

UNIVERSIDAD
PANAMERICANA

Machine Learning

(<https://leonpalafox.github.io/mlclase/>)

Leon F. Palafox PhD

Intro

- **Introduccion**

- En esta clase estudiaremos las diferentes técnicas de Machine Learning y sus aplicaciones. Se verán técnicas básicas y el estado del arte, así como técnicas para evaluar cada algoritmo con diferentes sets de datos.

- **Objetivo**

- Los estudiantes serán capaces de implementar y validar diferentes técnicas de Machine Learning en diferentes sets de datos relacionados a distintos campos. los estudiantes aprenderán la diferencia entre aprendizaje supervisado y no supervisado

Logistics

- Horario:
 - Diario de 7:00 pm a 10:00 pm
- Course Website: <https://leonpalafox.github.io/mlclase/>
- Horas de Oficina: Antes de clase, o con cita a lpalafox@up.edu.mx

Evaluación

- La evaluación consistirá en:
- El proyecto final será el 60% de la evaluación final.
 - El proyecto final consistirá en el uso de una técnica de Machine Learning en un set de datos de su preferencia, con su correspondiente diseño y evaluación.
 - Pueden hacer equipos de hasta tres personas.
 - Necesitan hacer un reporte de 3-5 páginas sobre el set de datos, el diseño y las variables usadas.
- El restante 40% será distribuido de la siguiente forma:
 - Dos exámenes.
 - Dos tareas.
 - Participación en clase

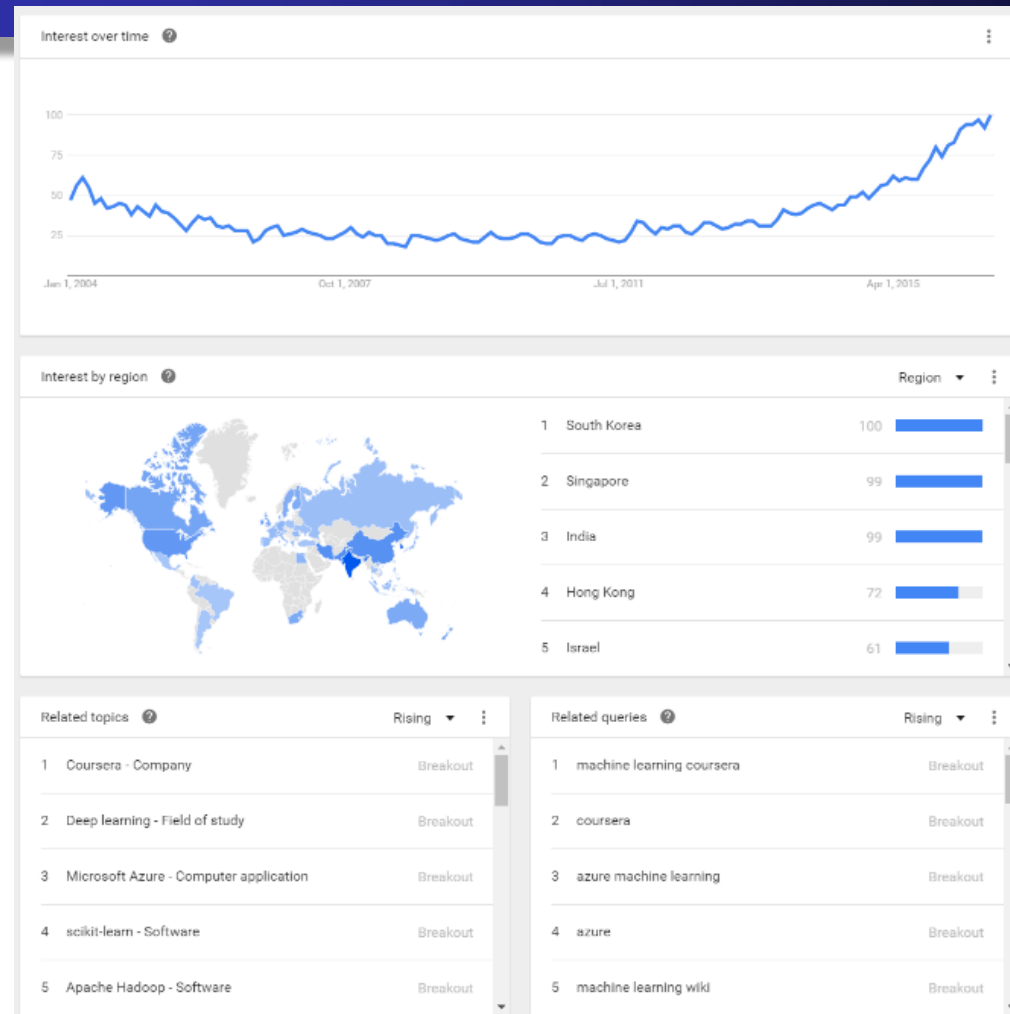
Preguntas?

Que es Machine Learning?

- Durante los 80s – 90s, mucho de lo que hoy llamamos Machine Learning se denominaba Inteligencia Artificial. IA era un termino sombrilla para todo lo que implicaba un entrenamiento usando datos.
 - Redes Neuronales
 - Algoritmos genéticos
 - Lógica difusa
 - Modelos probabilísticos

Renacimiento de AI-ML

- A finales de 1990s, mucha gente comenzó a utilizar herramientas más formales para el aprendizaje, mucha gente de matemáticas y estadística comenzaron a involucrarse con la comunidad de Machine Learning.
- IA se renombró Machine Learning, y muchos algoritmos clásicos de IA fueron adoptados por la comunidad de ML
 - Maquinas de Soporte Vectorial
 - K-Means
 - Regresión Lineal
 - Inferencia Bayesiana



ML ha logrado lo que se pensaba imposible

Google's Computer Program Beats Lee Se-dol in Go Tournament

By CHOE SANG HUN MARCH 15, 2016



Lee Se-dol with his daughter Lee Hye-lim on his way to the last Go match with Google's AlphaGo artificial intelligence program in Seoul, South Korea. Kim Hong-Ji/Reuters

SEOUL, South Korea — Ending what was billed as the match of the century, a [Google](#) computer program defeated a South Korean master of Go, an ancient board game renowned for its complexity, in their last face-off on Tuesday.

The program AlphaGo's 4-1 victory was a historic stride for computer

Todo mundo está haciendo ML

Apple acquires machine learning startup Turi, formerly known as GraphLab and Dato

JORDAN NOVET / AUGUST 5, 2016 1:51 PM

TAGS: APPLE, DATO, GRAPHLAB, MACHINE LEARNING, TOP STORIES, TURI



Image Credit: Mr. Gray/Photo

Apple has acquired Turi, a machine learning software startup. The startup formerly went by the names GraphLab and Dato.

Apple provided no information other than its standard boilerplate message for confirming acquisitions. "Apple buys smaller technology companies from time to time, and we generally do not discuss our purpose or plans," an Apple spokesperson told VentureBeat in an email. (Hat tip to [Geekwire](#) for breaking

Press Releases



HyTrust Unveils Enhanced Workload Security Solutions for the New Multi-Cloud World



DeepVista Powers 50+ VMs to

Crystal Ball for Corn Crop Yields Will Revolutionize Commodity Trading

TellusLabs is using NASA imagery, machine learning, and expert knowledge about vegetation to deliver accurate, in-season agricultural yield estimates.

by Elizabeth Woyke August 9, 2016



Deriving financial insights from satellite images isn't a new idea, but

TellusLabs is putting a twist on it. The Boston startup analyzes satellite imagery from NASA as well as weather data from the National Oceanic and Atmospheric Administration and seasonal, crop-growing information from the U.S. Department of Agriculture. It then uses machine-learning algorithms to generate intelligence about natural resources, such as predicting agricultural yields.

The strategy might sound similar to that of other satellite imagery analysis companies like **Descartes Labs** and **Orbital Insight**. However, TellusLabs plans to differentiate itself by applying scientific expertise in vegetation and climatology to its analysis, maintaining a narrow focus on natural resources, and quickly rolling out new products. Its goal is to be “a Bloomberg terminal for Earth signals.” “There’s a broad base of people who have to make tough decisions around natural resources, and we want to give them quality data, quickly,” says TellusLabs CEO and cofounder David Potere.

Libros

- Bishop, Christopher M. *Pattern recognition and machine learning*. Springer, 2006.
- Rogers, Simon, and Mark Girolami. *A first course in machine learning*. CRC Press, 2011.
(<http://www.dcs.gla.ac.uk/~srogers/firstcourseml/>)
- James, Gareth, et al. *An introduction to statistical learning*. New York: Springer, 2013. (<http://www-bcf.usc.edu/~gareth/ISL/>)
- Petersen, Kaare Brandt, and Michael Syskind Pedersen. *The matrix cookbook*. Technical University of Denmark 7 (2008): 15.
(<https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf>)

Preguntas?

Actividad Grupal

1st Message

- Mafia
- Poder
- Corrupción
- Complot
- México

2nd Message

- Gol
- México
- Tirititito
- Fútbol

Como supieron?

- Las palabras están asociadas a cada persona
- Su cerebro correlaciona las palabras con las personas.
- Su cerebro calcula las probabilidades conjuntas de que la persona esté asociada al mensaje.

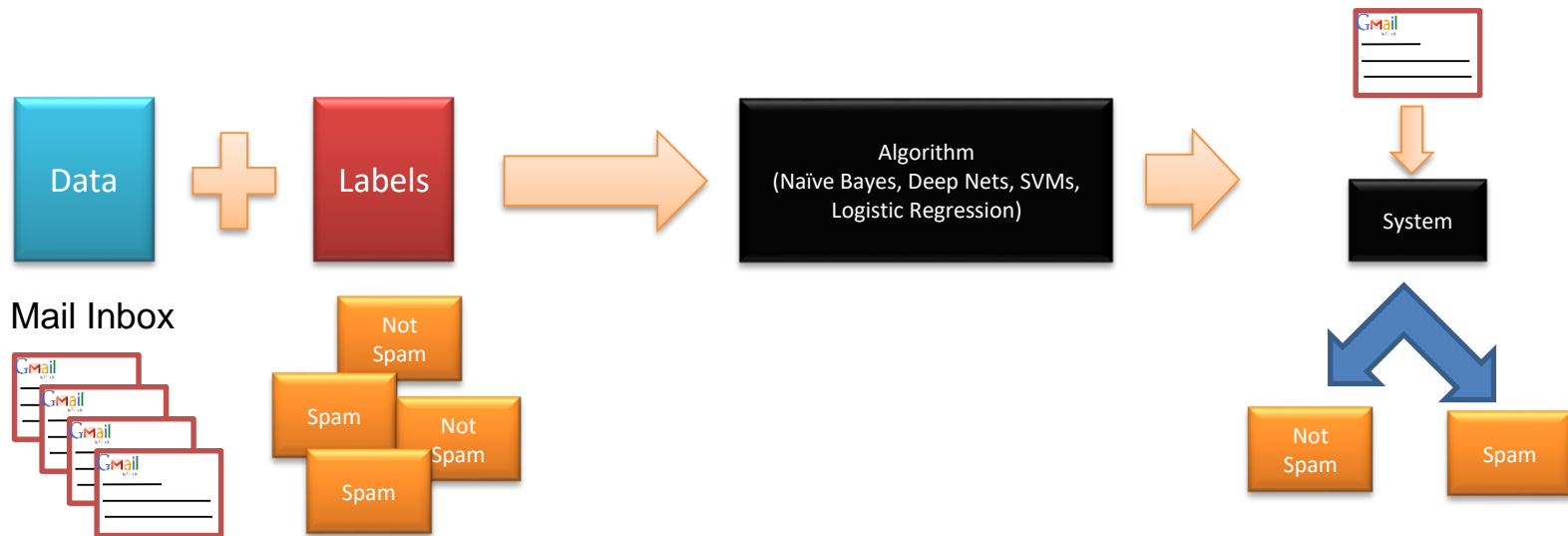
Reglas del juego

- Datos:
 - Documentos - > Texto
 - Imagenes - > Pixeles
 - Canciones -> notas, tonos
- Features, Características:
 - Textos -> Cadenas: hi, ho, amigo, ayuda
 - Imagenes -> RGB, DN, grayscale, flotantes
 - Tonos -> Flotantes que representen el tono.

Aprendizaje Supervisado

- Set de datos etiquetados
 - Set de emails con spam/not spam.
 - Reviews de Amazon (Estrellas)
 - Facebook like/not like.
 - Stock Market - > Volumen
- Algoritmo
 - Regresión Lineal
 - Regresión Logística
 - Maquinas de Soporte Vectorial
 - Deep Learning (Neural Networks and Convolutional NN)

Aprendizaje Supervisado



Cada categoría tendrá features que lo van a caracterizar

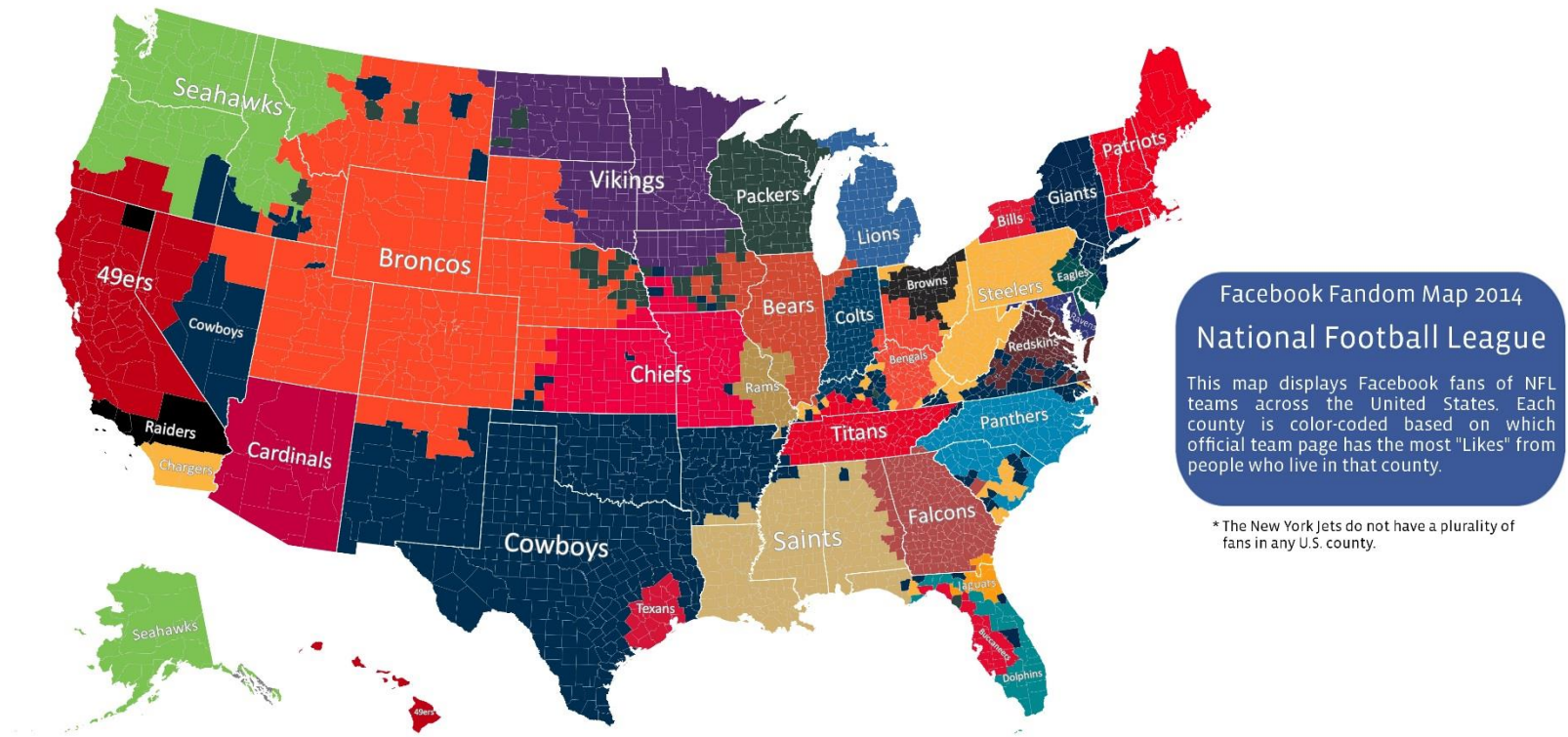
Spam: Offer, Viagra, medicine, Free, Conference in China

Not Spam: UP, Machine Learning, Evento, Mia, Mónica

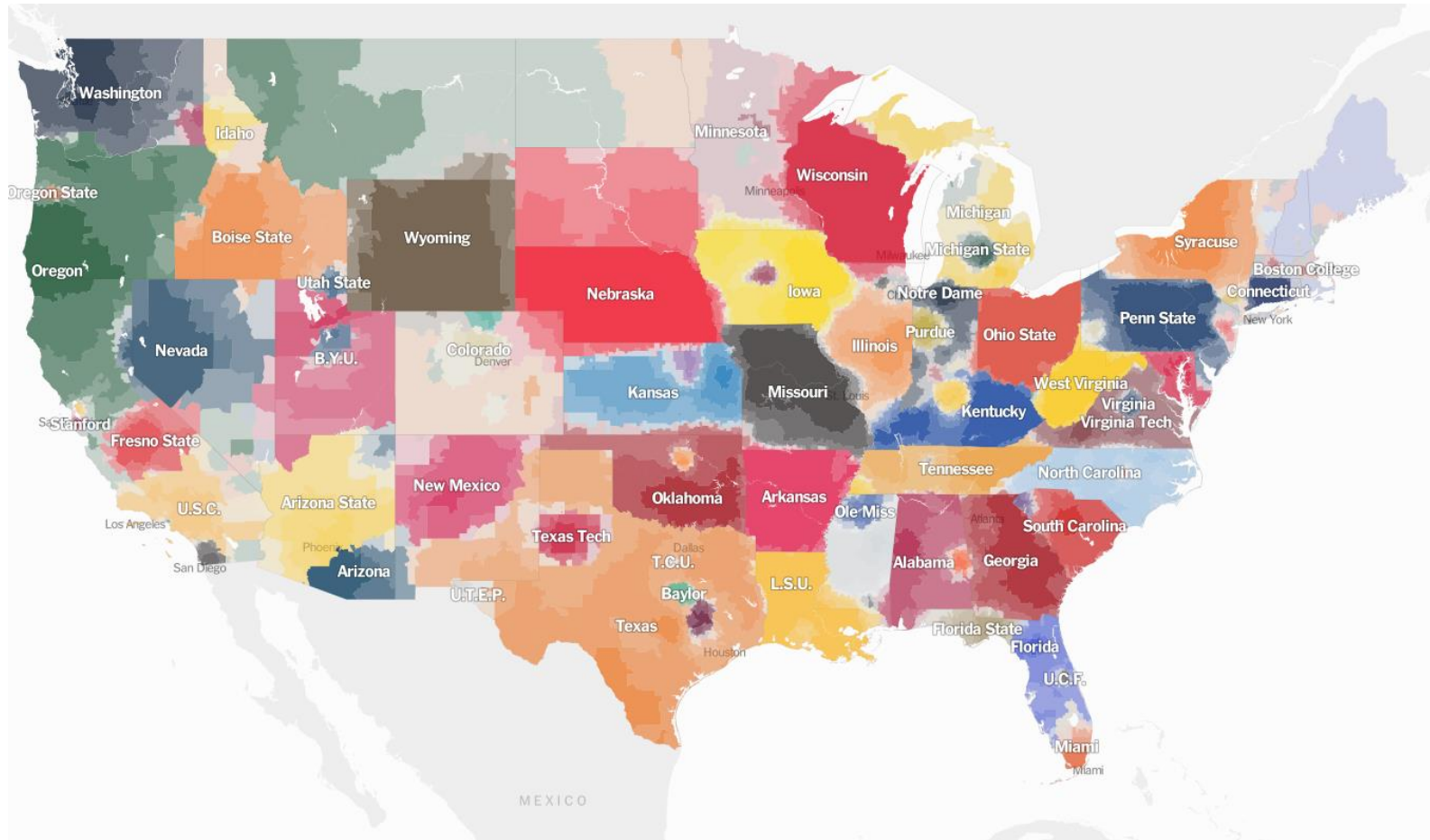
Métodos de Validación

- Cross Validation
 - Prevenir overfitting.
 - Encontrar el major set de parámetros.
- Bias-Variance Analysis
 - *“The needs of the Many outweigh the needs of the few”*
 - Spock
 - No le quieres decir a alguien que tiene cancer, pero **en verdad** no le quieres decir a alguien que no tiene, en caso de que si tenga.

Actividad Grupal



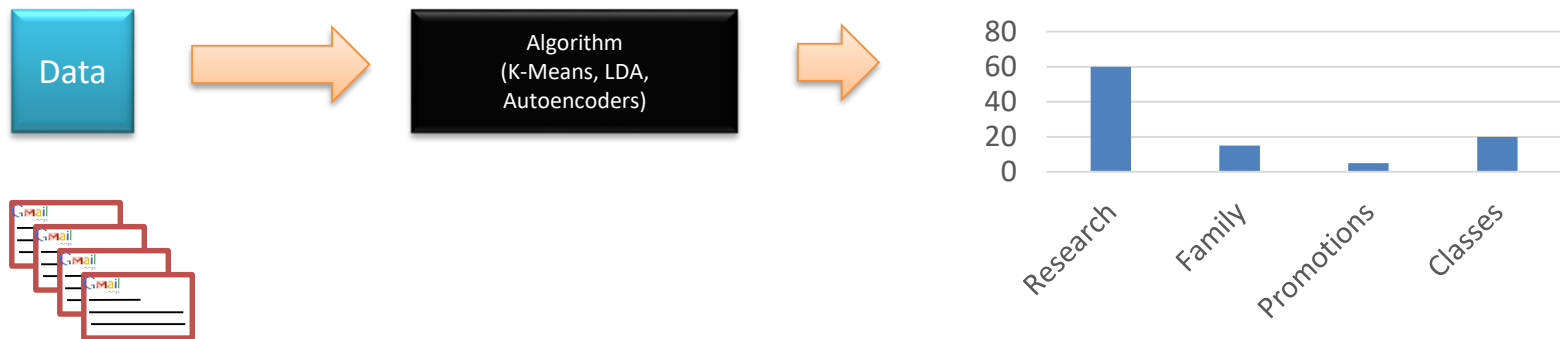
Actividad Grupal



Descubrimiento de conocimiento

- No necesitamos etiquetas
- Los datos se organizan solos
- La mayoría de los algoritmos descubren solos esa organización

Aprendizaje no Supervisado



Inbox

Los elementos que describen cada datum, en este caso son las palabras en cada email.

Cada tópico tendrá características que los separen del resto.

Investigación: NLP, Propuesta, Machine Learning, Deep Nets, Bayesian

Family: Mia, Casa, Mexico

Classes: Calificaciones, Tarea, Extensión, Horas de oficina

Preguntas?

A GUIDE TO
INTEGRATION BY PARTS:

GIVEN A PROBLEM OF THE FORM:

$$\int f(x)g(x)dx = ?$$

CHOOSE VARIABLES u AND v SUCH THAT:

$$u = f(x)$$

$$dv = g(x)dx$$

NOW THE ORIGINAL EXPRESSION BECOMES:

$$\int u dv = ?$$

WHICH DEFINITELY LOOKS EASIER.

ANYWAY, I GOTTA RUN.

BUT GOOD LUCK!