

BOURBAKI

COLEGIO DE MATEMÁTICAS



Causalidad en Machine Learning

Los principios del cálculo Do

- Dudas y examen
- Paradoja de Simpson
- Criterio de Back-Door
- Efectos directos





MÓDULO IV

¿ D U D A S ?





Caso de uso

EQUIDAD ALGORÍTMICA



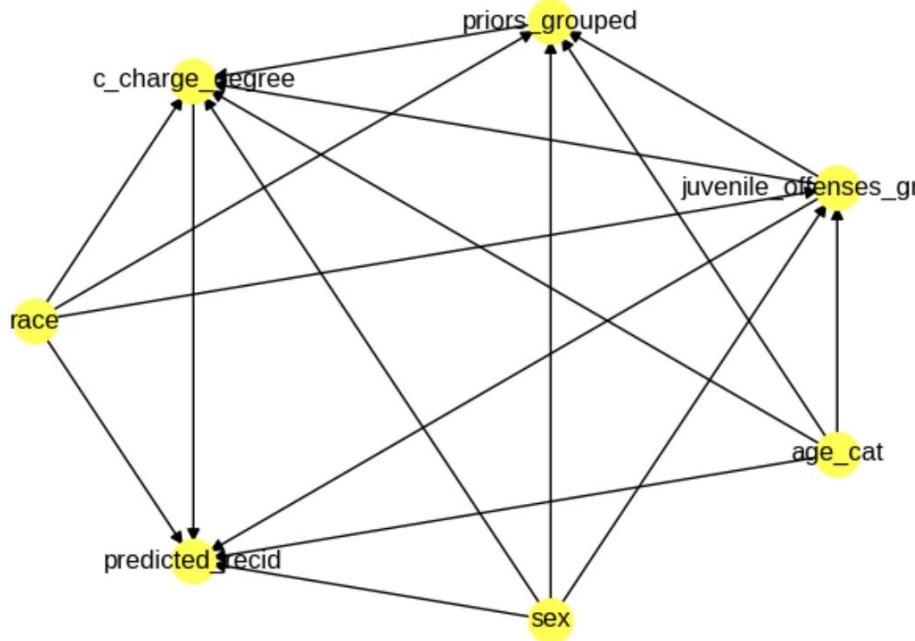


Bernard Parker, left, was rated high risk; Dylan Fuggett was rated low risk by the same software.

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.





Un sesgo en contra de
personas negras



COLEGIO DE MATEMÁTICAS





Paradoja

S I M P S O N



Sesgo entre hombres y
mujeres en Berkeley





PRÉSTAMOS
vs
AUMENTO LÍNEA
DE CRÉDITO





P O B L A C I Ó N:
casa o automóvil





OFERTA

Préstamo

Si o no



OFERTA

LÍNEA DE CRÉDITO

Si o no



Préstamo

Hipoteca

SÍ O NO

Préstamo

Automotriz

SÍ O NO





Línea de crédito

Hipoteca

SÍ O NO

Línea de crédito

Automotriz

SÍ O NO





PARADOJA DE SIMPSON

OFERTA

CONTRATO

Préstamo



Línea de Crédito



78%

No

86%

No

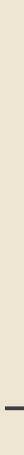




¿CORRELACIÓN?

$y = \text{contrato}$

contrato



oferta = x

préstamo

línea de crédito

CLIENTE

OFERTA

CONTRATO

Hipoteca 	Préstamo 	78%
Auto 		
Hipoteca 		
Auto 		
Hipoteca 	Línea de Crédito 	86%
Auto 		
Hipoteca 		
Auto 		

No

86%

No

CLIENTE

OFERTA

CONTRATO

		Hipoteca	Préstamo	78%
		Auto		No
A ₁				
A ₂	A ₁	Hipoteca		
	a ₁	Auto		
B ₁		Hipoteca	Línea de Crédito	86%
B ₂	b ₁	Auto		
	b ₂	Hipoteca		
b ₂	a ₂	Auto		

$$78\% = \frac{A_1 + a_1}{A_2 + a_2} < \frac{B_1 + b_1}{B_2 + b_2} = 86\%$$

Por ejemplo:

$$A_1 = 81 \quad B_1 = 234$$

$$a_1 = 87 \quad b_1 = 240$$

$$A_2 = 192 \quad B_2 = 55$$

$$a_2 = 263 \quad b_2 = 80$$

CLIENTE

OFERTA

CONTRATO

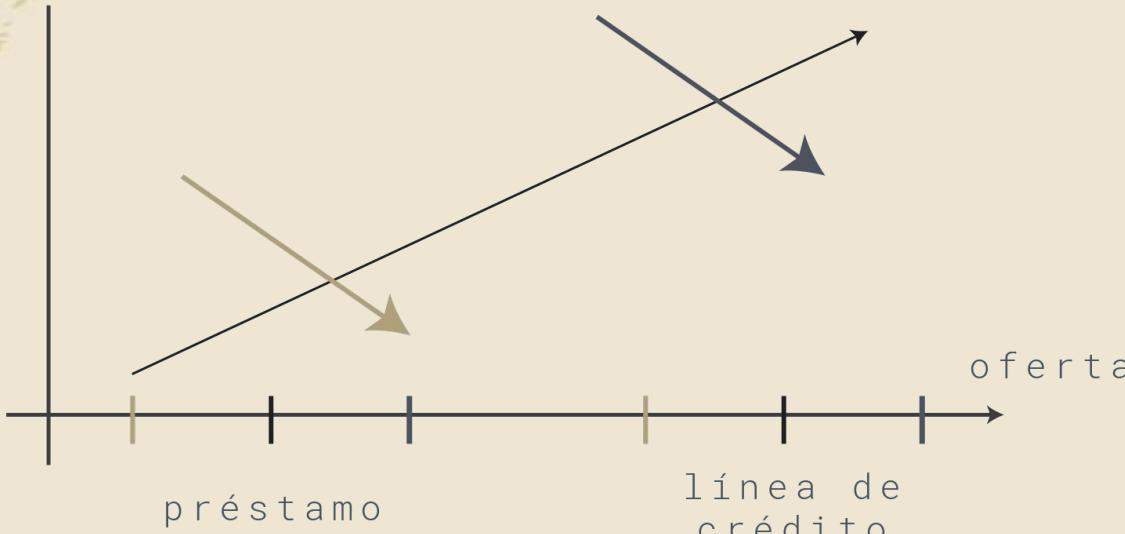
A_1	Hipoteca 		93%
A_2	Préstamo 	No	
a_1	Auto 	73%	
a_2		No	
	Hipoteca 	87%	
	Línea de Crédito 	No	
	Auto 	69%	
		No	

$$93\% = \frac{A_1}{A_2} > \frac{B_1}{B_2} = 87\%$$

$$43\% = \frac{a_1}{a_2} > \frac{b_1}{b_2} = 69\%$$

$z = \text{cliente}$

contrato



¿ P A R A D O J A ?

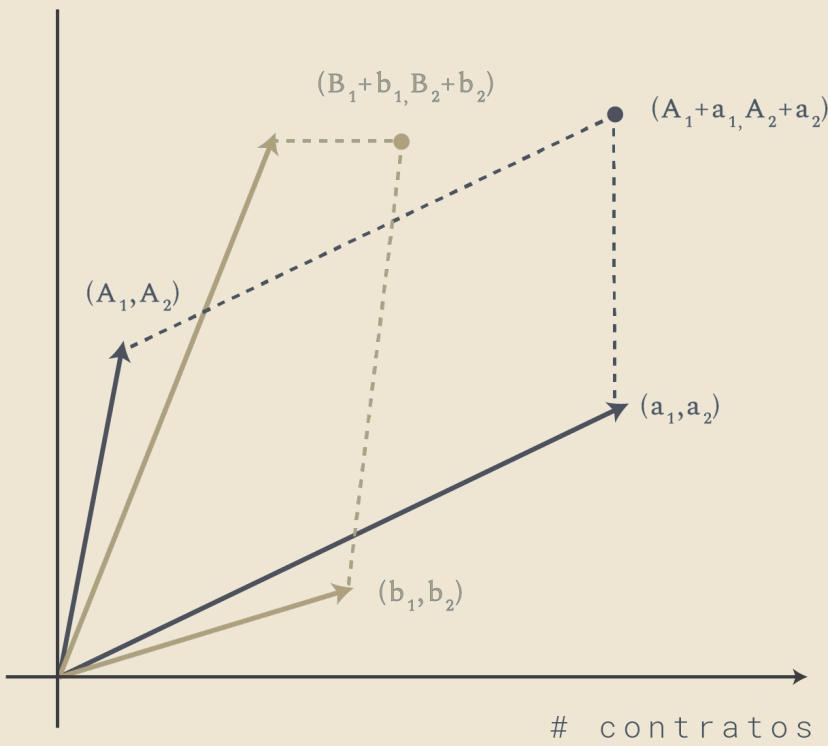


COLEGIO DE MATEMÁTICAS



VECTORIALMENTE

muestras



E X P L I C A C I Ó N
S A T I S F A C T O R I A



COLEGIO DE MATEMÁTICAS





SOLUCIÓN

Paradoja de Simpson

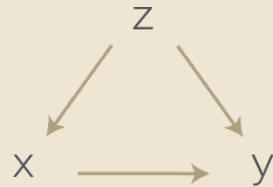
- Teorema de imposibilidad
- Distinguir entre los dos casos

T E O R E M A D E
I M P O S I B I L I D A D



ESTUDIO CAUSAL

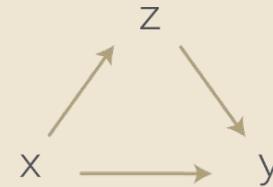
Simpson es posible



$$\mathbb{P}(z, x, y) = \mathbb{P}(z) \cdot \mathbb{P}(x | z) \cdot \mathbb{P}(y | x, z)$$

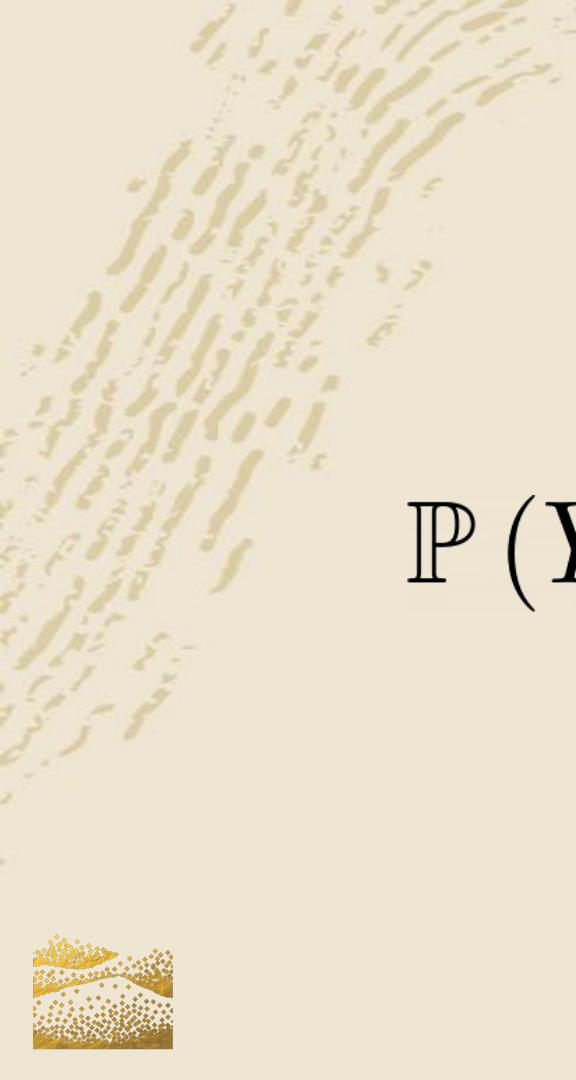
(datos)

Simpson es posible



$$\mathbb{P}(z | xy) = \mathbb{P}(z | x) \cdot \mathbb{P}(x) \cdot \mathbb{P}(y | zx)$$

(datos)



SIMPSON

$$\mathbb{P}(Y|X) > \mathbb{P}(Y|-X)$$



SIMPSON

$$\mathbb{P}(Y|X, Z) < \mathbb{P}(Y| - X, Z)$$

$$\mathbb{P}(Y|X, -Z) < \mathbb{P}(Y| -X, -Z)$$



SIMPSON

¿Se puede evitar con
el operador Do?





Simpson es imposible
si calculamos relaciones
condicionales causales

NUEVA HIPÓTESIS

$$\mathbb{P}(Z|do(X)) = \mathbb{P}(Z|do(-X))$$

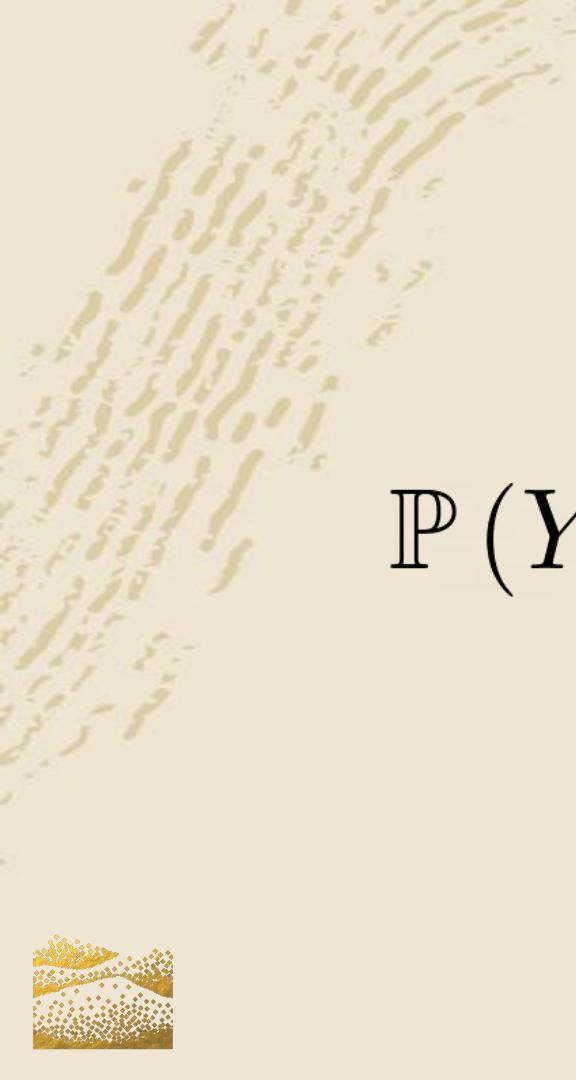
¿DAG?

z

x

y





¿ES REALISTA?

$$\mathbb{P}(Y|do(X)) = \mathbb{P}(Y|X)$$



SURE-THING PRINCIPLE

Si $P(\text{Contrato} \mid \text{do}[\text{Línea de crédito}], \text{Hipoteca}) > P(\text{Contrato} \mid \text{do}[\text{Préstamo}], \text{Hipoteca})$
 $P(\text{Contrato} \mid \text{do}[\text{Línea de crédito}], \text{Automotriz}) > P(\text{Contrato} \mid \text{do}[\text{Préstamo}], \text{Automotriz})$
 $\Rightarrow P(\text{Contrato} \mid \text{do}[\text{Línea de crédito}]) > P(\text{Contrato} \mid \text{do}[\text{Préstamo}])$

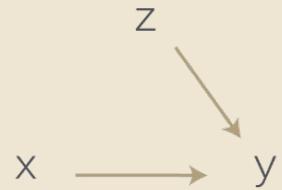
D I S T I N G U I R
E N T R E A M B O S C A S O S





¿Debemos tomar en
cuenta la variable z ?

ESTUDIO ALEATORIZADO

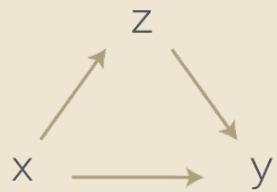


Confounding

Backdoor criterion



CASOI



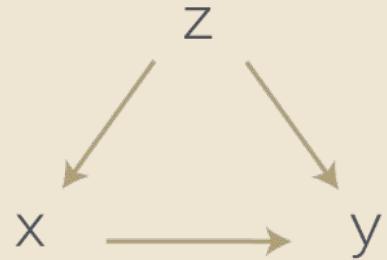


SOLUCIÓN

Paradoja de Simpson

- No se cumple el criterio back-door
- No debemos de condicionar en este caso

CASO II





SOLUCIÓN

Paradoja de Simpson

- Sí se cumple el criterio back-door
- Debemos de condicionar en este caso



B O U R B A K I

Contacto

tel. +52 56 2141 7850

info@colegio-bourbaki.com

colegio-bourbaki.com



B O U R B A K I