# The Predictive Paradigm

## Big Data y Machine Learning para Economía Aplicada

Ignacio Sarmiento-Barbieri

Universidad de los Andes

# Agenda

# ¿Qué entendemos por Big Data y ML?

- ¿Que es Big Data?
  - Big n, es solo parte de la historia
  - Big también es big k, muchos covariates, a veces $n << k$
  - Vamos a entender Big también como datos que no surgen de fuentes tradicionales (cuentas nac., etc)
    - Datos de la Web, Geográficos, etc.
- Machine Learning
  - Cambio de paradigma de estimación a predicción

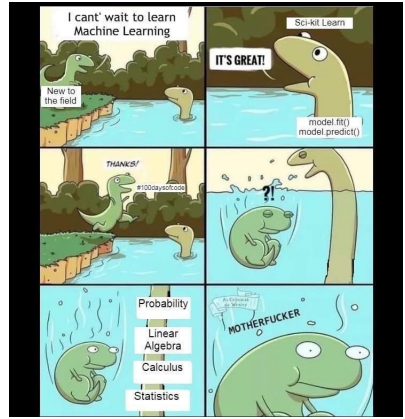
# Agenda

1. ¿Qué entendemos por Big Data y ML?
2. About the Course
3. Machine learning is all about prediction
4. ML Tasks
5. Prediction vs Causality
6. Getting serious about prediction
   - The basic logic of prediction
   - Minimizing our losses
7. Regression and the Best Linear Prediction Problem
8. Review

# Agenda

# Lenguajes

- Estadística y Econometría
- Inglés
- Código
  - Elijan el que quieran:
    - `R`, `Python`, o cualquier otro
    - no hay restricción
    - nosotros usaremos `R`
  - Github
  - Slack
- Materiales en BN
- Aprender haciendo y mucha prueba y error!



# Agenda

# Machine learning is all about prediction

- Machine learning is a branch of computer science and statistics, tasked with developing algorithms to predict outcomes $Y$ from observable variables $X$.
- The learning part comes from the fact that we don't specify how exactly the computer should predict $Y$ from $X$.
- This is left as an empirical problem that the computer can "learn".
- In general, this means that we abstract from the underlying model, the approach is pragmatic

# Machine learning is all about prediction

- Machine learning is a branch of computer science and statistics, tasked with developing algorithms to predict outcomes $Y$ from observable variables $X$.
- The learning part comes from the fact that we don't specify how exactly the computer should predict $Y$ from $X$.
- This is left as an empirical problem that the computer can "learn".
- In general, this means that we abstract from the underlying model, the approach is pragmatic

**"Whatever works, works...."**

# “Whatever works, works....”

# “Whatever works, works....”????

- In many applications, ML techniques can be successfully applied by data scientists with little knowledge of the problem domain.
- For example, Kaggle competitions

# “Whatever works, works....”????

- Much less attention has been paid to the limitations of pure prediction methods.
- When ML applications are used “off the shelf” without understanding the underlying assumptions or ensuring that conditions like stability are met, then the validity and usefulness of the conclusions can be compromised.

# Agenda

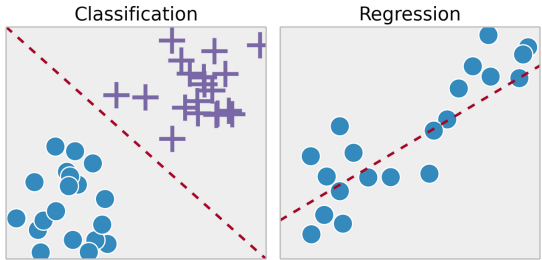
# Agenda

# ML branches

- ML tasks can (¿?) be divided into two main branches:
  1. Supervised Learning

# ML branches

- Supervised Learning
  - for each predictor $x_i$ a 'response' is observed $y_i$.
  - everything we have done in econometrics is supervised



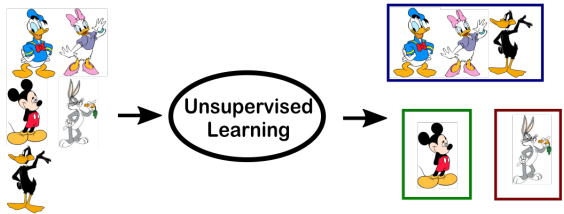

Source: shorturl.at/opqKT

# ML branches

- ML tasks can (¿?) be divided into two main branches:
  1. Supervised Learning
  2. Unsupervised Learning

# ML branches

- Unsupervised Learning
  - observed $x_i$ but no response.
  - example: cluster analysis




Source: shorturl.at/opqKT

# Agenda

# Policy Prediction Problems

- Empirical policy research often focuses on causal inference.
- Since policy choices seem to depend on understanding the counterfactual there's a tight link
- While this link holds in many cases, there are also many policy applications where causal inference is not central, or even necessary.

# The Causal Paradigm

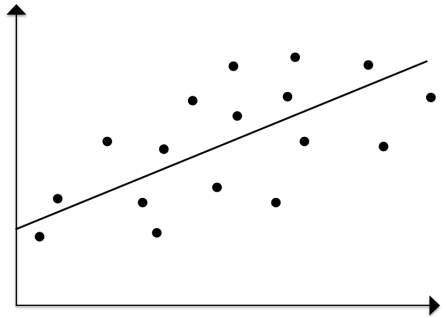
$$y = f(X) + u \tag{1}$$

- Interest lies on inference
- "Correct" $f()$ to understand how $y$ is affected by $X$
- Model: Theory, experiment
- Hypothesis testing (std. err., tests)

# Prediction vs. Causality: Target

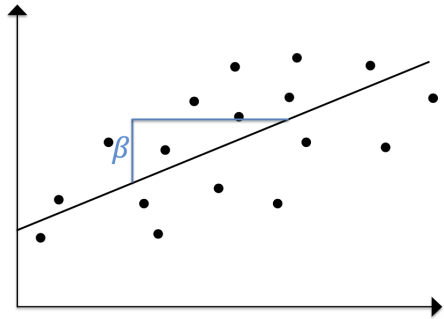
$$y = f(x) + \epsilon \tag{2}$$

$$y = \alpha + \beta x + \epsilon \tag{3}$$



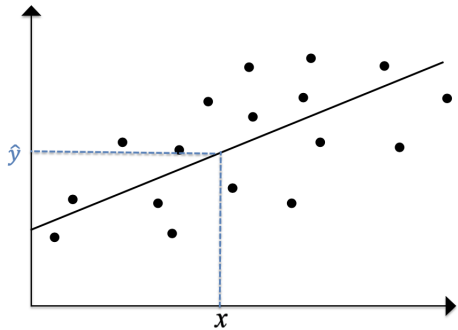
# Prediction vs. Causality: Target

$$y = \alpha + \beta x + \epsilon \tag{4}$$



# Prediction vs. Causality: Target

$$y = \underbrace{\alpha + \beta x}_{\hat{y}} + \epsilon \tag{5}$$



# The Predictive Paradigm

$$y = f(X) + u \tag{6}$$

- Interest on predicting $y$
- "Correct" $f()$ to be able to predict (no inference!)
- Model? We treat $f()$ as a black box, and any approximation $\hat{f}()$ that yields a good prediction is good enough (*Whatever works, works.*).

# Prediction vs. Causality: The garden of the parallel paths?

- We've seen that prediction and causality
  - Answer different questions
  - Serve different purposes
  - Seek different targets
- Different strokes for different folks, or complementary tools in an applied economist's toolkit?

# Agenda

# Agenda

1. ¿Qué entendemos por Big Data y ML?
2. About the Course
3. Machine learning is all about prediction
4. ML Tasks
5. Prediction vs Causality
6. Getting serious about prediction
   - The basic logic of prediction
   - Minimizing our losses
7. Regression and the Best Linear Prediction Problem
8. Review

# The basic logic of prediction

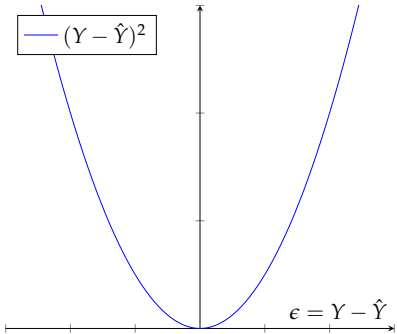
# Agenda

1. ¿Qué entendemos por Big Data y ML?
2. About the Course
3. Machine learning is all about prediction
4. ML Tasks
5. Prediction vs Causality
6. Getting serious about prediction
   - The basic logic of prediction
   - Minimizing our losses
7. Regression and the Best Linear Prediction Problem
8. Review

# Prediction error

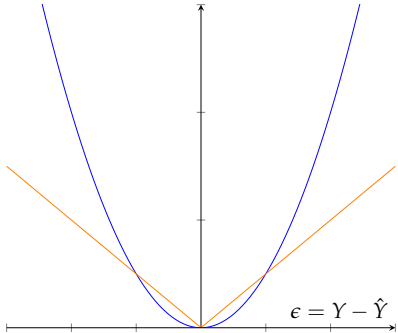
$$\epsilon = Y - \hat{Y} \tag{7}$$

$$L(Y, \hat{Y}) \tag{8}$$



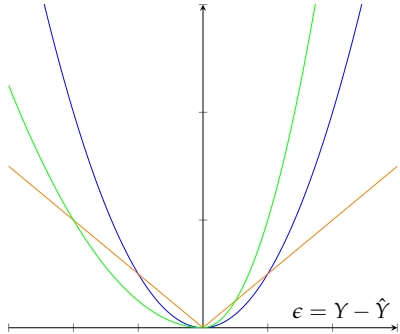
# Minimizing our losses

$$L(Y, \hat{Y}) \tag{9}$$



# Minimizing our losses

$$L(Y, \hat{Y}) \tag{10}$$



# Agenda

# Regression and the Best Linear Prediction Problem

- $Y$ scalar random variable: outcome
- $X = (X_1, \ldots, X_k)$ vector of covariates
- Goal: Construct the BLP

# Best Linear Approximation Property

$$E[(Y - \beta' X)X] = 0 \tag{11}$$

- by LIE

$$E[(E(Y|X) - \beta' X)X] = 0 \tag{12}$$

# From Best Linear Predictor to Best Predictor

- $W$ are "raw regressors"

$$X = T(W) \tag{13}$$

- dictionary

# From Best Linear Predictor to Best Predictor

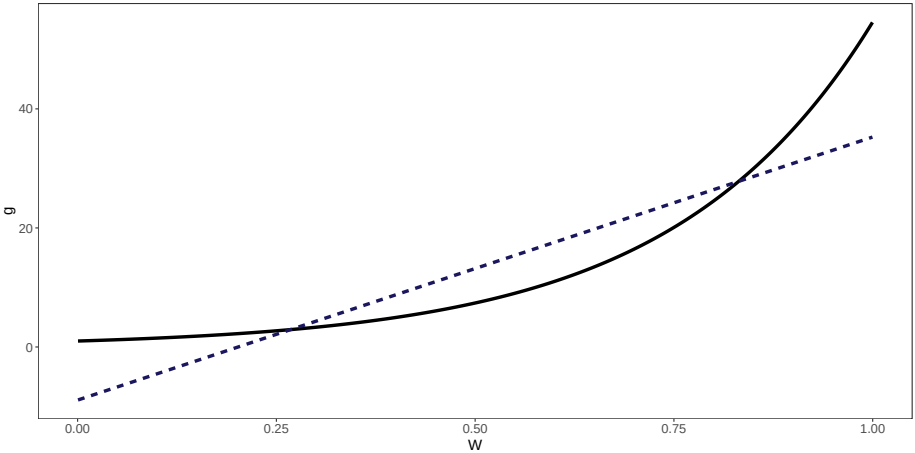
- In the population, the best predictor of Y given W

$$g(W) = E[Y|W] \tag{14}$$

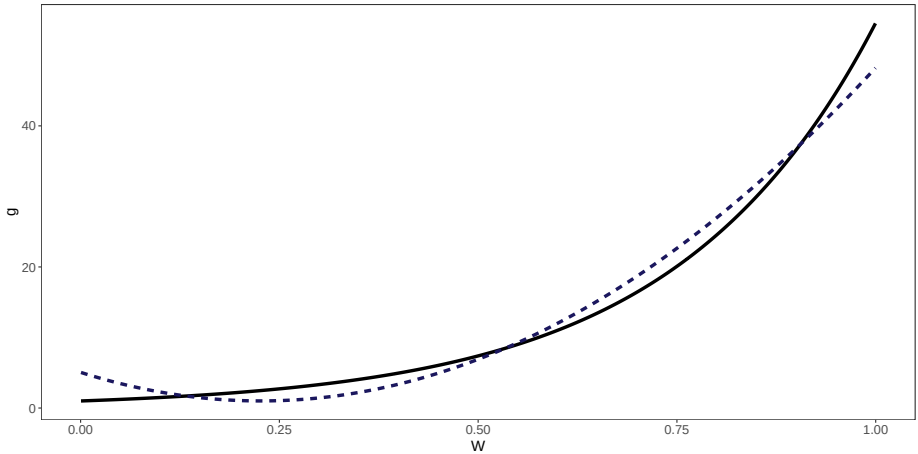
# From Best Linear Predictor to Best Predictor

$$E[(Y - b'T(W))^2] = E[(g(W) - b'T(W))^2] + E[(y - g(W)^2] \tag{15}$$

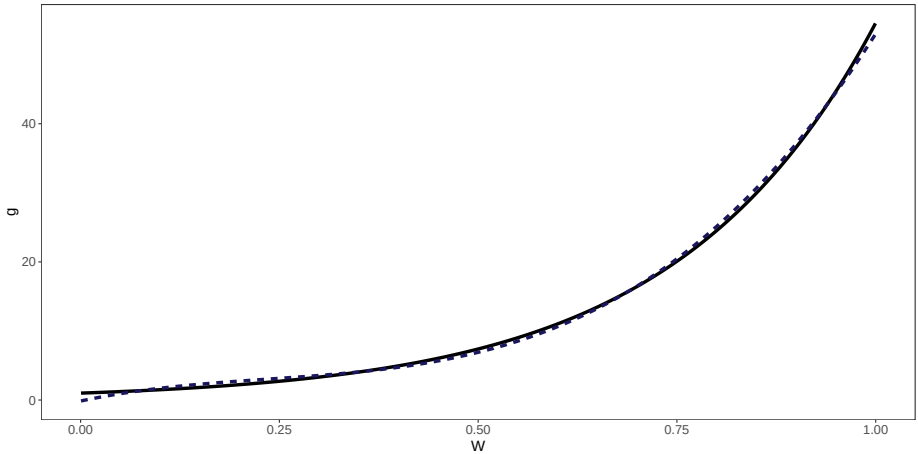
# From Best Linear Predictor to Best Predictor



# From Best Linear Predictor to Best Predictor



# From Best Linear Predictor to Best Predictor



# BLP problem in finite sample

$$(Y_i, X_i)_{i=1}^n = ((Y_1, X_1), \ldots, (Y_n, X_n)) \tag{16}$$

# Analysis of Variance

- Involves the decomposition of the variation of $Y$ into explained and unexplained parts.
- Explained variation is a measure of the predictive performance of a model.
- Can be conducted both in the population and in the sample.


photo from https://www.dailydot.com/parsec/batman-1966-labels-tumblr-twitter-vine/

# Agenda

# Review

- This Week: The predictive paradigm and linear regression
  - Machine Learning is all about prediction
  - ML targets something different than causal inference, they can complement each other
  - Linear Regression can approximate $E(y|X)$
  - BLP
- Next Week: Inner workings of linear regression, Quiz