

# Big Data y Aprendizaje Automático: Ideas y Desafíos para Economistas

Walter Sosa-Escudero

Universidad de San Andrés y CONICET, Argentina

[wsosa@udesa.edu.ar](mailto:wsosa@udesa.edu.ar) • [@wsosaescudero](https://twitter.com/wsosaescudero) • [waltersosa.weebly.com](http://waltersosa.weebly.com)

# Big data?

- Datos masivos, interacciones electronicas.
- *Construccion* de modelos (aprendizaje).
- Flexibilidad demanda muchos datos.
- Induccion y descripcion.

# Econometria vs. aprendizaje automatico

$$y = f(x) + u$$

	<b>Econometria</b>	<b>Aprendizaje</b>
<b>Actor</b>	$f(x)$	$y$
<b>Accion</b>	Estimar	Predecir
<b>Modelo</b>	Exogeno	Endogeno
<b>Bueno</b>	Insesgado, varianza minima, etc.	Capacidad predictiva fuera de la muestra

# Ejemplo: LASSO

$$L(\beta) = \sum (y_i - x_i\beta)^2 + \lambda \beta^2$$

- $\lambda = 0$ : MCO,  $\lambda = \infty : \hat{\beta} = 0$
- $\lambda \neq 0$ : sesgado, pero....
- ... **siempre** le gana a MCO en prediccion.
- Sesgo: pecado mortal en econometria, no en aprendizaje.
- **Aprendizaje**: sesgo puede reducir drasticamente la varianza.
- $\lambda$  es un **hiperparametro** que se 'aprende' en la estimacion.

Esto es ridge. LASSO reemplaza  $\beta^2$  por  $|\beta|$ .

- 1 Aprendizaje, estadística y econometría
- 2 Regresion para la predicción
- 3 Clasificacion
- 4 Aprendizaje no supervisado: clusters, reduccion de dimensionalidad.