction.
- The dot product of two vectors x and y is defined by:
  - $x \cdot y = |x||y|\cos(\theta)$, where θ is the angle between the two vectors.
  - $x.y = x_1 y_1 + x_2 y_2 + x_3 y_3 + \ ...$

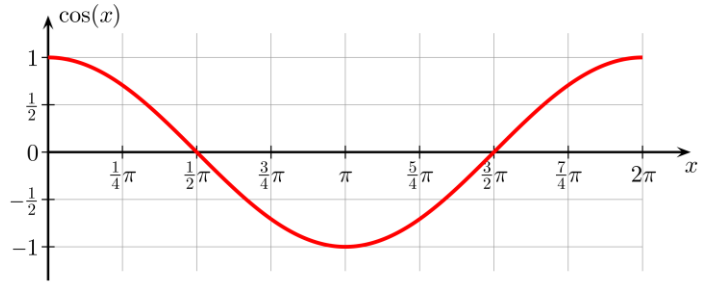
- What is the value of $A.B$?



Cosine Similarity

- Evaluates similarity in direction, regardless of magnitude.
- Commonly used in text analysis, recommendation systems, and information retrieval.
- Formulation:
  - $\cos(\theta) = \frac{x.y}{|x||y|}$
- Dimensions $= \begin{bmatrix} hi \\ hello \\ world \end{bmatrix}$
- $A = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},$
- $Cos - Sim(A, B) = \frac{1}{2}$





How to find a norm of a vector?

- The norm of a vector can be found using NumPy function np.linalg.norm():

- print("Norm of a vector v is", np.linalg.norm(v))
  - Norm refers to the L2 norm = $\sqrt{a^2 + b^2}$

Spam Classifier

Spam	Prize	Award
Yes	1	2
Yes	2	2
Yes	0	2
No	0	1
Yes	2	2
No	1	0

What threshold do we need to consider for the classifier to be considered as a model that does spam classifier?

Matrix Multiplication

Spam	Prize	Award
Yes	1	2
Yes	2	2
Yes	0	2
No	0	1
Yes	2	2
No	1	0

Best Threshold?

$Pred = y > \mathbf{1.5}$

a		w	$y = a.w$
1	2	$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\boldsymbol{y = \mathrm{a}.w = 3}$
0	1	$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\boldsymbol{y = \mathrm{a}.w = 1}$
1	0	$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\boldsymbol{y = \mathrm{a}.w = 1}$

Model

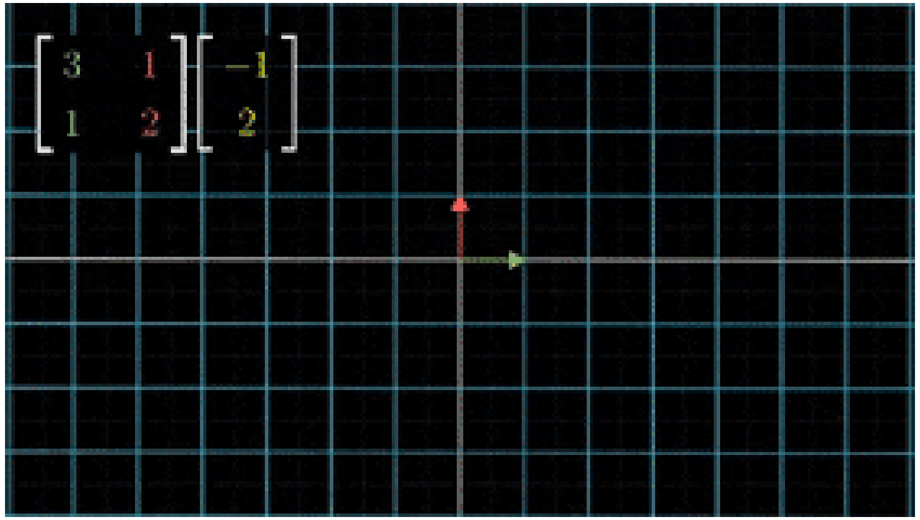
Linear Transformations

- A transformation is a function from one vector space to another that respects the underlying (linear) structure of each vector space.
- Linear Transformations
  - A transformation T is said to be linear if the following two properties are true for any scalar k and any input vectors u and v:
    - $T(kv) = k.T(v)$
    - $T(u + v) = T(u) + T(v)$

Linear Transformations as Matrices

- A= $\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$ $Vector = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ ->in

- Matrix Multiplication, A @ Vector = $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ ->Out



Eigen Vectors and Eigen Values

- Given a vector v, apply linear transformation with square matrix A.

$$A.v = ?$$

$$A.v = \lambda v \rightarrow \text{[Special Case]}$$

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

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

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

$$\lambda_1 = 3,\ v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\lambda_2 = -1,\ v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Eigen Values and Eigen Vectors

- An eigenvalue of a matrix represents the scaling factor that the matrix applies to a vector in its eigenvector direction.
- If v is an eigenvector of a matrix A with eigenvalue λ, then $Av = \lambda v$.
- Matrix A stretches or compresses the vector v by a factor of λ in its eigenvector direction.
- Eigenvalues and eigenvectors can be found using the NumPy function np.linalg.eig().
  - It returns a tuple consisting of a vector and an array.
  - The vector contains the eigenvalues.
  - The array contains the corresponding eigenvectors, one eigenvector per column

Quick Quiz 1

- Dimensions $= \begin{bmatrix} hi \\ hello \\ world \end{bmatrix}$

- What is the Cosine Similarity between the following pairs of sentences:
  - "hi hi" and "hello world"
  - "hello world" and "hello world"

References

- https://learning.oreilly.com/library/view/hands-on-machine-learning/9781098125967/ch01.html#what_is_machine_learning

Thank you!

- Any questions?

Thank You

Youssef Abdelkareem

yabdelkareem@conestogac.on.ca

** Credits for Dr. Anasuya Bhima for assisting in the slides content*