Systems of Equations

04/03/2025

What is the formula of a linear equation?

$y = w \cdot x + b$ — w: weight, b: bias

weight: wind speed → power output

$y = w_1 \cdot x_1 + w_2 \cdot x_2 + b$

wind speed + temperature → power output

How do we formulate linear regression in ML? (as multiple terms)

$$w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n + b = y$$

A dataset will have lots of rows, therefore ⇒

How do we formulate system of linear equations in ML?

$$\begin{aligned}
w_1 \cdot x_1^{(1)} + w_2 \cdot x_2^{(1)} + \dots + w_n x_n^{(1)} + b &= y^{(1)} \\
w_1 \cdot x_1^{(2)} + w_2 \cdot x_2^{(2)} + \dots + w_n x_n^{(2)} + b &= y^{(2)} \\
w_1 \cdot x_1^{(3)} + w_2 \cdot x_2^{(3)} + \dots + w_n x_n^{(3)} + b &= y^{(3)} \\
&\vdots \\
w_1 \cdot x_1^{(m)} + w_2 \cdot x_2^{(m)} + \dots + w_n x_n^{(m)} + b &= y^{(m)}
\end{aligned}$$

What are the values we aim to find so that we can get as close as possible to the best fitting line?

Now we have a lot of equations, a system of equations, and we aim to find values for weights and bias that gets us as close as possible to solve all these equations at once.

How do we formulate linear regression in ML (as vectors and matrices)?

$$W \cdot X + b = \hat{y}$$

W: Vector of weights

$$[w_1 \; w_2 \; w_3 \dots w_n]$$

X: Matrix of features

$$\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots & x_n^{(2)} \\ \vdots & & & & \\ x_1^{(m)} & x_2^{(m)} & x_3^{(m)} & \dots & x_n^{(m)} \end{bmatrix}$$

b: bias (a constant)

$\hat{y}$: Vector of Target Variables

$$[y^{(1)} \; y^{(2)} \; y^{(3)} \dots y^{(m)}]$$

When we apply a linear regression model, we actually try to solve a system of equations in a way that our formula ($W \cdot X + b = \hat{y}$) would give us the best solution.

How do we convert a system of equations into vectors and matrices?

Example:

① Linear algebra score added to your calculus score minus your probability score was 6.

② Your algebra score minus your calculus score plus double your probability score was 4.

③ Four times your linear algebra score minus double your calculus score added to your probability score was 10

$$+1a \quad +1c \quad -1p = 6$$
$$+1a \quad -1c \quad +2p = 4$$
$$+4a \quad -2c \quad +1p = 10$$

$W = [a \quad c \quad p]$

$X = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 2 \\ 4 & -2 & 1 \end{bmatrix}$

$\hat{y} = [6 \quad 4 \quad 10]$

$b = 0$

What is a system of sentences?

System 1	System 2	System 3
. The dog is black	. The dog is black	. The dog is black
. The cat is orange	. The dog is black	. The dog is white
"Complete"	"Redundant"	"Contradictory"
Non-singular system	Singular system	Singular system

What type of systems are there based on completeness and singularity?

A non-singular system carries as many pieces of information as sentences.

"System of equations is basically a system of sentences with numbers"

Sentences with numbers ⇒ equations

· You bought an apple and a banana for $10. Your wife bought an apple and two bananas for $12. How much each fruit cost?

$$a + b = 10$$
$$a + 2b = 12$$

Non-singular systems are complete. They carry as many pieces of information as equations, and they have one solution. Singular systems carry either redundant information (infinite number of solutions) or contradictory information (no solutions).

① You bought an apple and a banana for $10.
② You bought two apples and two bananas for $20.

Redundant → INFINITE NUMBER OF SOLUTIONS

· How much does each cost?

· Any two numbers that add to 10 are solutions.

① You bought an apple and a banana for $10.
② You bought two apples and two bananas for $24.

Contradictory → NO SOLUTIONS

How much each cost?

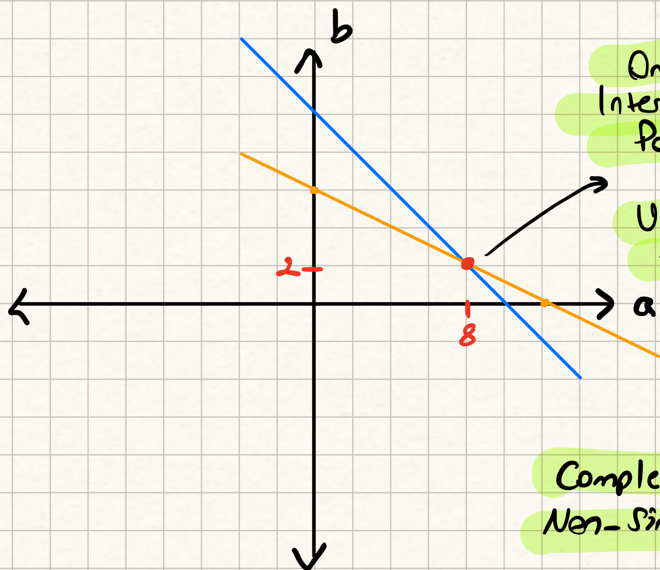
Graphical representation of system of equations

Two-variable linear equations can be visualized as lines in the coordinate plane. (Three-variables → planes in space, more variables → high-dimensional things we won't worry about now!)

Linear Equations ⟶ Lines

System of Linear Equations ⟶ Arrangements of lines

$a + b = 10$
$a + 2b = 12$

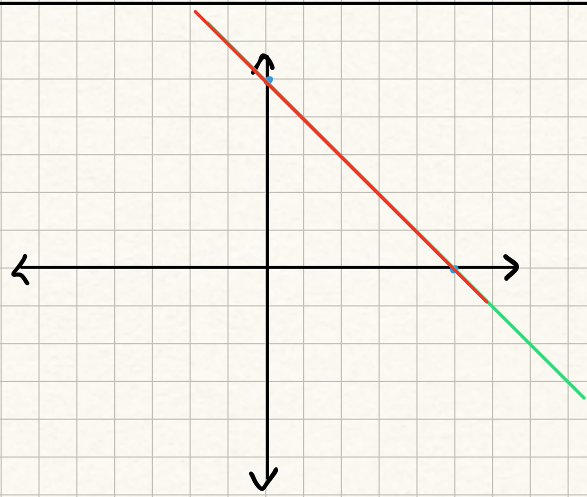


One Interception Point
↓
Unique Solution

Complete
Non-Singular

$a + b = 10$

$2a + 2b = 20$



Overlapping Lines

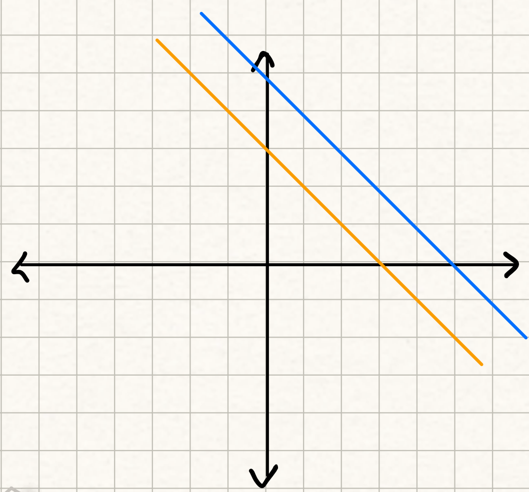
⇓

Infinitely many Solutions

Redundant Singuler

$a + b = 10$

$2a + 2b = 12$

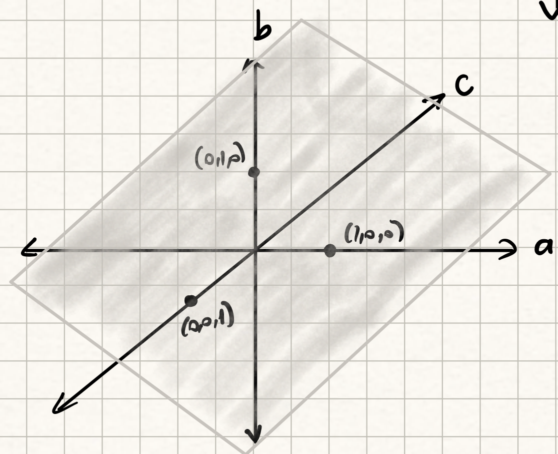


Parallel Lines

⇓

No Solutions

Contradictory Singular



$a + b + c = 1 \Rightarrow$

$1 + 0 + 0 = 1$

$0 + 1 + 0 = 1$

$0 + 0 + 1 = 1$

Do constants matter for singularity?

In ML, we care about if a system is non-singuler or not. If not, we don't care much about the reason (redundant / contradict therefore we can neglect the constants in the equations.

What is "linear dependency" in the context of matrices?

SYSTEM 1

$$\underbrace{\begin{aligned} a + b &= 0 \\ a + 2b &= 0 \end{aligned}}_{\text{Non-Singular system}} \qquad \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}}_{\text{Non-Singular Matrix}}$$

* No equation is a multiple of the other one

* No row is a multiple of the other one

Rows are linearly independent

SYSTEM 2

$$\underbrace{\begin{aligned} a + b &= 0 \\ 2a + 2b &= 0 \end{aligned}}_{\text{Singular System}} \qquad \underbrace{\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}}_{\text{Singular Matrix}}$$

* Second equation is a multiple of the first one

* Second row is a multiple of the first one

Rows are linearly dependent

$$\begin{aligned} a &= 1 \\ b &= 2 \\ a + b &= 3 \end{aligned} \quad + \quad =$$

$$\begin{aligned} 1a + 0b + 0c &= 1 \\ 0a + 1b + 0c &= 2 \\ 1a + 1b + 0c &= 3 \end{aligned} \quad + \quad = \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

Row 3 depends on rows 1 and 2.

$$\begin{aligned} a + b + c &= 0 \\ a + b + 2c &= 0 \\ a + b + 3c &= 0 \end{aligned} \qquad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 1 & 3 \end{bmatrix}$$

Row 2 is the average of the rows 1 and 3.

$\Rightarrow$ Row 2 depends on rows 1 and 3.

$$\begin{aligned} a + b + c &= 0 \\ a + 2b + c &= 0 \\ a + b + 2c &= 0 \end{aligned} \qquad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

No linear relations between rows

$\Rightarrow$ Non-Singular System

What is the determinant and how do we calculate it?

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \underbrace{a \cdot d - b \cdot c}_{\text{the determinant of the matrix}}$$

$= 0 \Rightarrow$ singular

$\neq 0 \Rightarrow$ non-singular

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \qquad \begin{aligned} &(a \cdot e \cdot i) + (b \cdot f \cdot g) + (c \cdot d \cdot h) \\ &- (c \cdot e \cdot g) - (f h a) - (i \cdot b \cdot d) \end{aligned} \qquad \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

$= 0 \Rightarrow$ singular

$\neq 0 \Rightarrow$ non-singular

#1 ↓ #2 ↓ #3 + #2 ↓ #3 ↓ #1 + #3 ↓ #1 ↓ #2

− #3 ↓ #2 ↓ #1 − #2 ↓ #1 ↓ #3 − #1 ↓ #3 ↓ #2

How do we present and solve linear systems as matrices in NumPy?

$$-x_1 + 3x_2 = 7$$
$$3x_1 + 2x_2 = 1$$

$$A = \begin{bmatrix} -1 & 3 \\ 3 & 2 \end{bmatrix} \qquad b = [7 \;\; 1]$$

```
A = np.array([[-1,3],[3,2]], dtype = np.dtype(float))
b = np.array([7,1], dtype = np.dtype(float))

x = np.linalg.solve(A,b)
print(x)   # Output : [-1, 2]

d = np.linalg.det(A)
print(d)   # Output : -11.00

A_system = np.hstack((A, b.reshape((2,1))))
```

we need this because:

b.shape was (2,)

we need (2,1)

* Determinant tells us if the matrix is non-singular (np.linalg.det(A) ≠ 0) or singular (np.linalg.det(A) = 0)

* np.linalg.solve(A,b) returns an array of solutions for X_1 and X_2 (or throws an LinAlgError if A is singular.

Solving Systems of Linear Equations

05/03/2025

How do we solve a system of equations with two variables?

Method 1: Manipulating equations so that we can get rid of one of the variables.

$3(5a + b = 17)$ $\Rightarrow$ $15a + 3b = 51$
$4a - 3b = 6$ $\quad$ $+\ 4a - 3b = 6$

$19a + 0b = 57$
$a = 3$

$5a + b = 17$
$5 \cdot 3 + b = 17$
$b = 2$

Method 2: Re-organize one of the equations so that you can define one of the variables in terms of other.

$5a + b = 17 \Rightarrow b = 17 - 5a$
$4a - 3b = 6 \rightarrow 4a - 3(17 - 5a) = 6$

$4a - 51 + 15a = 6$

$19a = 57$
$a = 3$

$5a + b = 17$
$5 \cdot 3 + b = 17$
$b = 2$

What if the system is redundant?

$(a + b = 10) \cdot -2$ $\quad$ $-2a - 2b = -20$
$2a + 2b = 20$ $\quad$ $+\ 2a + 2b = 20$

$0 = 0$

Solved System:
$a = x$
$b = 10 - x$

"Solution has 1 degree of freedom, which is x. The solutions form a line."

$(a + b = 10) \cdot -2$ $\quad$ $-2a - 2b = -20$
$2a + 2b = 24$ $\quad$ $2a + 2b = 24$

$0 = -4$
CONTRADICTION!

Solved System
N/A

Original System

$5a + b = 17$

$4a - 3b = 6$

Intermediate System

$a + 0.2b = 3.4$

$b = 2$

Solved System

$a = 3$

$b = 2$

Original Matrix

$$\begin{bmatrix} 5 & 1 \\ 4 & -3 \end{bmatrix}$$

Upper Diagonal Matrix

$$\begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix}$$

Row Echolon Form

Diagonal Matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Reduced Row Echolon Form

Row Echolon Form Rules

* Main diagonal can be 0s or 1s.
* Below the main diagonal, everything must be 0.
* Any number can exist on the right-side of 1s.
* Only zeros are allowed on the right side of zeros.

$$\begin{matrix} 1 & n & n & n & n \\ 0 & 1 & n & n & n \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

The first non-zero element of a row is called the pivot. For a matrix to be in row echelon form:

1. All rows without pivots must be at the bottom
2. All pivots have to be at the right-side of the pivot of the row above.

*Some texts say that the pivots also have to be reduced to 1. Even if this is not a "must" it's very handy because we must do it for the reduced row echelon form anyways.

What are elementary row operations?

① Switching rows

$$\begin{bmatrix} 5 & 1 \\ 4 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 4 & 3 \\ 5 & 1 \end{bmatrix}$$

② Multiplying a row by a non-zero scalar

$$\begin{bmatrix} 5 & 1 \\ 4 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 50 & 10 \\ 4 & 3 \end{bmatrix}$$

③ Adding a row to another row

$$\begin{array}{r} \begin{bmatrix} 5 & 1 \\ 4 & 3 \end{bmatrix} \\ + \\ \hline \begin{bmatrix} 9 & 4 \end{bmatrix} \end{array} \longrightarrow \begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}$$

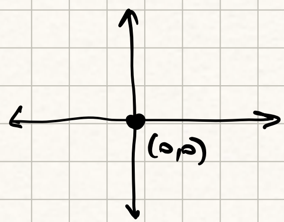
System 1	System 2	System 3
The dog is black.	The dog is black.	The dog
The cat is orange.	The dog is black.	The dog
Two Sentences	Two sentences.	Two sentences
Two pieces of information.	One piece of info.	Zero pieces of info*
Rank = 2	Rank = 1	Rank = 0

*About the color.

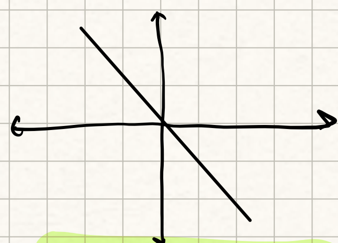
When we want to turn a matrix into a row echelon form and/or reduced row echelon form, we can perform the following elementary operations:

① Switch the order of rows
② Multiply the elements of a row by a non-zero scalar.
③ Add a row to another.

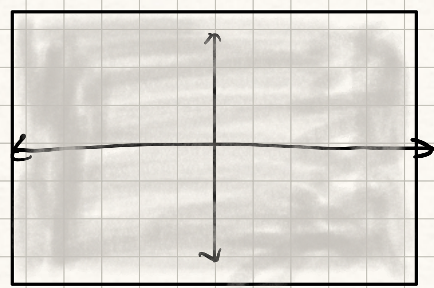
System 1	System 2	System 3
$a+b=0$	$a+b=0$	$0a+0b=0$
$a+2b=0$	$2a+2b=0$	$0a+0b=0$
Rank = 2	Rank = 1	Rank = 0
$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$



Dimension of solution space = 0 (it's a point)



Dimension of solution space = 1 (it's a line)



Dimension of solution space = 2 (It's a plane)

$$\text{Rank}^* = 2 - \left(\begin{array}{c}\text{Dimension of} \\ \text{solution space}\end{array}\right)$$

*for 2x2

If the rank is equal to the number of rows, the matrix is non-singular. B/c this means it carries as many as information as the number of equations it has.

System 1	System 2	System 3	System 4
$a+b+c=0$	$a+b+c=0$	$a+b+c=0$	$0a+0b+0c=0$
$a+2b+c=0$	$a+b+2c=0$	$2a+2b+2c=0$	$0a+0b+0c=0$
$a+b+2c=0$	$a+b+3c=0$	$3a+3b+3c=0$	$0a+0b+0c=0$
Rank 3	Rank 2	Rank 1	Rank 0

The number of pivots in a row echelon matrix is called its rank. If the rank equals to the number of rows, the matrix is non-singular.

$$\begin{bmatrix} 5 & 1 \\ 4 & -3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0.2 \\ 4 & -3 \end{bmatrix}$$

Let's turn 5 into 1 b/c main diag. can't have 5.

* 4 has to be converted into zero b/c bottom of the main diag. has to be zero. How?

$$\longrightarrow \begin{bmatrix} 1 & 0.2 \\ 1 & -0.75 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0.2 \\ 0 & -0.95 \end{bmatrix}$$

"first, make it 1", then subtract the first row from the second

* Now we need 1 on the bottom right. So we divide the row by -0.95

$$\rightarrow \begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix} \Rightarrow \text{Row Echolon Form} \qquad \text{Rank} = 2$$

(2) $$\begin{bmatrix} 5 & 1 \\ 10 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0.2 \\ 1 & 0.2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0.2 \\ 0 & 0 \end{bmatrix} \qquad \text{Rank} = 1$$

(3) $$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{Rank} = 0$$

* The rank of the matrix is the sum of the values in the main diagonal of the row echolon form!

*For 2x2

$$\begin{matrix} 2 & n & n & n & n \\ 0 & 1 & n & n & n \\ 0 & 0 & 3 & n & n \\ 0 & 0 & 0 & -5 & n \\ 0 & 0 & 0 & 0 & 1 \end{matrix}$$

Rank 5

(1) Rows with only zeros must go to the bottom.

(2) The left-most non-zero entry of a row is called a pivot. Every pivot must be to the right of the pivots on the rows above.

(3) Rank of the matrix is the number of pivots. (a general rule, not just for 2x2)

$$\begin{matrix} 3 & n & n & n & n \\ 0 & 0 & -1 & n & n \\ 0 & 0 & 0 & -4 & n \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix} \qquad \begin{matrix} \div 3 \\ \div -1 \\ \div -4 \\ \\ \\ \end{matrix}$$

Rank = 3

* We can make all pivots ones if we want.

$$\begin{matrix} 1 & n & n & n & n \\ 0 & 0 & 1 & n & n \\ 0 & 0 & 0 & 1 & n \\ 0 & 0 & 0 & 0 & n \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

Rank = 3

n's are now different of course, but the number of pivots is the same.

Reduced Row Echolon form

* The matrix must be in row echolon form
* Pivots must be converted to 1s.
* Any number above a pivot must be converted to 0.

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$$

* We need this 2 to be 0. So we multiply the second row by 2 and subtract it from the first row

$$\begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$$

Now we must get rid of this -5. Lets multiply 3^{rd} row by 5 and add it to the 1st row.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$$

Now let's multiply 3^{rd} row by 4 and subtract from the second.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

This is our reduced row echolon form.

The Gaussian Elimination Algorithm

$$2a - b + c = 1$$
$$2a + 2b + 4c = -2$$
$$4a + b + 0c = -1$$

$$\left[\begin{array}{ccc|c} 2 & -1 & 1 & 1 \\ 2 & 2 & 4 & -2 \\ 4 & 1 & 0 & -1 \end{array}\right]$$

Augmented Matrix

1. Turn R1 into a 1 by $R1 = R1/2$

$$\left[\begin{array}{ccc|c} 1 & -1/2 & 1/2 & 1/2 \\ 2 & 2 & 4 & -2 \\ 4 & 1 & 0 & -1 \end{array}\right]$$

2. Turn R2 into a 0 by: $R2 = R2 - (R1 \cdot 2)$

$$\left[\begin{array}{ccc|c} 1 & -1/2 & 1/2 & 1/2 \\ 0 & 3 & 3 & -3 \\ 4 & 1 & 0 & -1 \end{array}\right]$$

3. Turn R3 into a 0 by: $R3 = R3 - (R1 \cdot 4)$

$$\left[\begin{array}{ccc|c} 1 & -1/2 & 1/2 & 1/2 \\ 0 & 3 & 3 & -3 \\ 0 & 3 & -2 & -3 \end{array}\right]$$

4. Turn R2 into 1 by $R2 = R2/3$

$$\left[\begin{array}{ccc|c} 1 & -1/2 & 1/2 & 1/2 \\ 0 & 1 & 1 & -1 \\ 0 & 3 & -2 & -3 \end{array}\right]$$

5. Turn R3 into 0 by $R3 = R3 - (R2 \cdot 3)$

$$\left[\begin{array}{ccc|c} 1 & -1/2 & 1/2 & 1/2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & -5 & 0 \end{array}\right]$$

6. Turn R3 into 1 by

$R3 = R3 / -5$

$$\underbrace{\left[\begin{array}{ccc|c} 1 & -1/2 & 1/2 & 1/2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 0 \end{array}\right]}_{\text{Row Echelon Form}}$$

⇒ Back substitution

7. Turn R2 into 0 by

$R2 = R2 - R3$

$$\left[\begin{array}{ccc|c} 1 & -1/2 & 1/2 & 1/2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

8. Turn R1 into 0 by

$R1 = R1 - (1/2\, R3)$

$$\left[\begin{array}{ccc|c} 1 & -1/2 & 0 & 1/2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

9. Turn R1 into 0 by

$R1 = R1 + (1/2\, R_2)$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{array}\right] \begin{array}{l} = a \\ = b \\ = c \end{array}$$

Identity matrix (only 1s in the diagonal)

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 5 \\ 0 & 0 & -7 & 9 \\ 0 & 0 & 0 & X \end{array}\right]$$

$X = 0$ ⇒ Infinite # of solutions

$X \neq 0$ ⇒ No solutions

When we have a reduced row echelon of an augmented matrix, the last column is the vector of constants.

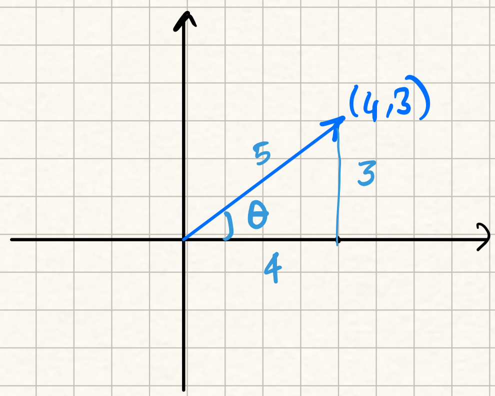
If we have a column with all-zero elements:

1. If the augmented column is also zero, the system has infinite solutions.
2. If the augmented column is not zero, the system has no solutions.

Vector Algebra

11.03.2025

Norm and direction



$$L_1 = |(a,b)|_1 = |a| + |b| = 7$$

$$L_2 = |(a,b)|_2 = \sqrt{a^2+b^2} = 5$$

$$\tan(\theta) = \frac{3}{4}$$

$$\theta = \arctan(3/4) = 0.64 = 36.87^\circ$$

Notation

Row Vector

$$x = [x_1 \ x_2 \ \dots \ x_n]$$

Column Vector

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

$$\|x\|_1 = |x_1| + |x_2| + \dots + |x_n|$$

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

Vector Operations

$u = (4,1)$

$v = (1,3)$

$u + v = (5,4)$

$u - v = (3,-2)$

$x = [x_1 \ x_2 \ \dots \ x_n]$ $\quad$ $y = [y_1 \ y_2 \ \dots \ y_n]$

$$x + y = [x_1+y_1 \ \ x_2+y_2 \ \dots \ x_n+y_n]$$

$$x - y = [x_1-y_1 \ \ x_2-y_2 \ \dots \ x_n-y_n]$$

$$\lambda . x = [\lambda x_1 \ \ \lambda x_2 \ \dots \ \lambda x_n]$$

* Every vector must have a magnitude and direction. Magnitude can be calculated in 2 different ways, L1 norm (aka taxicab) and L2 norm (Euclidean).
* We can add and subtract vectors from each other and multiply a vector with a scaler as if we're doing the math operations on regular numbers.

The Dot Product

$x = [x_1 \ x_2 \ \dots \ x_n]$ $\qquad$ $y = [y_1 \ y_2 \ \dots \ y_n]$

$$x \cdot y = (x_1 \cdot y_1) + (x_2 \cdot y_2) + \dots + (x_n \cdot y_n)$$

$$\langle x, y \rangle = x \cdot y$$

Transposing a vector / matrix

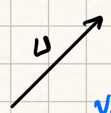
$$x = \begin{bmatrix} 1 & 4 & 8 \end{bmatrix} \Rightarrow x^T = \begin{bmatrix} 1 \\ 4 \\ 8 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \Rightarrow A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

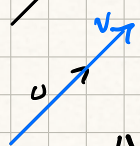
Geometric Dot Product

* Orthogonal vectors have dot product 0

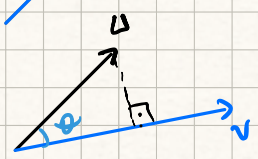
$$\langle u, v \rangle = 0$$



$$\langle u, u \rangle = |u|^2 = |u| \cdot |u|$$



$$\langle u, v \rangle = |u| \cdot |v|$$



$$\langle u, v \rangle = |u| \cdot |v| \cdot \cos(\theta)$$

$$\langle u, v \rangle > 0 \Rightarrow 0 < \theta < 90$$

$$\langle u, v \rangle < 0 \Rightarrow 90 < \theta < 180$$

The dot product is a mathematical operation that takes two vectors and returns a scalar, representing the projection of one vector onto another, or, equivalently the sum of the products of their corresponding components.

Multiplying a matrix by a vector

Equations as dot product

$a + b + c = 10$

$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 10$

$a + 2b + c = 15$

$\begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 15$

$a + b + 2c = 12$

$\begin{bmatrix} 1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 12$

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 10$$

$$\begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 15$$

$$\begin{bmatrix} 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 12$$

$$\underset{3\times 3}{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}} \cdot \underset{3\times 1}{\begin{bmatrix} a \\ b \\ c \end{bmatrix}} = \underset{3\times 1}{\begin{bmatrix} 10 \\ 15 \\ 12 \end{bmatrix}}$$

$3 \xleftrightarrow{=} 3$

Number of colums of the matrix has to match the number of rows of the vector.

$$\underset{4\times 3}{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{bmatrix}} \cdot \underset{3\times 1}{\begin{bmatrix} a \\ b \\ c \end{bmatrix}} = \underset{4\times 1}{\begin{bmatrix} 10 \\ 15 \\ 12 \\ 13 \end{bmatrix}}$$

To multiply a matrix and a vector, # of columns of the matrix must be equal to the # of rows of the vector. The i-th element of the resulting vector is the dot product of the i-th row of the matrix and the input vector.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

Linear Transformations

11.03.2025

Linear Transformation

= function

input vector $\rightarrow \boxed{f} \rightarrow$ output vector

1. Lines must remain lines
2. Origin must remain fixed in place.

Linear transformation as a matrix-vector multiplication

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x \cdot \begin{bmatrix} a \\ c \end{bmatrix} + y \begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

$\begin{bmatrix} a \\ c \end{bmatrix}$ is where $\hat{\imath}$ lands

$\begin{bmatrix} b \\ d \end{bmatrix}$ is where $\hat{\jmath}$ lands

$\begin{bmatrix} x \\ y \end{bmatrix}$ is the input vector

"We applied a linear transformation on the $[x \;\; y]$ vector. The linear transformation we applied can be shown as a matrix: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$ tells us that:

① $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ goes to $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$

② $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ goes to $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

Linear transformations as matrix multiplication

Let's add a new transformation to the new basis

$$\begin{bmatrix} 2 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix} \qquad \begin{bmatrix} 2 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$$

* A linear transformation is a function between two vector spaces that preserves the operations of vector addition and scalar multiplication.
  * $T(u+v) = T(u) + T(v)$
  * $T(c \cdot u) = c \cdot T(u)$
* Linear transformation maintains the "straightness" of lines and keep the origin fixed.
* We can show linear transformations as matrices. Linear Transformations can be applied on vectors as well as matrices.

This means $\Rightarrow$

$$\underbrace{\begin{bmatrix} 2 & -1 \\ 0 & 2 \end{bmatrix}}_{\text{SECOND! Transformation Matrix}} \cdot \underbrace{\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}}_{\text{FIRST! Transformation Matrix}} = \begin{bmatrix} 5 & 0 \\ 2 & 4 \end{bmatrix}$$

$$\Updownarrow$$

How do we perform matrix multiplication?

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

# of rows don't need to match.
# of columns of matrix 1 has to match # of rows of the matrix 2

$$\begin{bmatrix} 3 & 1 & 4 \\ 2 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 & 1 & -2 \\ 1 & 5 & 2 & 0 \\ -2 & 1 & 4 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 9 & 21 & -6 \\ 1 & -3 & 8 & -4 \end{bmatrix}$$

$2 \times 3 \quad = \quad 3 \times 4 \qquad\qquad 2 \times 4$

What happens if we use an identity matrix as the transformation matrix?

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix}$$

What is an inverse matrix?

$$\begin{bmatrix} \text{Matrix} \end{bmatrix} \cdot \begin{bmatrix} \text{Inverse Matrix} \end{bmatrix} = \begin{bmatrix} \text{Identity Matrix} \end{bmatrix}$$

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \cdot \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}^{-1}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$3a + 1c = 1$$
$$3b + 1d = 0$$
$$1a + 2c = 0$$
$$1b + 2d = 1$$

$$\left[\begin{array}{cccc|c} 3 & 0 & 1 & 0 & 1 \\ 0 & 3 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 \end{array}\right]$$

$a = 2/5$
$b = -1/5$
$c = -1/5$
$d = 3/5$

Matrices and linear transformations in Natural Language Processing

Spam	Lottery	Win
Yes	1	1
Yes	2	1
No	0	0
Yes	0	2
No	0	1
No	1	0
Yes	2	2
Yes	2	0
Yes	1	2

Scores:

Lottery = ... points / occurance

Win = ... points / occurance

Rule:

If the number of points of the sentence is bigger than ......, then the email is spam.

Goal:

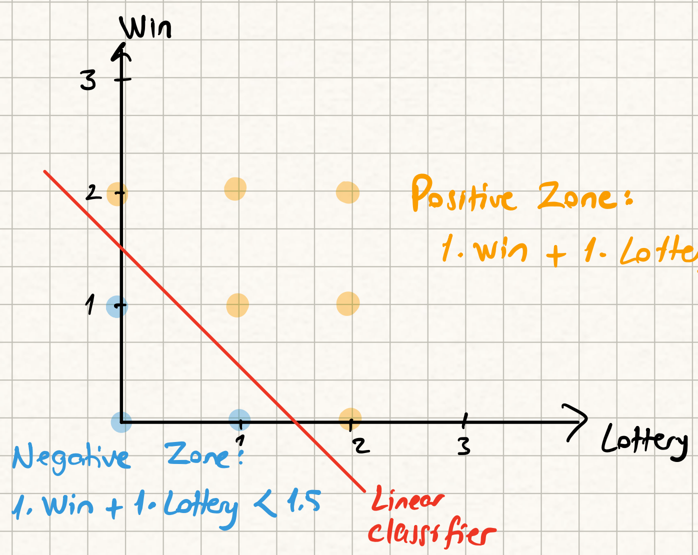
Find the best points and threshold.

Inverse matrix as a linear transformation transforms the plane to its original state. Only non-singular matrices have inverses. They are also called invertible.

a) Lottery: 1 pt, Win: 2 pts, Threshold: 3 pts.

b) Lottery: 1 pt, Win: 1 pt, Threshold: 1.5 pts.

c) Lottery: 2 pts, Win: 3 pts, Threshold: 1.5 pts.



One-Layer Neural Network Model

Input Data

$$\begin{bmatrix} \#\text{word Lottery} & \#\text{word Win} \end{bmatrix}$$

Model

$$\begin{bmatrix} \text{Pts. for Lottery} \\ \text{Pts. for win} \end{bmatrix} > \text{Threshold?}$$

$$\begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3 > 1.5 \quad \text{(spam)}$$

$$\begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \not> 1.5 \quad \text{(not spam)}$$

Data · Model = Product > 1.5? Target

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 0 & 0 \\ 0 & 2 \\ 0 & 1 \\ 1 & 0 \\ 2 & 2 \\ 2 & 0 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 0 \\ 2 \\ 1 \\ 1 \\ 4 \\ 2 \\ 3 \end{bmatrix} > 1.5? \quad \begin{bmatrix} \text{YES} \\ \text{YES} \\ \text{NO} \\ \text{YES} \\ \text{NO} \\ \text{NO} \\ \text{YES} \\ \text{YES} \\ \text{YES} \end{bmatrix}$$

$1 \cdot \text{Win} + 1 \cdot \text{Lottery} > 1.5$

$1 \cdot \text{Win} + 1 \cdot \text{Lottery} - 1.5 > 0$

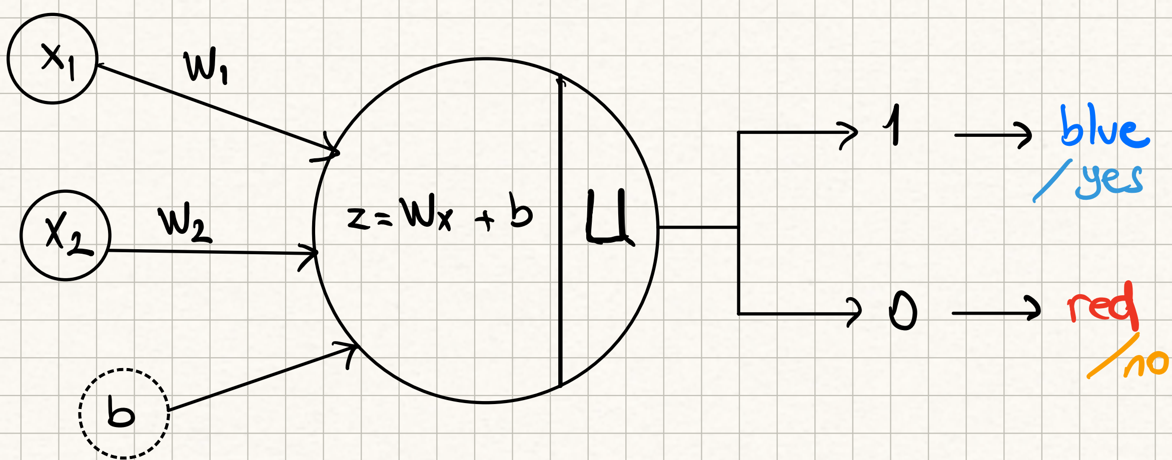
↳ BIAS

"AND" logic with linear tranformation.

AND	x	y
No	0	0
No	1	0
No	0	1
Yes	1	1

$$\cdot \begin{bmatrix} \text{Model} \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \text{Dot Product} \\ 0 \\ 1 \\ 1 \\ 2 \end{bmatrix} > 1.5 \,?$$

How does a single perceptron neural network work?



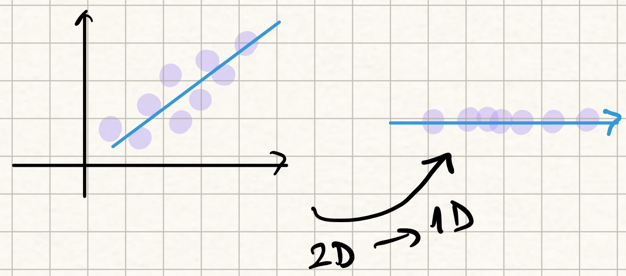
A perceptron is the simplest type of artificial neural network, designed to model a single neuron. It takes multiple inputs, each with an associated weight; it calculates a weighted sum of these inputs; it applies an activation function (typically a step function) to the weighted sum; and outputs a binary value indicating the category the input belongs to. Perceptrons use the concept of a linear transformation, which can be expressed as a matrix multiplication, to calculate its output and make decisions.

Determinants and Eigenvectors

13.03.2025

What is our motivation for learning determinants and eigenvectors?

Principal Component Analysis (PCA)



Goal

- ↓ dimensions of d. set
- Preserv. as much info as. psble.

NON-SINGULAR TRANSFORMATION

Non-singular matrix → Linear Transform. → Covers all the plane

SINGULAR TRANSFORMATION

Singular Matrix → Linear Transform. → A line or a dot! Doesn't cover all the plane

Rank of the lin. transf. MX = # dimensions of the result (the image)

Determinant of the lin. transf. MX = Area of the image

Non-singular lin transf. mx → Det ≠ 0

Singular lin. transf. mx → Det = 0

Determinant of a product

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ -3 & 3 \end{bmatrix}$$

det = 5   det = 3   det = 15

$5 \times 3 = 15$

$\det(A \cdot B) = \det(A) \cdot \det B$

Determinant of an inverse matrix

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 0.4 & -0.2 \\ -0.2 & 0.6 \end{bmatrix}$$

det = 5   det = 1/5

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

- We are going to use determinants and eigenvectors in order to perform a Principal Component Analysis (PCA), which will help us to reduce the dimensions of a dataset as much as possible without losing significant amount of data.

What are bases?	(1) Define the fundamental directions of a vector space (2) Allow any point in that space to be reached through linear combinations of those vectors. (3) Must be linearly independent (4) Must span the entire vector space (5) Must be the minimum number of vectors that span the entire space
What's the relationship between the bases and the dimension of the vector space?	* If we have more vectors than the dimension of the space we're trying to span, we will always have a linearly dependent group. Ex: If we have a 2D plane, we can span all of it with 2 linearly independent vectors. If we add a 3rd vector, it will always be able to reached by the other two.

Let $v_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $v_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $v_3 = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$

Two dimensions
Three vectors

$$\alpha \cdot v_1 + \beta v_2 = v_3$$

$$\alpha \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$$

$$\begin{aligned} -\alpha + 2\beta &= -5 \\ \alpha + \beta &= 3 \end{aligned} \Rightarrow \begin{aligned} \beta &= -2/3 \\ \alpha &= 11/3 \end{aligned} \Rightarrow v_3 = \frac{11 \cdot v_1 - 2 v_2}{3}$$

Basis is a set of linearly independent vectors that can "build" any other vector in a vector space through linear combinations. It's like the core set of directions we need to navigate the space.

Span is the set of all possible vectors that can be created by taking linear combinations of a given set of vectors. It's the "area" or "volume" covered by those vectors.

What is an eigenbasis?

A special basis where the linear transformation acts by simply stretching the basis vectors. This means the output parallelogram's sides are parallel to the input parallelogram's sides.

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

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad A \cdot v_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \Rightarrow A \cdot v_1 = \lambda_1 \cdot v_1$$

$$\lambda_1 = 2$$

Eigenvalues

$$v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad A \cdot v_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \Rightarrow A \cdot v_2 = \lambda_2 \cdot v_2$$

Eigenvectors

$$\lambda_2 = 3$$

$$v_3 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad A \cdot v_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix} \Rightarrow A \cdot v_3 \neq \lambda_3 \cdot v_3$$

If $A \cdot v_1 = \lambda_1 \cdot v_1 \Rightarrow$ LESS WORK!

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\text{8 multiplications}} = \underbrace{2 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\text{2 multiplications}}$$

A special basis where the linear transformation acts by simply stretching the basis vectors. This means the output parallelogram's sides are parallel to the input parallelogram's sides.

A eugenbasis is a special basis where the linear transformation acts by simply stretching the basis vectors. This means the output parallelogram's sides are parallel to the input parallelogram's sides.

Even if $A \cdot v_3 \neq \lambda_3 \cdot v_3$ might be less work.

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \left[-3 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right] \Rightarrow$$

$$A \cdot v_3 = A \cdot \left[-3 \cdot v_1 + 2 \cdot v_2 \right]$$

$$A \cdot v_3 = -3 \cdot \underbrace{A \cdot v_1}_{\lambda_1 \cdot v_1} + 2 \cdot \underbrace{A \cdot v_2}_{\lambda_2 \cdot v_2} \Rightarrow$$

$$A \cdot v_3 = -3 \cdot 2 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \cdot 3 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow$$

$$= -6 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 6 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

* $A \cdot v = \lambda \cdot v \Rightarrow \lambda =$ Eigenvalue , $v =$ Eigenvector
* Eigenvectors : Direction of the stretch.
* Eigenvalues : How much stretch?
* Eigenbasis : The set of a matrix's eigenvectors. Can be arranged as a matrix with one eigenvector in each column.
* They save work and characterize a transformation.

If λ is an eigenvalue, then

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$ for infinitely many (x,y)

$$\Rightarrow \begin{bmatrix} 2-\lambda & 1 \\ 0 & 3-\lambda \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$ has infinitely many solutions

$$\Rightarrow \det \begin{bmatrix} 2-\lambda & 1 \\ 0 & 3-\lambda \end{bmatrix} = 0 \Rightarrow (2-\lambda)\cdot(3-\lambda) - 1\cdot 0 = 0$$

$\lambda_1 = 2$

$\lambda_2 = 3$

$$\Rightarrow \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 2 \cdot \begin{bmatrix} x \\ y \end{bmatrix} \Rightarrow \begin{matrix} 2x + y = 2x \\ 0x + 3y = 2y \end{matrix} \Rightarrow \begin{matrix} x = 1 \\ y = 0 \end{matrix} \quad \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 3 \begin{bmatrix} x \\ y \end{bmatrix} \Rightarrow \begin{matrix} 2x + y = 3x \\ 0x + 3y = 3y \end{matrix} \Rightarrow \begin{matrix} x = 1 \\ y = 1 \end{matrix} \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

* If matrix is not square $\Rightarrow$ No eigenvalues / eigenvectors !

* $\det(A - \lambda I) = 0$ is called the characteristic equation. (det = determinant, A: transformat. matrix, λ: Eigenvalues, I: Identity matrix. We use it to find eigenvalues.

* Then we find the eugenvectors with this formula: $A \cdot v_i = \lambda_i \cdot v_i$

* Only square matrices can have eugenbases.

14.03.2025

If we have repeated eigenvalues, can we make assumptions about the # of eugen vectors?

NO!

$$A = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 4 & -0.5 \\ 0 & 0 & 2 \end{bmatrix} \qquad \det(A - \lambda I) = 0 \Rightarrow$$

$$(2-\lambda)(4-\lambda)(2-\lambda) + 0.(-0.5).0 + 0.(-1).0 - 0.4.0 + 0.(-1).2 + 2.(0.5).0 = 0$$

$\lambda_1 = 2$ $\quad \lambda_2 = 4$ $\quad \lambda_3 = 2$

$$\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

2 EIGEN VALUES
3 EIGEN VECTORS

point diff. directions

$$A = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 4 & -0.5 \\ 4 & 0 & 2 \end{bmatrix}$$

$\lambda_1 = 2$ $\quad \lambda_2 = 4$ $\quad \lambda_3 = 2$

$$\begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 0.5 \\ 2 \end{bmatrix}$$

2 EIGEN VALUES
2 EIGEN VECTORS

$\Downarrow$

3 Dimensions 2 Eigen vectors
NO EIGENBASIS
(cannot cover the whole space)

$[0 \ k \ 4k] \Rightarrow$ same direction

2x2

$\lambda_1 \neq \lambda_2 \Rightarrow$ 2 eigenvectors

$\lambda_1 = \lambda_2$ → 1 eigenvector
→ 2 eigenvectors

3x3

$\lambda_1 \neq \lambda_2 \neq \lambda_3 \Rightarrow$ 3 eigenvectors

$\lambda_1 = \lambda_2 \neq \lambda_3$ → 2 eigenvectors
→ 3 eigenvectors

$\lambda_1 = \lambda_2 = \lambda_3$ → 1 eigenvector
→ 2 eigenvectors
→ 3 eigenvectors

Why do we perform dimensionality reduction?

Dimensionality Reduction → Same # of rows, less # of columns.

Why?

- Smaller datasets
- Easier to visualize

How?

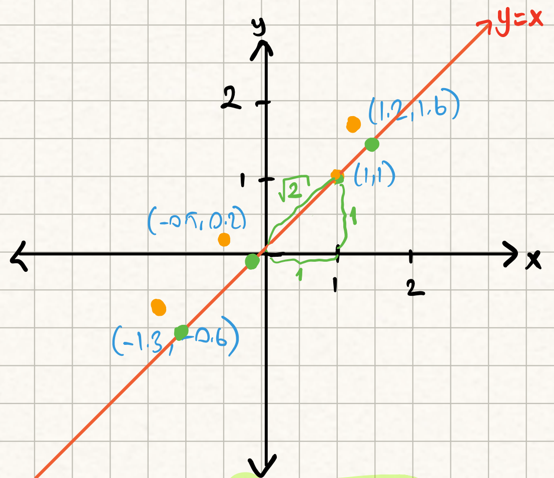
Why do we choose PCA to reduce dimensions?

- Delete columns? – Easy but probably loss of useful info.
- PCA – Harder but avoids losing too much useful info.

What is projection?

Projection: Moving our data points into a vector space with fewer dimensions.

	x	y
V_1	1.0	1.0
V_2	1.2	1.6
V_3	-0.5	0.2
V_4	-1.3	-0.6



projecting:

● → ●

(1,1) → $\sqrt{2}$

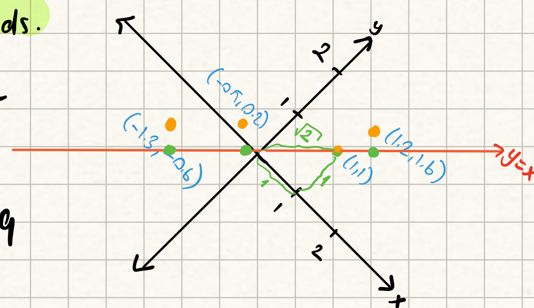
Final Coords.

$$V_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \frac{1+1}{\sqrt{2}} = 1.4142$$

$$V_2 \begin{bmatrix} 1.2 \\ 1.6 \end{bmatrix} \Rightarrow \frac{1.2+1.6}{\sqrt{2}} = 1.9799$$

$$V_3 \begin{bmatrix} -0.5 \\ 0.2 \end{bmatrix} \Rightarrow \frac{-0.5+0.2}{\sqrt{2}} = -0.2121$$

$$V_4 \begin{bmatrix} -1.3 \\ -0.6 \end{bmatrix} \Rightarrow \frac{-1.3-0.6}{\sqrt{2}} = -1.344$$



We reduce dimensions of our dataset by projecting our data points onto vectors from a vector space with fewer dimensions than our dataset.

What's the formula for projection?

To project a matrix A onto a vector v:

$$A_p = A \cdot \frac{v}{\|v\|_2}$$

$$\underset{r\times 1}{A_p} = \underset{r\times c}{A} \cdot \underset{c\times 1}{\frac{v}{\|v\|_2}}$$

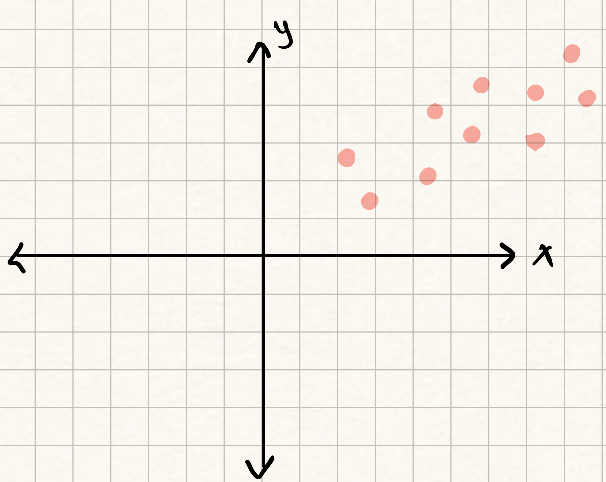
To project a matrix A onto vectors v_1 and v_2

$$A_p = A \cdot \underbrace{\begin{bmatrix} \frac{v_1}{\|v_1\|_2} & \frac{v_2}{\|v_2\|_2} \end{bmatrix}}_{} \rightarrow \text{matrix } V$$

$$r\times 2 \qquad r\times c \qquad c\times 2$$

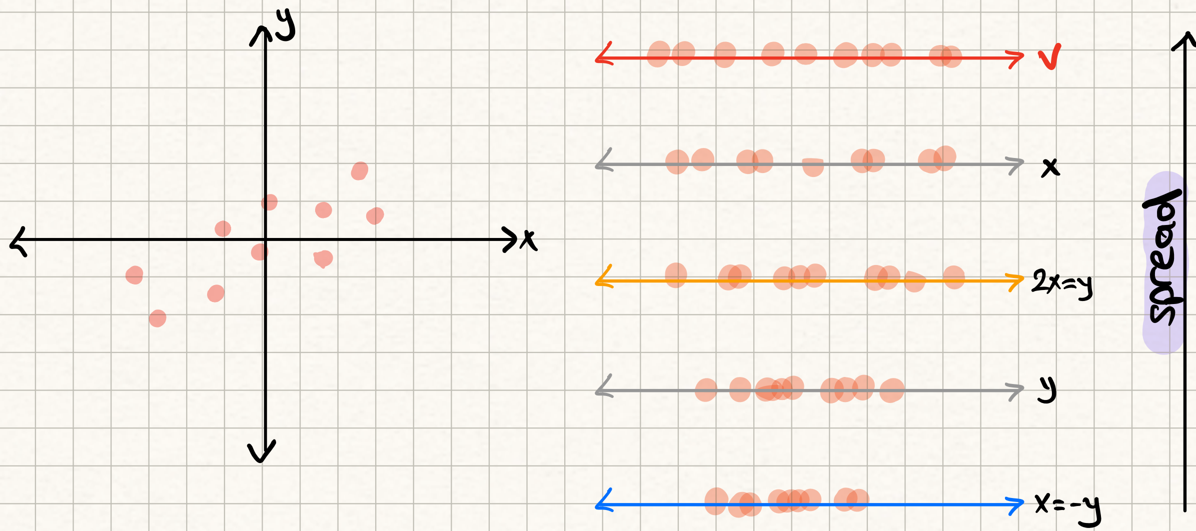
$$\boxed{A_p = A \cdot V}$$

How do we pick the vectors to project onto?



This is the dataset we want to apply PCA.

Let's start with centering our dataset around the origin:



more spread → preserving more info.

so how do we find v ?

spread = variance = "The average squared distance from the mean"

* μ = mean

$$\mathrm{var}(x) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu_x)^2$$

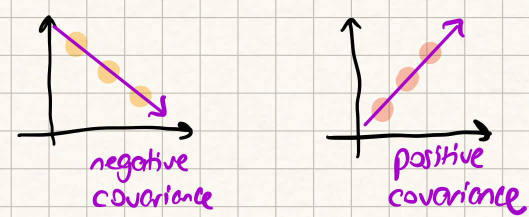
* $x_i - \mu_x$ gives coords for the centered data

Problem !

Both datasets have the same variance !

Solution :

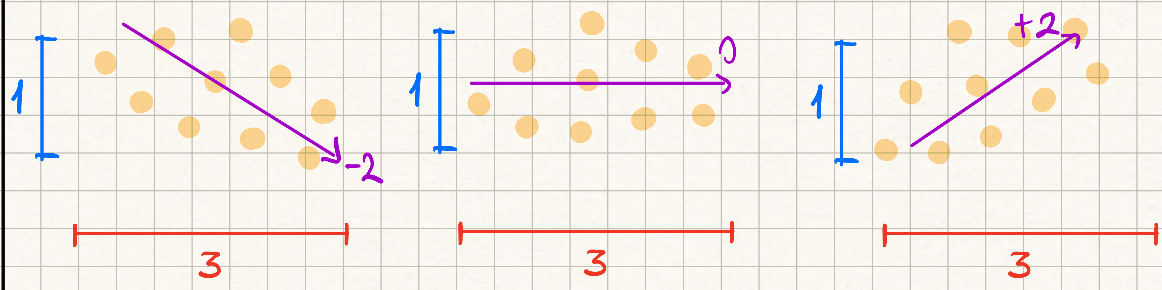
Covariance !



Covariance tells us the direction of the relationship between two variables.

$$\mathrm{Cov}(x,y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu_x) \cdot (y_i - \mu_y)$$

· Our goal is to choose a projection matrix that will give us the maximum spread when we project data onto it, because maximum spread means minimum loss of information.



All datasets have the same variances on the x-axis (3) and y-axis (1). However, their covariances differ.

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

$$C = \begin{bmatrix} var(x) & cov(x,y)^* \\ cov(y,x)^* & var(y) \end{bmatrix} \Rightarrow$$

$*\, cov(x,y) = cov(y,x)$

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

$*\, var(a) = cov(a,a)$

$$A = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_x & \mu_y \\ \mu_x & \mu_y \\ \vdots & \vdots \\ \mu_x & \mu_y \end{bmatrix} \Rightarrow$$

$$C = \frac{1}{n-1} (A - \mu)^T (A - \mu) \Rightarrow$$

A covariance matrix is a square matrix that summarizes the variances and covariances of features in a dataset. The eugen vectors for the n number of largest eugen values for dataset is used to find the projection matrix that will give us the largest spread (and minimum data loss).

$$= \frac{1}{n-1}\left(\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix} - \begin{bmatrix} \mu_x & \mu_y \\ \mu_x & \mu_y \\ \vdots & \vdots \\ \mu_x & \mu_y \end{bmatrix}\right)^T \cdot \left(\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix} - \begin{bmatrix} \mu_x & \mu_y \\ \mu_x & \mu_y \\ \vdots & \vdots \\ \mu_x & \mu_x \end{bmatrix}\right)$$

$$= \frac{1}{n-1}\begin{bmatrix} x_1-\mu_x & y_1-\mu_y \\ x_2-\mu_x & y_2-\mu_y \\ \vdots & \vdots \\ x_n-\mu_x & y_n-\mu_y \end{bmatrix}^T \cdot \begin{bmatrix} x_1-\mu_x & y_1-\mu_y \\ x_2-\mu_x & y_2-\mu_y \\ \vdots & \vdots \\ x_n-\mu_x & y_n-\mu_y \end{bmatrix}$$

$$= \frac{1}{n-1}\begin{bmatrix} x_1-\mu_x & x_2-\mu_x & \cdots & x_n-\mu_x \\ y_1-\mu_y & y_2-\mu_y & \cdots & y_n-\mu_y \end{bmatrix} \cdot \begin{bmatrix} x_1-\mu_x & y_1-\mu_y \\ x_2-\mu_x & y_2-\mu_y \\ \vdots & \vdots \\ x_n-\mu_x & y_n-\mu_y \end{bmatrix}$$

$$(x_1-\mu_x)(x_1-\mu_x)+(x_2-\mu_x)(x_2-\mu_x)+\dots+(x_n-\mu_x).(x_n-\mu_x)$$

$$\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i-\mu_x)^2 = var(x)$$

$$(x_1-\mu_x)(y_1-\mu_y)+(x_2-\mu_x)(y_2-\mu_y)+\dots+(x_n-\mu_x)(y_n-\mu_y)$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y) = cov(x,y)$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\mu_y)(x_i-\mu_x) = cov(y,x)$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\mu_y)^2 = var(y)$$

$$A-\mu = \begin{bmatrix} x_1-\mu_x & y_1-\mu_y \\ \vdots & \vdots \\ x_n-\mu_x & y_n-\mu_y \end{bmatrix} \qquad C = \frac{1}{n-1}(A-\mu)^T(A-\mu)$$

$$\underset{(2\times 8)}{A} = \begin{bmatrix} 10 & 5 \\ 12 & 3 \\ 6 & 9 \\ 6 & 4 \\ 5 & 11 \\ 14 & 2 \\ 8 & 1 \\ 3 & 13 \end{bmatrix} \Rightarrow A-\mu = \begin{bmatrix} 2 & -1 \\ 4 & -3 \\ -2 & 3 \\ -2 & -2 \\ -3 & 5 \\ 6 & -4 \\ 0 & -5 \\ -5 & 8 \end{bmatrix} \Rightarrow$$

$\mu_x = 8 \quad \mu_y = 6$

$$C = \frac{1}{8-1} \cdot \begin{bmatrix} 2 & 4 & -2 & -2 & -3 & 6 & 0 & 5 \\ -1 & -3 & 3 & -2 & 5 & -4 & -5 & 8 \end{bmatrix} \cdot \begin{bmatrix} 2 & -1 \\ 4 & -3 \\ -2 & 3 \\ -2 & -2 \\ -3 & 5 \\ 6 & -4 \\ 0 & -5 \\ -5 & 8 \end{bmatrix}$$

$$C = \begin{bmatrix} 14 & -11.86 \\ -11.86 & 19.71 \end{bmatrix}$$

1. Arrange data with a different feature in each column
2. Calculate column averages
3. Subtract each average from their respective column $(A-\mu)$
4. Apply : $\frac{1}{n-1} \cdot (A-\mu)^T \cdot (A-\mu)$

What are the 6 steps of PCA?

Covariance Matrix → PCA
Eigenvalues / Eigenvectors → PCA
Projections → PCA

We have n observations of 5 variables and our goal is to reduce our dataset to 2 variables.

1) Create Matrix

$$X = \begin{bmatrix} x_{1,1} & x_{2,1} & \cdots & x_{5,1} \\ x_{2,1} & x_{2,2} & \cdots & x_{5,2} \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,5} \end{bmatrix}$$

2) Center the Data

$$X-\mu = \begin{bmatrix} x_{1,1}-\mu_1 & x_{1,2}-\mu_2 & \cdots & x_{1,5}-\mu_5 \\ x_{2,1}-\mu_1 & x_{2,2}-\mu_2 & \cdots & x_{2,5}-\mu_5 \\ \vdots & \vdots & & \\ x_{n,1}-\mu_1 & x_{n,2}-\mu_2 & \cdots & x_{n,5}-\mu_5 \end{bmatrix}$$

3) Calculate the Covariance Matrix

$$C = \frac{1}{n-1} \cdot (X-\mu)^T \cdot (X-\mu)$$

4) Calculate Eigenvectors / Eigenvalues

$$\det(A - \lambda I) = 0$$

largest ⟷ smallest

λ_1 λ_2 λ_3 λ_4 λ_5

v_1 v_2 v_3 v_4 v_5

5) Create Projection Matrix

$$V = \begin{bmatrix} \frac{v_1}{\|v_1\|_2} & \frac{v_2}{\|v_2\|_2} \end{bmatrix}$$

6) Project the centered data onto the projection matrix

$$X_{PCA} = (X-\mu) \cdot V$$

NO SUMMARY: Learn the 6-step process of PCA through the instructions on this page!

17.03.2025

What is a discrete dynamic system?

Discrete → B/c not continuous. We're looking at changes at specific time points (e.g. each day, each click)

Dynamic → B/c it changes

What is a Markov Matrix? (M)

① All entries are non-negative (e.g. probabilities)

AND

② The columns add up to 1 (e.g. probabilities)

What is the equilibrium state?

State vector : Current state of the system : X_0

Next state : $X_1 = M \cdot X_0$

$$\Rightarrow \quad X_{n+1} = M \cdot X_n$$

- Eventually $X_{n+1} \cong X_n$

$$\Rightarrow \quad \lim_{x \to \infty} X_n = X_{eq}$$

- So $M \cdot X_{eq} = X_{eq}$

$\Rightarrow$ X_{eq} = Eigenvector of the matrix M !
(Equilibrium vector)

What is a transition matrix?

Markov matrix in a discrete dynamic system = Transition Matrix (P)