Lecture 1: Introducción

Aprendizaje y Minería de Datos para los Negocios

Ignacio Sarmiento-Barbieri

Universidad de los Andes

October 20, 2021

Agenda

1. Aprendizaje de máquinas es todo sobre predicción
   - Ejemplos para Motivarnos
2. Sobre el curso
3. Tipos de Aprendizaje
4. A nadar
   - Regresión Lineal
   - Precisión del modelo
5. Recap
6. Break
7. `R para ML`

1. Aprendizaje de máquinas es todo sobre predicción
   - Ejemplos para Motivarnos
2. Sobre el curso
3. Tipos de Aprendizaje
4. A nadar
   - Regresión Lineal
   - Precisión del modelo
5. Recap
6. Break
7. `R para ML`

¿Qué es la máquina de aprender?

- El aprendizaje de máquinas es una rama de la informática y la estadística, encargada de desarrollar algoritmos para predecir los resultados y a partir de las variables observables X.
- La parte de aprendizaje proviene del hecho de que no especificamos cómo exactamente la computadora debe predecir y a partir de X.
- Esto queda como un problema empírico que la computadora puede ”aprender”.
- En general, esto significa que nos abstraemos del modelo subyacente, el enfoque es muy pragmático

¿Qué es la máquina de aprender?

- El aprendizaje de máquinas es una rama de la informática y la estadística, encargada de desarrollar algoritmos para predecir los resultados y a partir de las variables observables X.
- La parte de aprendizaje proviene del hecho de que no especificamos cómo exactamente la computadora debe predecir y a partir de X.
- Esto queda como un problema empírico que la computadora puede ”aprender”.
- En general, esto significa que nos abstraemos del modelo subyacente, el enfoque es muy pragmático

“Lo que sea que funciona, funciona...”

“Lo que sea que funciona, funciona...”

Motivación

La primera victoria y derrota de ML

- Contexto ¿similar? al de hoy: Epidemia de la gripe A en 2009
- En EEUU la forma de monitorear es a través de reportes de la CDC
- La CDC agrega a nivel de ciudad, condado, estado, región y a nivel nacional
- Todo esto llevaba aproximadamente 10 días $\rightarrow$ demasiado tiempo para una epidemia

Motivación

Google se ha unido a la conversación

- Google propuso un mecanismo ingenioso: **Google Flu Trends**
- Punto de partida:
  - Proporción de visitas semanales por Gripe A en hospitales
  - 9 regiones $\times$ 5 años (2003-2007) = $2,340$ datos
  - Estos son los datos que tomaban 10 dias en elaborarse (comparemos con la Colombia de 2009)
- Google cruzó estos datos con las búsquedas sobre la gripe A
- Con estos datos, construyeron un modelo para predecir intensidad de gripe A

Motivación

Google se ha unido a la conversación

- ¿Un sólo modelo?

Motivación

Google se ha unido a la conversación

- ¿Un sólo modelo?
- Los investigadores de Google estimaron **450 millones** de modelos
- Eligieron el que mejor predice sobe la intensidad de búsqueda
- Les permite tener información diaria, semanal o mensual para cualquier punto de EEUU y el mundo
- A Google le toma 1 día lo que a la CDC 10!

Motivación

Google se ha unido a la conversación

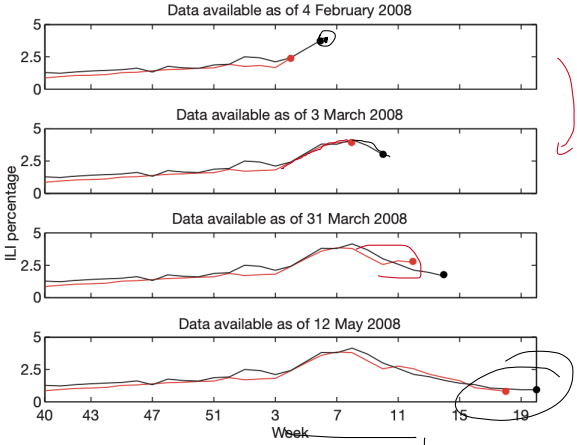


Figure 3 | ILI percentages estimated by our model (black) and provided by the CDC (red) in the mid-Atlantic region, showing data available at four points in the 2007-2008 influenza season. During week 5 we detected a sharply increasing ILI percentage in the mid-Atlantic region; similarly, on 3

Motivación

El rey ha muerto, larga vida al rey

- ¿Qué tienen en común Google Flu y Elvis?
  - Abanderados de la revolución
  - Definió y redefinió las reglas sistemáticas para hallar la solución a un problema
  - Éxito rotundo → Publicación en Nature!
    https://www.nature.com/articles/nature07634
  - Pero como a Elvis el éxito fue efímero
  - La predicciones comenzaron a sobre-estimar considerablemente la incidencia de la gripe A
  - Google Flu esta ahora archivado (disponible al público)
  - Continúa recolectando datos pero solo algunas instituciones científicas tienen accesso

Motivación

- Otro ejemplo, los algoritmos de reconocimiento de cara:
  - no son reglas fijas basadas en que los humanos entendemos por rostros y a partir de ello buscar combinaciones de pixeles.
  - son algoritmos que usan datos de fotos etiquetadas con un rostro y estiman una función $f(x)$ que predice si es un rostro o no a partir de pixeles x.
- El aprendizaje de maquinas se hizo una realidad cuando los investigadores dejaron de afrontarlo de manera teórica y lo hicieron empíricamente.
- Las similitudes con la econometría plantea interrogantes:
  - ¿Estos algoritmos están simplemente aplicando técnicas estándar a nuevos y grandes conjuntos de datos?
  - Si hay herramientas empíricas fundamentalmente nuevas, ¿cómo encajan con lo que conocemos?
  - Como economistas empíricos, ¿cómo podemos utilizarlas?

1. Aprendizaje de máquinas es todo sobre predicción
   - Ejemplos para Motivarnos
2. Sobre el curso
3. Tipos de Aprendizaje
4. A nadar
   - Regresión Lineal
   - Precisión del modelo
5. Recap
6. Break
7. `R para ML`

Sobre el curso

- El aprendizaje automático en economía es muy nuevo y dinámico.
- Estas 7 clases darán un *"snapshot"* de este campo en evolución.
- Estudiaremos ML a través de ejemplos, centrándonos en algunas aplicaciones y algoritmos de ML.
  - Primera parte de la clase será teórica
  - Break
  - Segunda parte trabajo en `R`

Sobre el curso

- Un aspecto adicional de las clases es que intentaré resaltar cómo encaja el ML en (y dónde se diferencia de) los métodos cuantitativos existentes en la microeconomía aplicada.
- Lo que no haré es centrarme demasiado en los detalles de algoritmos y problemas computacionales.

The master-economist must possess a rare combination of gifts...He must be mathematician, historian, statesman, philosopher and data scientist – in some degree.

adaptado de *Keynes (1924), Economic Journal*

Lecturas

1. El libro de texto principal (nivel de maestría): James, Witten, Hastie and Tibshirani, An Introduction to Statistical Learning with Applications in R [ISLR] link

2. También es útil (nivel de primer año de doctorado): Hastie, Tibshirani, Friedman, Elementos de aprendizaje estadístico [ESL]. link

3. Para los inquietos, estos artículos de introducción general en el Journal of Economic Perspectives estan muy interesantes:
   - Varian (2014), "Big Data: New Tricks for Econometrics" link
   - Mullainathan y Spiess (2017), "Machine Learning: An Applied Econometric Approach" link

. . . y otros que voy a mencionar en clase y los compartiremos en Slack

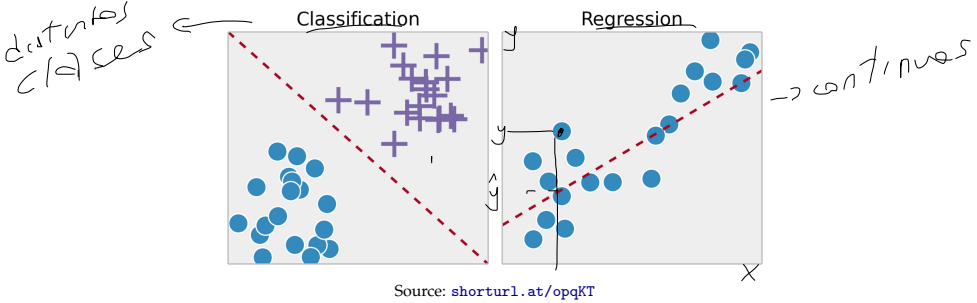
1. Aprendizaje de máquinas es todo sobre predicción
   - Ejemplos para Motivarnos
2. Sobre el curso
3. Tipos de Aprendizaje
4. A nadar
   - Regresión Lineal
   - Precisión del modelo
5. Recap
6. Break
7. `R para ML`

Tipos de Aprendizaje

- ML se divide en dos (¿?) ramas principales:
  1. Aprendizaje supervisado: Tenemos datos tanto sobre un resultado y como sobre las variables explicativas X.
     - Esto es lo más cercano al análisis de regresión que conocemos. De hecho, OLS es un ejemplo de aprendizaje supervisado.
     - Si y es discreto, también podemos ver esto como un problema de clasificación.
     - La mayoría de los ejemplos que veremos, y la mayor parte de nuestra discusión metodológica, caen en esta clase.

Tipos de Aprendizaje

- Aprendizaje supervisado
  - para cada predictor x_i hay una 'respuesta' observada y_i.
  - lo que hacemos en econometría cae en esta rama



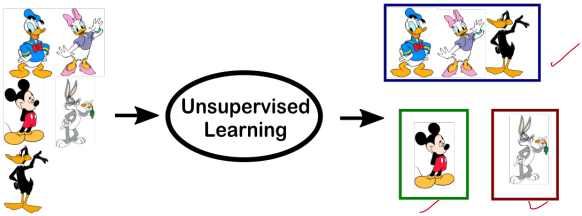
Source: shorturl.at/opqKT

Tipos de Aprendizaje

- ML se divide en dos (¿?) ramas principales:
  1. Aprendizaje supervisado: Tenemos datos tanto sobre un resultado y como sobre las variables explicativas X.
     - Esto es lo más cercano al análisis de regresión que conocemos. De hecho, OLS es un ejemplo de aprendizaje supervisado.
     - Si y es discreto, también podemos ver esto como un problema de clasificación.
     - La mayoría de los ejemplos que veremos, y la mayor parte de nuestra discusión metodológica, caen en esta clase.
  2. Aprendizaje no supervisado: No tenemos datos sobre y, sólo sobre X.
     - Queremos agrupar estos datos (sin especificar qué agrupar).
     - Permite reducir la dimensionalidad y explorar datos
     - Algunos algoritmos destacados: PCA, y K-medias

Tipos de Aprendizaje

- Aprendizaje no supervisado
  - Observamos X pero no las respuetas
  - Ejemplo: agrupar datos



Source: shorturl.at/opqKT

1. Aprendizaje de máquinas es todo sobre predicción
   - Ejemplos para Motivarnos
2. Sobre el curso
3. Tipos de Aprendizaje
4. A nadar
   - Regresión Lineal
   - Precisión del modelo
5. Recap
6. Break
7. `R para ML`

Un modelo estadístico general

- Asumamos que la relación entre la variable a predecir y y los predictores X esta dada por

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

- donde f es una función fija pero desconocida de X, con $E(u) = 0$ y $E(X'u) = 0$
- f no está restringida de ninguna manera; puede ser una función completamente arbitraria y compleja.

El paradigma clásico

$$Y = f(X) + u \tag{2}$$

- El interés esta en la inferencia
- Tests de hipótesis (std. err., tests)
- La $f(.)$ "correcta" nos ayuda a entender como y es afectado por X
- Modelo: surge de la teoría

$$\frac{\partial y}{\partial X}$$

El paradigma predictivo

$$Y = f(X) + u \tag{3}$$

- El objetivo del aprendizaje supervisado es predecir y basada en las características X.
- Debido a que nos ayuda a hacer una predicción, es útil estimar $f(.)$
- Entonces la $f(.)$ "correcta" es la que es capaz de predecir (no hacer inferencia!)
- Modelo?

El paradigma predictivo

$$Y = f(X) + u \tag{3}$$

- El objetivo del aprendizaje supervisado es predecir y basada en las características X.
- Debido a que nos ayuda a hacer una predicción, es útil estimar $f(.)$
- Entonces la $f(.)$ "correcta" es la que es capaz de predecir (no hacer inferencia!)
- Modelo?
- Como en predicción **no** nos interesa $f(.)$ en si misma, podemos tratarla como una *black box*, cualquier aproximación que nos da una buena predicción funciona. Recordemos:

El paradigma predictivo

$$Y = f(X) + u \tag{3}$$

- El objetivo del aprendizaje supervisado es predecir y basada en las características X.
- Debido a que nos ayuda a hacer una predicción, es útil estimar $f(.)$
- Entonces la $f(.)$ "correcta" es la que es capaz de predecir (no hacer inferencia!)
- Modelo?
- Como en predicción **no** nos interesa $f(.)$ en si misma, podemos tratarla como una *black box*, cualquier aproximación que nos da una buena predicción funciona. Recordemos:

"Lo que sea que funciona, funciona..."

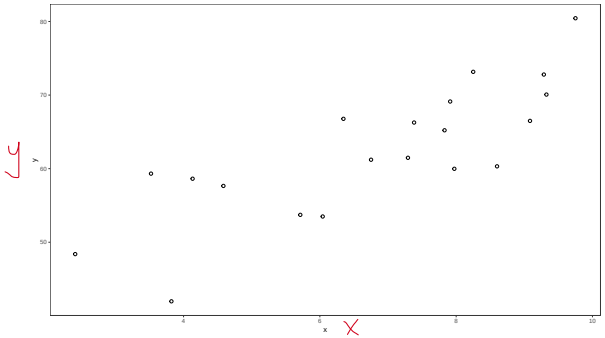
El paradigma predictivo

- A lo largo del curso vamos a explorar varios algoritmos lineales y no lineales que nos permitan estimar $f(.)$
- Estos métodos podemos resumirlos en dos grandes ramas
  1. Modelos paramétricos
  2. Modelos no paramétricos

Como estimamos $f(.)$?

Métodos paramétricos

- Supongamos que tenemos los siguientes datos



Fuente: datos simulados

Como estimamos $f(.)$?

Métodos paramétricos

- Los modelos paramétricos generalmente envuelven un enfoque de 2 pasos:
  1. Primero hacemos un supuesto sobre la forma funcional de $f(.)$. La mas común es que es lineal en X

$$f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \tag{4}$$

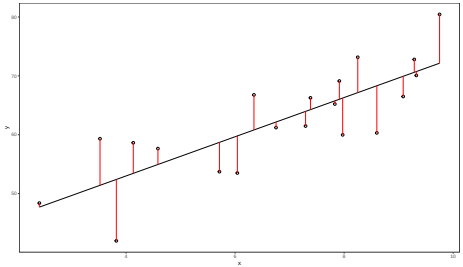
  2. Luego de seleccionar el modelo, necesitamos un procedimiento que nos ayude a "ajustar" o "entrenar" el modelo

- Este enfoque se conoce como paramétrico porque reduce el problema de estimar $f(.)$ a estimar un conjunto de parámetros

Como estimamos $f(.)$?

Métodos paramétricos

$$f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \tag{5}$$



Fuente: datos simulados

Como estimamos $f(.)$?

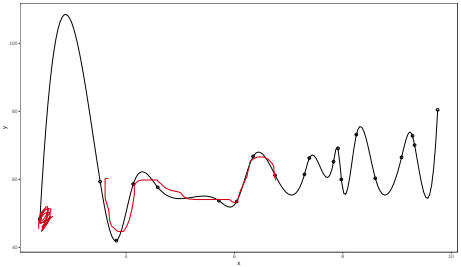
Métodos no paramétricos

- No hacen supuestos explícitos sobre la forma funcional de $f(.)$
- Buscan estimar $f(.)$ que se acerque lo más posible a los puntos sin ser lo demasiada rígida u ondulada
- Tienen la ventaja que se pueden adaptar a más formas posibles de $f(.)$ PEro suelen necesitar un número mas alto de observaciones y en algunos casos ser poco robustos.

Como estimamos $f(.)$?

Métodos no paramétricos

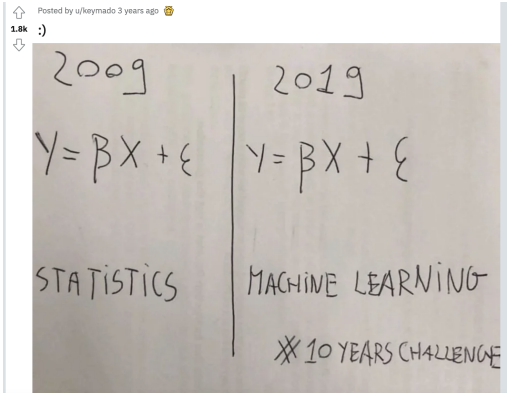
$$f(X) = g(X) \text{donde g es un spline} \tag{6}$$



Fuente: datos simulados

1. Aprendizaje de máquinas es todo sobre predicción
   - Ejemplos para Motivarnos
2. Sobre el curso
3. Tipos de Aprendizaje
4. A nadar
   - Regresión Lineal
   - Precisión del modelo
5. Recap
6. Break
7. `R para ML`

Regresión Lineal



Fuente: https://www.reddit.com/r/datascience/comments/ah0q69/_/

Regresión Lineal

- El problema es:

$$y = f(X) + u \tag{7}$$

- proponemos que:

$$f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \tag{8}$$

- El problema se reduce a encontrar los βs
  - Method de Momentos
  - MLE
  - OLS: minimiza SSR ($e'e$)
    - donde $e = y - \hat{y} = y - X\hat{\beta}$

Como medimos: "Lo que sea que funciona, funciona..."

- Con los parámetros podemos entonces recobrar la predicción

$$\hat{y} = \hat{f}(X) \tag{9}$$

$$= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p \tag{10}$$

- Como sabemos que funcionó bien?

Como medimos: "Lo que sea que funciona, funciona..."

- Con los parámetros podemos entonces recobrar la predicción

$$\hat{y} = \hat{f}(X) \tag{9}$$
$$= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p \tag{10}$$

- Como sabemos que funcionó bien?
- En estos problemas la medida más utilizada es el *error cuadrático medio (MSE)*

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{f}(X))^2 \tag{11}$$

Como medimos: "Lo que sea que funciona, funciona..."

- Notemos que esto no es otra cosa que la suma de los residuales al cuadrado

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(X))^2 \tag{12}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2 \tag{13}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(e)^2 \tag{14}$$

$$= RSS \tag{15}$$

- Esta medida nos da una idea de *lack of fit* que tan mal ajusta el modelo a los datos

Como medimos: "Lo que sea que funciona, funciona..."

- Un problema del RSS es que nos da una medida absoluta de ajuste de los datos, y por lo tanto no esta claro que constituye un buen RSS.
- Una alternativa muy usada en economía es el R^2
- Este es una proporción (la proporción de varianza explicada),
  - toma valores entre 0 y 1,
  - es independiente de la escala (o unidades) de y

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{16}$$

$$= 1 - \frac{RSS}{TSS} \tag{17}$$

1. Aprendizaje de máquinas es todo sobre predicción
   - Ejemplos para Motivarnos
2. Sobre el curso
3. Tipos de Aprendizaje
4. A nadar
   - Regresión Lineal
   - Precisión del modelo
5. Recap
6. Break
7. `R para ML`

Quick Recap antes de ir a R

- Vamos a ver distintos algoritmos donde el objetivo es predecir bien
- *"Lo que sea que funciona, funciona..."*
- La medida más utilizada en problemas de regresión es el MSE

Volvemos en 15 mins con R

6 45

R para ML

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