

Curso de Introducción al Pensamiento Probabilístico

David Aroesti



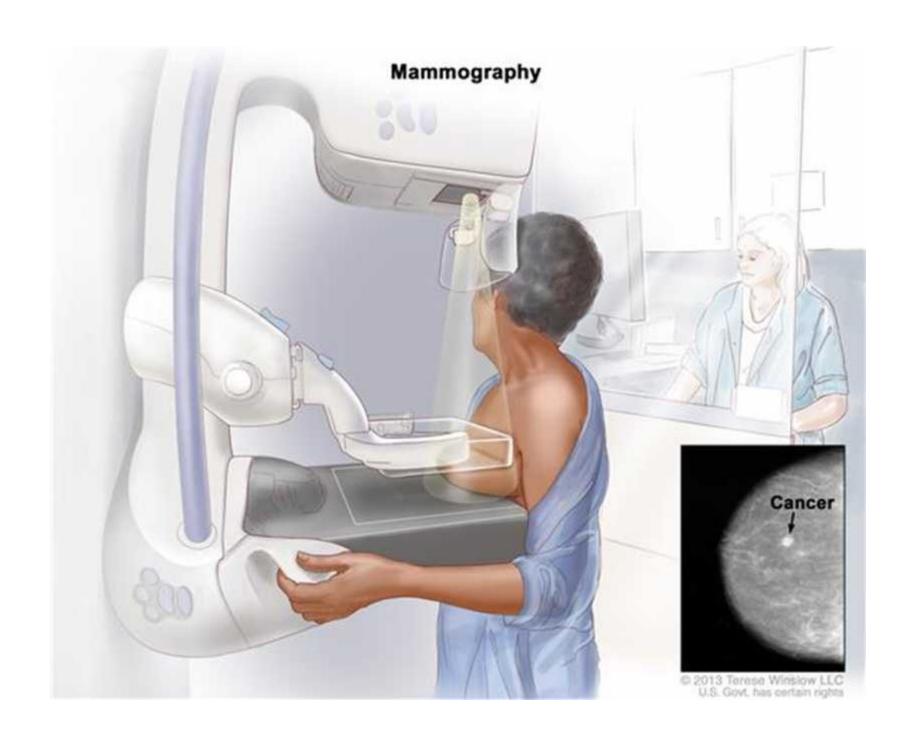
Introducción a la programación probabilística



- La programación probabilística utiliza probabilidades y modelos probabilísticos para ejecutar cómputos.
- Se utiliza en una gran cantidad de campos: investigación científica, inteligencia artificial, medicina, etc.
- Existen lenguajes y librerías especializadas para ejecutar este tipo de cómputo, como Pyro de Uber.

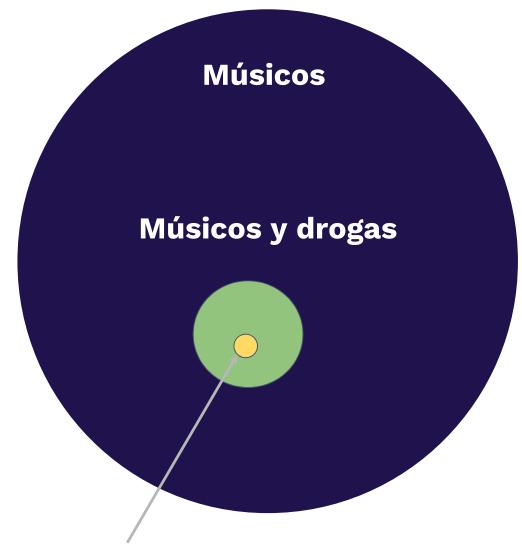










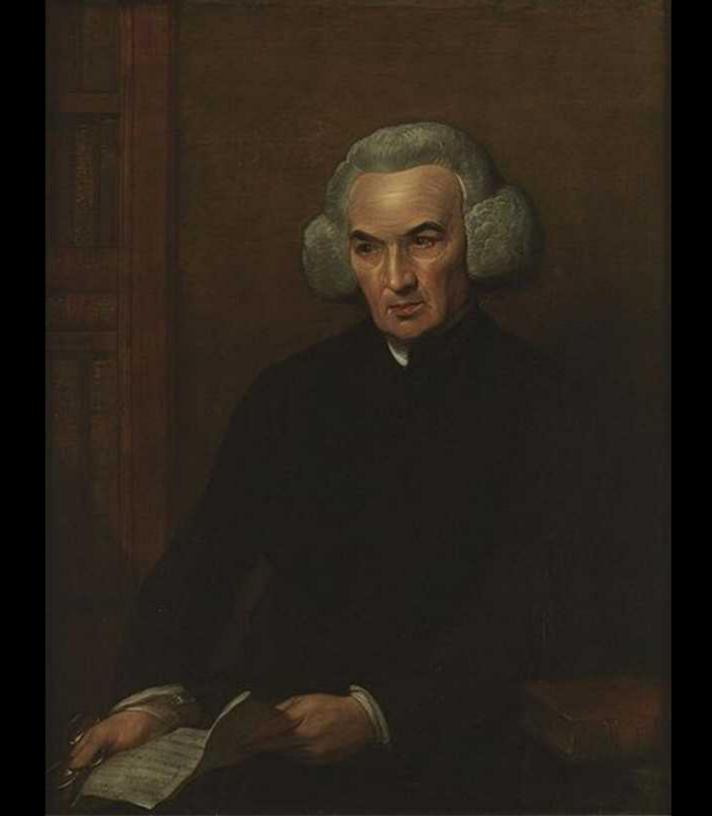


Músicos y drogas y Pokemón



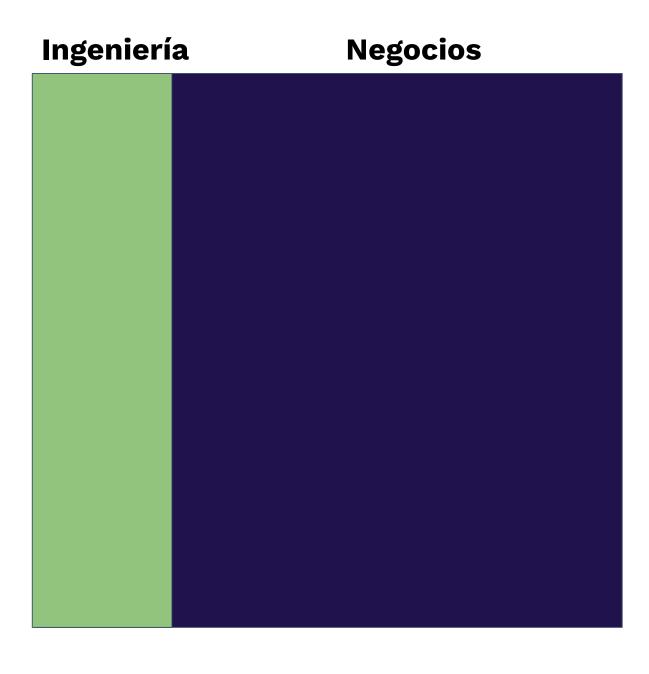


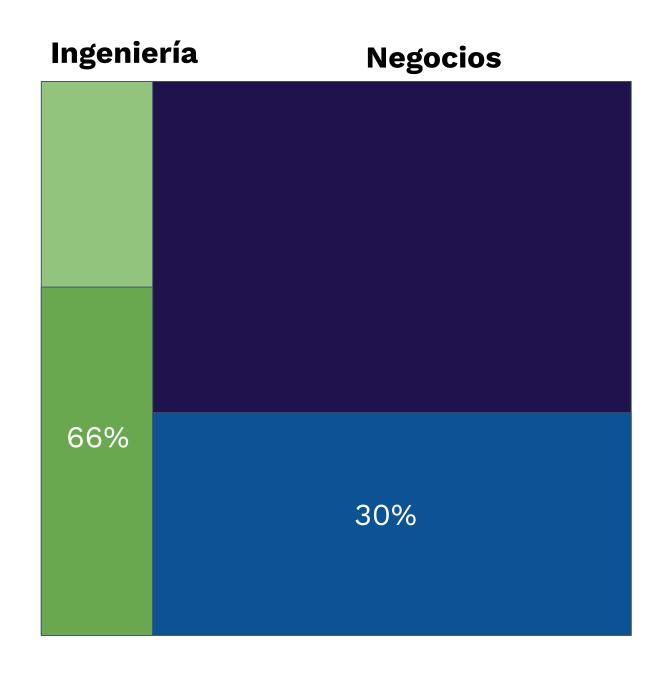




studying ...















 $P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$

Análisis de Síntomas



Symptom	Cano	Total		
Symptom	No	Yes	Total	
No	99989	0	99989	
Yes	10	1	11	
Total	99999	1	100000	

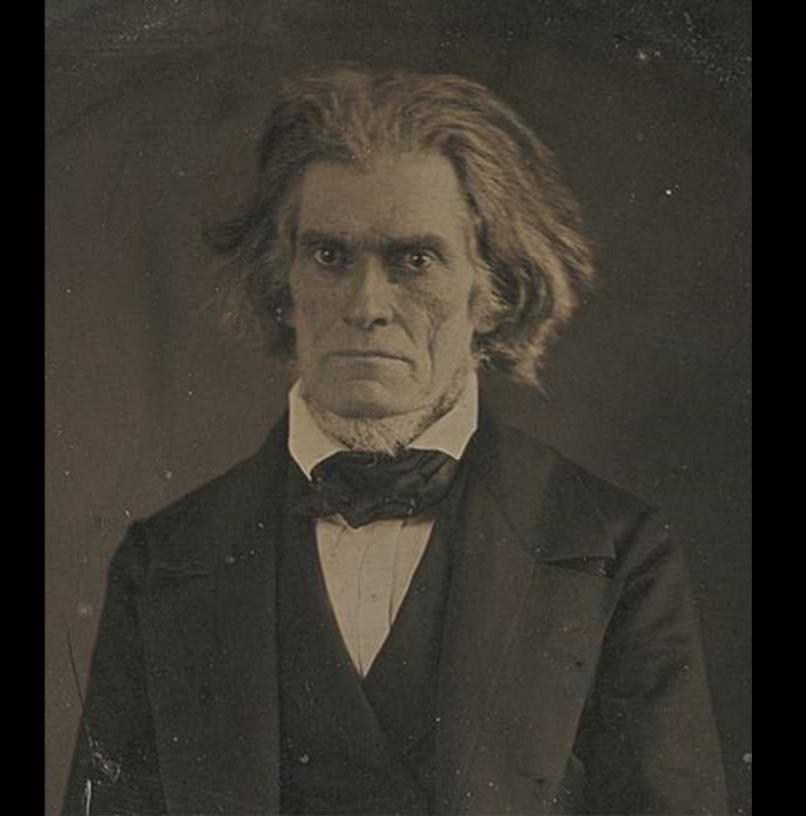
Garbage in, garbage out (GIGO)

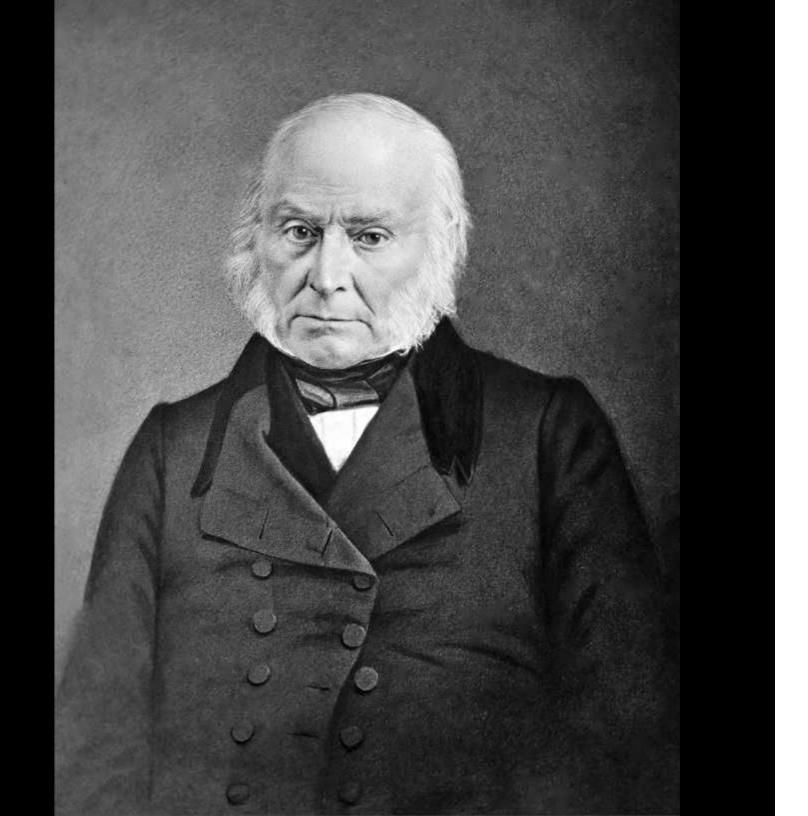
- La calidad de nuestros datos es igual de fundamental que la precisión de nuestros cómputos.
- Cuando los datos son errados, aunque tengamos un cómputo prístino nuestros resultados serán erróneos.
- En pocas palabras: con datos errados las conclusiones serán erradas.

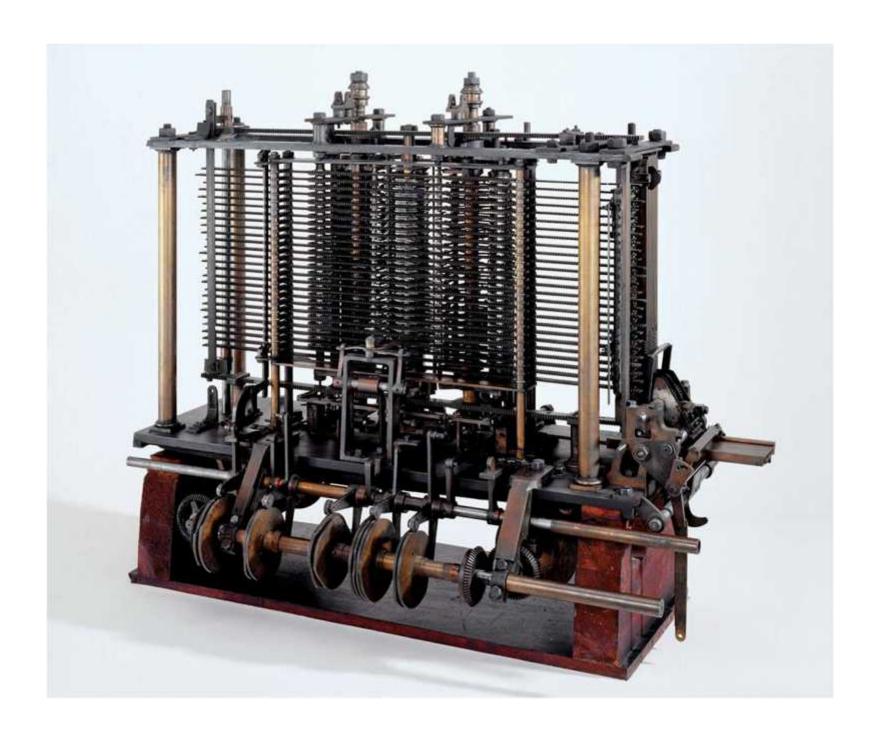


SCHEDULE 4. Prod	sho	160	numer	ated b	уп	ie, on t	he _	11	10%	1	day	of .	L	12	1	/_	1850	
-	Aures of						_	ook, Japa	Ist, 1860.			1		print.	h	ndun di	ering the	
10 41 Tel. 1		7		tt [1			1				15
Hans of Dener, Agent, or Manager of the Parm.	. 1	4	1	H	-	1 1		1			4	1	1	1	- market	4	3	
The Later of	1	1	1	ii l	1	1 0	-	Dame Co	+	1	11	i	3	1	Open he	1	1	ŀ
11.47. 110.20	1	8	3	20	4	4 9	8	-	4		13	14	15	4	17	018	10	
13 121 117 116 11	911		A 400			7 1 6	•	10	11	19		-	-	16	=	1	-	1
pre Breman	13	371	150		1	/		-/		1	20		-	L U		7		ľ
grate Delchura	1	150	3501	25		2	2	/	-	6	50				400	-	3	H
trick . Malony	200	200	4000	150	2		F 5533	10	16	30	819	500		300	11/0	,	-	n
Sward Golache	3	157	300			1	2	-	-		60		-			X	15	E
mas Helany	60	124	200	55	2	4				10	175	100		50	200	-	77	R
for Filmer	20	150	150			2			-		25		-			1		k
comst. M Som	10	190	400			1	1		-		12	-	-	-		2	-	
Jamas Culon	30	170	530	10		1	2				60	-		Project.		X		li di
frick Dum	20	300	550	40		2	2	3		9	25	-		100		* / a	-	1
med Conden	11	150	355	65	TAN	1	2				65	90		- 0	= -		-	1
Wirk Secule	15	145	350	AFF		1		1		-	20	40			- 1	X	1000	
trick Delawy	20	110	400	15			2		احد	53	50	-1			T -	- X		11
ichart Sullivan	12	253	850	55	,1	1		550	1		140	1				X		*
trick Shearman	11	100	750			2	VA	12	31		25	-				X		
maid Dum	10	150	1000	1.3	3	2	N	1			25	30	14	-31		X		
minal Hagar	. 50	140	130	30	1		4	2	15:	3	20%	212		100	50	41.0	5.5	
una Galy	BURNEY.	2/2	1. 1. 1. 3	20		2		1	ye	3.	35	130		4.7	150	E	Sin 5	A
Wrick Thell	1	211	1000	75	2	.2	3			10	140	10%	215	300	70	21		
her Day		160	300	400	1					2	50	1	55	434		Y		
Villani Clayer	100		- 40	20		2	2		4	15	95	300		200		y.	13	
Lyunas James		- 5		20	4	3	6	1	13	14	265	200	1	200		X		
ance Mer	100		1	15		4	2	1	2	4	65	120		100		X	0	
broket A South	100		100	75	.2	- /		/		2	125					×	91	
0. 11	***	020	1000	120		4	6	12	. 18	10	280	800		20	50			
Sandie Comith	50	1 P. F.	500	TOTAL S	2	2			16	10	150	150		7.550.0	20		- 5	9
1 11	24		600	40	1	2	6	-	22	7	215	100	10	400	1		i lo	9
egypte & Melso	20	10.00	150	45	Í,		2	1	20	1	100	1	- 3	-		X.	-	9
ania Mane		40	100	40	Í,	2		2			130	200		100	150	-	1	
Martin Mister	40	50	2000	20	5	1	-				260	Jackson representation of the last section of		60		E-1 130		١,
lygam Maument	***	.7	3000		_	2	1	6		9		The second second	-		150			
genes Deofen	100	100	2000		12		1	100000		3	67	35		-	100	200	-27	1
ance Mins		-	-	10		2	1122	-		4		ECC10 O	8		30	01091		1
Latel Sinny	7.5	58	U/10285-00	36		1100	1	1		55	550	300	E	E. T. C.	200			
good Chad		F	8500		100		1	-		- 200	1000	STUDIOS.			150		1	
Tighard Coll	Committee of the committee of		1200	90		. /	1000	-		15	125	300			12		-	
Las Minten		他の はらばか かんき	1000	105	COLUMN TO SERVICE	1 - 1	1	3		10	175				3	3.714		
Jennis Stattery	35	275	1100	50	3	-/	11.	-	-	11	140	300	-		20			1
Mr. B. Cochran	-	100	1 4	25	-	2		3	-11	14	125	_	_	200		X		
agreed Garrender	60	320	100	30	3		2			100		300		- 77	533	h		-
sage Gilecter	20	1000		50	/	1	1.74	/	3	15	Carlo Carlo Carlo	650		777	20	4		1
Hartin Donne por	35	125		20		1	-	-	1	-	12	295		100	25	-	111	-
Mr. P. Johnson	55	105	35057	-36	1	1.2	10	5		1	5.976	174		2.96	4 3 4	60	1	4









66

En dos ocasiones me han preguntado (miembros del parlamento) 'Disculpe, Sr. Babbage, si introduce en la máquina números incorrectos, la respuesta correcta saldrá'. Me cuesta trabajo apreciar la confusión de ideas que puede provocar dichas preguntas.



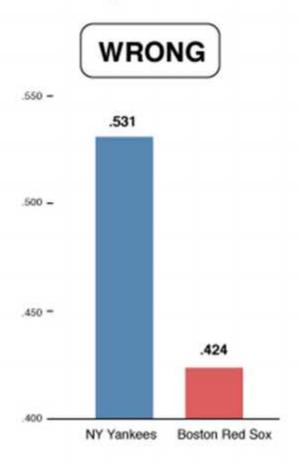
Charles Babbage

Imágenes •

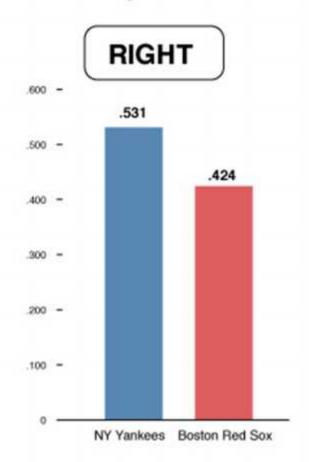


- Las visualizaciones son muy importantes para entender un conjunto de datos.
- Sin embargo, cuando se juega con la escala se puede llegar a conclusiones incorrectas.
- Nunca se debe confiar en una gráfica sin escalas o etiquetas

Percentage of victories



Percentage of victories



Cum Hoc Ergo Propter Hoc

- Dos variables están positivamente correlacionadas cuando se mueven en la misma dirección y negativamente correlacionadas cuando se mueven en direcciones opuestas.
- Correlación no implica causalidad.
- Pueden existir variables escondidas que generen la correlación
- Después de esto, eso; entonces a consecuencia de esto, eso.

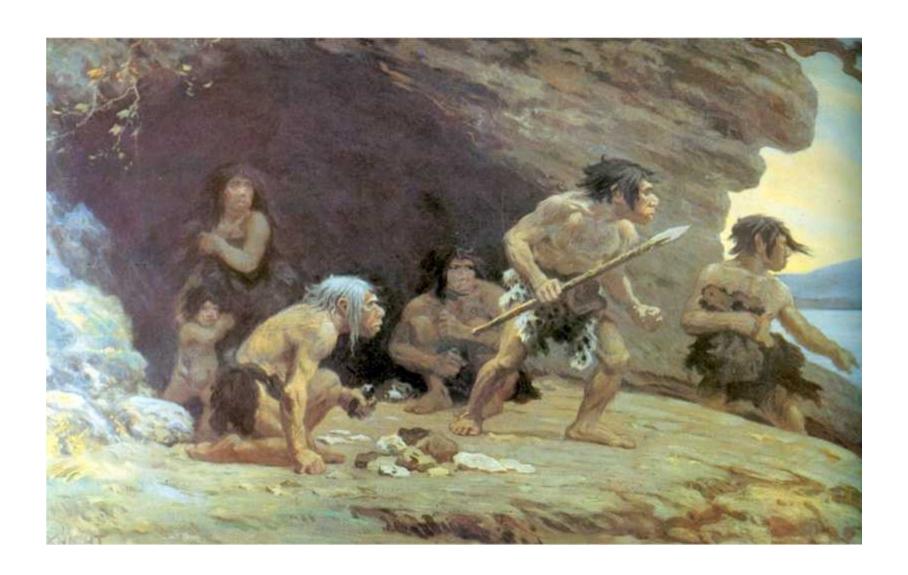






Prejuicio en el muestreo

- Para que un muestreo pueda servir como base para la inferencia estadística tiene que ser aleatorio y representativo.
- El prejuicio en el muestreo elimina la representatividad de las muestras.
- A veces conseguir muestras es difícil, por lo que se utiliza a la población de más fácil acceso (caso estudios universitarios).





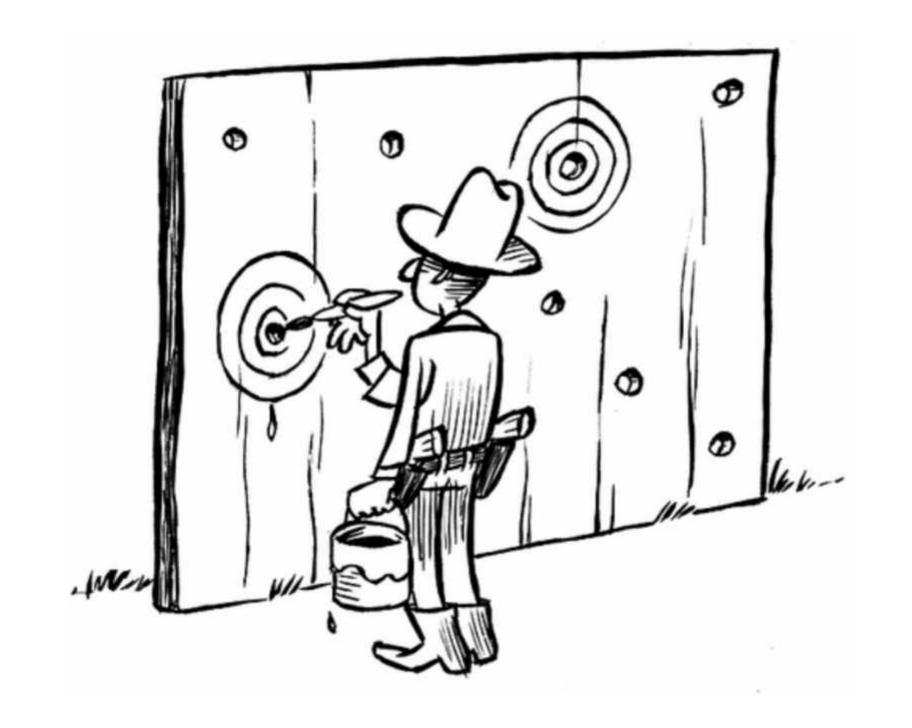
Falacia del

francotirador

de Texas



- Esta falacia se da cuando no se toma la aleatoriedad en consideración.
- También sucede cuando uno se enfoca en la similitudes e ignora las diferencias.
- Cuando fallamos al tener una hipótesis antes de recolectar datos estamos en alto riesgo de caer en esta falacia (muy común en Data Science).







Porcentajes confusos

- Cuando no sabemos la cuenta total del cual se obtiene un porcentaje tenemos el riesgo de concluir falsos resultados.
- Siempre es importante ver el contexto.
- Los porcentajes, en vacío, no significan mucho.

- Escuela A incrementó su rendimiento en 25%
- Escuela B incrementó su rendimiento en 10%
- Escuela C incrementó su rendimiento en 5%

	Rendimiento 2018	Rendimiento 2019	Incremento	Incremento porcentaje
Escuela A	20	25	5	25%
Escuela B	50	55	5	10%
Escuela C	95	100	5	5%

- En 1970, 12.5 millones de jóvenes vivían con sus padres
- En 2015 esta cifra se incrementó a 18.6 millones
- Esto representa un incremento del 48%?

	Jóvenes viviendo con sus padres (millones)	Población del país (millones)	Porcentaje de jóvenes viviendo con sus padres
1970	12.5	234.38	5.33%
2015	18.6	309.38	6.01%
Diferencia	¿48?		NO 48%



Falacia regresión



- Muchos eventos fluctúan naturalmente, por ejemplo, la temperatura promedio de una ciudad, el rendimiento de un atleta, los rendimientos de un portafolio de inversión, etc.
- Cuando algo fluctúa y se aplican medidas correctivas se puede creer que existe un vínculo de causalidad en lugar de una regresión a la media.







Introducción Machine Learning

66

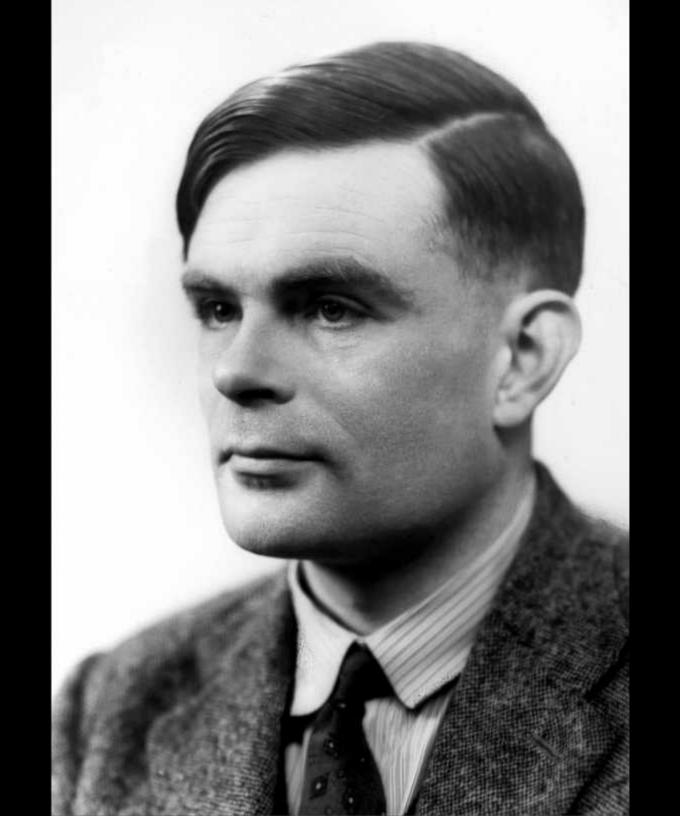
Es el campo de estudio que le da a las computadoras la habilidad de aprender sin ser explícitamente programadas.

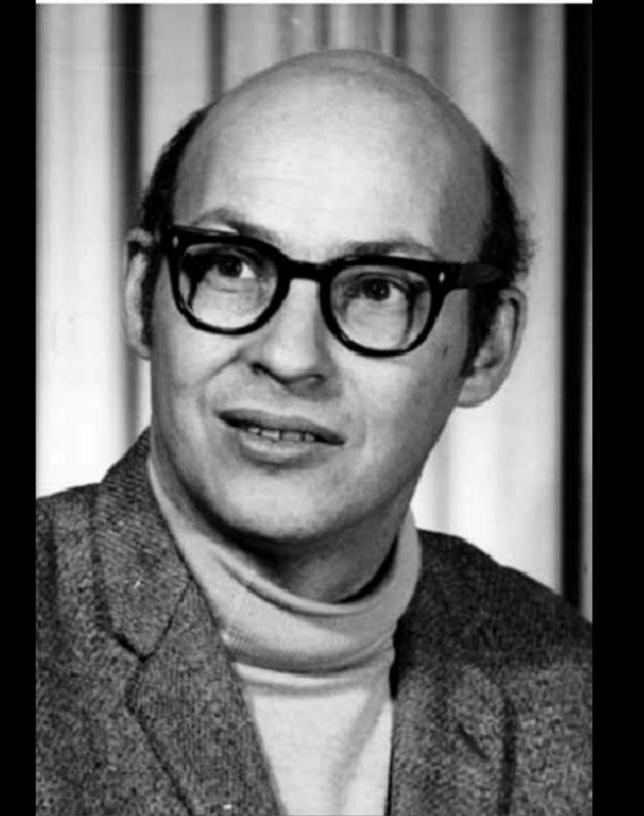


Arthur Samuel, 1959



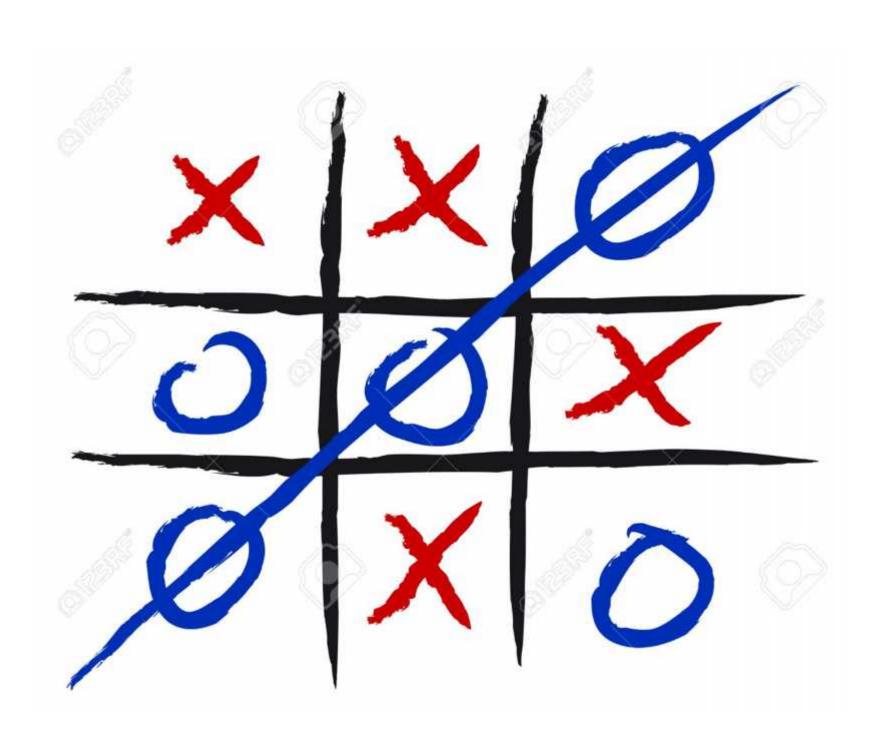
Sent Mail $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$













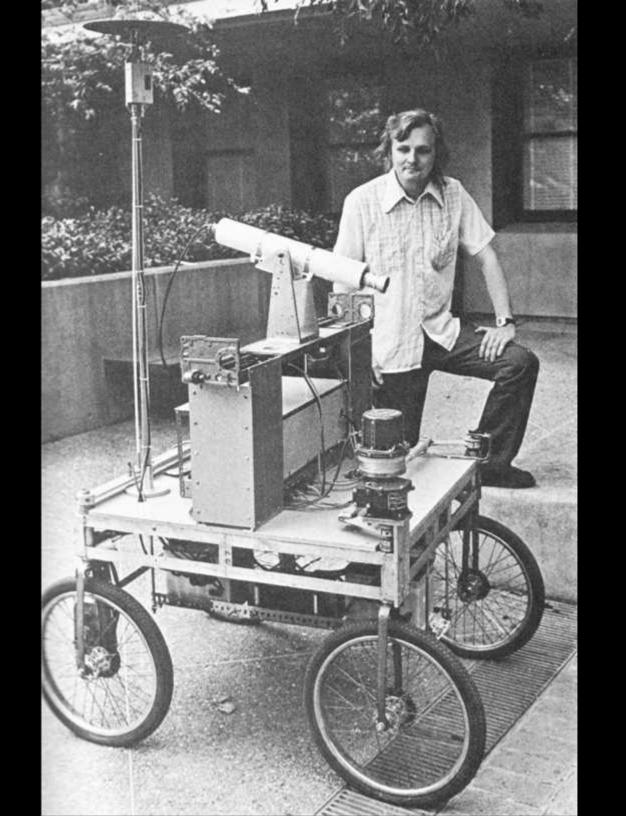
Expanded Edition



Perceptrons



Marvin L. Minsky Seymour A. Papert





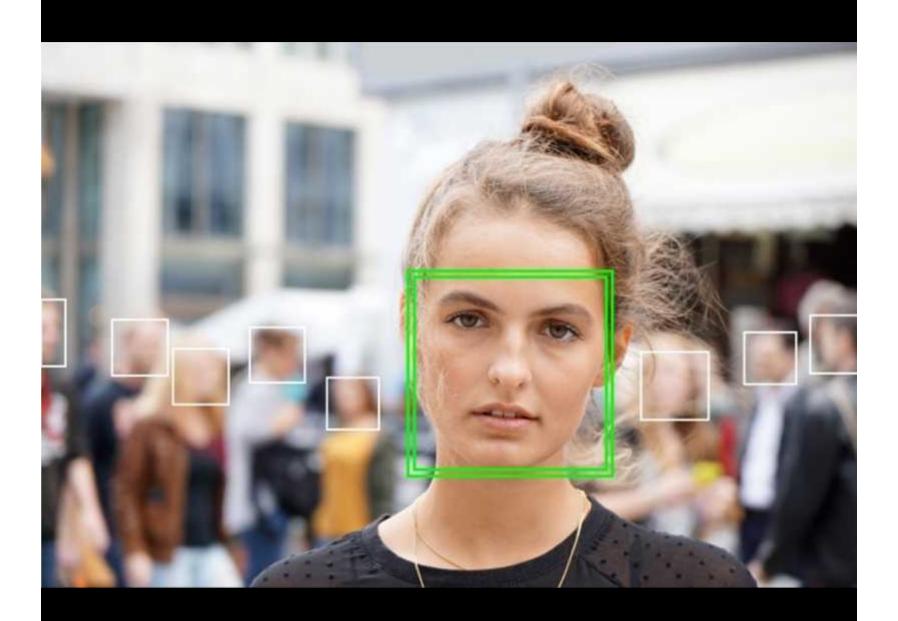


/ | | | / 1 | / 7 1 | / / / / ファチ17ァフフフフフファ













Machine learning se utiliza cuando:

- O Programar un algoritmo es imposible
- O El problema es muy complejo o no se conocen algoritmos para resolverlo
- O Ayuda a los humanos a entender patrones (data mining)
- Aprendizaje supervisado vs no supervisado vs semisupervisado
- Batch vs online learning

Feature vectors

- Se utilizan para representar características simbólicas o numéricas llamadas features.
- Permiten analizar un objeto desde una perspectiva matemática.
- Los algoritmos de machine learning típicamente requieren representaciones numéricas para poder ejecutar el cómputo.
- Uno de los feature vectors más conocidos es la representación del color a través de RGB

O color = [R, G, B]

- Procesamiento de imágenes:
 - O Gradientes, bordes, áreas, colores, etc.
- Reconocimiento de voz:
 - O Distancia de sonidos, nivel de ruido, razón ruido / señal, etc.
- Spam:
 - O Dirección IP, estructura del texto, frecuencia de palabras, encabezados, etc.

Métricas de Distancia

- Muchos de los algoritmos de machine learning pueden clasificarse como algoritmos de optimización.
- Lo que desean optimizar es una función que en muchas ocasiones se refiere a la distancia entre features.

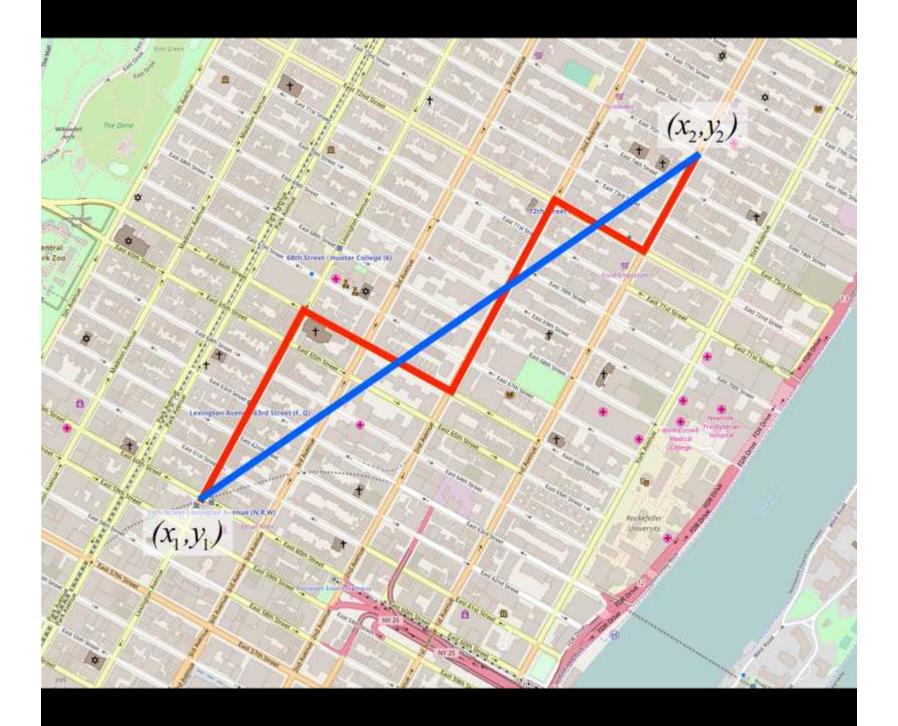
$$x = (a, \mathbf{b}) \qquad y = (c, \mathbf{d})$$

Distancia euclidiana

$$\sqrt{(a-c)^2+(\pmb{b}-\pmb{d})^2}$$

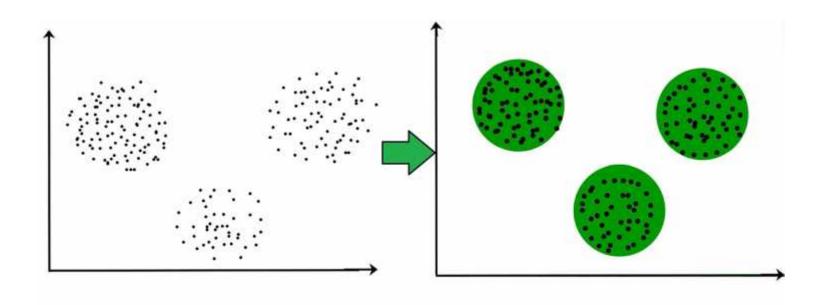
o Distancia de Manhattan

$$|a-c|+|b-d|$$



Introducción al agrupamiento (clustering)

- Es un proceso mediante el cual se agrupan objetos similares en clusters que los identifican.
- Se clasifica como aprendizaje no supervisado ya que no requiere la utilización de etiquetas.
- Permite entender la estructura de los datos y la similitud entre los mismos.
- Es utilizado en motores de recomendación, análisis de redes sociales, análisis de riesgo crediticio, clasificación de genes, riesgos médicos, etc.



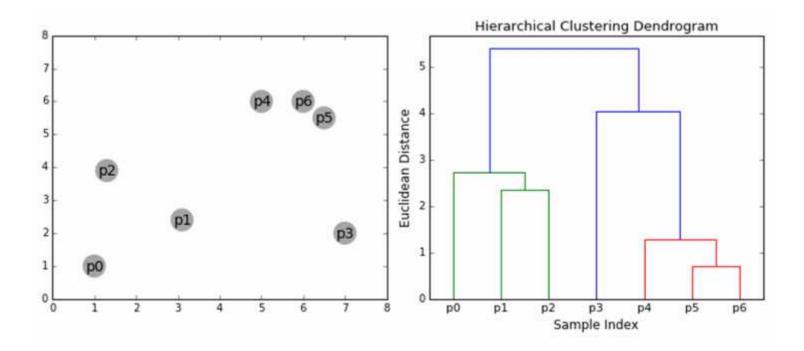


Agrupamiento

jerárquico

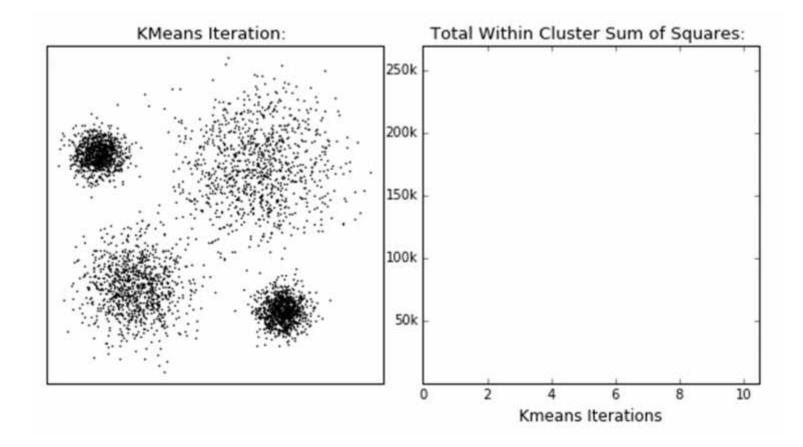


- Es un algoritmo que agrupa objetos similares en grupos llamados clusters
- El algoritmo comienza tratando a cada objeto como un cluster individual y luego realiza los siguientes pasos de manera recursiva:
 - O Identifica los dos clusters con menor distancia (los más similares)
 - O Agrupa los dos clusters en uno nuevo
- El output final es un dendrograma que muestra la relación entre objetos y grupos.
- Es importante determinar qué medida de distancia vamos a utilizar y los puntos a utilizar en cada cluster (linkage criteria)



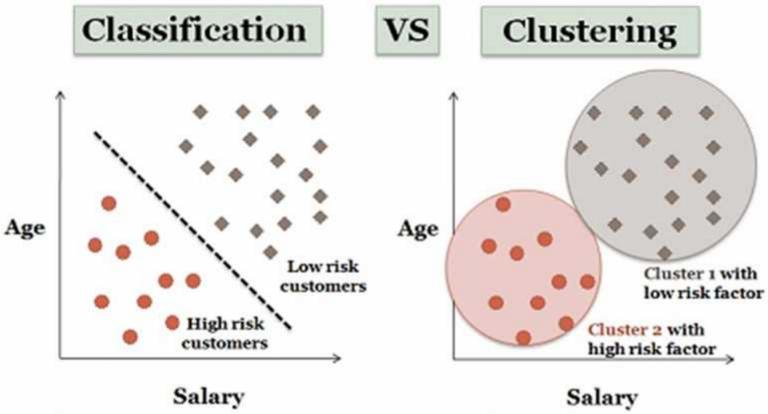
Agrupamiento por K-means >

- Es un algoritmo que agrupa utilizando centroides.
- El algoritmo funciona asignando puntos al azar (K define el número inicial de clusters) y después:
 - O En cada iteración el punto se ajusta a su nuevo centroide y cada punto se recalcula con la distancia con respecto de los centroides
 - O Los puntos se reasignan al nuevo centro
 - O El algoritmo se repite de manera iterativa hasta que ya no existen mejoras



Introducción a la clasificación

- Es el proceso mediante el cual se predice la clase de cierto dato.
- Es un tipo de aprendizaje supervisado ya que para que funcione, se necesitan etiquetas con los datos (labels).
- Se utiliza en muchos dominios, incluyendo la medicina, aprobación crediticia, reconocimiento de imágenes, vehículos autónomos, entre otros.
- Sigue dos pasos: aprendizaje (creación del modelo) y clasificación.

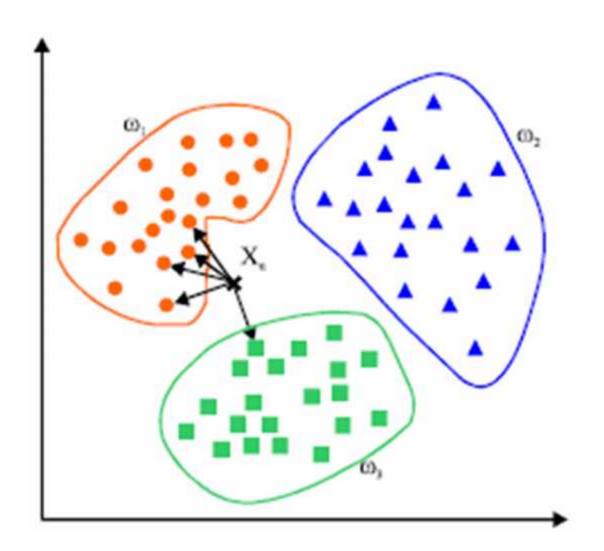


Risk classification for the loan payees on the basis of customer salary



por K-nearest neighbors

- Parte del supuesto de que ya tenemos un conjunto de datos clasificado.
- Trata de encontrar los "vecinos más cercanos".
- K se refiere a la cantidad de vecinos que se utilizarán para clasificar un ejemplo que aún no ha sido clasificado.
- Es sencillo de implementar y tiene aplicaciones en medicina, finanzas, agricultura, etc.
- Es computacionalmente muy costoso y no sirve con datos de alta dimensionalidad.





Conclusiones

- El pensamiento probabilístico permite evaluar un mundo que es por su propia naturaleza no determinista.
- También permite incorporar evidencia para ajustar el modelo probabilístico.
- Los números no mienten, pero las personas que los usan pueden usarlos para mentir.
- Machine learning permite que las computadoras generen modelos matemáticos sin necesidad de ser programadas explícitamente.

- Las ciencias de la computación es un área de conocimiento basta y que ha revolucionado al mundo.
- Computadoras, lenguajes de programación, algoritmos, estructuras de datos, técnicas de programación (OOP, dinámica, estocástica, probabilística, funcional, etc.), inferencia estadística, machine learning, etc.

- Ahora ya tienes un panorama para que puedas escoger qué área te gustaría profundizar.
- La serie de pensamiento computacional te dio herramientas no sólo para ser un mejor ingeniero, sino para fortalecer tu capacidad de razonamiento lógico.