Recommender Systems:

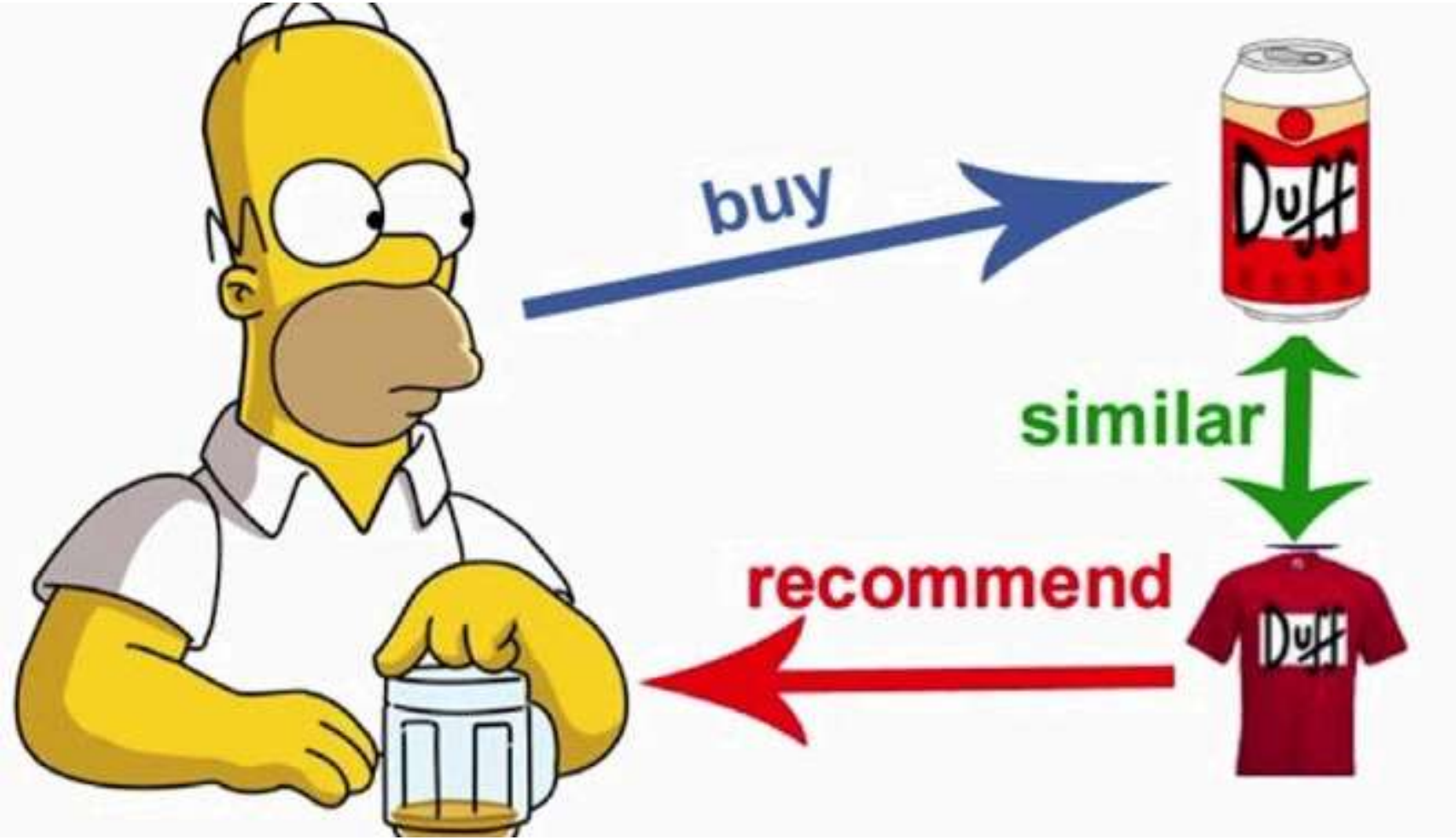
How Machines Discover Your Thoughts and Preferences

About This Presentation

This presentation originated with the lecture series entitled "Puzzling Problems in Science and Technology," devised for a ten-week, one-unit, freshman seminar course by Behrooz Parhami, Professor of Computer Engineering at University of California, Santa Barbara. The material can be used freely in teaching and other educational settings. Unauthorized uses, including any use for financial gain, are prohibited. © Behrooz Parhami

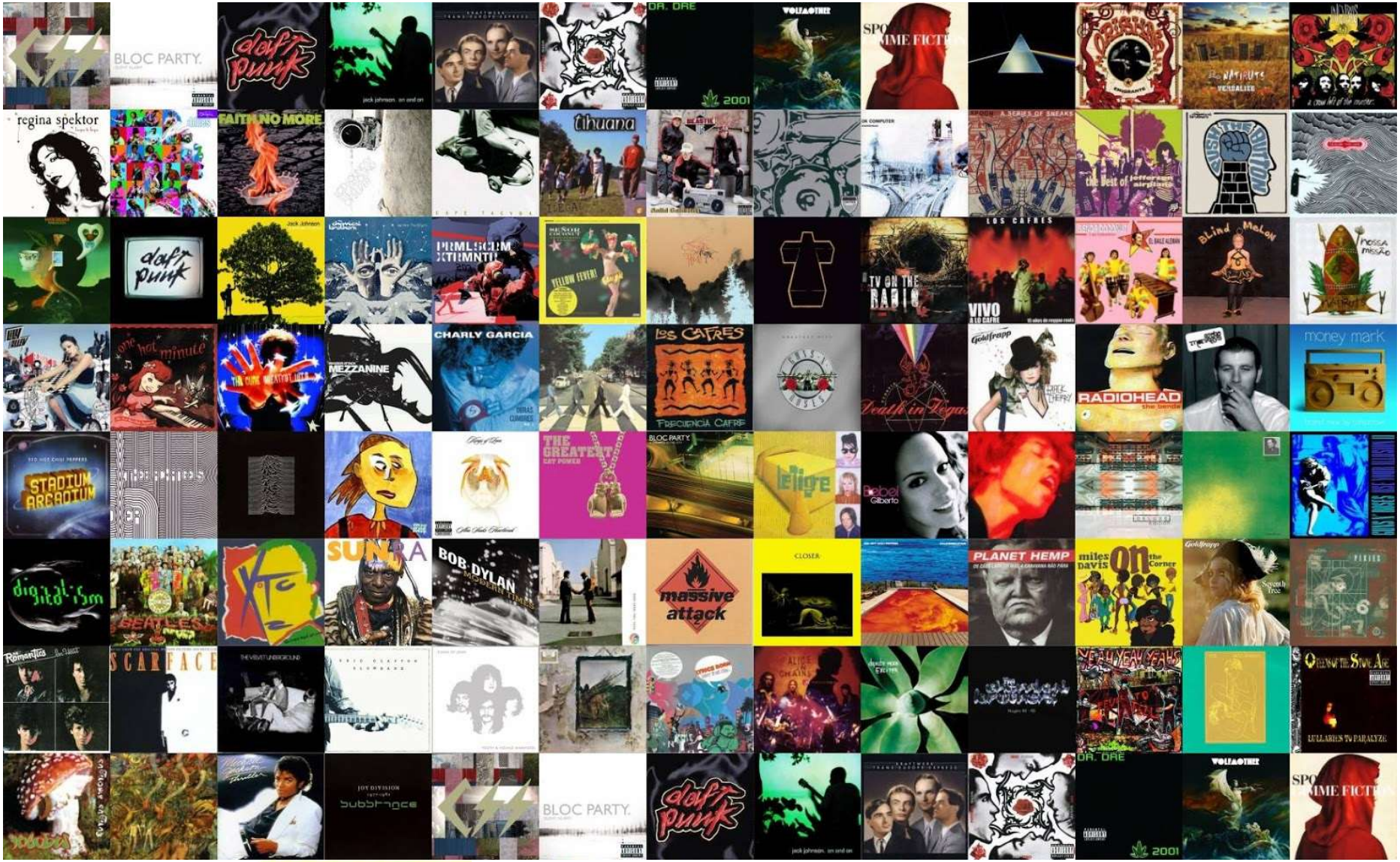
Edition	Released	Revised	Revised	Revised	Revised
First	Oct. 2016	Oct. 2018	Sep. 2020		

Discovering What You Want: Your Own Past Behavior



Discovering What You Want: Other People's Behaviors

Movies in a Collection that Likely Appeal to You



Trade-offs in Number/Type of Recommendations

Too many recommendations that you end up not choosing:
You lose confidence in the quality of recommendations

Too few recommendations:
Loss of business/sales opportunities

Serendipity: Random recommendation may open up a path

A good recommender system may find items for you that you might have never discovered on your own

A recommender system may suggest a movie or book the like of which you have never watched/read

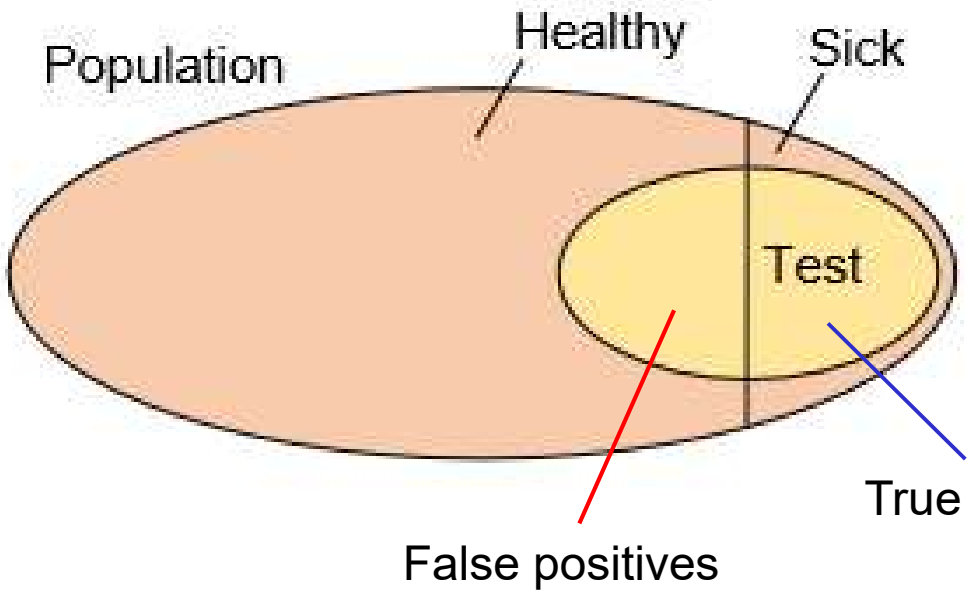
Brave New World: Recommendations based on your mood

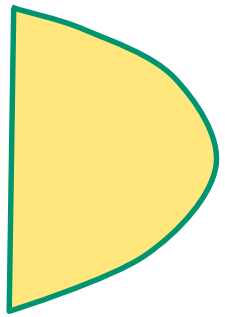
We Want Recommendations to Be Both Specific and Sensitive

Like medical diagnostic tests

Specificity: The fraction of healthy subjects correctly identified

Sensitivity: The fraction of sick subjects correctly identified





Ideal test:
Positive
= Sick

There is often a trade-off between specificity and sensitivity

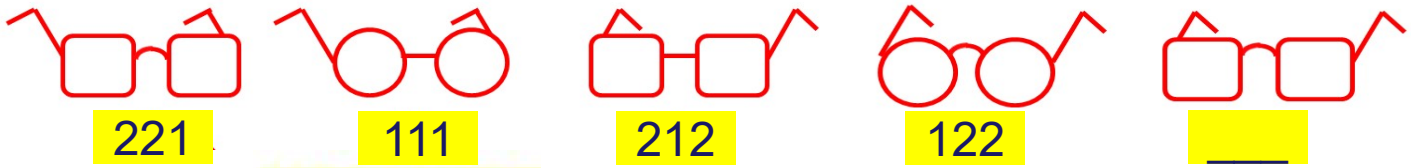
Find the Next Term in an Integer Sequence

1	2	3	4	__				
2	4	8	16	__				
1	2	2	3	3	3	4	4	__
1	3	5	7	9	__			
3	7	11	__					
10	15	19	22	24	__			
1	1	2	3	5	8	13	__	
1	4	9	16	__				
1	3	6	10	15	21	28	__	
0	1	2	…	8	0	1	2	__
1	2	3	4	9	27	512	__	[OEIS]

Online Encyclopedia of Integer Sequences: http://oeis.org/

Find Missing Term in an Arbitrary Sequence

Z	O	T	T	F	F	__		
J	F	M	__	M	J	J		
31	__	31	30	31	30	31	31	30
A	E	F	H	I	__			
3	3	5	4	4	3	5	__	
3	4	6	9	__	18	24		
1	3/2	__	7/8	9/16				
1	11	21	1211	111221		312211		__



Similarity Puzzles with Words

What do the following sets of words have in common?

assess; banana; dresser; grammar; potato; revive; uneven
(besides all having at least two repeating letters)

bulb; orange; angel; silver; month; revive; uneven

baobab; youngberry; hopscotch; yieldability; dachshund; dumbfounded

aquamarine; beloved; discrepancy; frangipani; freedom; gallipot; overflowing; pagoda; scrounger

Which Name Should Come Next?

Mark Susan Jeff Jenny Brad Marco Jill ___

Choose from: Donald Fereshteh Robin Bill Christy Elizabeth

John Shawn Suzy Bradley Dan Barney ___

Choose from: David Elvira Tommy Robert Camelia Betty

Candy Frank Irene Lauren Oren Rose ___

Choose from: David Cyrus Angelina Jose Uma Darin

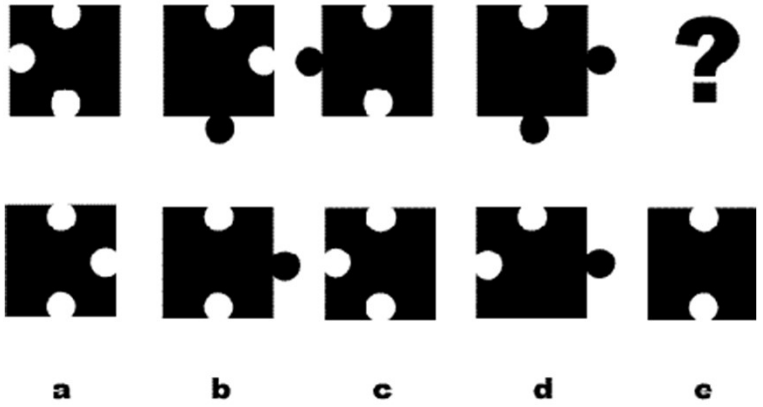
Charles Dion Stuart Kevin Joshua Sergio ___

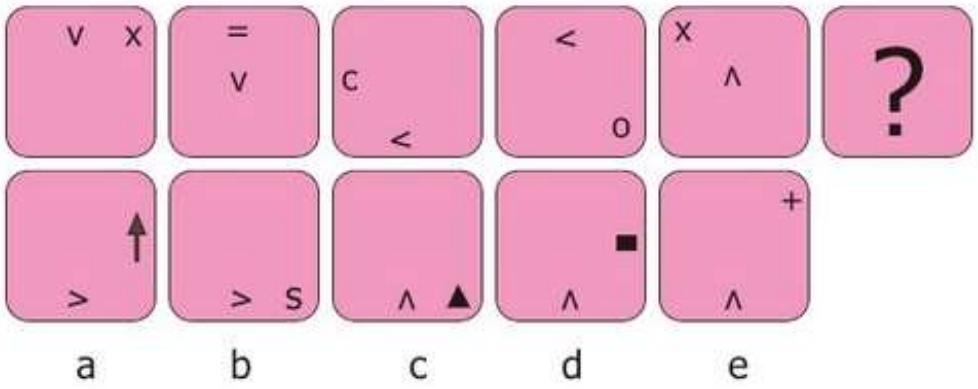
Choose from: Jeremy Shaun Thomas Duane Rupert Ulysses

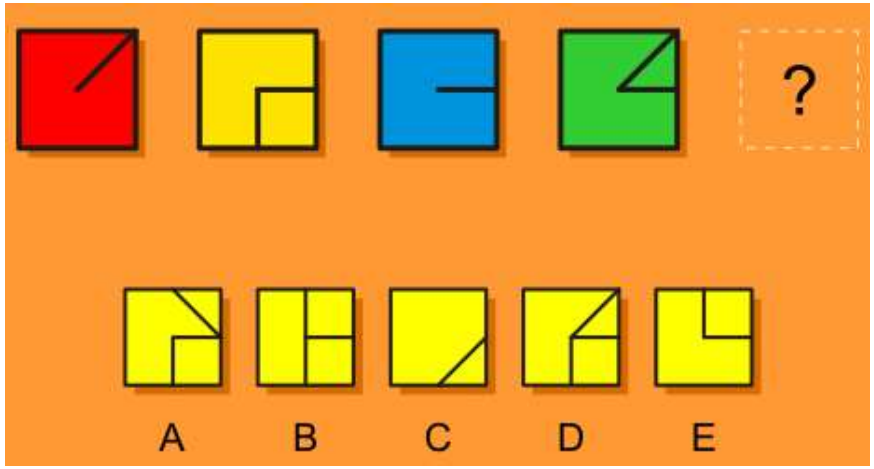
Parrot Pigeon Robin Sparrow ___

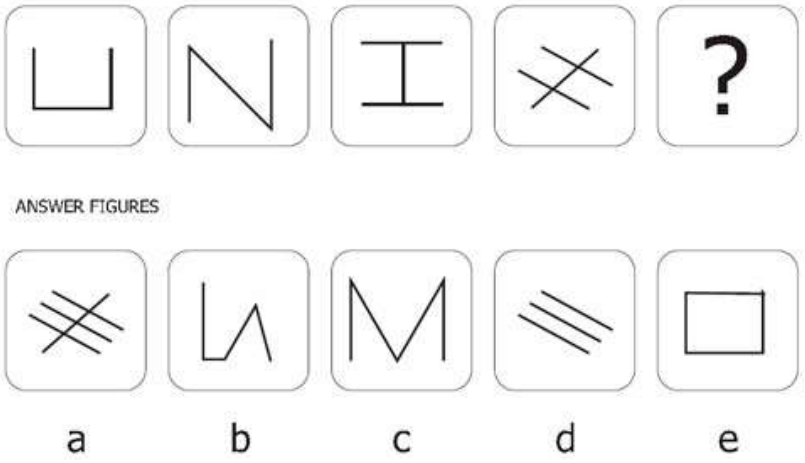
Choose from: Cardinal Oriole Lovebird Thrush Wren

Which Image Should Come Next? (Part 1)

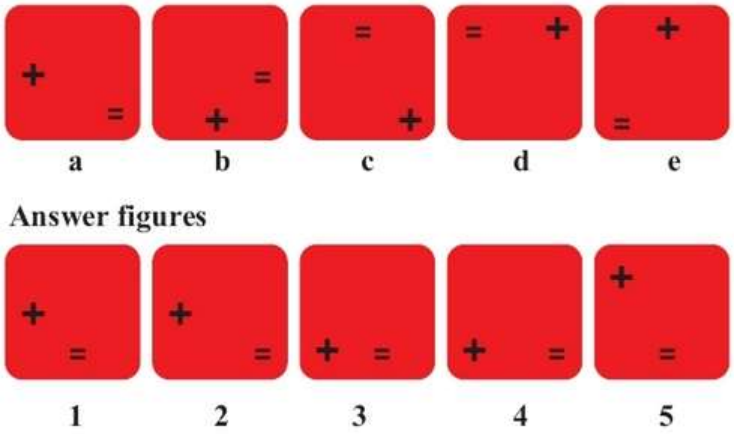


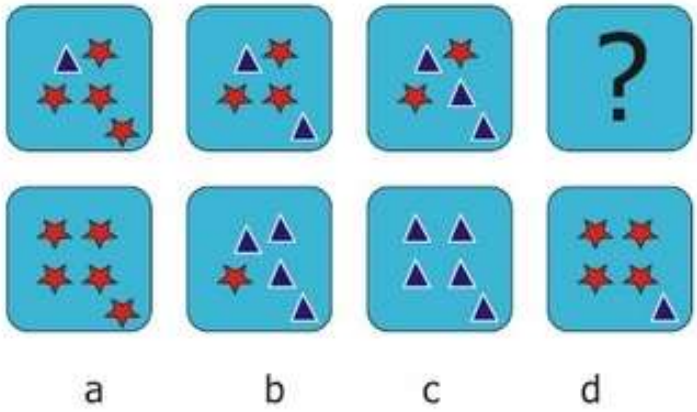


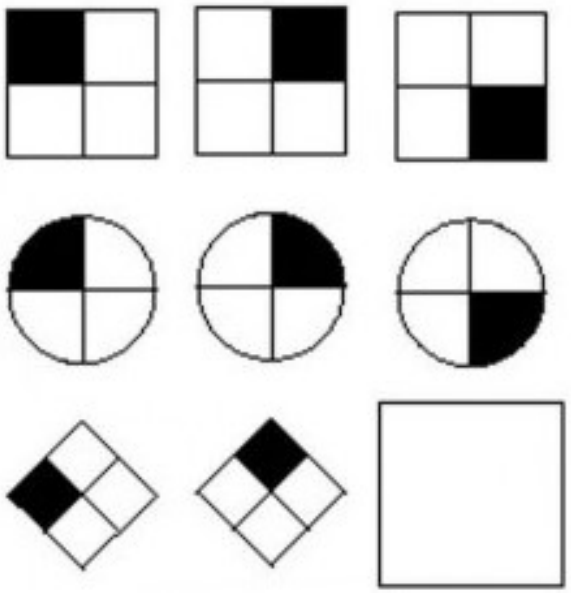


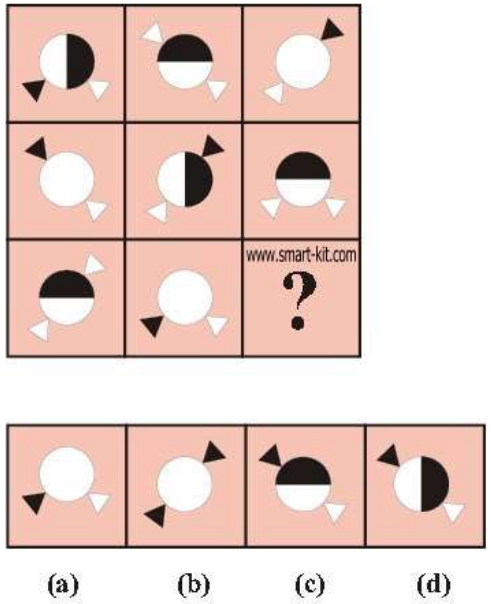


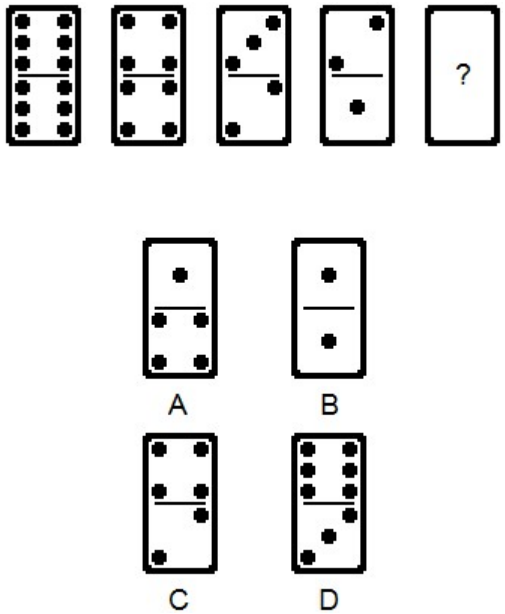
Which Image Should Come Next? (Part 2)







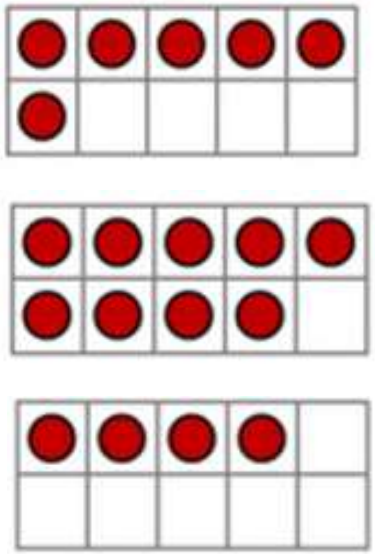


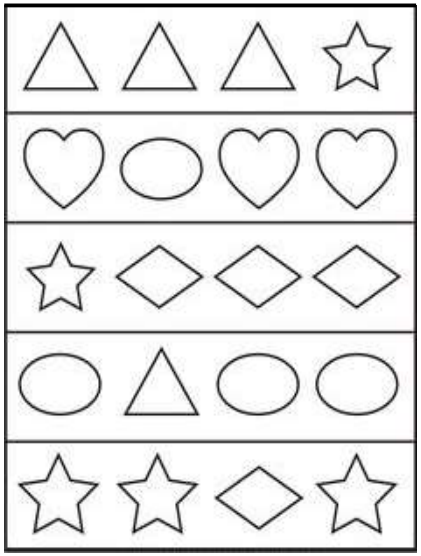


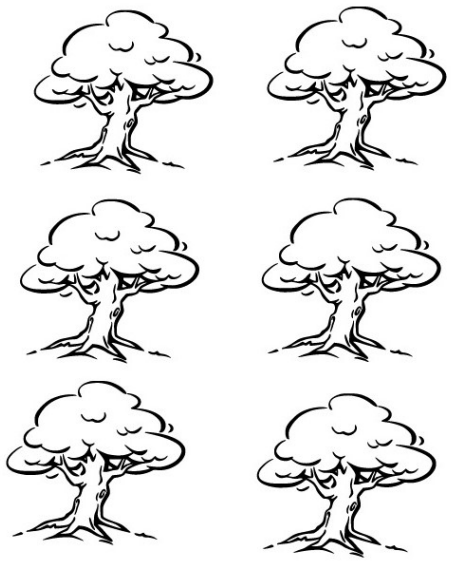
Which Term Isn't Like the Others?

A	H	I	M	N	O	V	W	X
0	3	6	7	8	9			
3	4	7	13	20	33	53	86	139

Big Fast Green Warm



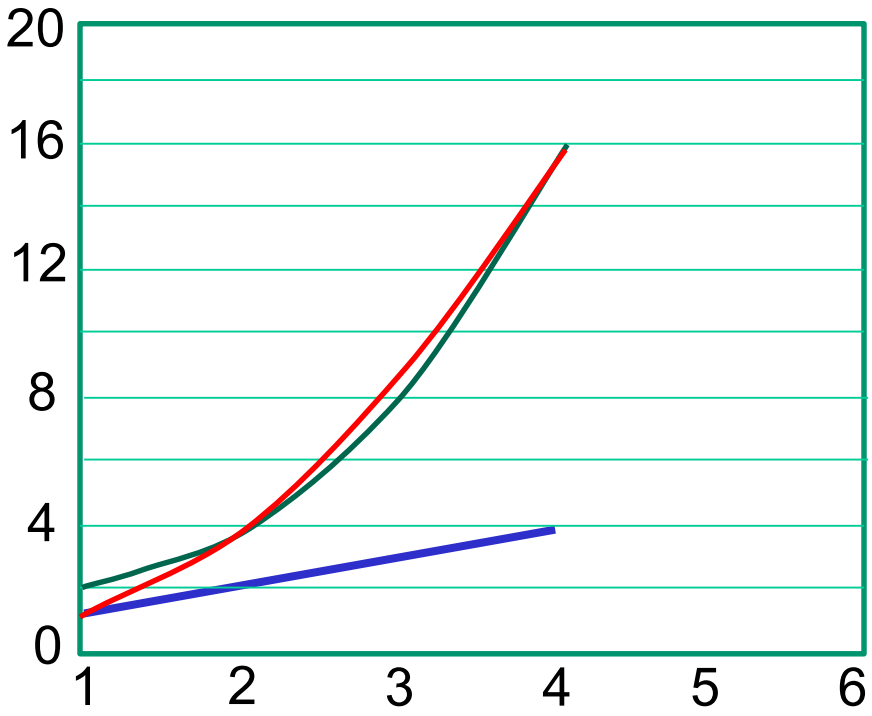




A Solution Method for Numerical Series

Polynomial interpolation:

You can pass a line through any two points, a hyperbola through any three points, a third-degree curve through any four points, and so on



1 2 3 4 __

1 4 9 16 __

$f(n) = an^3 + bn^2 + cn + d$

$n = 1$: $a + b + c + d = 1$

$n = 2$: $8a + 4b + 2c + d = 4$

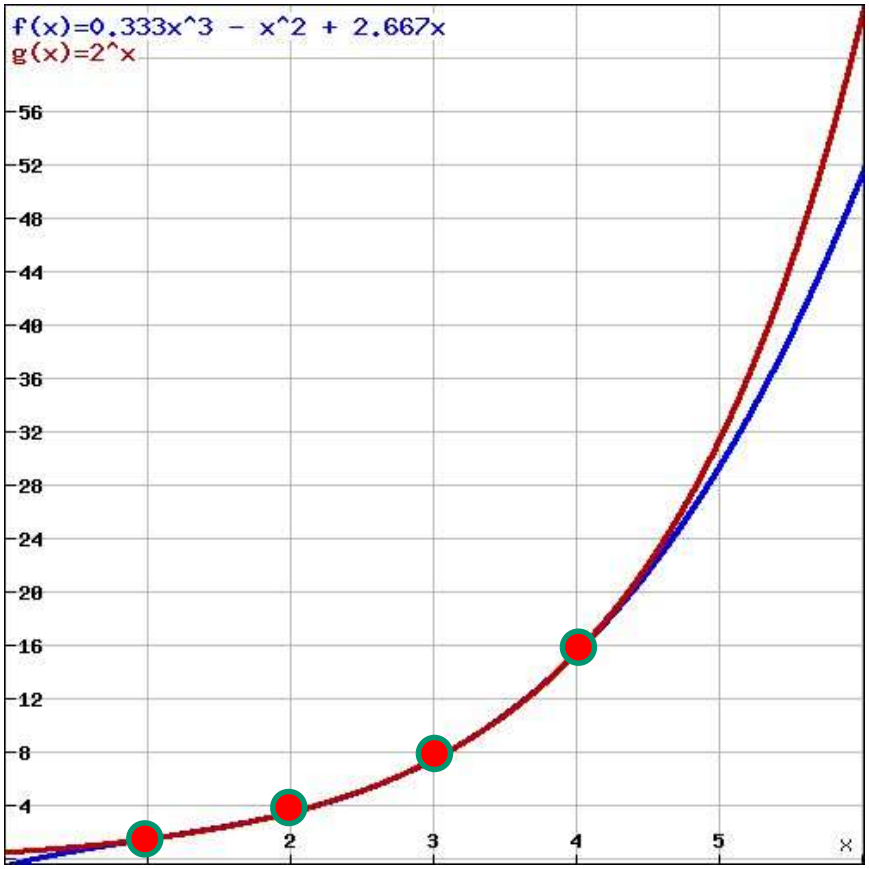
$n = 3$: $27a + 9b + 3c + d = 9$

$n = 4$: $64a + 16b + 4c + d = 16$

$b = 1$; $a = c = d = 0$; $f(n) = n^2$

2 4 8 16 __

When Several Answers Are Possible



2 4 8 16 __

Answer 1:
2 4 8 16 32
Reason: $f(n) = 2^n$

Answer 2:
2 4 8 16 30
Reason:
$f(n) = (1/3)n^3 - n^2 + (8/3)n$

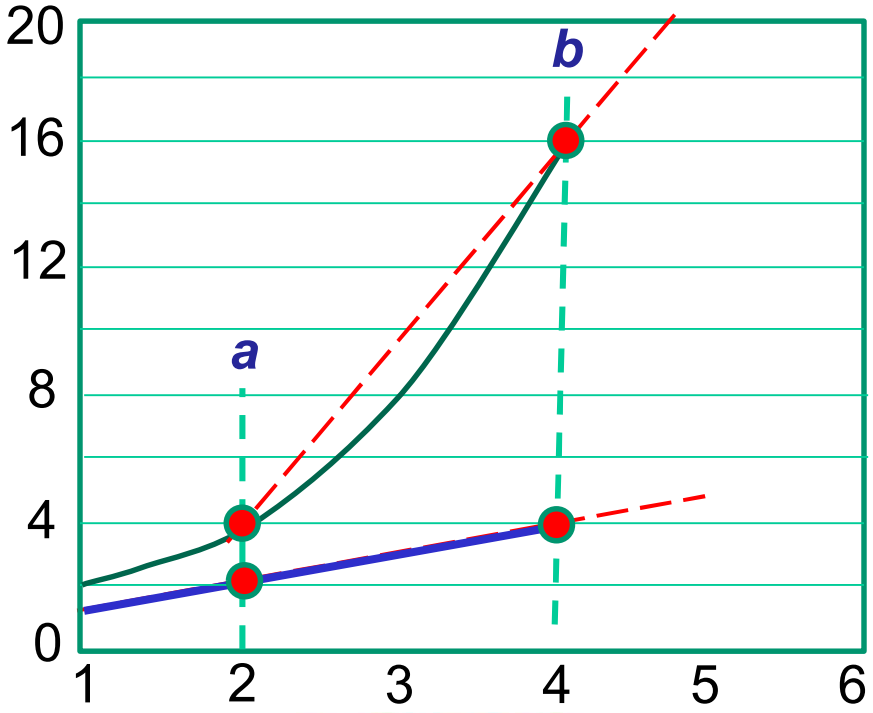
Which is the correct answer?

Challenge:
Why does $f(n)$ always yield an integer result for an integer n?

Interpolation and Extrapolation

Interpolation: Given the values of the function $f(n)$ at points a and b, find its value at some given point between a and b

Extrapolation: Given the values of the function $f(n)$ at some points between a and b, find its value at a given point before a or after b



1 2 3 4 __

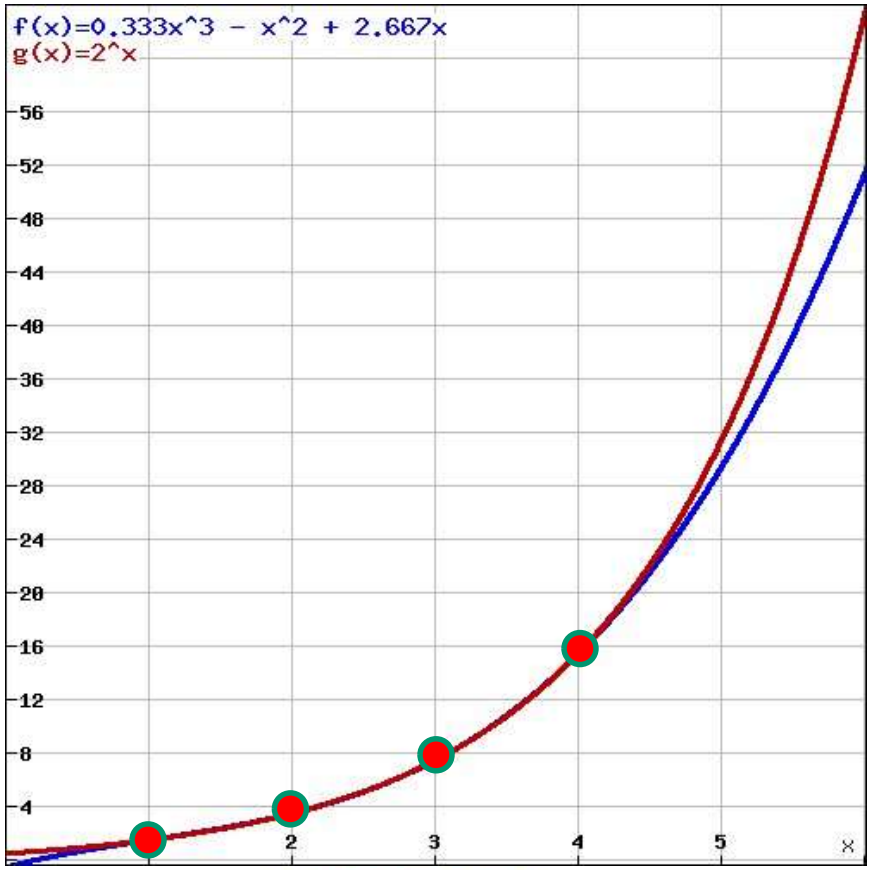
2 4 8 16 __

Khan-Academy/Pixar video illustrating the use of interpolation for animation:

https://www.khanacademy.org/partner-content/pixar/animate/ball/v/a2-quick

Polynomial Extrapolation Example

This exponential series, when solved via polynomial extrapolation, yields a different answer!



2 4 8 16 __

$f(n) = an^3 + bn^2 + cn + d$

$n = 1$: $a + b + c + d = 2$

$n = 2$: $8a + 4b + 2c + d = 4$

$n = 3$: $27a + 9b + 3c + d = 8$

$n = 4$: $64a + 16b + 4c + d = 16$

$a = 1/3$; $b = -1$; $c = 8/3$; $d = 0$;

$f(n) = (1/3)n^3 - n^2 + (8/3)n$

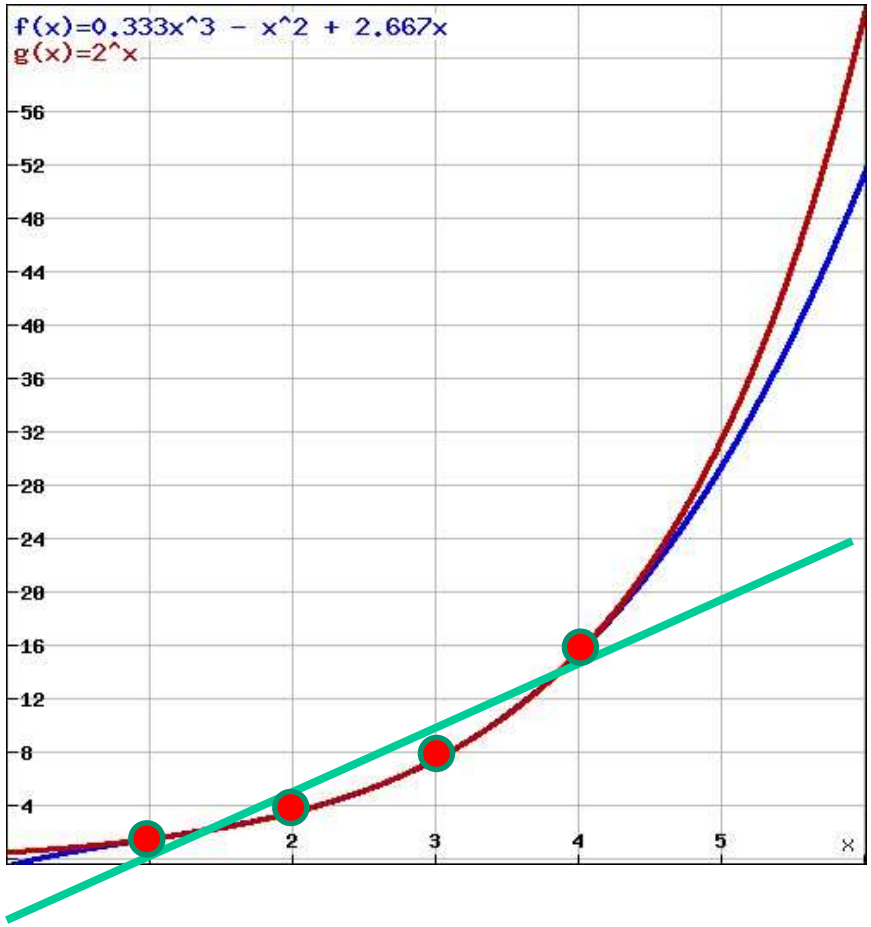
$f(5) = (1/3)125 - 25 + (8/3)5 = 30$

$f(6) = (1/3)216 - 36 + (8/3)6 = 52$

$f(30) = 8{,}180$

$2^{30} = 1{,}073{,}741{,}824$

Polynomial Curve-Fitting Example



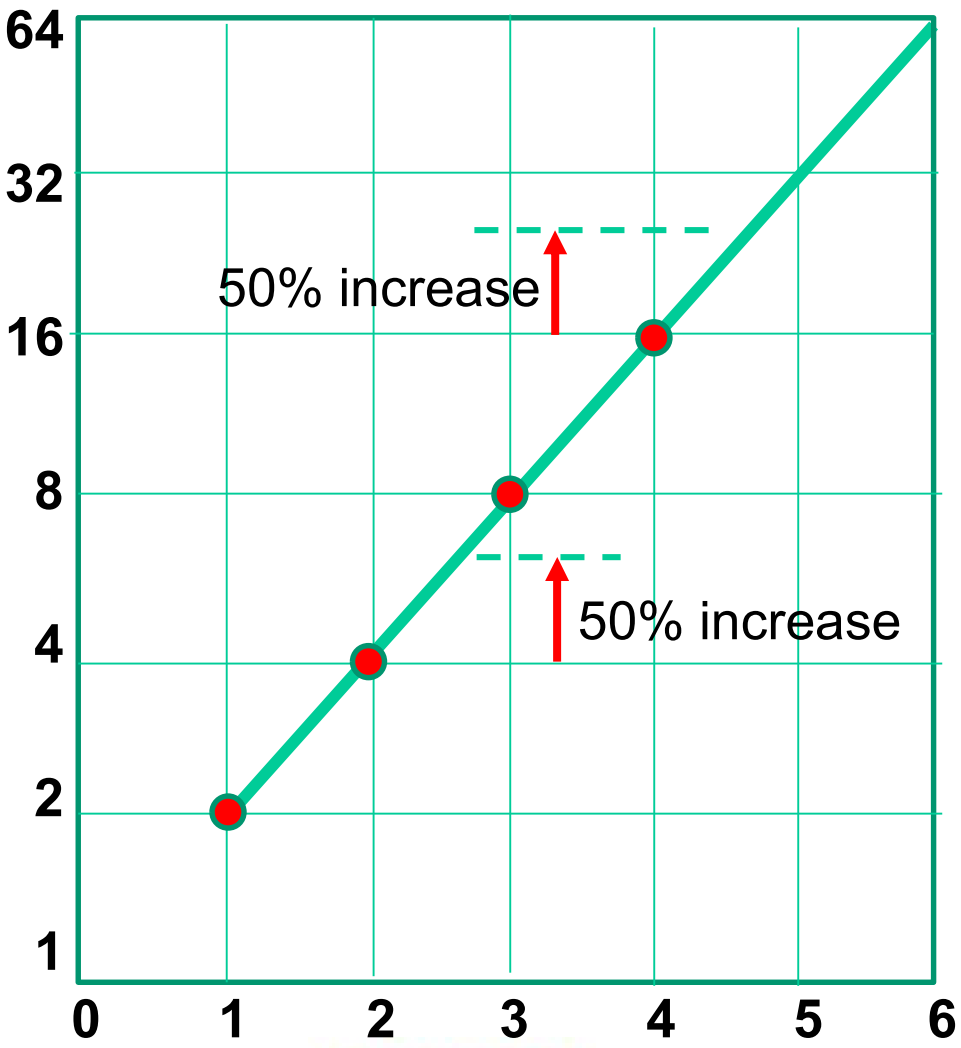
2 4 8 16 __

$f(x) = ax + b$
$x = 1$: $a + b$ vs. 2
$x = 2$: $2a + b$ vs. 4
$x = 3$: $3a + b$ vs. 8
$x = 4$: $4a + b$ vs.16

$$D(a, b) = (a + b - 2)^2 + (2a + b - 4)^2 + (3a + b - 8)^2 + (4a + b - 16)^2$$
$$= 30a^2 + 4b^2 + 20ab - 196a - 60b + 340$$

$dD/da = 0$ and $dD/db = 0$ yield
$a = 23/5$ and $b = -4$
$f(x) = 4.6x - 4$

Log-Scale Linearizes Exponential Trends



2 4 8 16 __

In log-scale, one unit of distance represents not a fixed increase but multiplication by a factor

It also allows us to focus on relative, rather than absolute, variations.

Question 1: Where is the place of zero on the vertical axis?

Question 2: Is 50% decrease represented by the same vertical distance as 50% increase?

Example: Two-Dimensional Feature Space

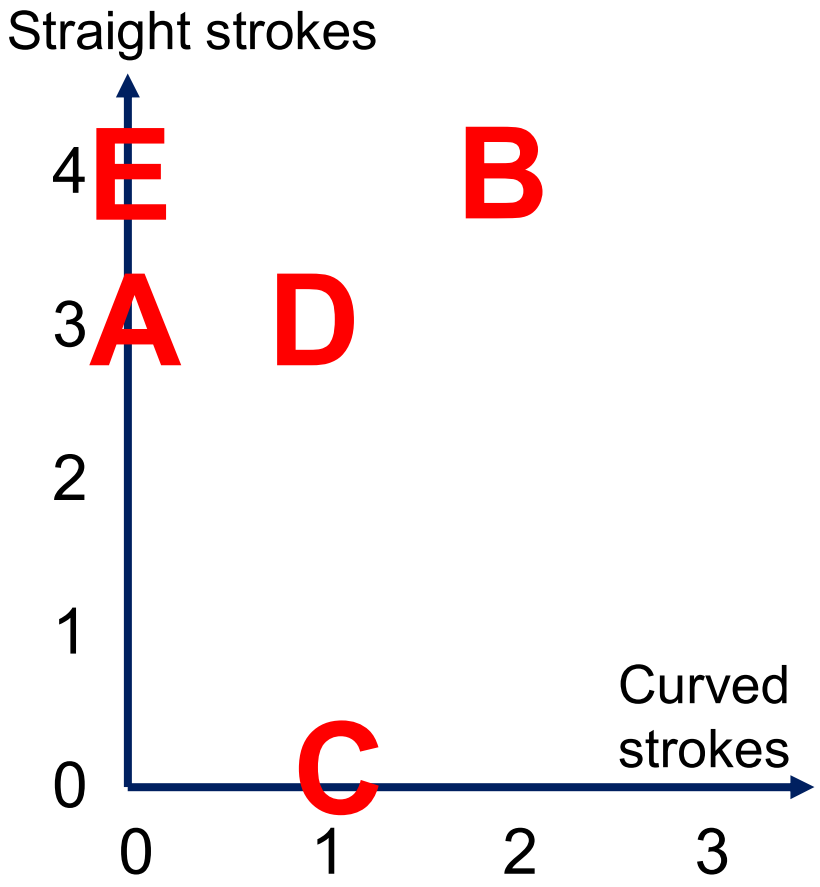
A, B, C, D, E

Example features:

x: Number of curved segments
y: Number of straight segments

Where would “F” fall?

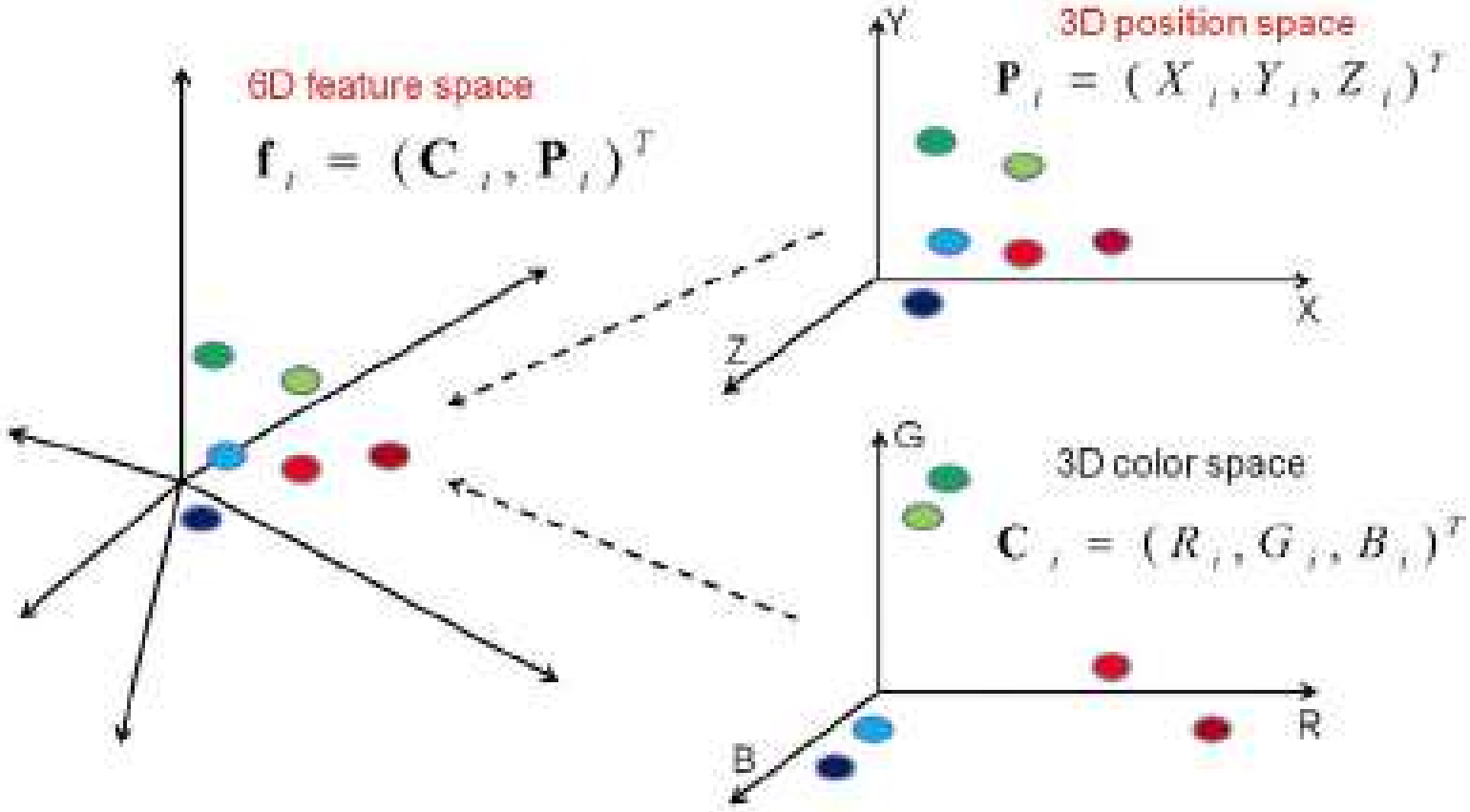
Suggest an additional feature



Higher-Dimensional Feature Spaces

The points in this diagram have six features: X, Y, Z, R, B, G

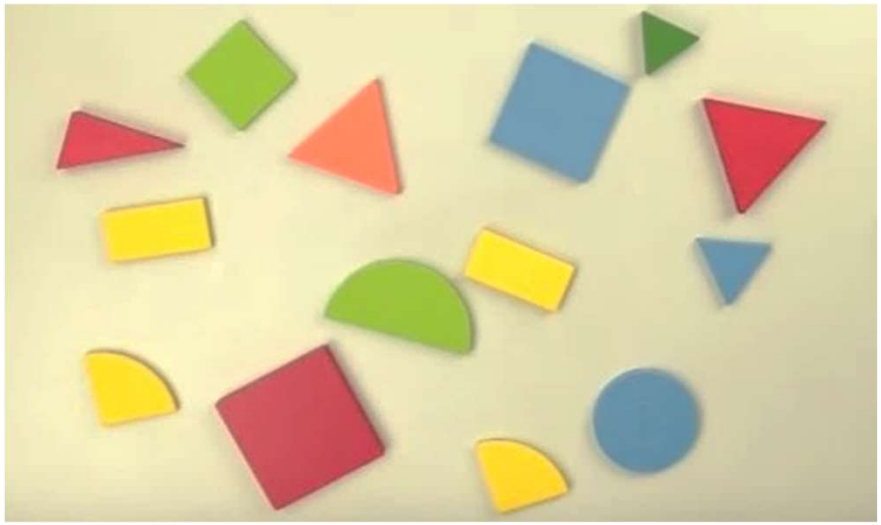
Position Color



Classifying by Color, Shape, or Other Features

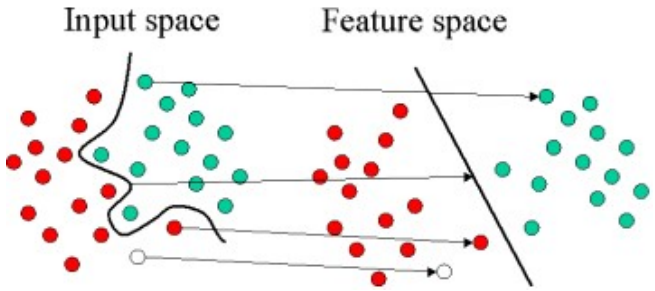
Very young kids are taught about classification by features
(2-minute video: http://www.youtube.com/watch?v=5bip0bcFlgo)

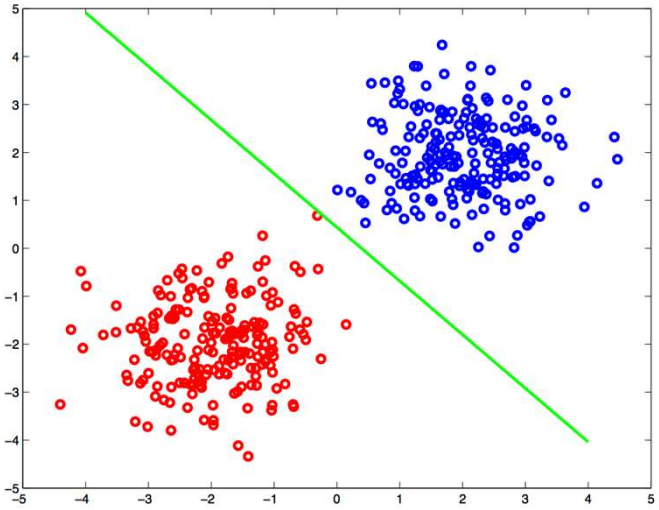
Possible features in the shapes shown in the video:
Color: Blue, Green, Orange, Yellow
Geometric shape: Square, Rectangle, Triangle, Circle, …
Curvature: Straight sides only, at least one curved side
Size: Large, Small (area)
Number of sides: 2, 3, 4
Triangleness: Yes, No
Thickness?
Material?
Weight?
Floats on water?

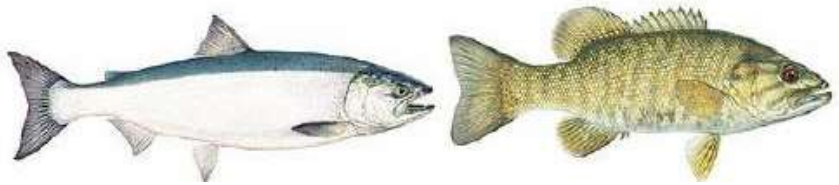


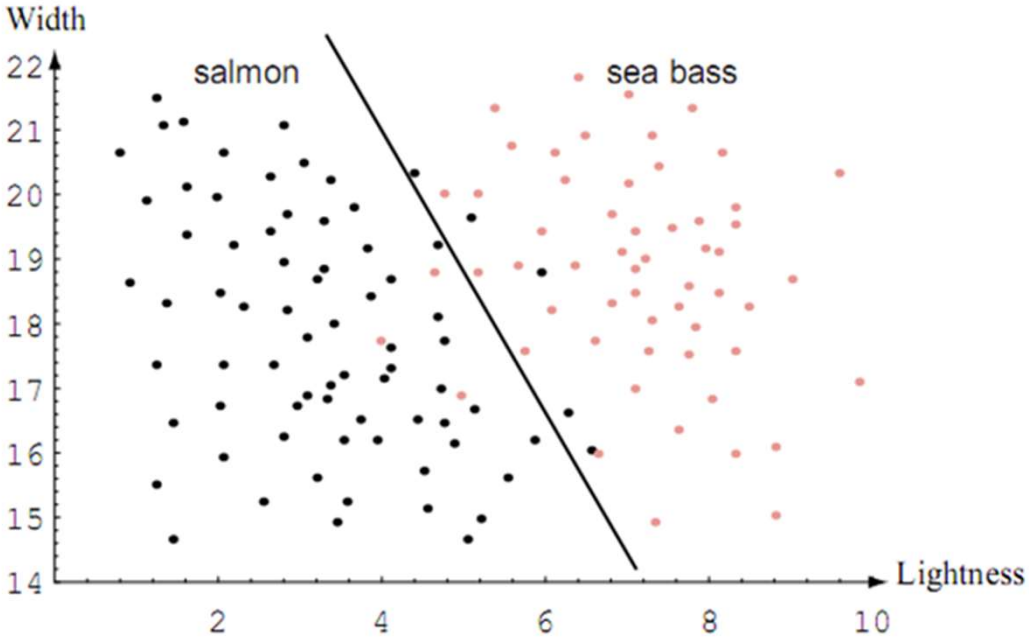
Pattern Classification

Extracting features from given inputs allows us to separate and classify the inputs according to desired categories

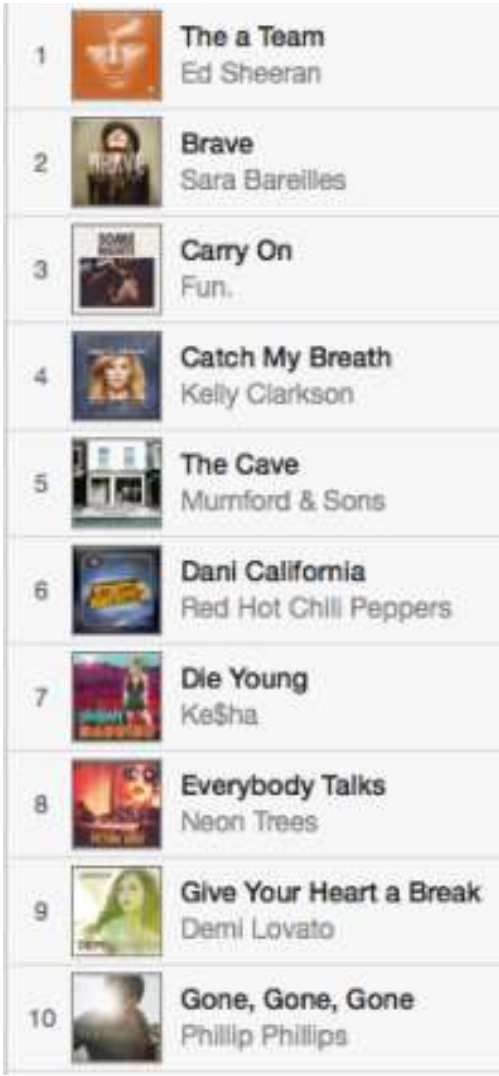




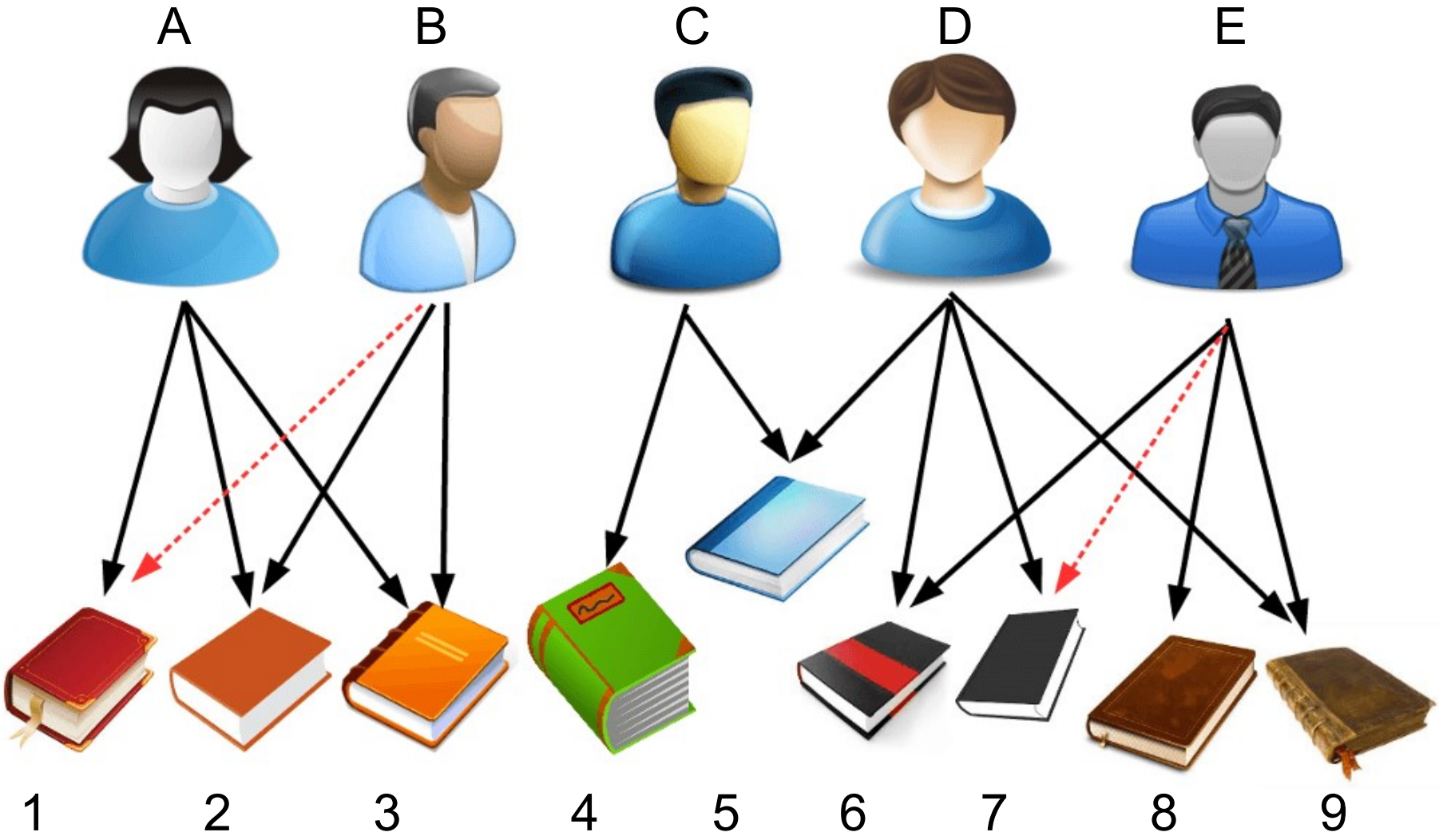




Which Book/Movie/Song Should Come Next?



Example: Book Recommendation



Fingerprint Classification and Matching

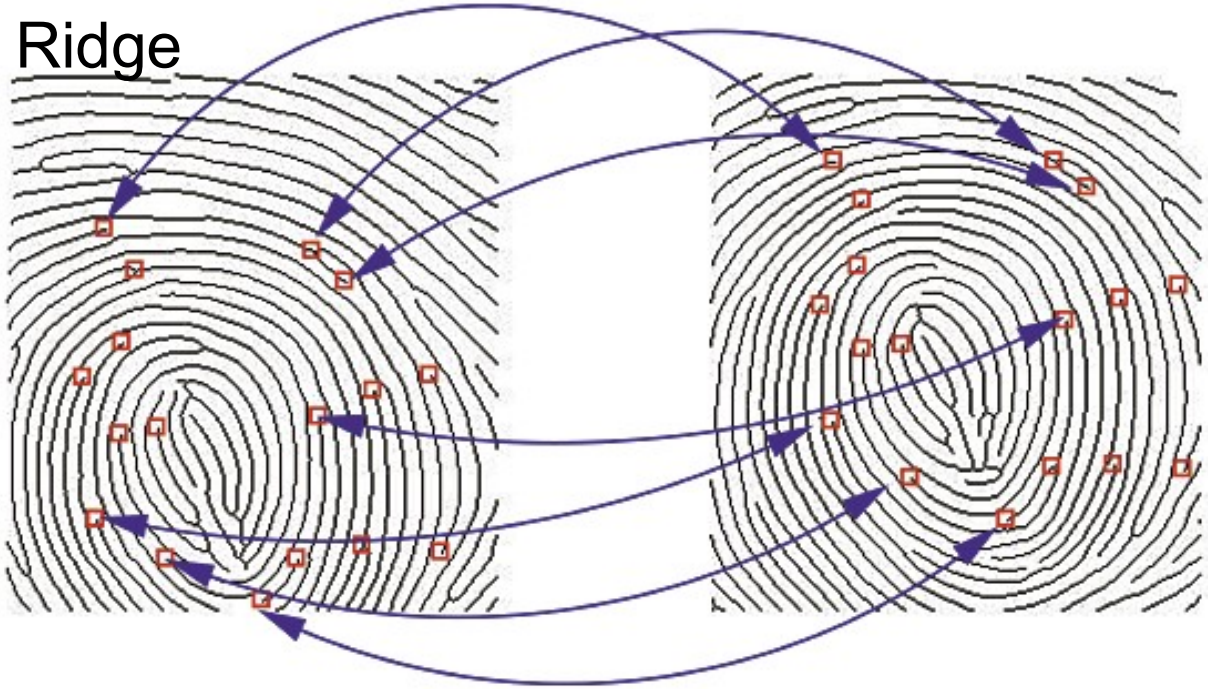
Needed for criminal investigations and biometric identification

Does a fingerprint match any of the prints in a criminal database?

Does the fingerprint match one recorded for an authorized user?

Human fingerprints tend to be unique

Even identical twins have different prints



The Basics of Comparing Fingerprints

(6-minute video: http://www.youtube.com/watch?v=IrpTqKkgygA)

Image Search

By keywords (when stored images have been indexed previously)

By photographer, location, etc. (image metadata)

By providing an image as key (Google image search)

https://images.google.com/

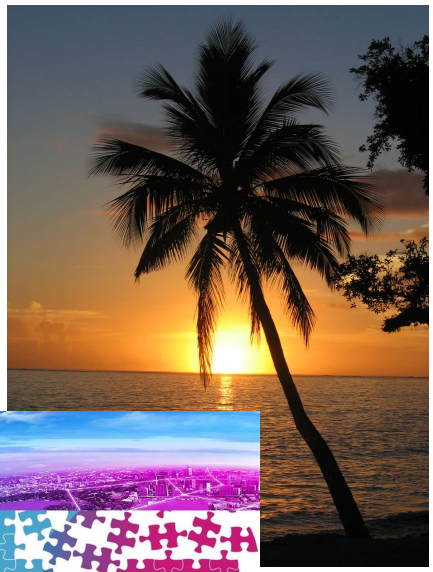
Example searches:

Sunset

UCSB

Soccer

INT 94TN



Searching the Worldwide Web

Google has indexed the entire contents of the Web

a	b	c	b	b	b	a	b	c	b	a	b	b	c	a	a	b	c	b	a	b	c	b	a	b	c	b	b
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27

a a b	14
a b b	10
a b c	0, 6, 15, 19, 23
b a b	5, 9, 18, 22
b b a	4
b b b	3
b b c	11
b c a	12
b c b	1, 7, 16, 20, 24
c a a	13
c b a	8, 17, 21
c b b	2, 25

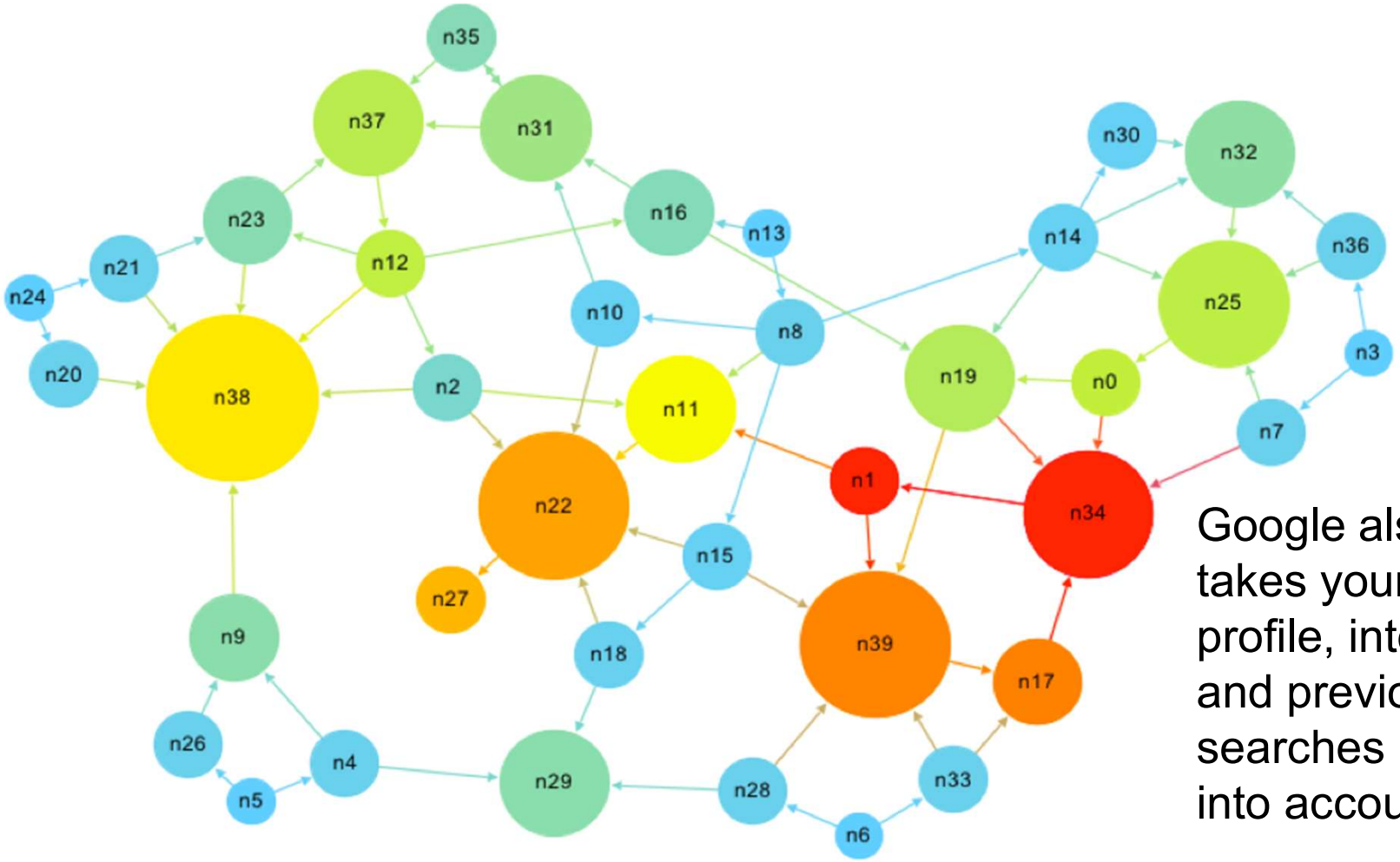
Find all occurrences of the pattern "**abcbab**"

a b c	0, 6, 15, 19, 23
b c b	1, 7, 16, 20, 24
c b a	8, 17, 21
b a b	5, 9, 18, 22

a b c	0, 6, 15, 19, 23
b c b	1, 7, 16, 20, 24
c b a	8, 17, 21
b a b	5, 9, 18, 22

Google's "Pagerank" Algorithm

Algorithm to rank the hits so that the most useful ones come first



Google also takes your profile, interests, and previous searches into account

Neural-Network Pattern Recognition

Train the system using known patterns, then use it on others
(4-minute video: http://www.youtube.com/watch?v=kGv-1it8Sac)

Facial Recognition Technology

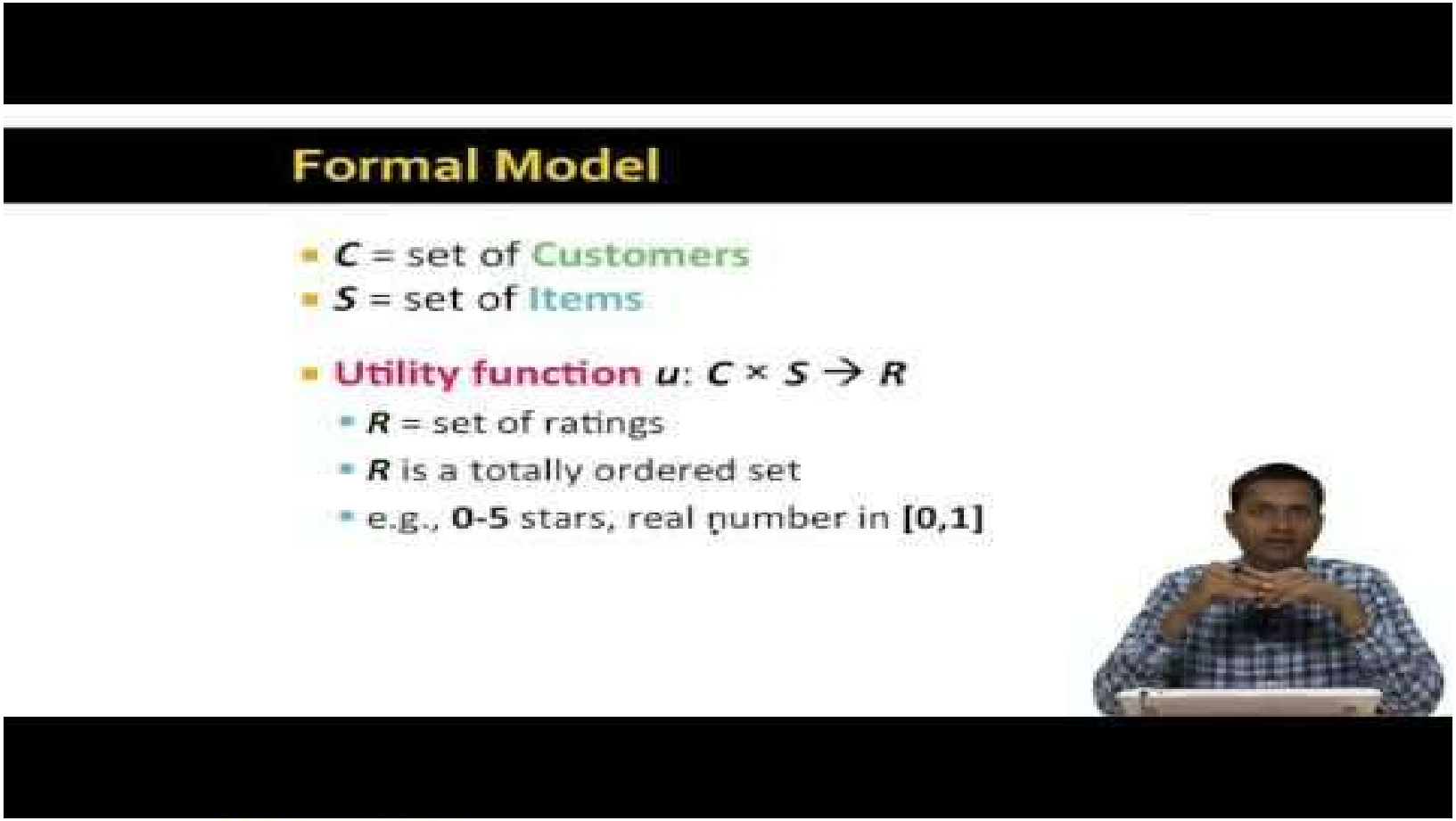
Train the system using known patterns, then use it on others
(1-minute video: http://www.youtube.com/watch?v=tZzIH4Qf5Y8)

Gender Classification by Neural Networks

Train the system using known faces, then use it on others
(2-minute video: http://www.youtube.com/watch?v=3jAqlu7HtnI)

Overview of Recommender Systems

Track activity, interactions, and ratings, combine with other data
(17-minute video: http://www.youtube.com/watch?v=1JRrCEgiyHM)



Questions?

To dig deeper, see:

J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. "Recommender Systems Survey," *Knowledge-Based Systems*, Vol. 46, pp.109-132, July 2013.

parhami@ece.ucsb.edu
www.ece.ucsb.edu/~parhami/