Welcome! We will begin shortly

Data Science Essentials - 101

The mathematics and statistics underlying the technology

Learning Outcomes

- Comprehend the importance of math and stats in data science
- Understand how linear algebra is used to represent data
- Understand how calculus is used to optimize solutions
- Understand how probability is used for decision making

Guidelines

Listen only mode

Ask questions at the interest of the larger audience

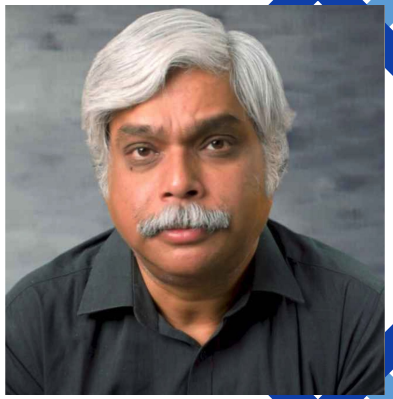
FAQ

Questions in the Q&A Box

Thank you

Kindly utilize the chat box for **subject-relevant questions only** to maximize your learnings from the session.
Your **questions** are being managed by the academic team, and they will be **answered**

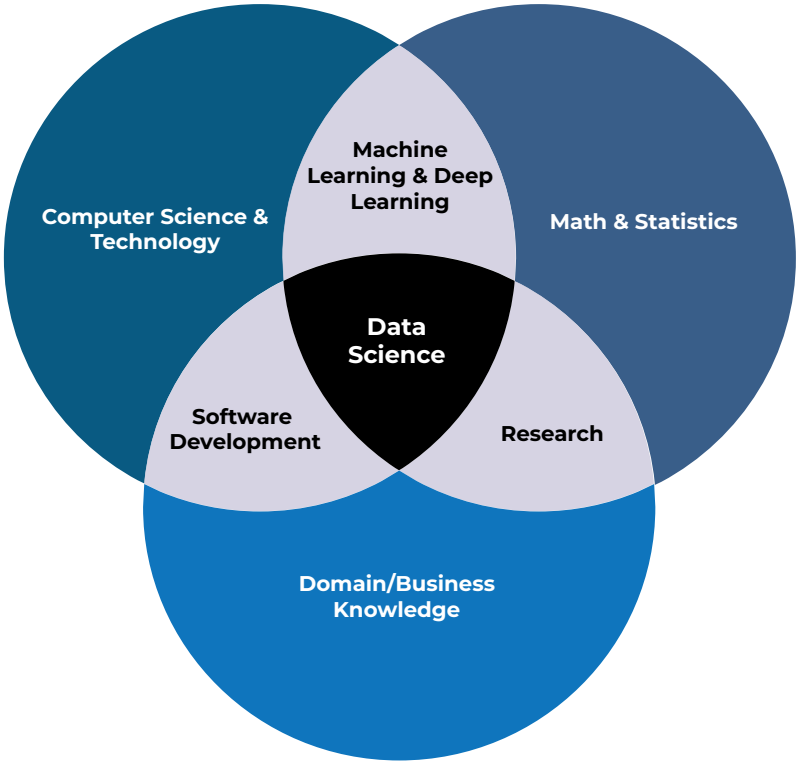
Meet Your Speaker

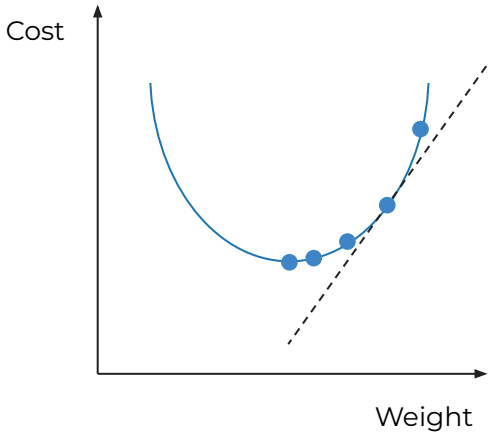


Dr. Abhinanda Sarkar
Academic Director at Great Learning

- Alumnus - Indian Statistical Institute, Stanford University
- Faculty - MIT, Indian Institute of Management, Indian Institute of Science
- Experienced in applying probabilistic models, statistical analysis and machine learning to diverse areas
- Certified Master Black Belt in Lean Six Sigma and Design for Six Sigma in GE

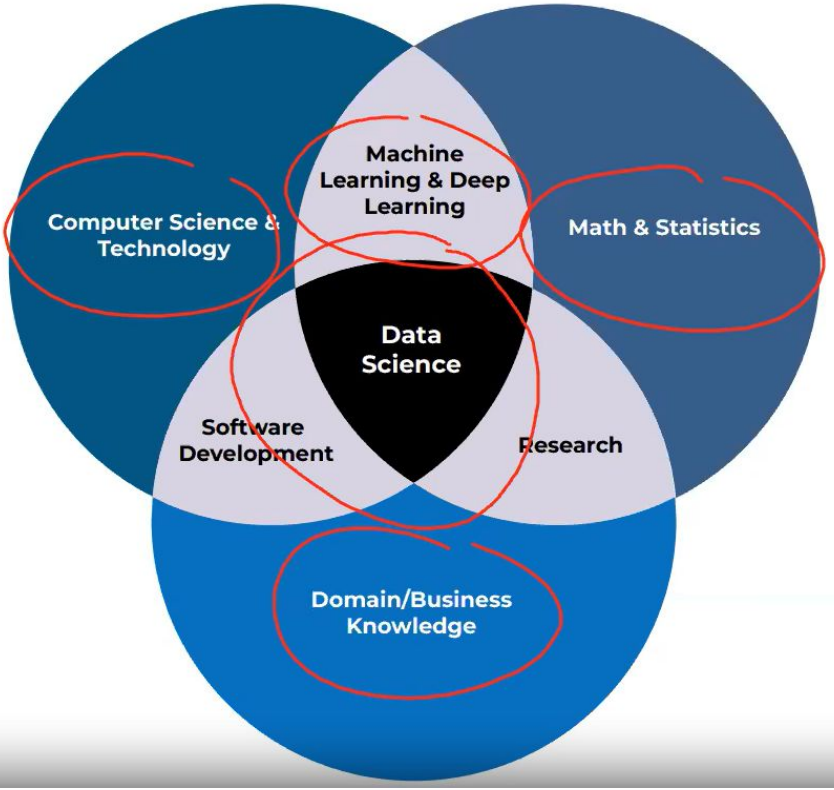
The need for math in data science

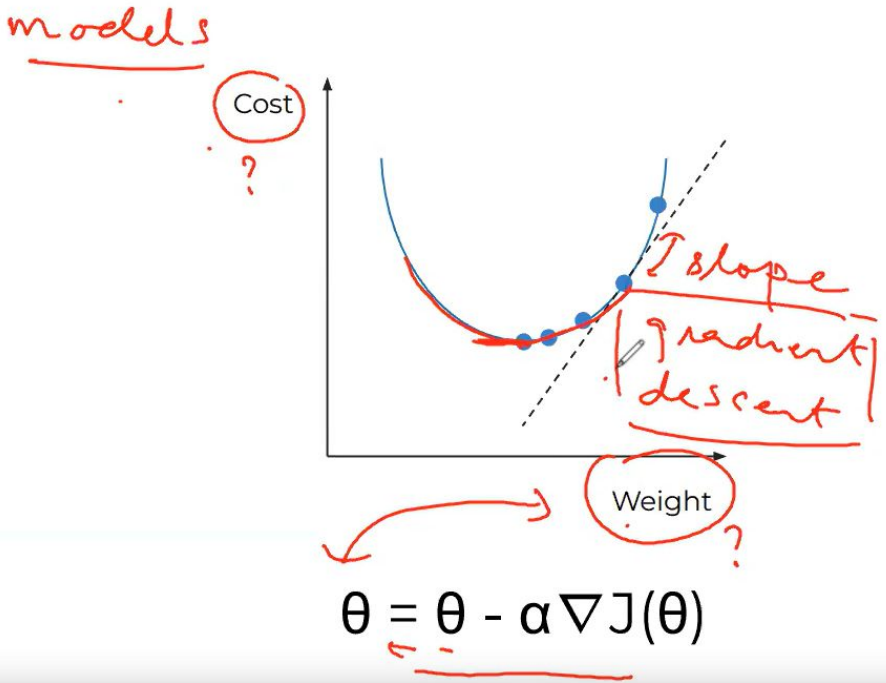




$$\theta = \theta - \alpha \nabla J(\theta)$$

The need for math in data science

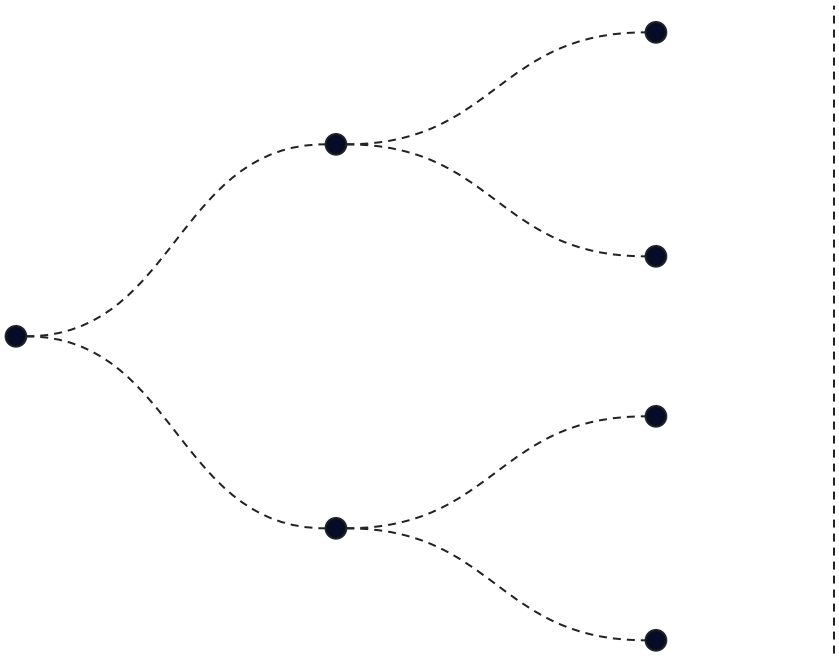




Probability & Descriptive Statistics

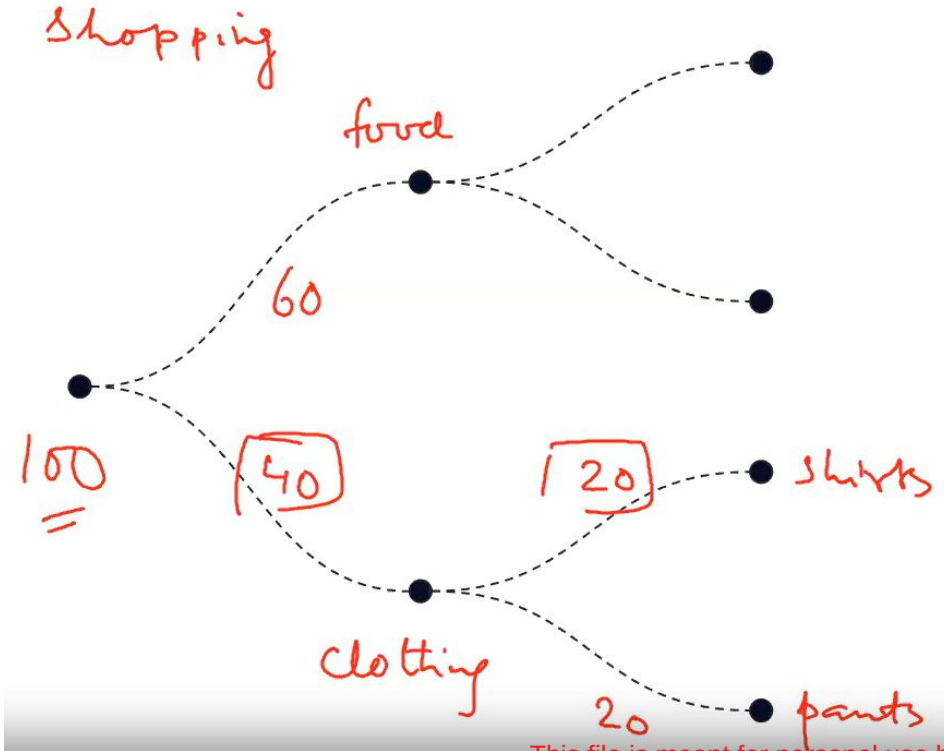
Bayes Rule

"Determine probability of a hypothesis based on prior knowledge and new evidence"



Bayes Rule

"Determine probability of a hypothesis based on prior knowledge and new evidence"



Prob (food) = 0.6

Prob (clothing) = 0.4

Prob (shirt | clothing) = 0.5

given

conditioning

Covid Example Demo

Covid test 95% accurate / correct

$P(+ \mid \text{disease}) = .95$

$P(- \mid \text{no disease}) = .95$

Given a +ve test result, what is prob that there is disease?

Prevalence 1%

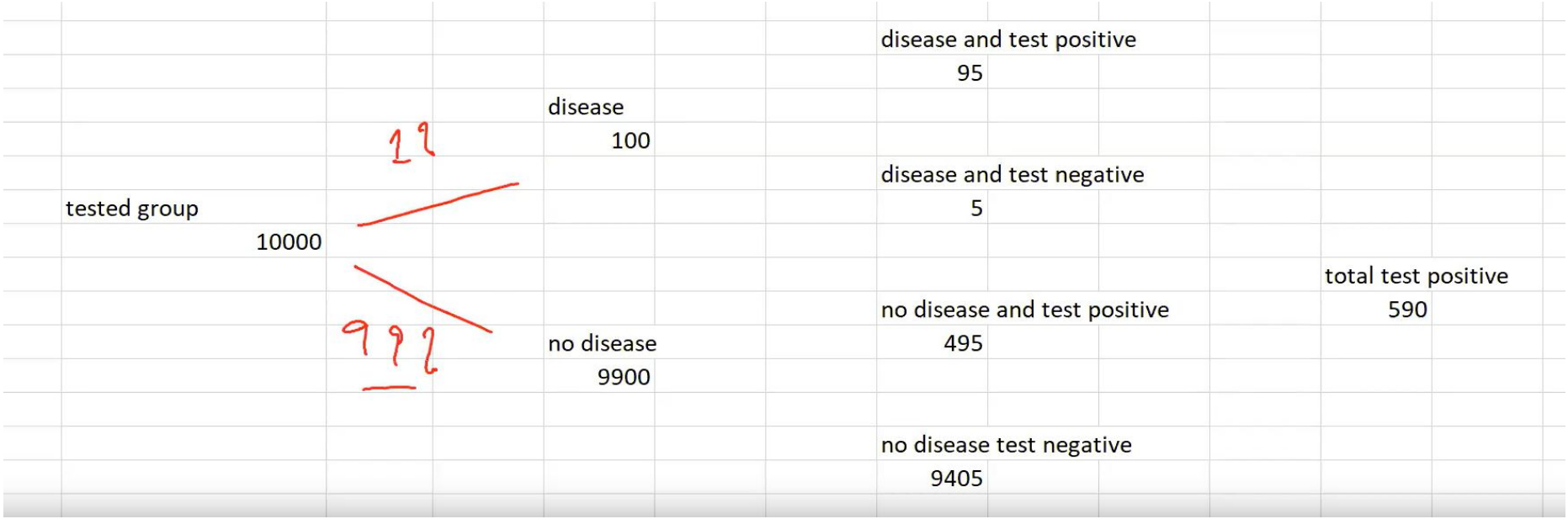
$P(\text{disease}) = .01$

$P(\text{disease} \mid +) = ?$

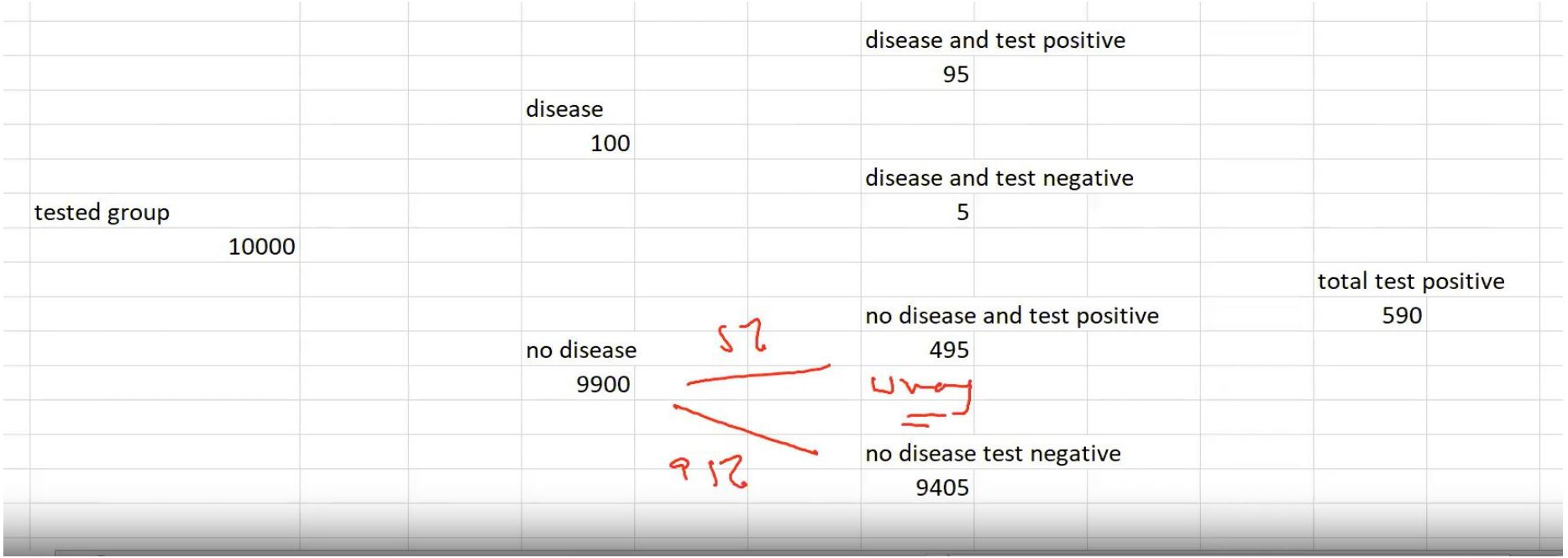
Covid Example Demo

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	P(test positive | disease)	sensitivity		0.95		P(disease | test positive) = ?				0.16102			
2	P(test negative | no disease)	specificity		0.95									
3	P(disease)	prevalence		0.01									
4													
5													
6								disease and test positive					
7								95					
8					disease								
9					100								
10								disease and test negative					
11		tested group						5					
12		10000											
13												total test positive	
14								no disease and test positive				590	
15					no disease			495					
16					9900								

Covid Example Demo



Covid Example Demo



Covid Example Demo

P(test positive | disease) sensitivity		0.95		P(disease | test positive) = ?		0.82609	
P(test negative | no disease) specificity		0.95					
P(disease) prevalence		0.2					
					disease and test positive		
					1900		
			disease				
			2000				
					disease and test negative		
	tested group				100		
		10000					
							total test positive
					no disease and test positive		2300
			no disease		400		
			8000				

Covid Example Demo

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	P(test positive | disease) sensitivity			0.95		P(disease | test positive) = ?				0.48969			
2	P(test negative | no disease) specificity			0.99									
3	P(disease) prevalence			0.01									
4													
5													
6								disease and test positive					
7								95					
8					disease								
9					100								
10								disease and test negative					
11		tested group						5					
12		10000											
13												total test positive	
14								no disease and test positive				194	
15					no disease			99					
16					9900								

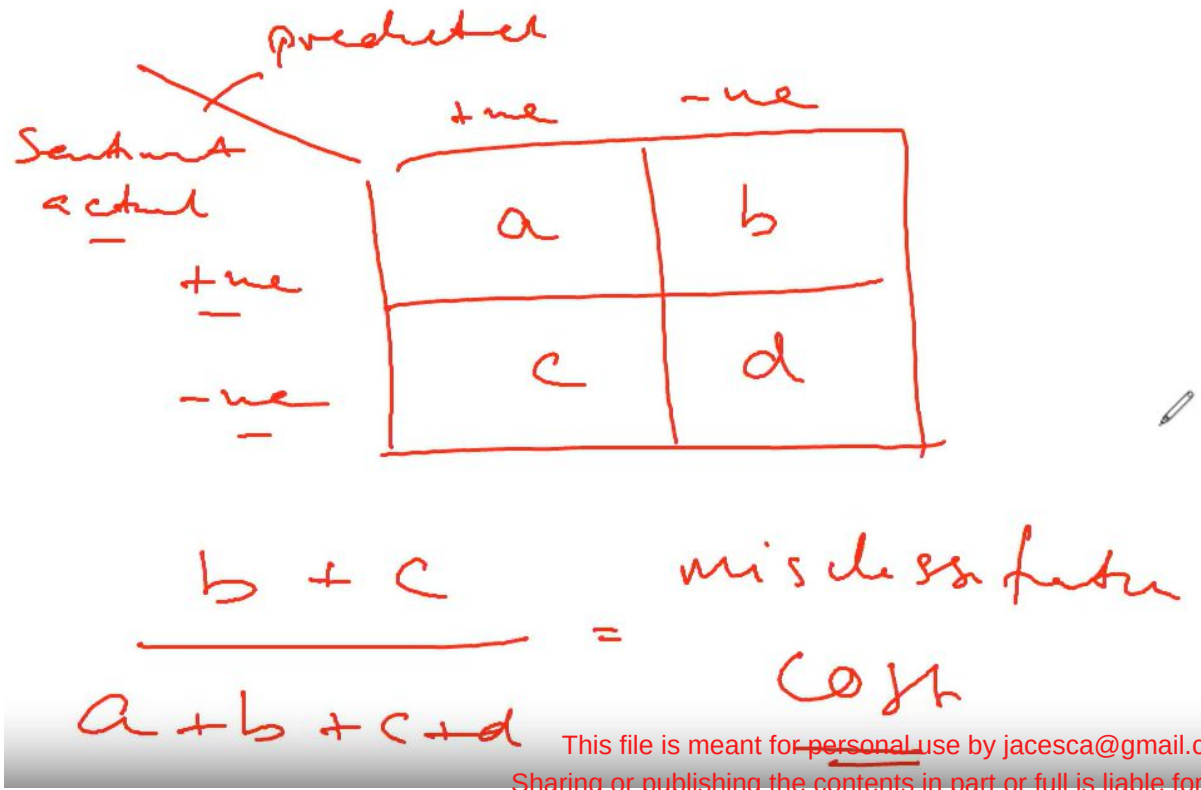
Covid Example Demo

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	P(test positive | disease) sensitivity			0.95		P(disease | test positive) = ?				0.90562			
2	P(test negative | no disease) specificity			0.999									
3	P(disease) prevalence			0.01									
4													
5													
6								disease and test positive					
7								95					
8					disease								
9					100								
10								disease and test negative					
11		tested group						5					
12		10000											
13												total test positive	
14								no disease and test positive				104.9	
15					no disease			9.9					
16					9900								

Covid Example Demo

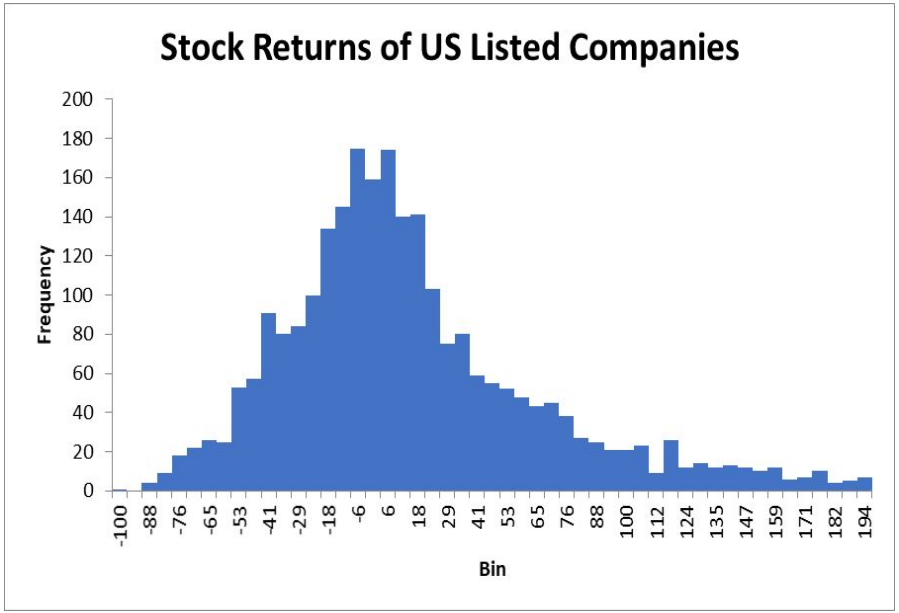
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	P(test positive | disease) sensitivity			0.95		P(disease | test positive) = ?				0.16102			
2	P(test negative | no disease) specificity			0.95									
3	P(disease) prevalence			0.01									
4													
5													
6								disease and test positive					
7								0.0095					
8					disease								
9					0.01								
10								disease and test negative					
11		tested group						0.0005					
12			1										
13												total test positive	
14								no disease and test positive				0.059	
15					no disease			0.0495					
16					0.99								

Questions?



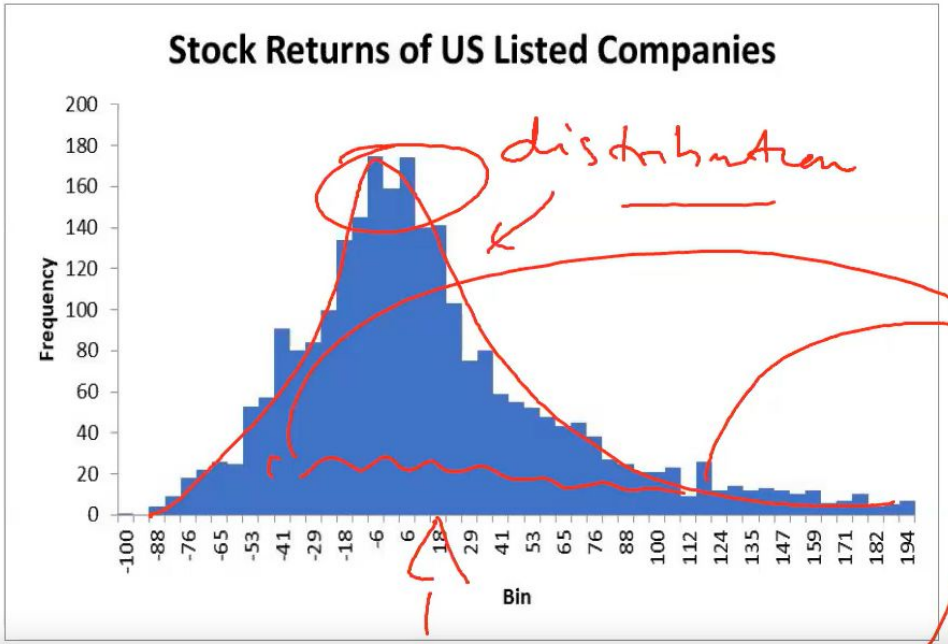
Descriptive Statistics

"Describing features of a dataset by generating summaries about data samples "



Descriptive Statistics

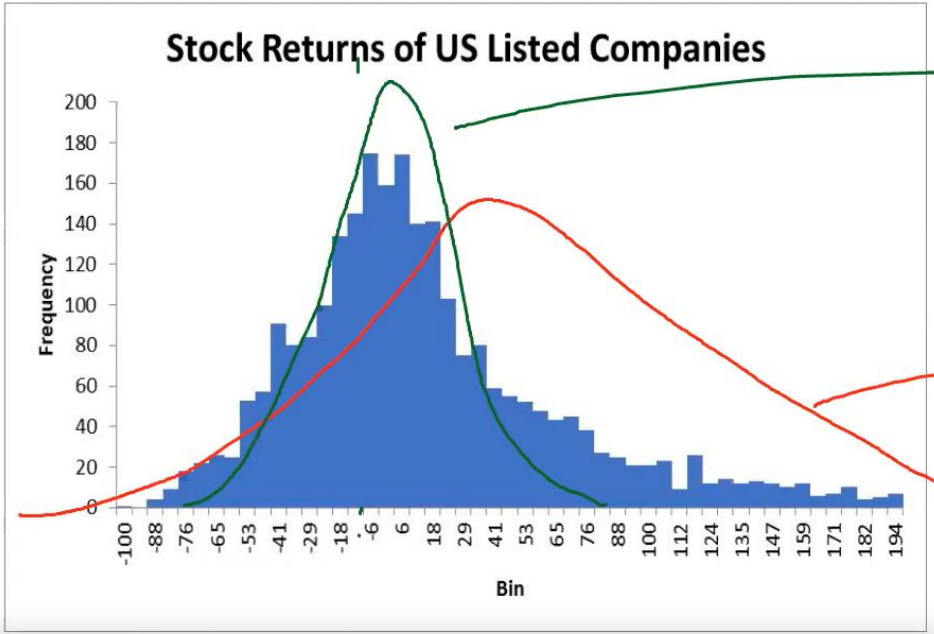
"Describing features of a dataset by generating summaries about data samples "



1:04:38 / 2:13:21

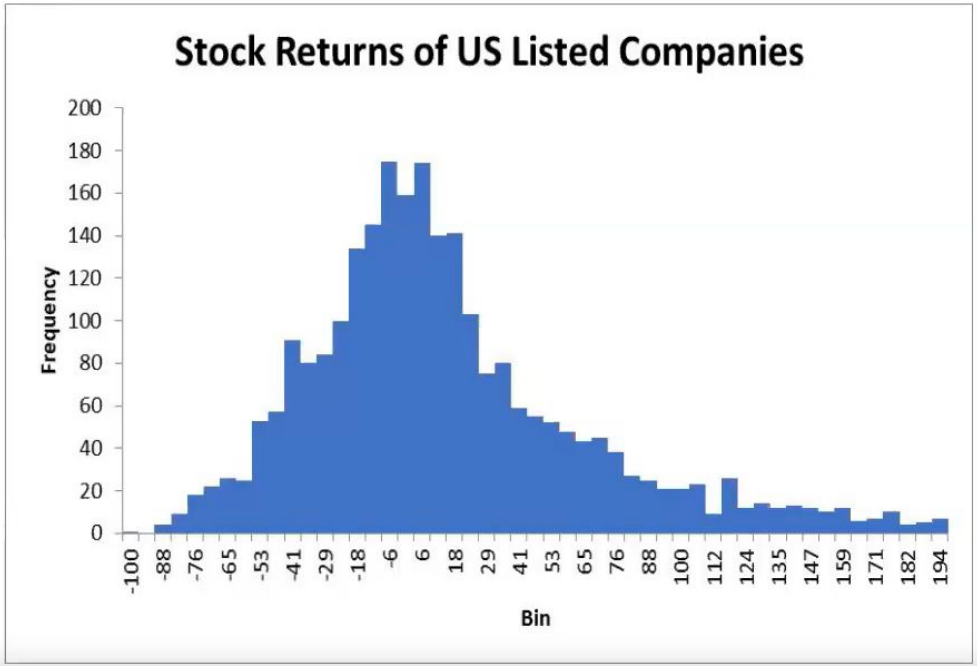
Descriptive Statistics

"Describing features of a dataset by generating summaries about data samples"



Descriptive Statistics

"Describing features of a dataset by generating summaries about data samples "



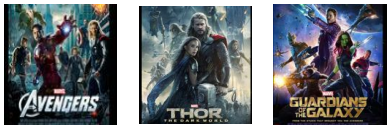
Matrices & Linear Algebra

A simple recommendation system

"Personalized recommendations to users based on their past behavior & preferences"

Approach - 1

You watched:



So I'll recommend:

Approach - 2

You are between age 20 to 30, prefer fantasy, action and science fiction

So I'll recommend:

A simple recommendation system

"Personalized recommendations to users based on their past behavior & preferences"

Approach - 1

You watched:

So I'll recommend:

Approach - 2

You are between age 20 to 30, prefer fantasy, action and science fiction

So I'll recommend:

items ← movies

users ← viewers

	it 1	it 2	. .
us 1	—	—	—
us 2	—	—	
1	—		

A simple recommendation system

"Personalized recommendations to users based on their past behavior & preferences"

Approach - 1

You watched:

So I'll recommend:

Approach - 2

You are between age 20 to 30, prefer fantasy, action and science fiction

So I'll recommend:

items ← movies

users ← viewers

	it 1	it 2	. .
us 1	—	—	—
us 2	—	—	
1	—		

A simple recommendation system

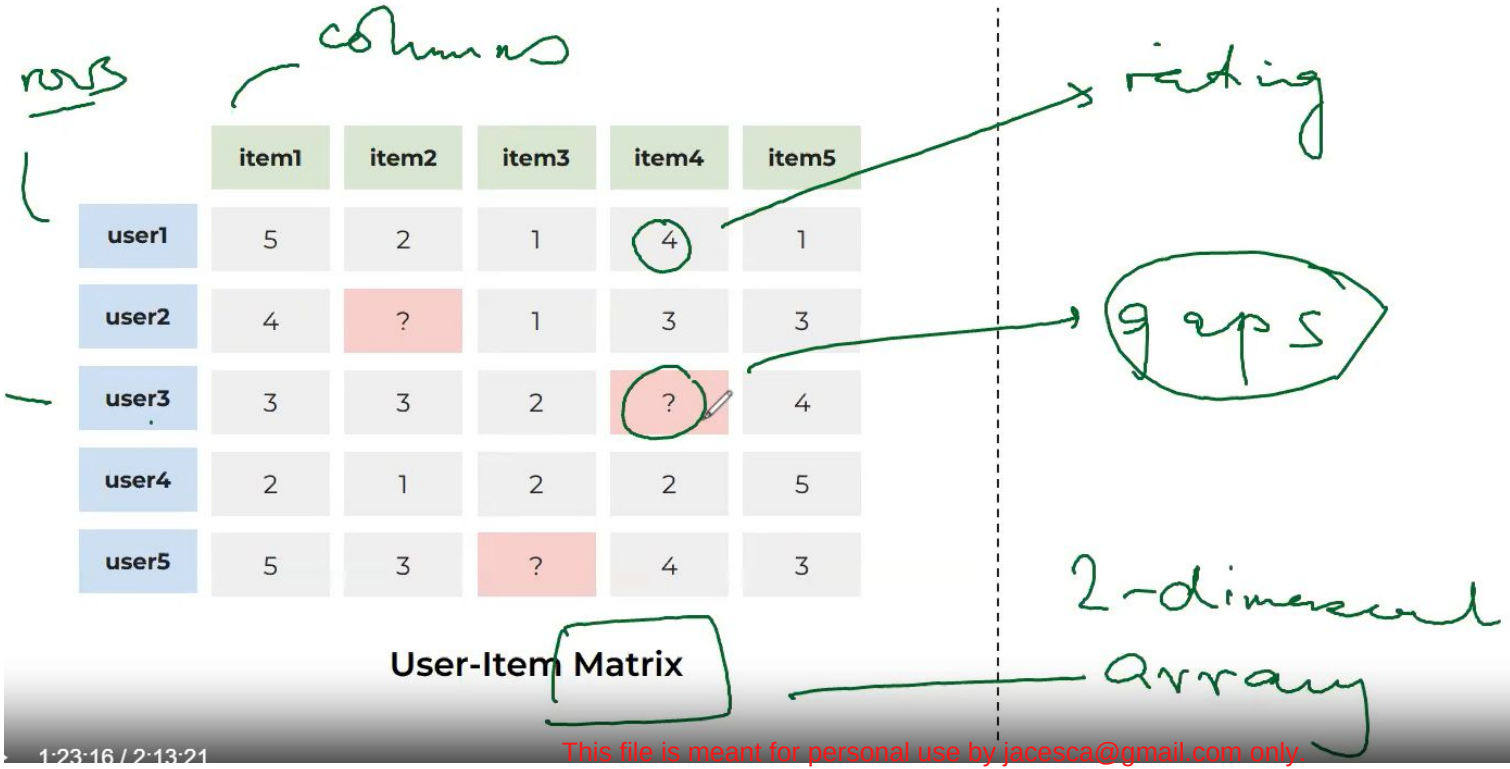
"Personalized recommendations to users based on their past behavior & preferences"

	item1	item2	item3	item4	item5
user1	5	2	1	4	1
user2	4	?	1	3	3
user3	3	3	2	?	4
user4	2	1	2	2	5
user5	5	3	?	4	3

User-Item Matrix

A simple recommendation system

"Personalized recommendations to users based on their past behavior & preferences"



	item1	item2	item3	item4	item5
user1	5	2	1	4	1
user2	4	?	1	3	3
user3	3	3	2	?	4
user4	2	1	2	2	5
user5	5	3	?	4	3

User-Item Matrix

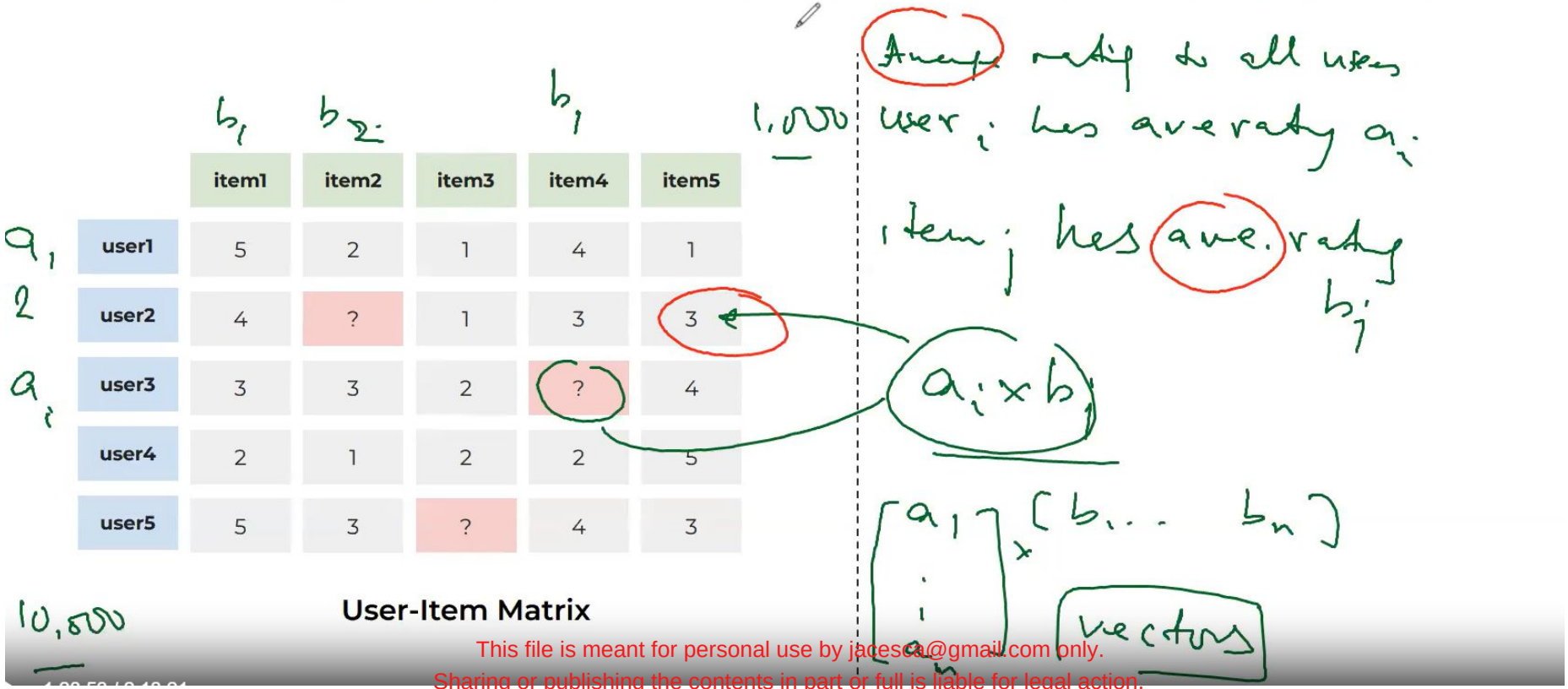
1:23:16 / 2:13:21

A simple recommendation system

"Personalized recommendations to users based on their past behavior & preferences"

	item1	item2	item3	item4	item5
user1	5	2	1	4	1
user2	4	?	1	3	3
user3	3	3	2	?	4
user4	2	1	2	2	5
user5	5	3	?	4	3

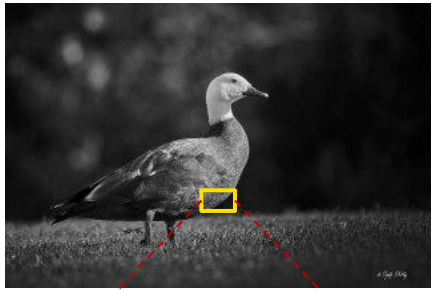
User-Item Matrix

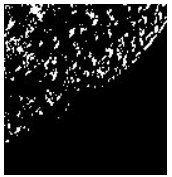


Questions?

Image Representation

"Convert an image into an array of numbers and each pixel has an intensity number"

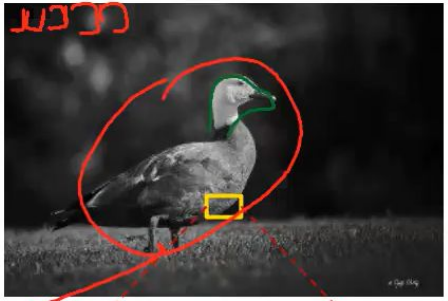




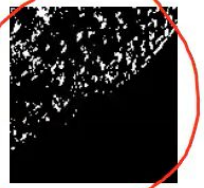
010	146	025
057	071	082
012	098	099

Image Representation

"Convert an image into an array of numbers and each pixel has an intensity number"



128 x 128



010	146	025
057	071	082
012	098	099

matrix

image is stored as a matrix

[ ]

128 x 128

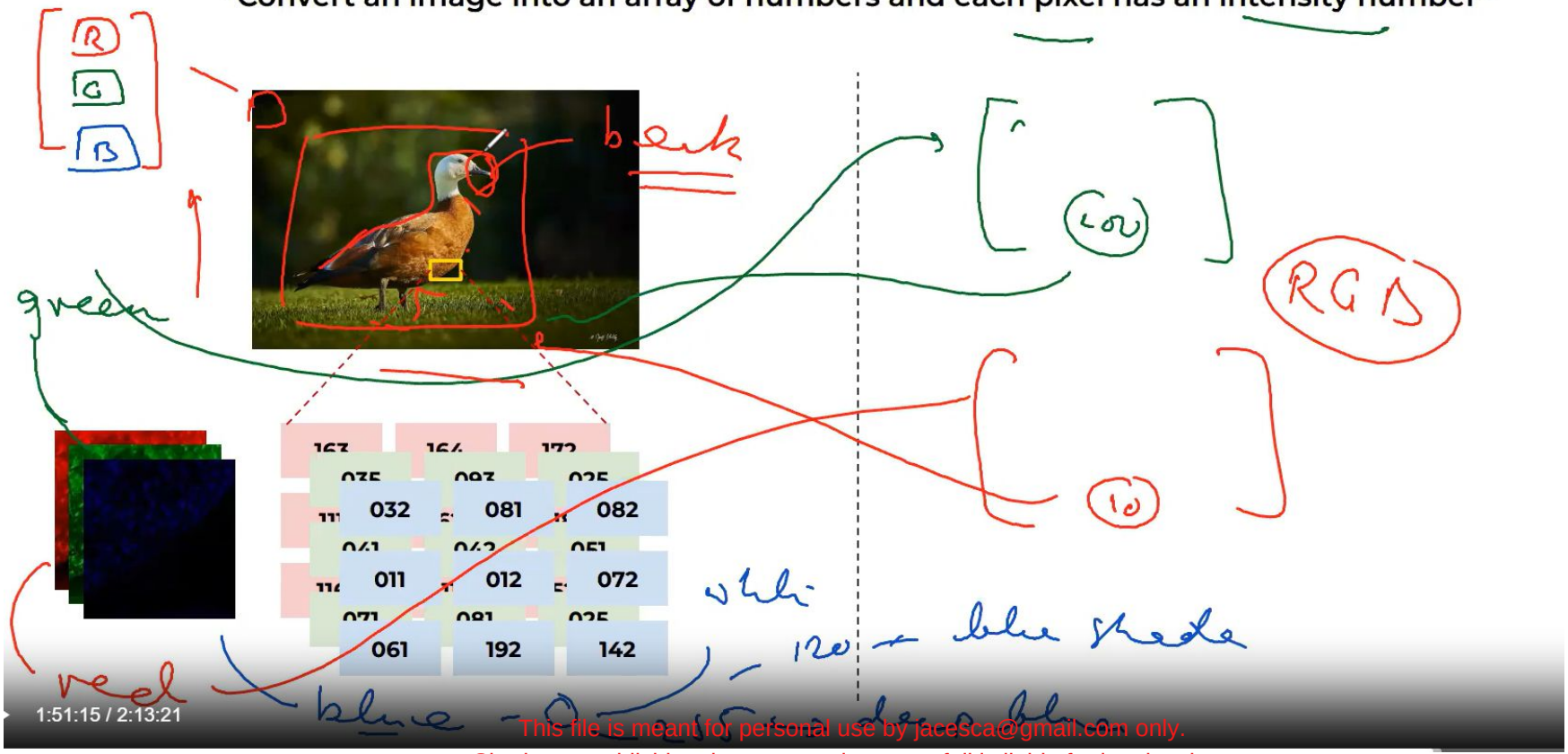
black → 0 - 255 ← white

Image Representation

"Convert an image into an array of numbers and each pixel has an intensity number"

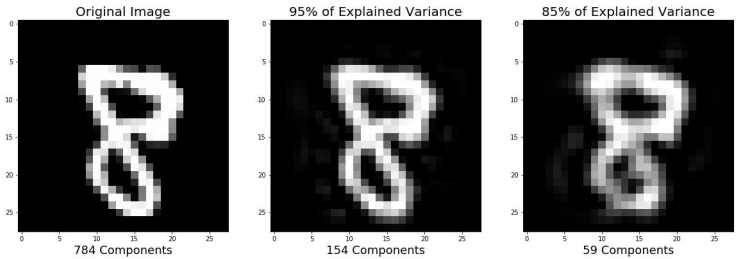
Image Representation

"Convert an image into an array of numbers and each pixel has an intensity number"



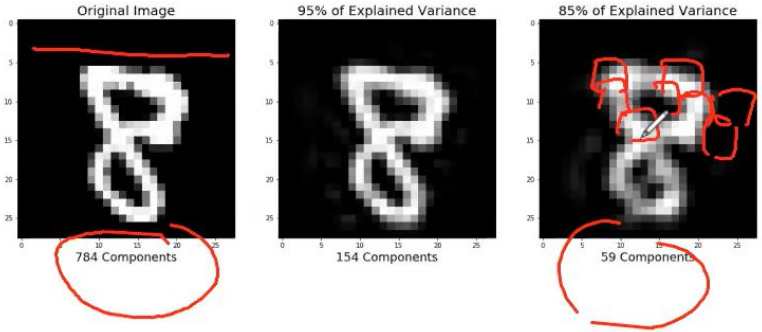
Dimension Reduction

"Increasing the interpretability of data while preserving the maximum amount of information"



Dimension Reduction

"Increasing the interpretability of data while preserving the maximum amount of information"



784 → 60 components

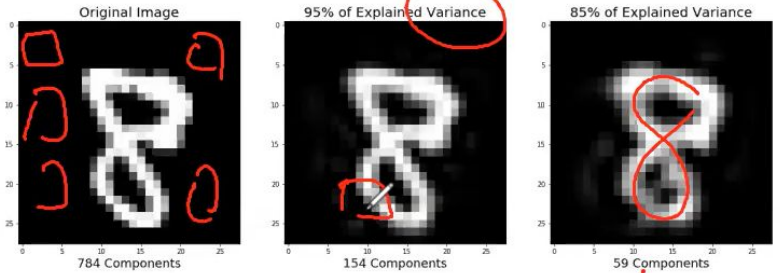
Which 60?

→ example.

[ ] replaces 8×8 by example

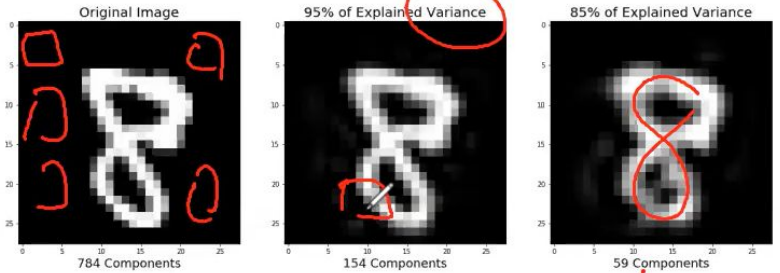
Dimension Reduction

"Increasing the interpretability of data while preserving the maximum amount of information"



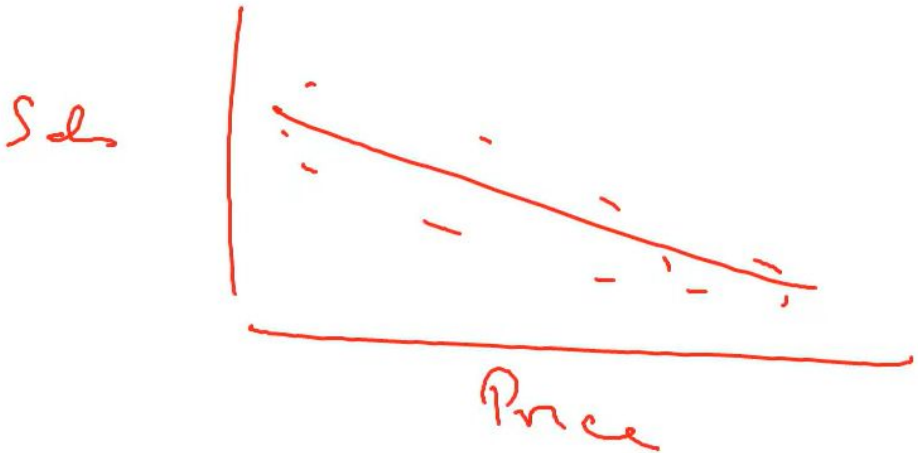
Dimension Reduction

"Increasing the interpretability of data while preserving the maximum amount of information"



Questions?

Questions?



$$Sales = 100 - 20 \times Price$$

Summary & Conclusion

Great Learning

Happy Learning !