

# Aprendizaje Automático

2-2018 UTFSM San Joaquín



# Contenidos

- Introducción conceptual-motivacional al área.
- Formalidades (syllabus).

# Motivación

- Qué es el aprendizaje automático (machine learning)?
- Para qué sirve o cuándo es útil?
- Cuáles son las dificultades más relevantes?

“Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed”



Arthur Samuel (1959)

# Definición Preliminar

El aprendizaje automático es esencialmente un conjunto de conceptos, modelos y algoritmos que permiten obtener una “máquina” (programa, sistema) que ejecuta una determinada tarea de interés sin definir explícitamente el algoritmo necesario para obtener el resultado deseado.

En vez de programar explícitamente la máquina para que ejecute la tarea de interés, un método de aprendizaje automático permite entrenarla a partir de ejemplos que reflejan el comportamiento que nos esperamos de ella. Decimos que la máquina aprende cuando logra reproducir el comportamiento que nos esperamos de ella a partir de los ejemplos que le damos.

# Motivación

**Ejemplo.** Filtrado de correo spam. Queremos un programa capaz de leer nuestros correos y filtrar automáticamente aquellos maliciosos o aquellos que simplemente eliminaríamos de nuestra casilla de entrada.



# Motivación

Asumamos que un “correo” se encuentra representado como una estructura de 3 campos (dirección de origen, título y cuerpo), donde el primero es una cadena de la forma name@domain y los últimos dos consisten en listas de palabras.

Por “*programar explícitamente una máquina*” para esta tarea, nos referimos a definir manualmente un conjunto de reglas que, operando sobre la representación disponible, permitan determinar si el correo debe ser filtrado o no. Por ejemplo:

**IF** email.source.domain == ‘utfsm’ **THEN** type=spam

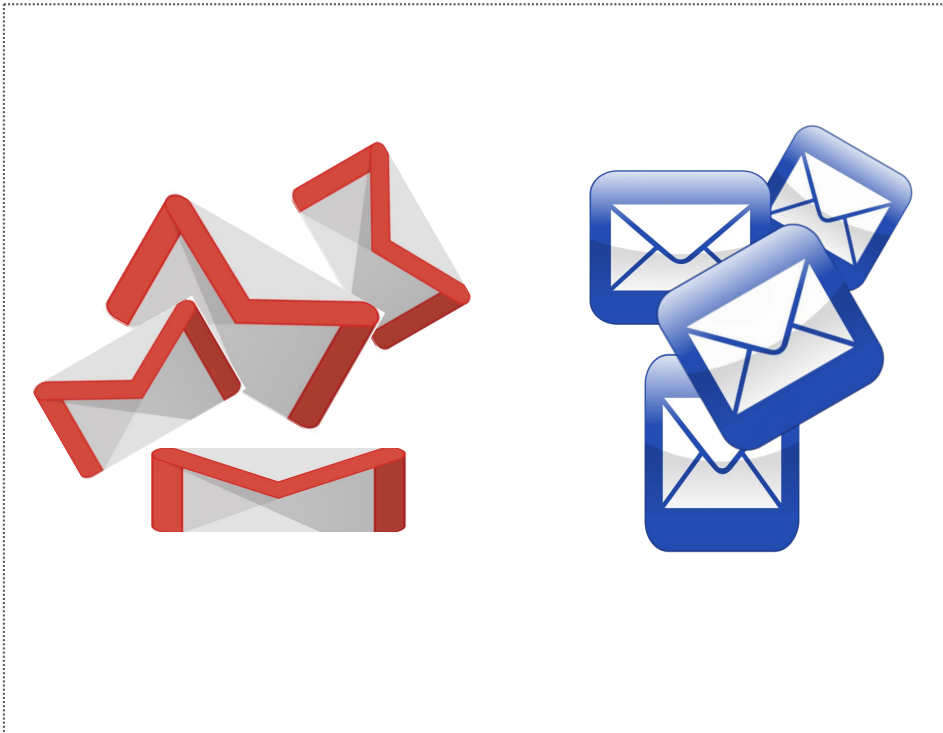
**IF** (‘tarea’ IN email.header) **THEN** type=spam

# Motivación

Por “*entrenar una máquina*” para esta tarea, nos referimos (por ejemplo) a dar a una máquina (programa) un conjunto de correos que hemos filtrado en el pasado (ejemplos positivos) y posiblemente un conjunto de correos que no deseamos sean filtrados (ejemplos negativos), de modo que la máquina infiera automáticamente las reglas que distinguen los correos no deseados de aquellos que si queremos recibir.



Ejemplos (datos)



Máquina (programa)



Programa deseado

```
IF email.source.domain == 'utfsm' THEN type=spam  
IF ('tarea' IN email.header) THEN type=spam
```

# Motivación

El punto de vista anterior es un poco “estática” y corresponde a un escenario que se denomina “batch setting” (por lotes).

Consideremos una definición alternativa de un libro clásico del área.



“A program is said to learn from experience **E** with respect to some task **T** and some performance measure **P**, if its performance on **T**, as measured by **P**, improves with experience **E**”

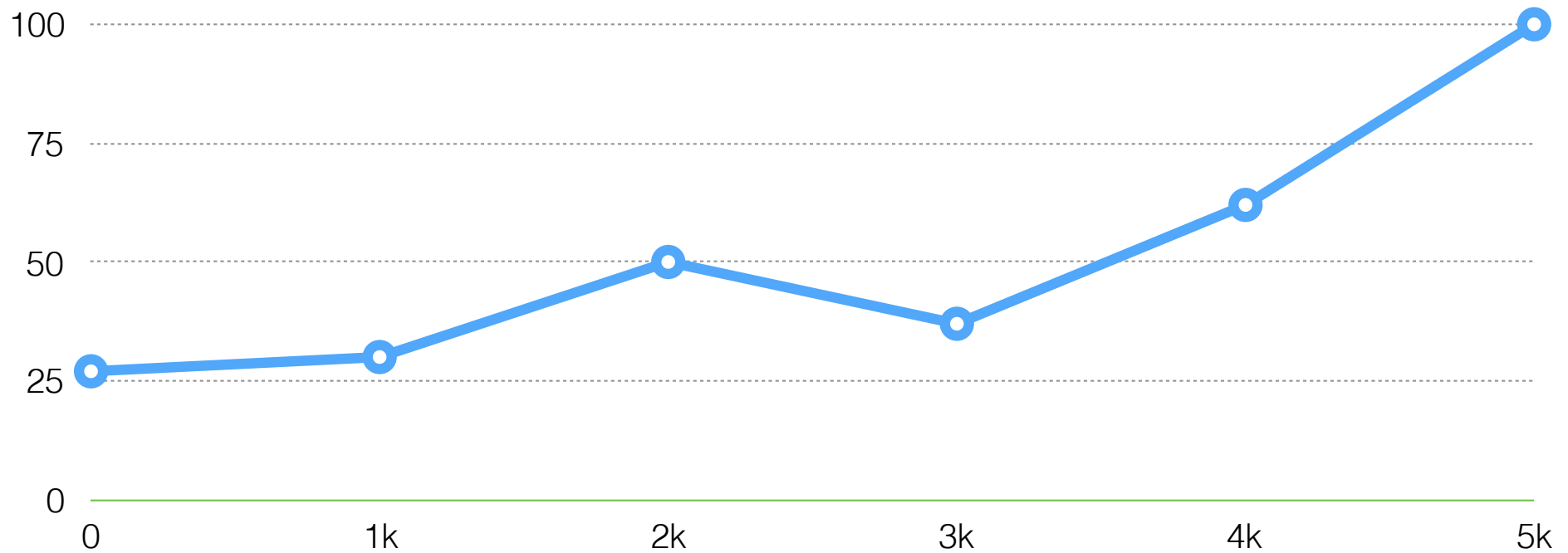
Tom Mitchell, *Machine Learning*, 1997.

# Motivación

Este punto de vista es más consistente con lo que denomina “aprendizaje online” (online learning), escenario en que no se asume que la máquina dispone de todos los ejemplos antes del aprendizaje sino que los observa de manera continua y aprende (mejora en la tarea de interés) de manera continua.



# Performance



Experience (es decir, datos)

# Cuándo es esto útil?

El aprendizaje automático es útil para resolver problemas en que resulta difícil especificar un algoritmo que produzca el comportamiento deseado, y es mucho más fácil recolectar ejemplos que reflejen el comportamiento deseado.

El aprendizaje automático es también útil para problemas en que se requiere una adaptación constante al cambios en el ambiente (por ejemplo, nuevos ataques spam).

El aprendizaje automático es también útil para problemas en que se requiere un alto grado de personalización de la solución (por ejemplo, filtrado de noticias depende del gusto del usuario).

# Es esto posible?

Sí, si tenemos un conjunto de datos suficientemente grande, conocemos el estado del arte y tenemos hardware suficientemente potente.

Sí, si podemos acotar suficientemente bien el problema y tenemos un conocimiento del problema suficientemente amplio que nos permita identificar los atributos que contienen la información relevante para resolver el problema.

Sí, si conocemos los problemas fundamentales del área y estamos disponibles a hacer mucha experimentación (prueba y error).

# Motivación

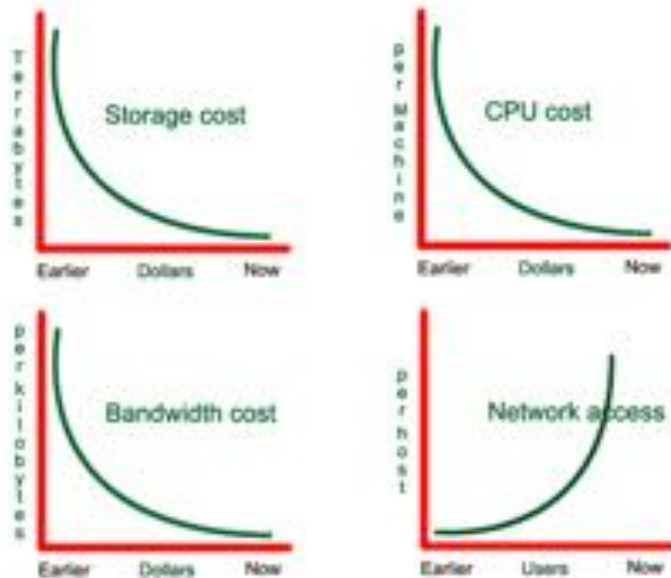
Los avances de los últimos años han hecho que el aprendizaje automático deje de ser una disciplina puramente académica y se convierta en tecnología ampliamente utilizada en la industria.





# Motivación

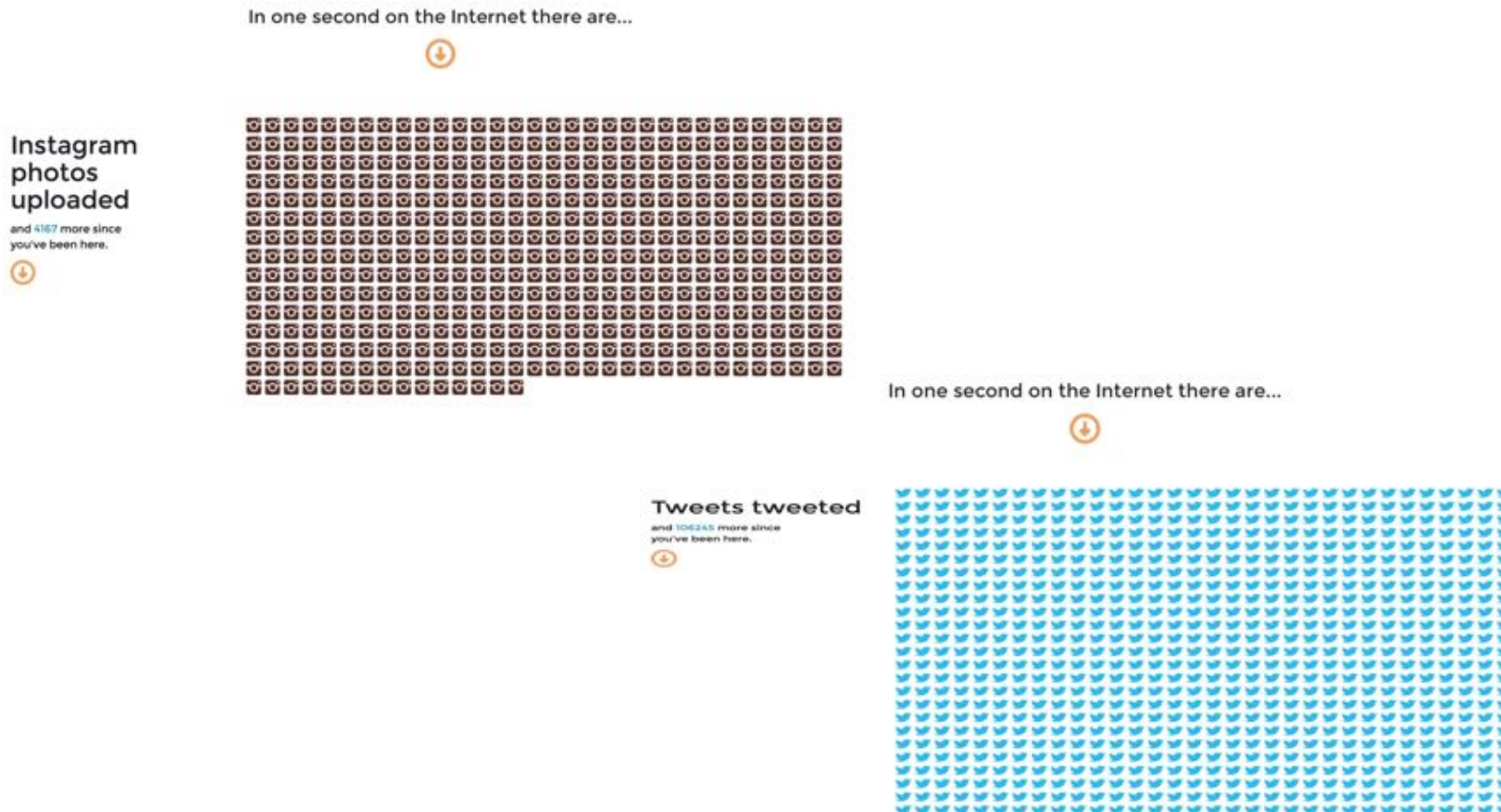
Parte de todos estos avances han sido posibles gracias al progreso en el hardware disponible (paralelismo, GPU's, etc) en el software disponible (librerías especializadas) y sobre todo en la cantidad de datos disponibles.





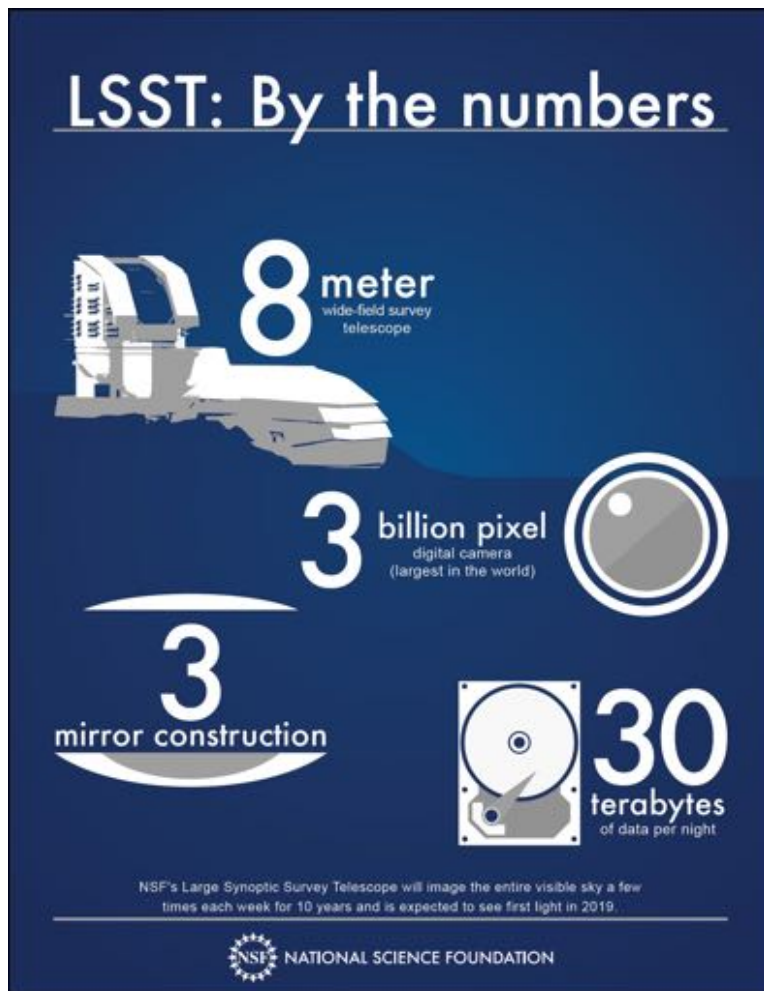
# Motivación

En efecto, esta sobre-abundancia de datos es lo que hace también necesarios los progresos que se están produciendo en el área.



# Motivación

En efecto, esta sobre-abundancia de datos es lo que hace también necesarios los progresos que se están produciendo en el área.

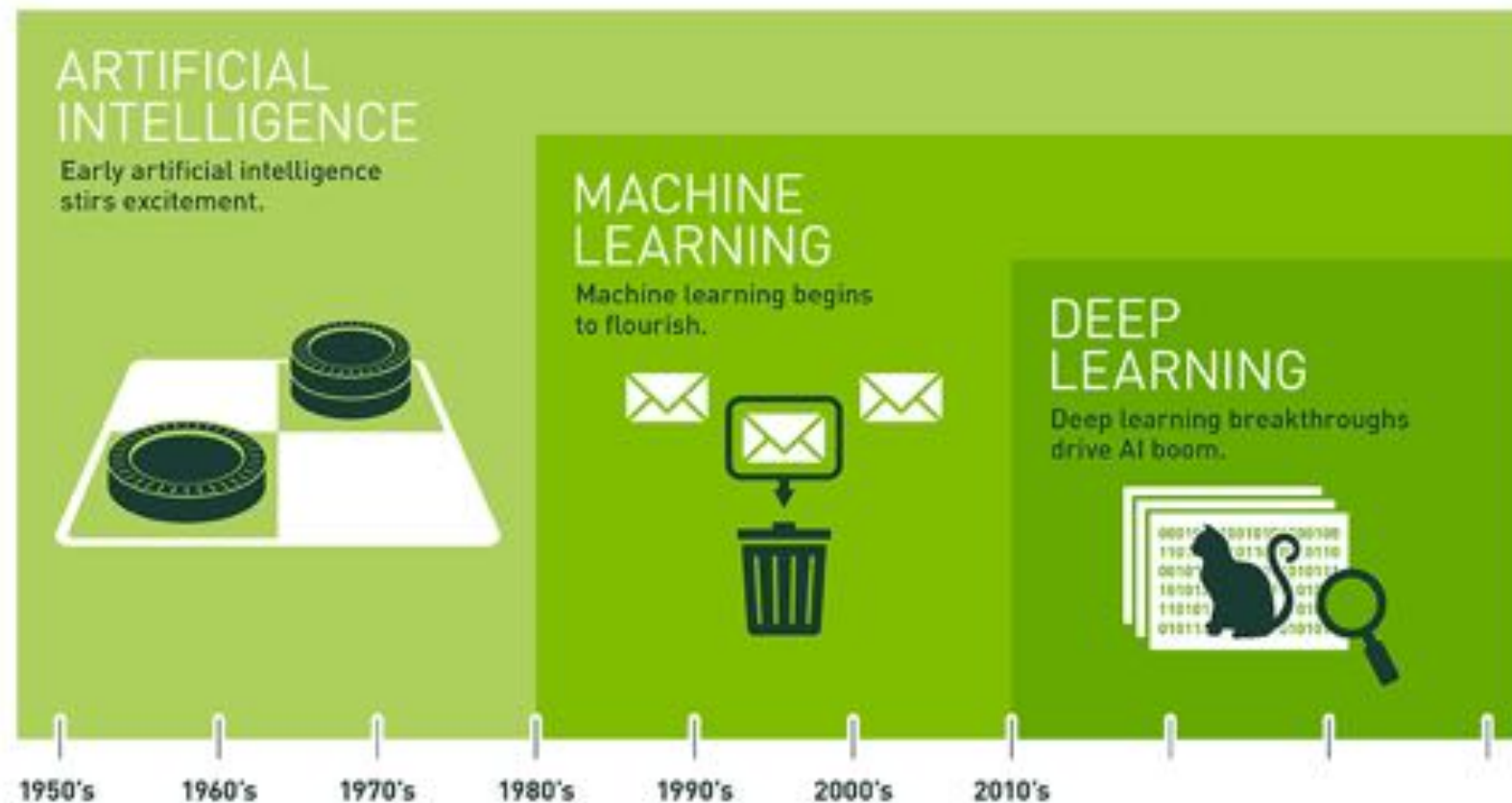


LSST (Large Synoptic Survey Telescope)



# Motivación

Los avances actuales también han provocado un re-florecimiento de las expectativas que se tienen con respecto a lo que la inteligencia artificial puede lograr en los próximos años.



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.



CADE METZ BUSINESS 12.09.13 03:14 PM

# FACEBOOK TAPS 'DEEP LEARNING' GIANT FOR NEW AI LAB

Twitter pays up to \$150M for Magic Pony Technology, which uses neural networks to improve images

Artificial Intelligence Chef Wants To Disrupt Factory Farming With Innovative Vegan Products

A Chilean startup is using a machine-learning algorithm that cooks plant-based products, in a bid to disrupt the factory-farm food system.

## FOOD TECH

**NOT**  
**MAYO**  
Sabor  
**ORIGINAL**



ROBERT MCMILLAN BUSINESS 03.13.13 06:30 AM

# GOOGLE HIRES BRAINS THAT HELPED SUPERCHARGE MACHINE LEARNING

## BIG DATA

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

## NEWS DESK

IS "DEEP LEARNING" A REVOLUTION IN ARTIFICIAL INTELLIGENCE?

Support The Guardian

Subscribe Find a job Sign in / Register Search

News

Opinion

Sport

Culture

Lifestyle

More

The Guardian International edition

## Google DeepMind pairs with NHS to use machine learning to fight blindness

'Deep learning' research company will use 1m anonymised eye scans to train a neural network to identify early signs of degenerative eye conditions

ARTIFICIAL INTELLIGENCE INTEL BIG DATA

## Intel is paying more than \$400 million to buy deep-learning startup Nervana Systems

The chip giant is betting that machine learning is going to be a big deal in the data center.



Science. Julio 2015.



Nature. Mayo 2017.



# Motivación

Desafío *Imagenet* (reconocer entre 22.000 tipos de objetos)



Accuracy humana: 94.5%

Estado del arte al 2010: 75%

**Algoritmo Resnet (Microsoft) al 2015: 96.43%\***

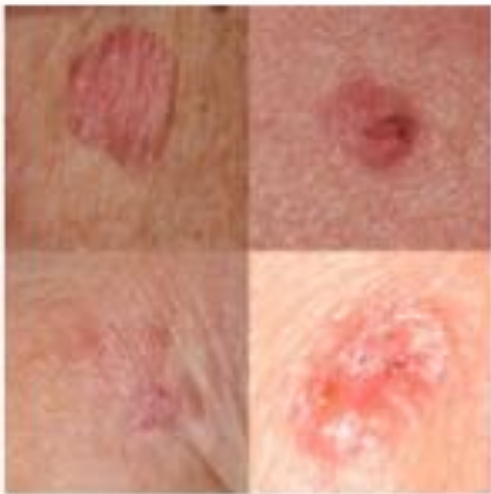
\* Ensamblado de redes neuronales profundas

# Motivación

Benign



Malignant



Classifier	Three-way accuracy
Dermatologist 1	65.6%
Dermatologist 2	66.0%
CNN	69.5% *
CNN - PA	<b>72.0%</b> **

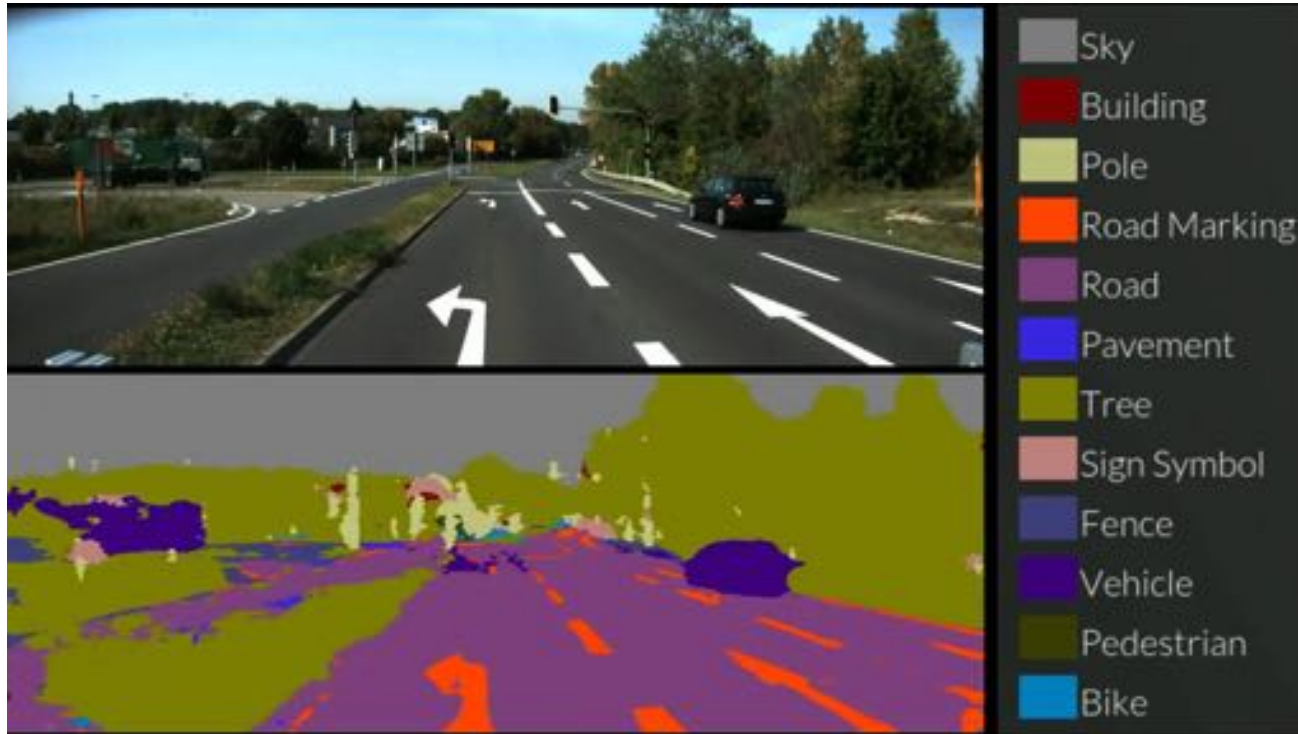
## Disease classes: three-way classification

- 0. Benign single lesions
- 1. Malignant single lesions
- 2. Non-neoplastic lesions

\* Red neuronal profunda entrenada sobre 3 clases.

\*\* Red neuronal profunda entrenada sobre 757 clases.

# Motivación

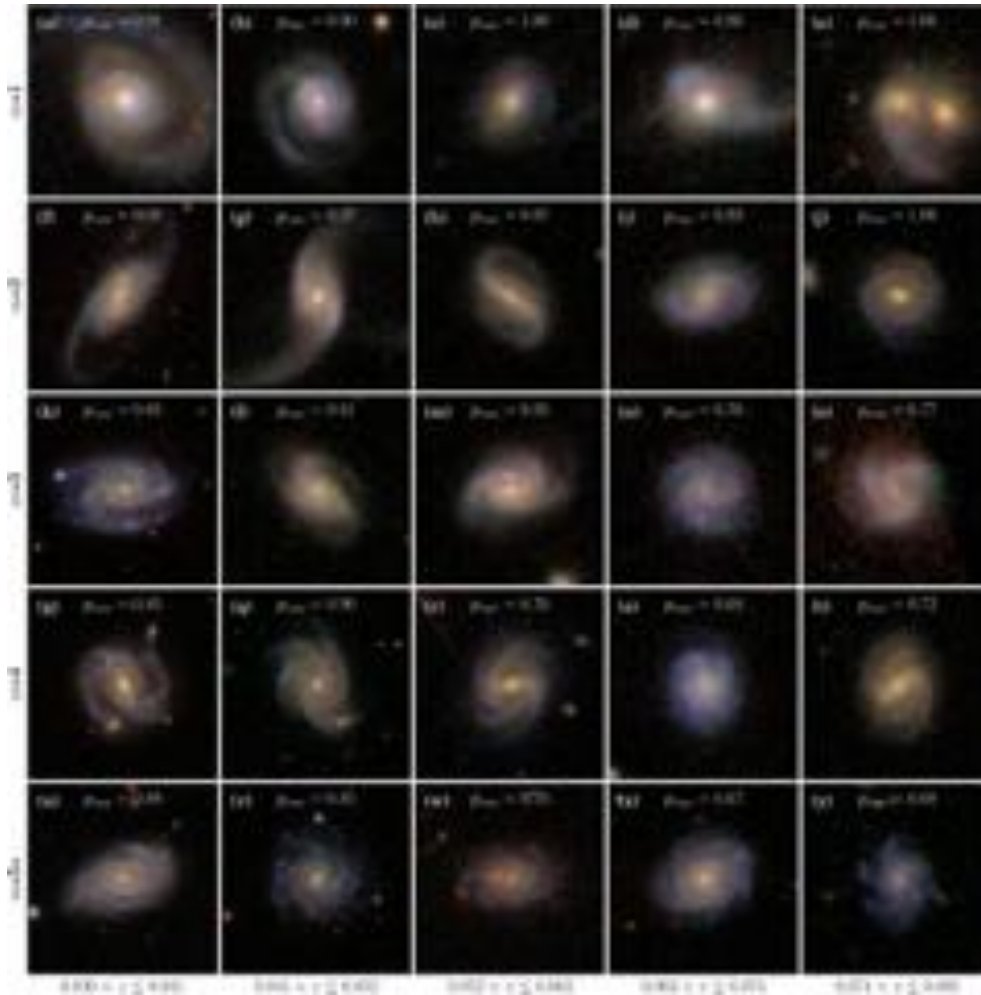


\* Deep net  
+ Reinforcement learning



# Motivación

\*



Accuracy SVM 88%

Accuracy Red Neuronal: 98%

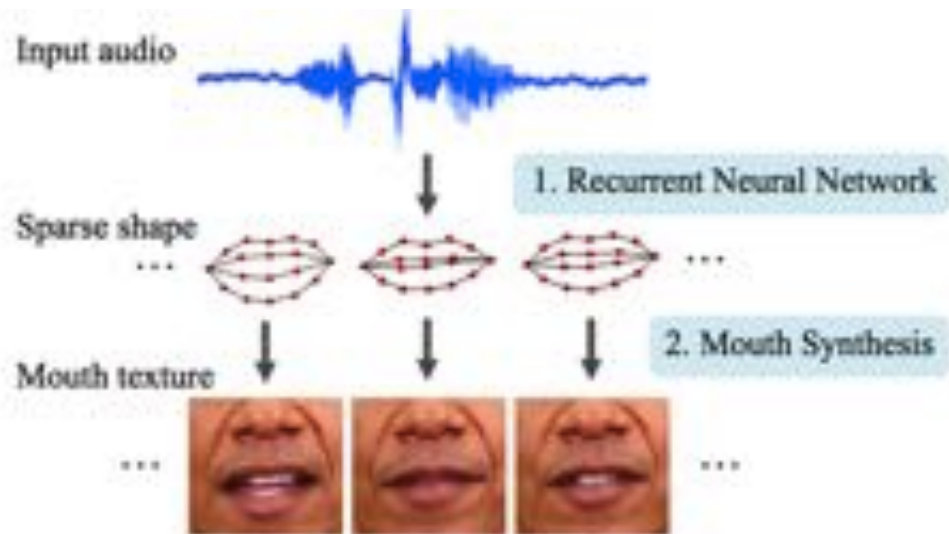
# Motivación

El texto que describe esta imagen fue generado por una red neuronal artificial.



A woman is throwing a **frisbee** in a park.

# Motivación



# Motivación

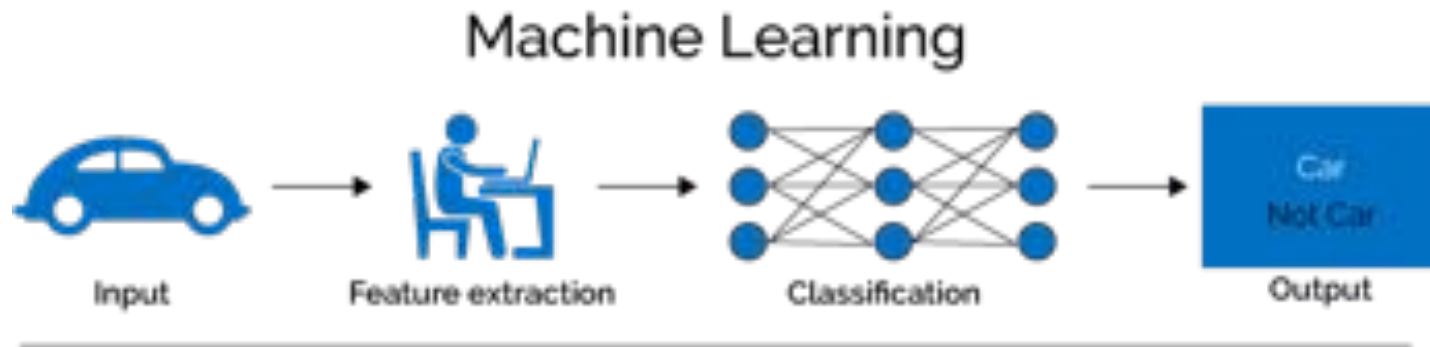


Deep learning + Reinforcement learning

# Warning

- Este es un curso de índole académica: no veremos las herramientas como cajas negras sino que intentaremos entender los aspectos teórico-conceptuales detrás de ellas que permitan al estudiante profundizar más adelante.
- Este es un curso introductorio al área. Veremos los métodos más clásicos en el área y una breve introducción a temas más avanzados. En particular, este no es un curso de deep learning. Un curso más especializado (aunque de todos modos introductorio) sobre ese tema es INF395 o INF477 (redes neuronales artificiales) que se dicta los primeros semestres de cada año.

# Deep learning vs Classic ML





# Deep learning vs Classic ML



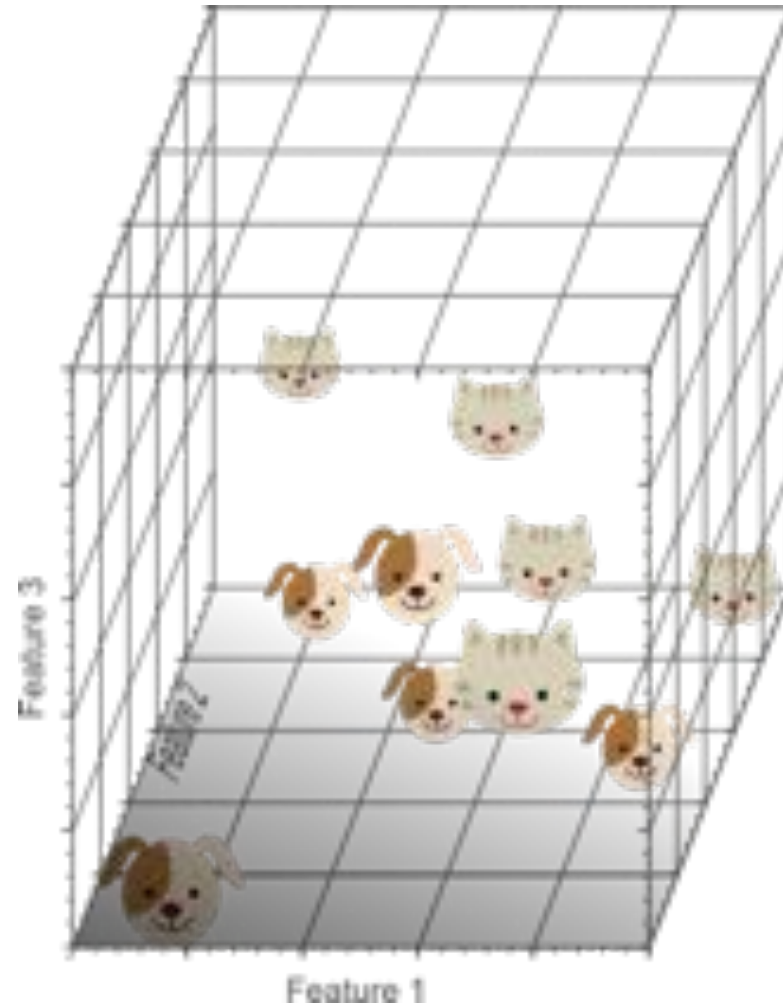
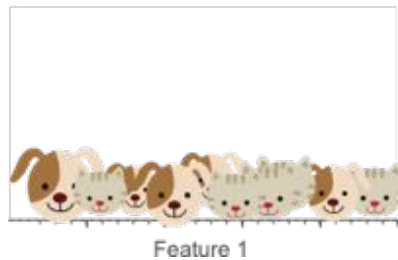
## A Few Useful Things to Know about Machine Learning

Pedro Domingos  
Department of Computer Science and Engineering  
University of Washington  
Seattle, WA 98195-2350, U.S.A.  
pedrod@cs.washington.edu

### **8. FEATURE ENGINEERING IS THE KEY**

At the end of the day, some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used. If you have many independent features that each correlate well with the class, learning is easy. On the other hand, if the class is a very complex function of the features, you may not be able to learn it. Often, the raw data is not in a form that is amenable to learning, but you can construct features from it that are. This is typically where most of the effort in a machine learning project goes. It is often also one of the most interesting parts, where intuition, creativity and “black art” are as important as the technical stuff.

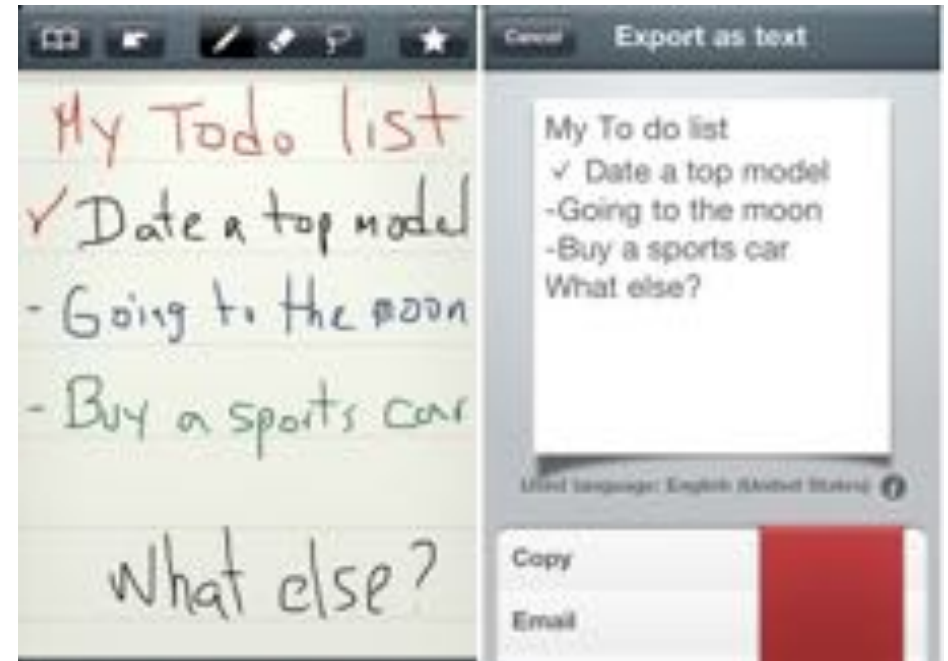
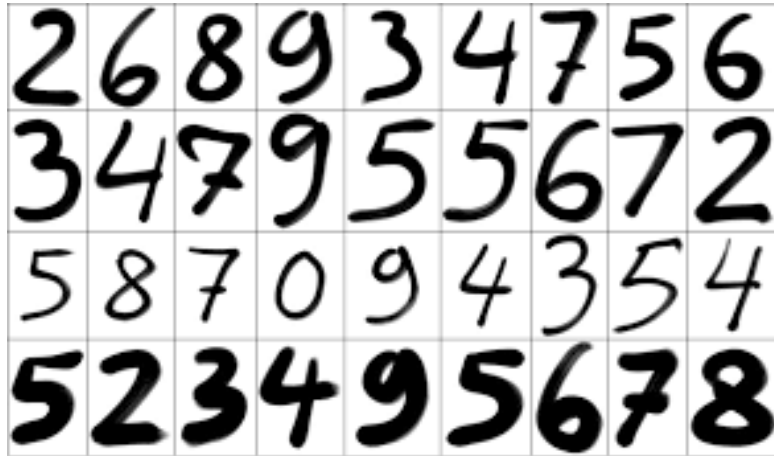
# Deep learning vs Classic ML





# Temas de Tareas

- Reconocimiento de dígitos o caracteres manuscritos.



# Temas de Tareas

**The Funny Racist™** @TheFunnyRacist 33m  
How do you start a foot race in Ethiopia? Roll a doughnut down the street.

**Niklaus** @YellowEyed\_Joe  
@WhyYouSObsessed lol cause im a loyal house nigga lol. I was thinkin bout throw a piece of chicken in **ethiopia** to watch the Hunger Games.

**ツ☆☆ Brya B.** @BryaMB  
I Wouldn't Let My Dude Look Like A **Starving** Nigga From **Ethiopia** If I Had One Cause I'd Feed His Ass. lol

**Yung Stunna (Promo)** @Y\_STUNNA1 61  
If Mitt Romney Wins Next Week Atlanta Gone Be Called "Lil **Ethiopia**" Lol Folks Gone Be Starving Without Dem **Food** Stamps #NoLie

**The Funny Racist™** @TheFunnyRacist 11h  
Why are Ethiopians so poor at negotiations? Because they never bring anything to the table.

**Sizakele Mthembu** @Syzaa 1 Nov  
Lol RT @Slick\_Langa: :( RT @Syzaa: Shame they must be hungry RT @SuperSportBlitz: CAF African Women Championship - Result: **Ethiopia** 0 -

**The Funny Racist™** @TheFunnyRacist 9h  
How many Ethiopians can you fit in a bath tub? As many as you want - they keep sliding down the drain.

**César** @SeniorCee  
About to order food from two locations and pay whoever gets here first. I feel like a **starving** kid in **Ethiopia**.  
[Expand](#) [Reply](#) [Retweet](#) [Favorite](#)

**Ria Forbes** @rienne1 [Retweet](#)  
If any Man Utd player was **starving** in **Ethiopia** & I was only person with a burger van, I would refuse them. Without a second tough #HATEU  
[Expand](#)

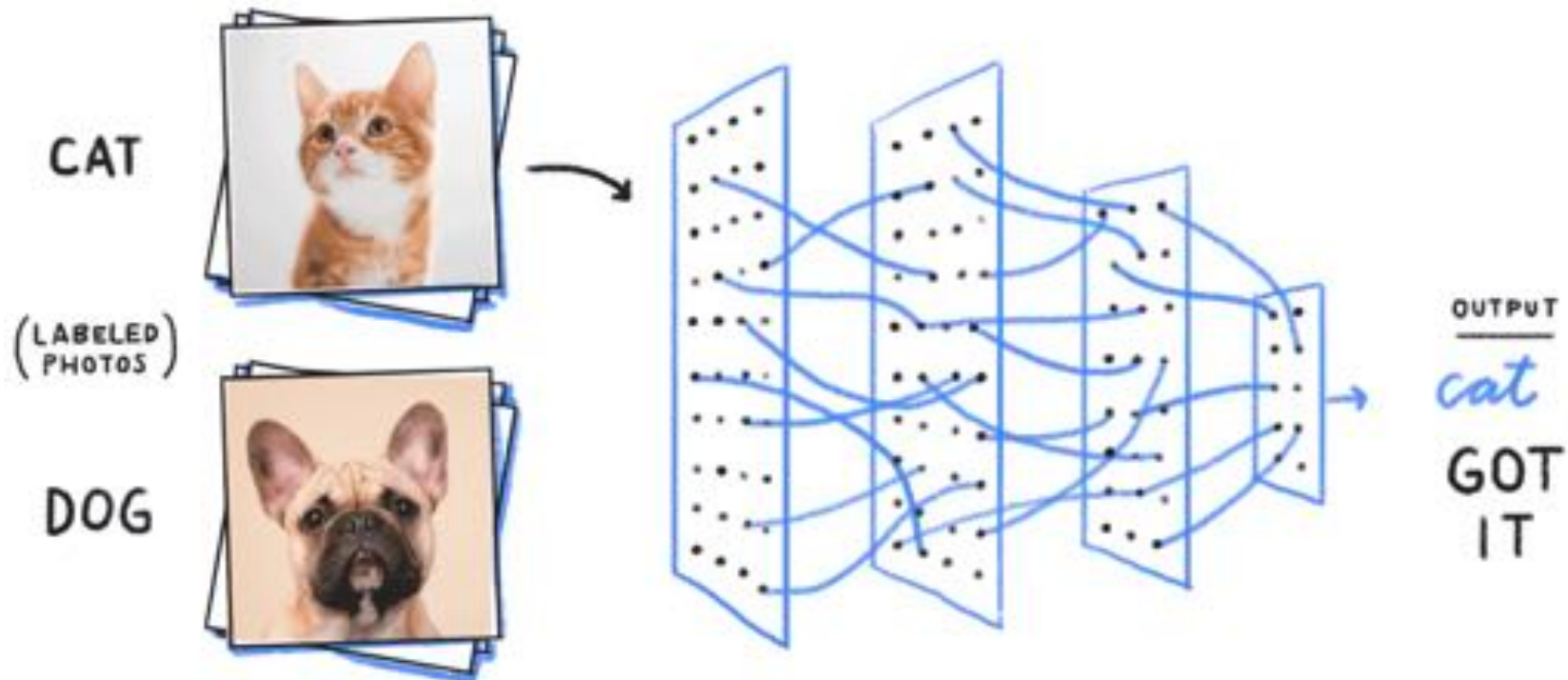
**Karleigh Reilly** @KarleighReilly  
Like I'm glad we aren't helping **starving** kids in **Ethiopia**. As long as people in America can eat and overcome their fears  
[Expand](#)

## Sentiment Analysis



# Temas de Tareas

- Reconocimiento de imágenes.



# Temas de Tareas

- Filtrado colaborativo

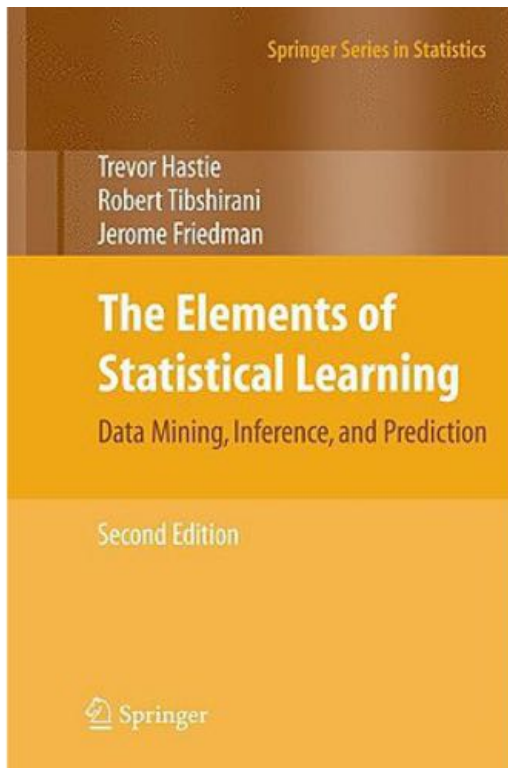


# Curso

- Aproximadamente 16 semanas de trabajo.
- Evaluado mediante tareas y quices.
- Tareas: aplicaciones lo más “realistas” usando los métodos que estudiamos, programación, análisis experimental de resultados.
- Exposiciones: equipos serán seleccionados al azar para presentar oralmente sus soluciones y resultados.,
- Quices (30 min) más teóricos o conceptuales.
- 60% tareas y 40% quices o viceversa dependiendo del perfil.

# Referencias

- ▶ T. Hastie et al. *The elements of statistical learning*.



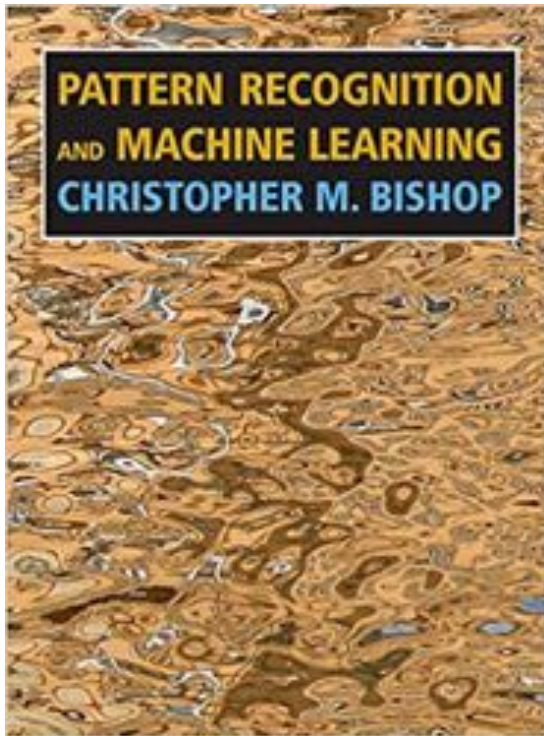
- ▶ E. Alpaydin. *Introduction to Machine Learning*.





# Referencias

- ▶ C. Bishop *Pattern Recognition and Machine Learning*.



- ▶ E. Alpaydin. *Introduction to Machine Learning*.

