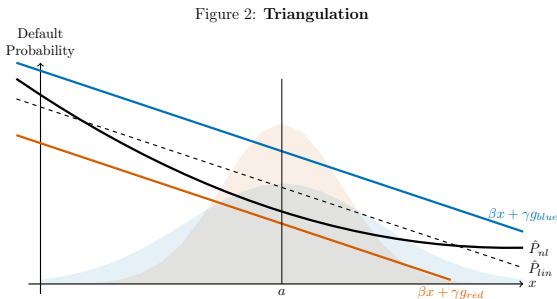


- Sub reporte: los datos son una muestra sesgada de la realidad.
- Equilibrio: las predicciones desconocen que pueden existir interacciones estratégicas.
- Privacidad: con suficiente información es posible identificar personas o atributos privados.
- Este es uno de los grandes retos en las aplicaciones de las matemáticas y aprendizaje de máquinas al mundo real.

- Ahora suponga que existe una característica (i.e., raza) que no puede ser utilizada para discriminar y tiene correlación cero con la variable observada.
- Supongamos que el verdadero modelo es de la forma $y = \beta x + \gamma g$.
- Una tecnología no lineal no puede, por definición, mejorar el anterior modelo.
- Sin embargo, en ausencia de la variable protegida en el modelo, sí puede ser mejor descubriendo el papel de la variable protegida.

Triangulación

- Una mejor tecnología tiene efectos negativos sobre el grupo azul (los más dispersos).



Predicción de incumplimiento

Table 1: Descriptive Statistics, 2009-2013 Originations

Group		FICO	Income	LoanAmt	Rate (%)	SATO (%)	Default (%)
Asian (N=574,812)	Mean	764	122	277	4.24	-0.07	0.42
	Median	775	105	251	4.25	-0.05	0.00
	SD	40	74	149	0.71	0.45	6.49
Black (N=235,673)	Mean	735	91	173	4.42	0.11	1.88
	Median	744	76	146	4.50	0.12	0.00
	SD	58	61	109	0.71	0.48	13.57
White Hispanic (N= 381,702)	Mean	746	90	187	4.36	0.07	0.99
	Median	757	73	159	4.38	0.07	0.00
	SD	52	63	115	0.71	0.47	9.91
White Non-Hispanic (N=7,134,038)	Mean	761	110	208	4.33	-0.00	0.71
	Median	774	92	178	4.38	0.02	0.00
	SD	45	73	126	0.69	0.44	8.37
Native Am, Alaska, Hawaii/Pac Isl (N=59,450)	Mean	749	97	204	4.39	0.04	1.12
	Median	761	82	175	4.45	0.04	0.00
	SD	51	65	123	0.70	0.46	10.52
Unknown (N=984,310)	Mean	760	119	229	4.38	0.00	0.79
	Median	773	100	197	4.50	0.02	0.00
	SD	46	78	141	0.68	0.44	8.85

Note: Income and loan amount are measured in thousands of USD. SATO stands for “spread at origination” and is defined as the difference between a loan’s interest rate and the average interest rate of loans originated in the same calendar quarter. Default is defined as being 90 or more days delinquent at some point over the first three years after origination. Data source: HMDA-McDash matched dataset of fixed-rate mortgages originated over 2009-2013.

Predicción de incumplimiento

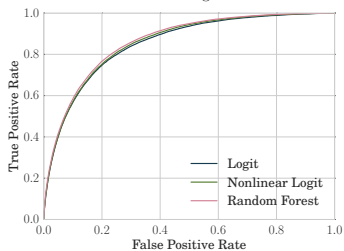
Table 2: **Variable List**

<i>Logit</i>	<i>Nonlinear Logit</i>
Applicant Income (linear)	Applicant Income (25k bins, from 0-500k)
LTV Ratio (linear)	LTV Ratio (5-point bins, from 20 to 100%; separate dummy for LTV=80%)
FICO (linear)	FICO (20-point bins, from 600 to 850); separate dummy for FICO<600)
(with dummy variables for missing values)	
<i>Common Covariates</i>	
Spread at Origination “SATO” (linear)	
Origination Amount (linear and log)	
Documentation Type (dummies for full/low/no/unknown documentation)	
Occupancy Type (dummies for vacation/investment property)	
Jumbo Loan (dummy)	
Coapplicant Present (dummy)	
Loan Purpose (dummies for purchase, refinance, home improvement)	
Loan Term (dummies for 10, 15, 20, 30 year terms)	
Funding Source (dummies for portfolio, Fannie Mae, Freddie Mac, other)	
Mortgage Insurance (dummy)	
State (dummies)	
Year of Origination (dummies)	

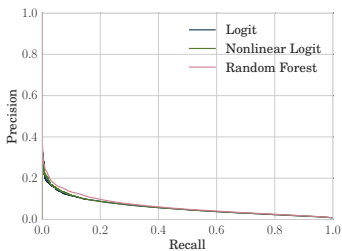
Note: Variables used in the models. Data source: HMDA-McDash matched dataset of conventional fixed-rate mortgages.

Predicción de incumplimiento

Figure 3: ROC and Precision-Recall Curves



Panel A: ROC



Panel B: Precision-Recall

Predicción de incumplimiento

Table 3: Performance of Different Statistical Technologies Predicting Default

Model	ROC AUC		Precision Score		Brier Score $\times 100$		R^2	
	(1) No Race	(2) Race	(3) No Race	(4) Race	(5) No Race	(6) Race	(7) No Race	(8) Race
Logit	0.8522	0.8526	0.0589	0.0592	0.7172	0.7171	0.0245	0.0246
Nonlinear Logit	0.8569	0.8573	0.0598	0.0601	0.7146	0.7145	0.0280	0.0281
Random Forest	0.8634	0.8641	0.0630	0.0641	0.7114	0.7110	0.0323	0.0329

Table 4: Performance of Different Statistical Technologies Predicting Race

Model	ROC AUC	Precision Score	Brier Score $\times 10$	R^2
Logit	0.7478	0.1948	0.5791	0.0609
Nonlinear Logit	0.7485	0.1974	0.5783	0.0622
Random Forest	0.7527	0.2110	0.5665	0.0813

Curvas: ROC y Precision-Recall

- La curva ROC (TPR o sensitivity vrs. FPR o 1-especificidad) no es una muy buena medida cuando las clases están muy desbalanceadas.
- En particular, si los positivos son muy poco (incumplidos) los falsos positivos son aún menos (piense en un modelo perfecto en donde los falsos positivos serían cero) y FPR es cercano a cero.

Curvas: ROC y Precision-Recall

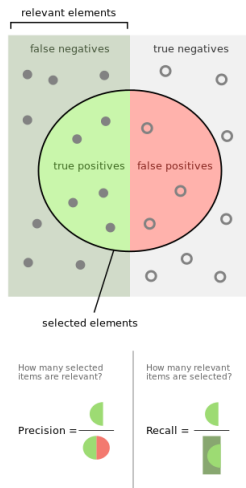
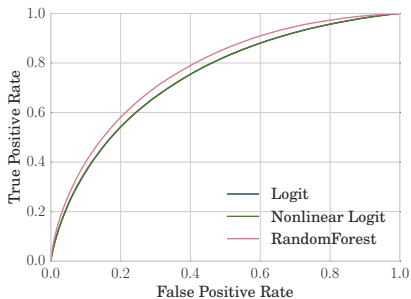


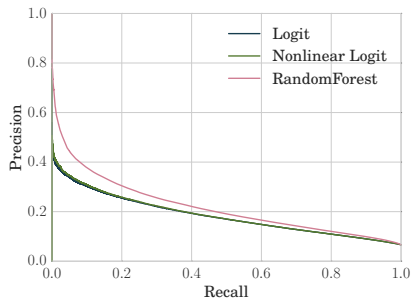
Figura 1: By Walber - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=36926283>

Obsérvese que $\text{TPR} = \text{Recall}$

Figure 4: ROC and Precision-Recall Curves of Predicting Race



Panel A



Panel B

Mejor tecnología vrs triangulación

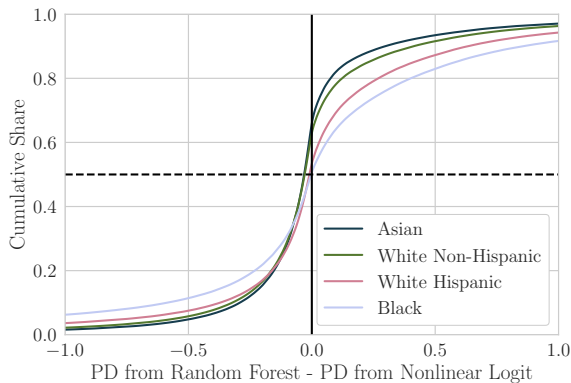
Table 5: **Decomposition of Performance Improvement**

	Race	Technology		Technology	Race
ROC-AUC	5.88	94.12	ROC-AUC	91.16	8.84
Precision	7.90	92.10	Precision	77.21	22.79
Brier	3.25	96.75	Brier	90.63	9.37
R^2	2.04	97.96	R^2	87.75	12.25

Panel A: Race Controls First Panel B: New Technology First

Efectos discriminatorios

Figure 6: **Comparison of Predicted Default Probabilities Across Models, by Race Groups**



Panel A

Outline

- 1 Introducción
- 2 Discriminación
- 3 Modelos Formales de Discriminación**
- 4 Teorema de Imposibilidad
- 5 Algoritmos derivados (aprox) no discriminatorios

Introducción

- Existen muchas definiciones formales de discriminación algorítmica.
- Vamos a estudiar formalmente tres:
 - 1 Paridad demográfica (*equalized odds*).
 - 2 Calibración y calibración condicional.
 - 3 Balance de clases positivas (igualdad de oportunidad).
 - 4 Balance de clases negativas.

- 1 Y es una variable binaria que queremos predecir.
- 2 \hat{Y} es la predicción.
- 3 La clase 1 se va considerar la decisión social preferida (otorgar un crédito, ser beneficiario de un programa social, etc.).
- 4 X son las variables observadas de cada individuo.
- 5 A es una variable binaria que representa la clase protegida (hombres, mujeres, blanco, negro, religioso, ateo, etc.)

Definition

Decimos que un algoritmo satisface priridad demográfica si:

$$P(\hat{Y} = 1 | A) = P(\hat{Y} = 1) \quad (1)$$

Definition

Decimos que un algoritmo satisface la propiedad de calibración con respecto a la variable protegida si:

- 1 El pronóstico está bien calibrado.
- 2 Condicional a la variable protegida el pronóstico está bien calibrado.

Definition

Decimos que un algoritmo satisface igualdad de oportunidad si:

$$P(\hat{Y} = 1 \mid Y = 1, A) = P(\hat{Y} = 1 \mid Y = 1) \quad (2)$$

Definition

Decimos que un algoritmo satisface balance de clases negativas:

$$P(\hat{Y} = 1 \mid Y = 0, A) = P(\hat{Y} = 1 \mid Y = 0) \quad (3)$$

Kleinber, Mullainathan y Raghavan

IMullgualdg, ad de oportunidades y balance de clases negativas se llama también *equalized odds*.

- Es imposible que, excepto en casos triviales, que un algoritmo satisfaga: equalized odds y la propiedad de calibración.

