Q355/Q590
Linear Algebra

What is linear algebra?

Which among the following are linear equations?

- ☑ $3x = 1$
- ☒ $x^2 + y = 5$
- ☑ $3x + 2y + z = -4$
- ☒ $x + xy = 9$
- ☒ $y = 1/x$

Linear algebra is the branch of mathematics concerning linear equations such as:

$$a_1x_1 + \cdots + a_nx_n = b,$$

linear maps such as:

$$(x_1, \ldots, x_n) \mapsto a_1x_1 + \cdots + a_nx_n,$$

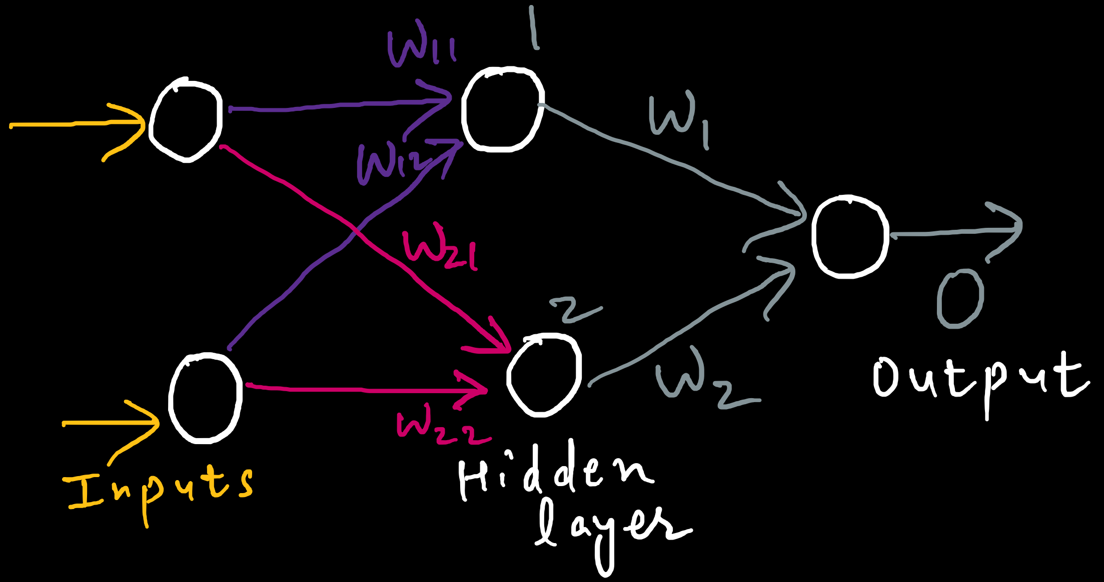
and their representations in vector spaces and through matrices.[1][2][3]

Wikipedia

(Alphabets represent variables)

Why do we care?

(in context of neural networks)



$\underbrace{y}_{\text{output of a neuron}}$ = weighted sum of inputs

$$H_1 = W_{11} I_1 + W_{12} I_2$$
$$H_2 = W_{21} I_1 + W_{22} I_2$$

Linear equations in variables I1, I2, H1, and H2

$$O = W_1 H_1 + W_2 H_2$$

Linear equation in variables O, H1, and H2

Notice that,

$$H_1 = W_{11} I_1 + W_{12} I_2$$
$$H_2 = W_{21} I_1 + W_{22} I_2$$

$\xrightarrow[\text{written as}]{\text{can be}}$

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$$

$$O = W_1 H_1 + W_2 H_2 \longrightarrow \begin{bmatrix} W_1 & W_2 \end{bmatrix} \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} = O$$

(output)

This is matrix representation of the equations.

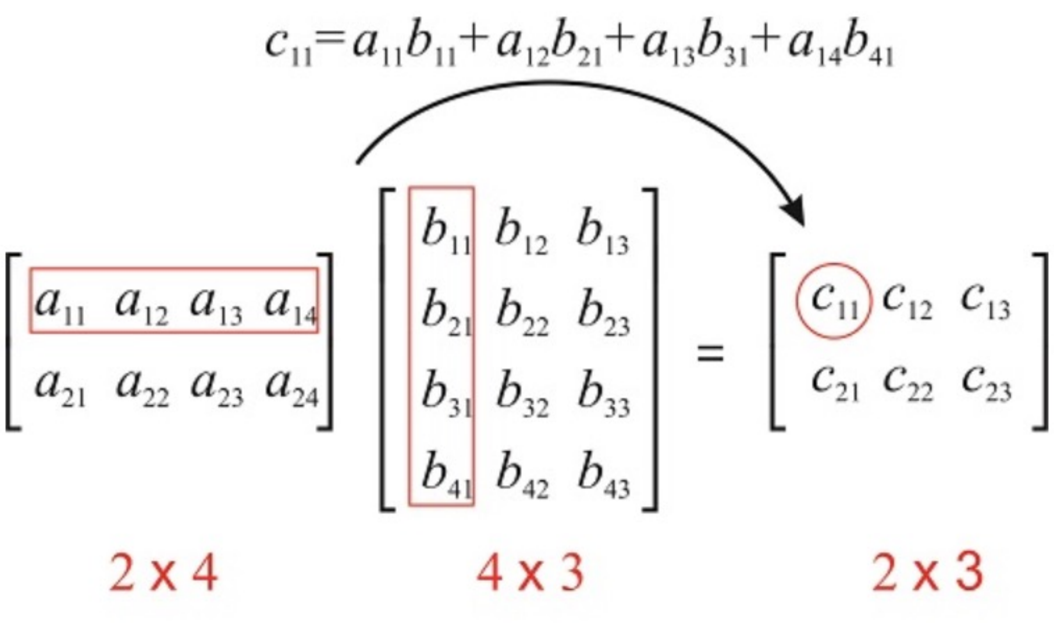
Matrix – Numbers arranged in rows and columns.

Rules for matrix multiplication?

Such one-dimensional matrices are often called <u>vectors</u>.

(Dimensions of a matrix = number of rows x number of columns

Matrix multiplication



- Entry at (i,j) where 'i' is the row number and 'j' is the column number in the product corresponds to multiplication of ith row of first matrix and jth column of second matrix.

- Matrix multiplication A x B can only take place if,

Number of columns in A = Number of rows in B

Why to use matrix representation?

You can say,

"I am someone who sells or grows flowers or studies or writes about flowers." OR "I am a florist"

Similarly, you can solve,

$$H_1 = W_{11} I_1 + W_{12} I_2$$

$$H_2 = W_{21} I_1 + W_{22} I_2$$

$$O = W_1 H_1 + W_2 H_2$$

OR

$$\begin{bmatrix} W_1 & W_2 \end{bmatrix} \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = O$$

LHS will soon get intractable as your network size grows

Different types of matrices

$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow$ identity matrix $\Rightarrow$ so called because $\underset{A}{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} \times \underset{I}{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} = \underset{A}{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}$

$\begin{bmatrix} a & b \\ c & d \end{bmatrix}_{2\times 2} \Rightarrow$ Square matrix

$\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}_{3\times 2} \Rightarrow$ Rectangular matrix $\Leftarrow \begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix}_{2\times 3}$

$\begin{bmatrix} a & b \\ b & d \end{bmatrix} \Rightarrow$ symmetric matrices $\Leftarrow \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$

Properties of matrices

Transpose of a matrix: $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$

$A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$ $A^T = \begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix}$

Transpose, of
A = [3 2
1 5] ?

Determinant of a matrix: $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $|A| = ad - bc$

If $|A| = 0 \Rightarrow$ Matrix is 'singular' or non-investible

Lets see what this means.

Determinant of
A = [3 2
1 5] ?

Inverse of a matrix

Consider, $ax + by = c$ — (1)

$dx + ey = f$ — (2)

From (1), $x = \frac{1}{a}(c - by)$, inserting this in (2):

$$\frac{d}{a}(c - by) + ey = f$$

$$y = \frac{af - dc}{ae - db} \qquad x = \frac{ce - bf}{ae - db}$$

Answers match :)

Solving using matrix

$$\underbrace{\begin{bmatrix} a & b \\ d & e \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} c \\ f \end{bmatrix}}_{k}$$

$$AX = k \Rightarrow X = A^{-1}k$$

$$A^{-1} = \frac{1}{|A|}\left(\text{cofactor}(A)\right)^T = \frac{1}{|A|}\begin{bmatrix} e & -b \\ -d & a \end{bmatrix}$$

$$X = A^{-1}k = \frac{1}{|A|}\begin{bmatrix} e & -b \\ -d & a \end{bmatrix}\begin{bmatrix} c \\ f \end{bmatrix} = \frac{1}{ae - db}\begin{bmatrix} ce - bf \\ af - dc \end{bmatrix}$$

Cofactor matrix

$$A = \begin{bmatrix} a & b \\ d & e \end{bmatrix} \Rightarrow$$

Let co-factor matrix of A

$$= \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

To get $C_{11} \Rightarrow \begin{bmatrix} \cancel{a} & \cancel{b} \\ \cancel{d} & e \end{bmatrix} \Rightarrow C_{11} = e$

To get $C_{21} \Rightarrow \begin{bmatrix} \cancel{a} & b \\ \cancel{d} & \cancel{e} \end{bmatrix} \Rightarrow C_{21} = -b$

For entries whose position indices add to an odd number, also add a negative in-front of them.
So, add a negative infront of C12 and C21 in case of 2x2 matrix.

Similarly other elements can be found, so that

$$\text{cofactor matrix of } A = \begin{bmatrix} e & -d \\ -b & a \end{bmatrix}$$

Cofactor matrix of
A = [3 2
1 5] ?

$|A| = 0$ scenario

Consider, $A = \begin{bmatrix} a & b \\ d & e \end{bmatrix}$

we know

$$|A| = ae - bd$$

$$|A| = 0 = ae - bd$$

$$\frac{a}{b} = \frac{d}{e}$$

This means,

$$\underbrace{\begin{bmatrix} a & b \end{bmatrix}}_{v1} = \beta \times \underbrace{\begin{bmatrix} d & e \end{bmatrix}}_{v2} \Rightarrow$$

If any row or a column can be written as linear combination of others, $|A| = 0$

In such a case, matrix is not invertible.

Rank of a matrix

- Indicates number of linearly independent rows OR columns in the matrix

- Number of linearly independent rows and columns in any matrix are always equal

- Rank of a non-singular n x n square matrix is ‘n’

- Rank of a m x n rectangular matrix can at-most be equal to min(m,n)

Norm

- How “long” is a vector?
- It depends on what “length” means.
- L1 norm = City block norm

- L2 norm = Euclidean Norm

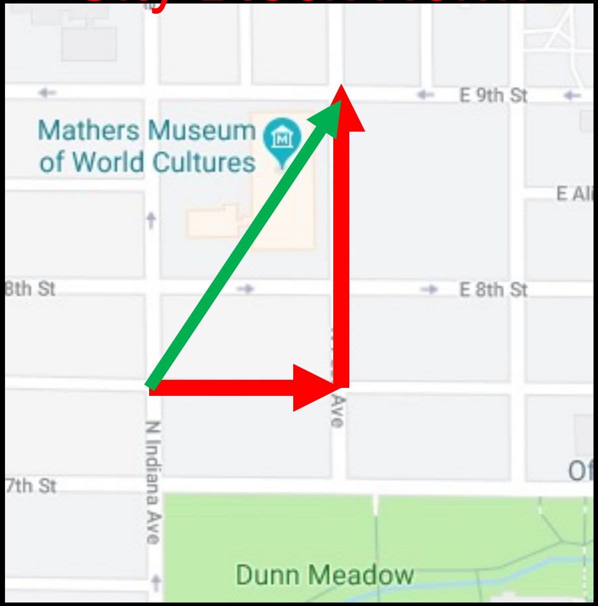
The Euclidean norm

$$\|A\|_E = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij})^2}$$

- In general, L(p) norm is:

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

City Block Norm



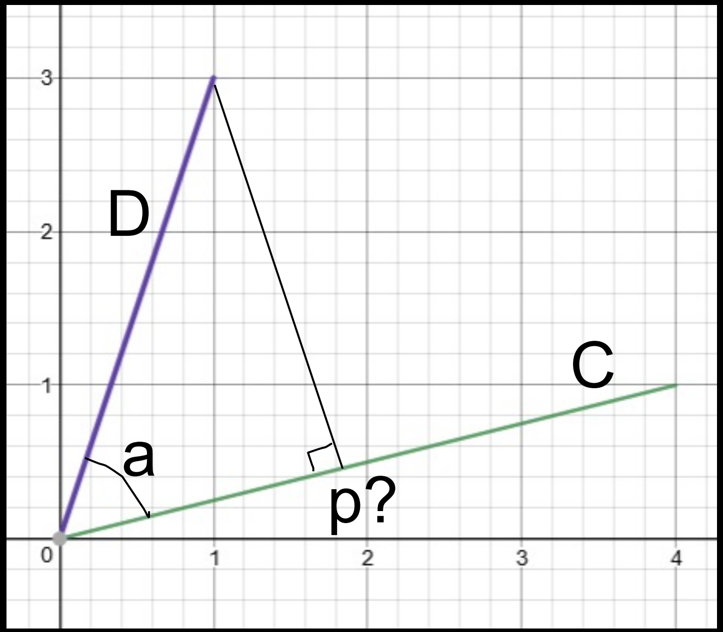
Euclidean Norm

Useful properties of matrix math

- We can add matrices:
  - $\begin{bmatrix} 1 & 4 \\ 2 & 9 \end{bmatrix} + \begin{bmatrix} 8 & 3 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} 9 & 7 \\ 1 & 13 \end{bmatrix}$
- The distributive property is valid for matrix math:
  - A(B+C) = AB + AC
- The associative property also works:
  - A(BC) = (AB)C

Dot product (aka inner product)

- Suppose we want to project a vector onto another, asking e.g. "How far out is p along C from the origin?"
- Vectors are one dimensional matrices, e.g. C=$\begin{bmatrix}4\\1\end{bmatrix}$ and D = $\begin{bmatrix}1\\3\end{bmatrix}$
- We can plot vectors as a line from the origin to the coordinates of the ector
- The dot product: C • D = $C^T D = [4\ 1] * \begin{bmatrix}1\\3\end{bmatrix}$=7
- In general, C • D = |C|*|D|*cos(a)
  =4.123*3.162*cos(71.57-14.04) = 7
  (This matters for neural network weights later on)
- Dot product of A^T = [1 1] and B^T = [1 – 1]?



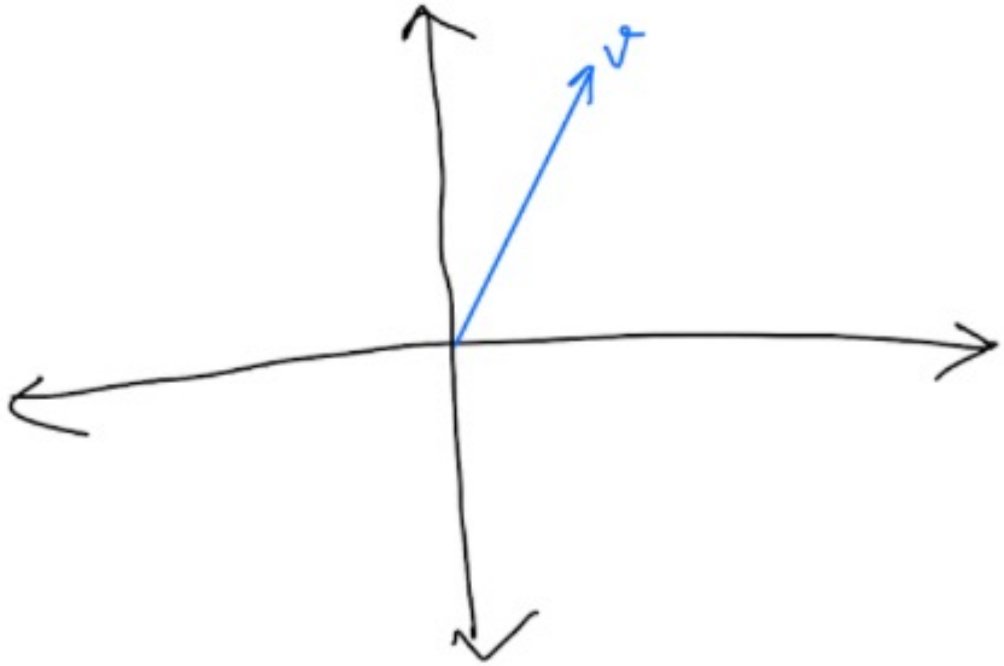
So p= |D|*cos(a)

Thus P= C•D / |C|
= 1.70

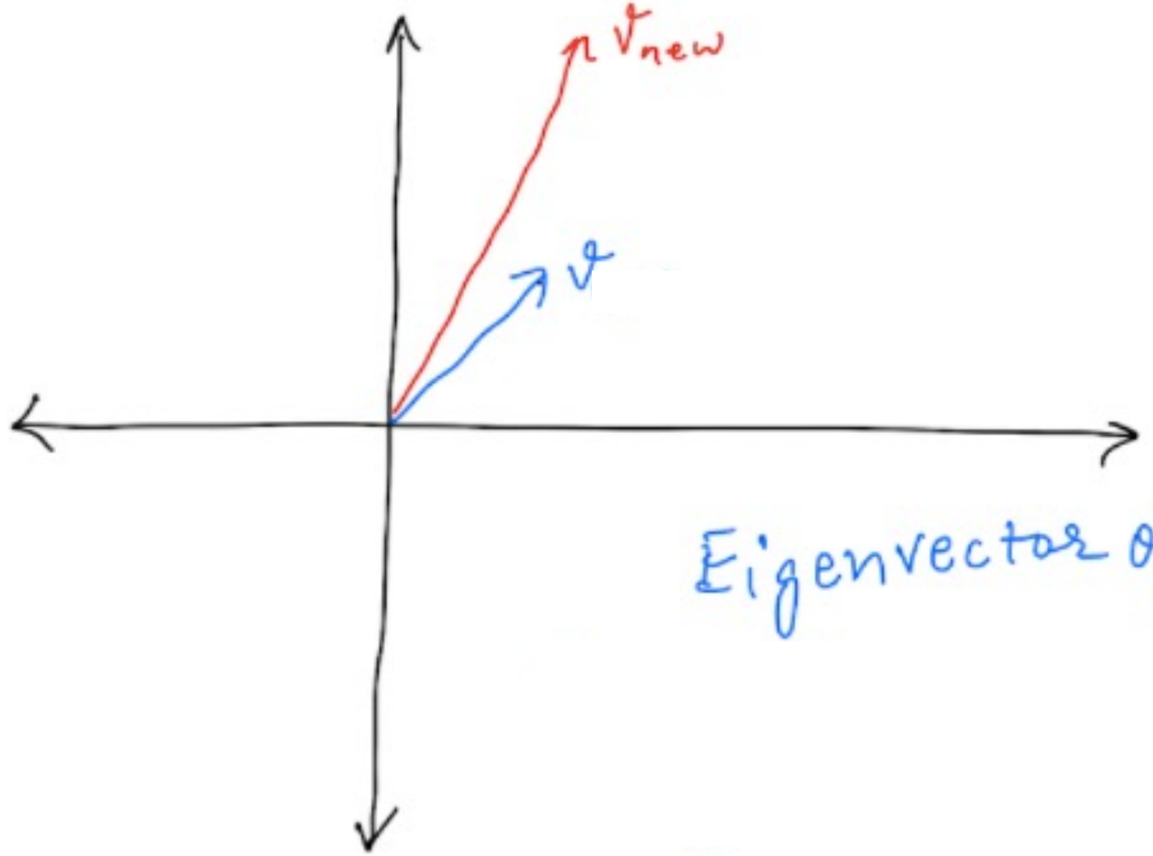
Think of a matrix as performing linear transformation

Consider $A = \begin{bmatrix} 5 & 3 \\ 1 & 7 \end{bmatrix}$ Let $v = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$

v can be represented as a 2-dimensional vector in x-y plane.



$$Av = \begin{bmatrix} 5 & 3 \\ 1 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 17 \\ 29 \end{bmatrix} = v_{new}$$



A can be thought of as performing rotation & stretching in XY plane

Eigenvector of A : Vector that only undergoes stretching but no rotation upon application of A.

Eigenvalue & Eigenvectors

v is called an eigenvector of matrix A if:

$A \times v = \lambda v$, λ: eigenvalue corresponding to v

$$Av = \lambda v \Rightarrow Av - \lambda v = 0 \Rightarrow (A - \lambda I)v = 0$$

Assuming $v \neq 0$, $|A - \lambda I| = 0$

Consider $A = \begin{bmatrix} 5 & 3 \\ 1 & 7 \end{bmatrix}$ $\quad |A - \lambda I| = \det\left(\begin{bmatrix} 5-\lambda & 3 \\ 1 & 7-\lambda \end{bmatrix}\right)$

$$= (5-\lambda)(7-\lambda) - 3 = 0$$

$$\lambda^2 - 12\lambda + 32 = 0$$

$$(\lambda - 8)(\lambda - 4) = 0$$

$$\lambda = 8, 4$$

Finding eigenvectors

$$Av = \lambda v \; ; \quad v = \begin{bmatrix} x \\ y \end{bmatrix}$$

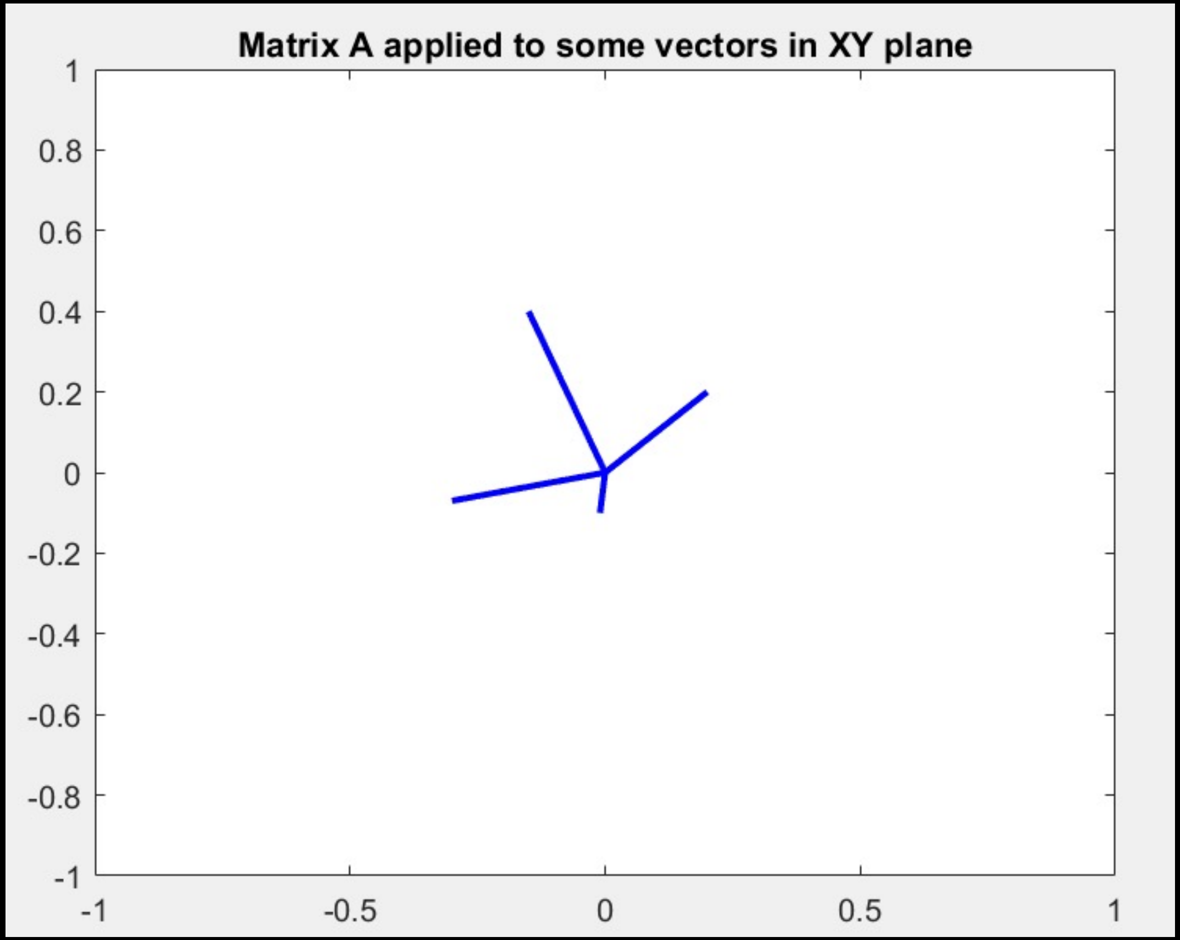
$$\begin{bmatrix} 5 & 3 \\ 1 & 7 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda y \end{bmatrix} \Rightarrow \begin{bmatrix} 5x + 3y \\ x + 7y \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda y \end{bmatrix}$$

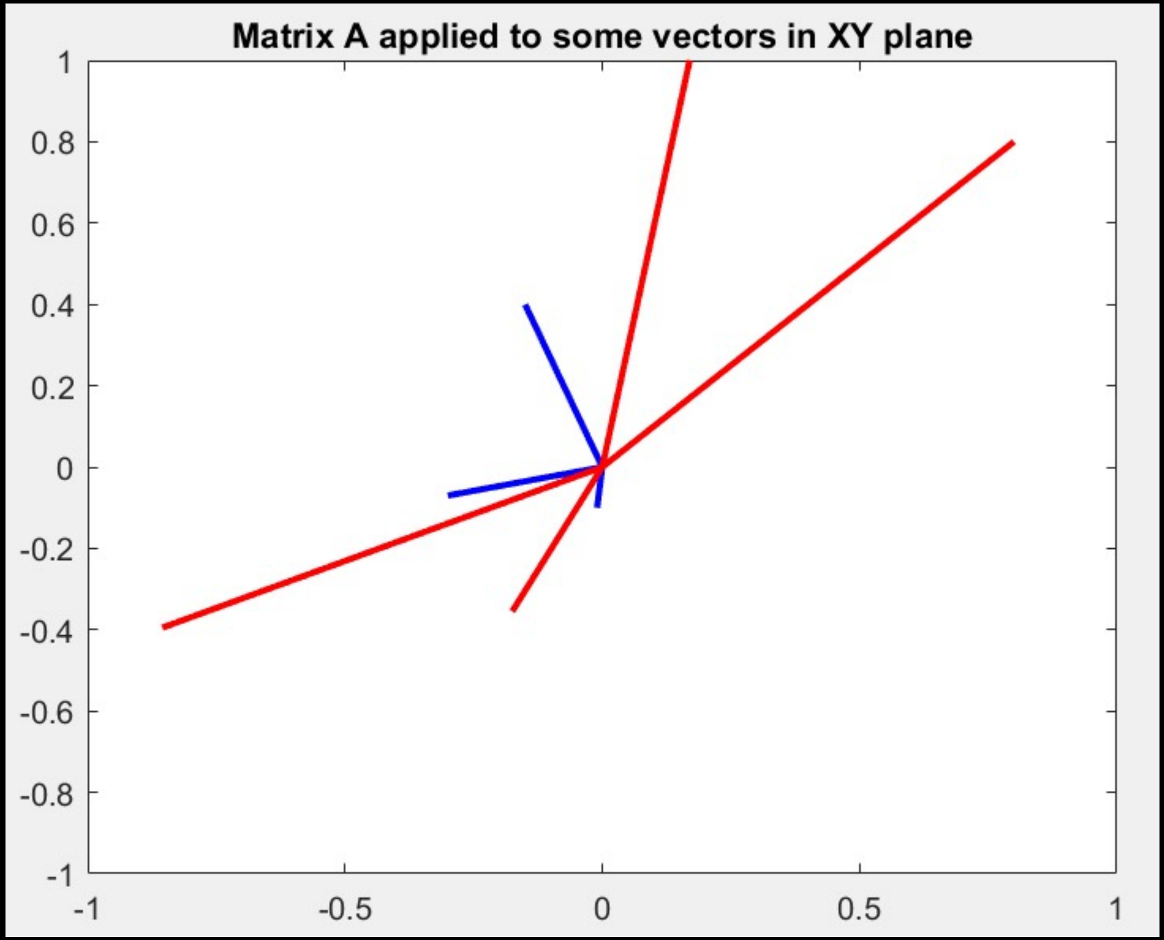
$$5x + 3y = \lambda x \Rightarrow y = \frac{(\lambda - 5)x}{3}$$

$$y/x = \frac{\lambda - 5}{3}$$

$$\left.\begin{array}{l} \lambda = 8 \Rightarrow y/x = 1 \\ \lambda = 4 \Rightarrow y/x = -1/3 \end{array}\right\}$$

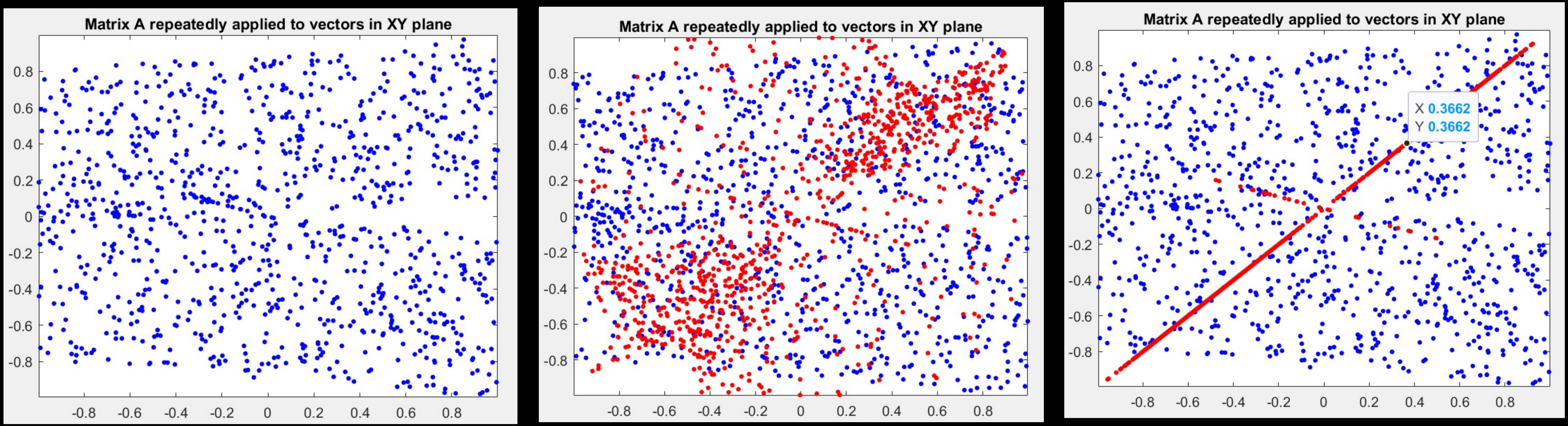
All vectors along these 2 directions are eigenvectors of matrix A.





Original data points: They have been joined with the origin (0,0) so that each data points looks like a vector.

Data points after matrix A was applied to them; Observe that three of the vectors (data points) got rotated and stretched, but one of them only got stretched (but was not rotated) → such a vector is called 'eigenvector of matrix A'



Blue: Original points (vectors); each dot represents the tip of the vector.
Red: Resultant vectors after application of matrix A

Observe that as A is applied repeatedly, the vectors get stretched and rotated and eventually settle along 2 directions → This are the directions of the eigenvectors of matrix A.

PCA (Principal Component Analysis)

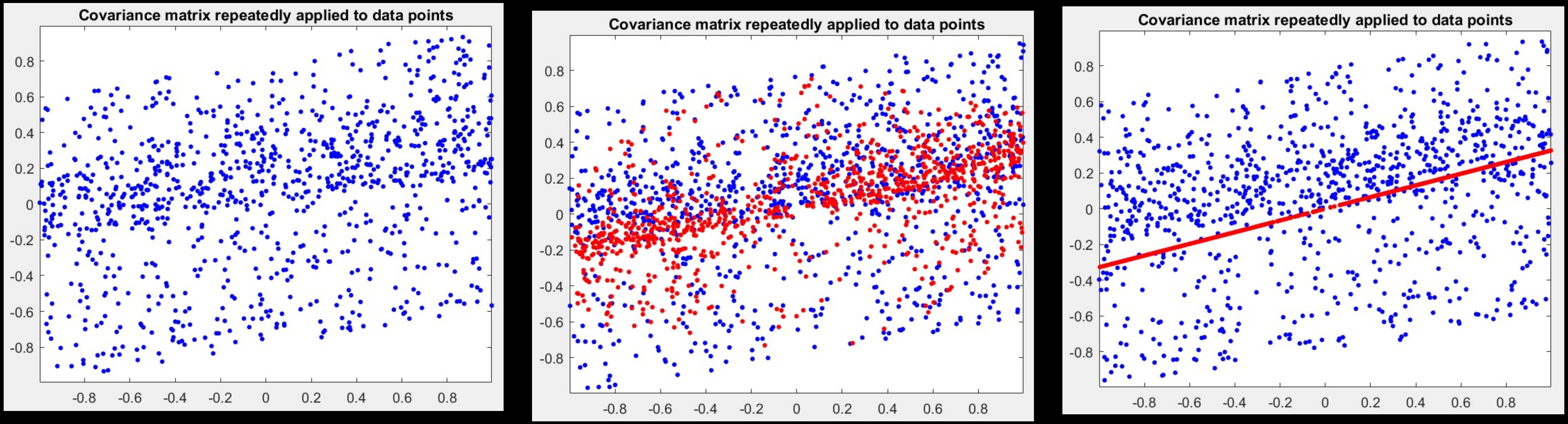
- when you have some data, $D = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \end{bmatrix}_{n \times 2}$

you can find its covariance matrix.

$$C = \begin{bmatrix} Var(x) & cov(x,y) \\ cov(x,y) & Var(x) \end{bmatrix}$$

- Eigenvectors of C : · Directions along which data has maximum variance.
  · 'Principal components'

- Project the data onto principal components.
  (can be used for dimensionality reduction)

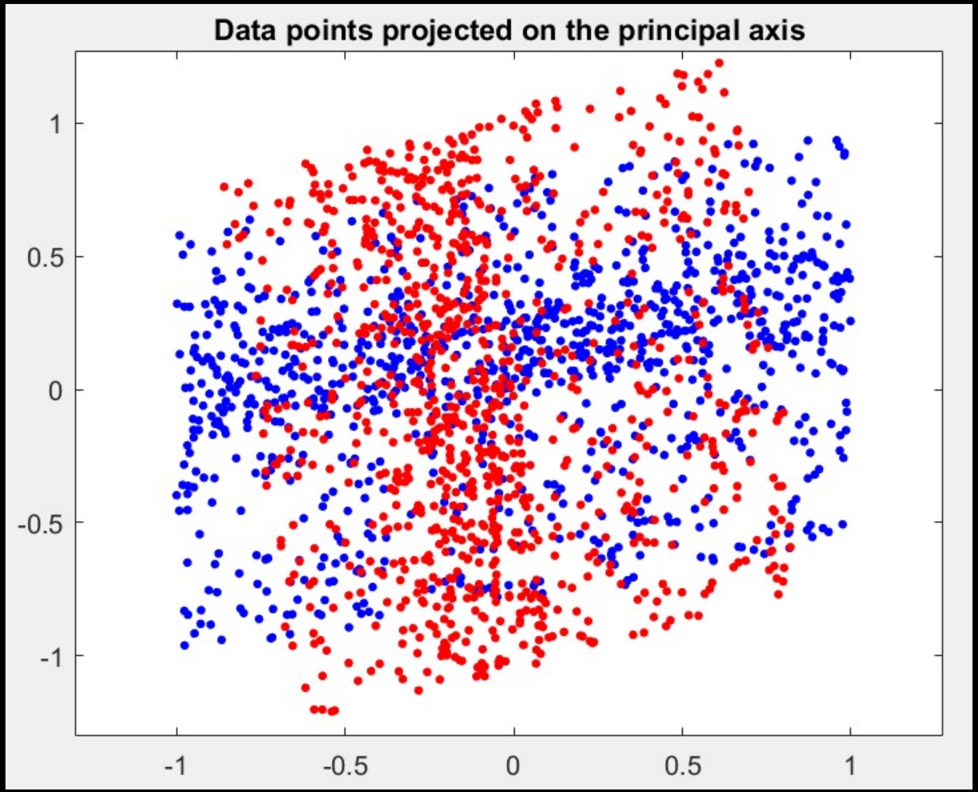


Blue: Data points (vectors); each dot represents the tip of the vector.

A covariance matrix is formed using these very data points and that is applied to these repeatedly.

Red: Resultant vectors after application of matrix A

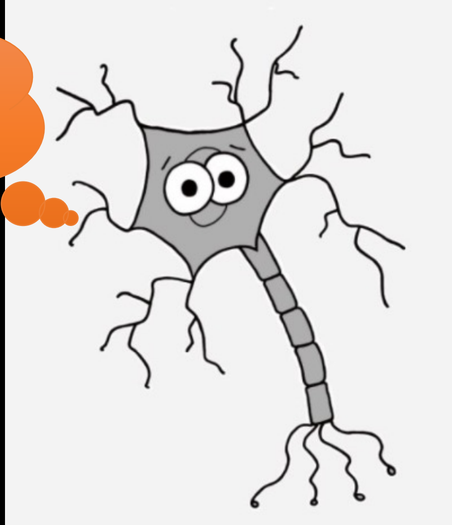
Observe that as the covariance matrix is applied repeatedly, the vectors get stretched and rotated and eventually settle along the directions of the eigenvectors of the covariance → One clearly visible is the direction of the eigenvector with highest eigenvalue →Also the direction along which the data has maximum variance.



Blue: Original data as also shown in previous slide

Data projected onto both the eigenvectors of covariance matrix; You can also project the data only along the eigenvector with highest eigenvalue → data will become one dimensional → dimensionality reduction

I have some questions!



Q1. What is the transpose of matrix, A = [5 3
3 2] ?

What can we say about symmetric matrices based on that?

A. $A^T = -A$ B. $A^T = A$ C. $A^TA=I$

Q2. Why does this system of equations not have finite number of solutions, in terms of matrix properties? A = [6 3
2 1]; C = $[9\ -7]^T$, AX = C

Q4. What are the eigenvalues of matrix, A = [5 0
0 3]?

A. Don't exist B. (5,3) C. (15,1)

Q3. Which of the following matrices is definitely non-invertible? (multiple correct answers)

A. Singular B. Non-singular C. det(A) is non-zero D. det(A) is zero

Q5. What are the eigenvectors of matrix A in Q4?

A. Does not have any B. [0 1], [1,0] C. All vectors in XY plane