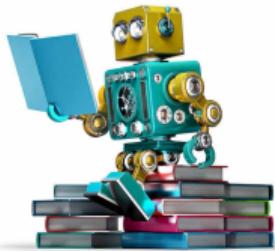


Aprendizaje de Maquina

Inteligencia Artificial



Marco Teran

2021 - Bogotá

Contenido

1 Introducción

2 Machine Learning

■ Aprendizaje supervisado

3 El proceso de aprendizaje automático

Introducción

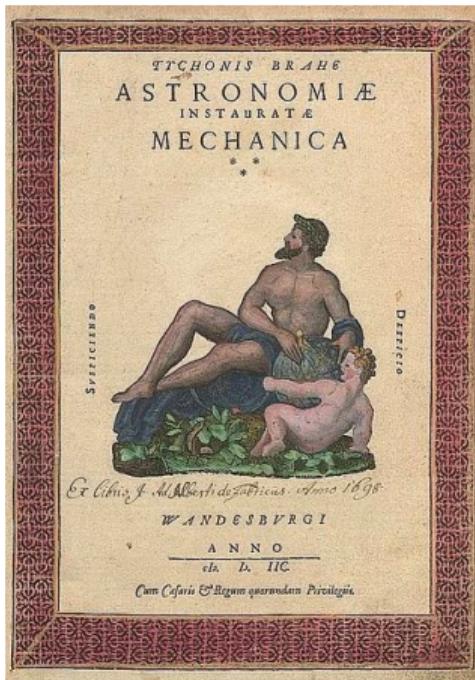
Observación y análisis



Observación y análisis



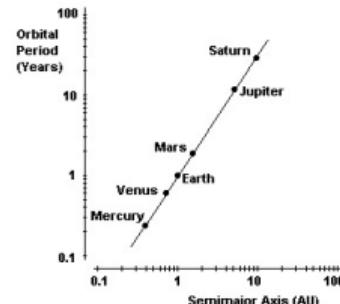
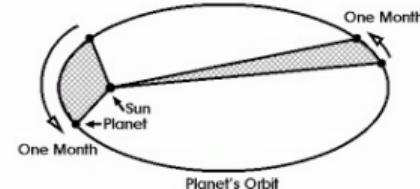
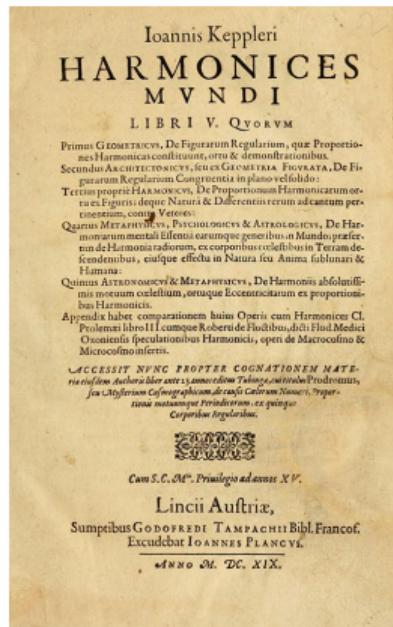
Tycho Brahe



Tycho Brahe

Date, Old Style				Longitude						Latitude				Mean Longitude			
	Year	Day	Month	H	M	D	M	S	Sign	D	M	S	D	M	S		
I	1580	18	November	1	31	6	28	35	Gemeni	1	40	N.	1	25	49	31	
II	1582	28	December	3	58	16	55	30	Cancer	4	6	N.	3	9	24	55	
III	1585	30	January	19	14	21	36	10	Leo	4	32	N.	4	20	8	9	
IV	1587	6	March	7	23	25	43	0	Virgo	3	41	N.	6	0	47	40	
V	1589	14	April	6	23	4	23	0	Scorpio	1	12	N.	7	14	18	26	
VI	1591	8	June	7	43	26	43	0	Sagitt.	4	0	S.	9	5	43	55	
VII	1593	25	August	17	27	12	16	0	Pisces	6	2	S.	11	9	49	31	
VIII	1595	31	October	0	39	17	31	40	Taurus	0	8	N.	1	9	55	4	
IX	1597	13	December	15	44	2	28	0	Cancer	3	33	N.	2	23	11	56	
X	1600	18	January	14	2	8	38	0	Leo	4	30	N.	4	4	35	50	
XI	1602	20	February	14	13	12	27	0	Virgo	4	10	N.	5	14	59	37	
XII	1604	28	March	16	23	18	37	10	Libra	2	26	N.	6	27	0	12	

Johannes Kepler

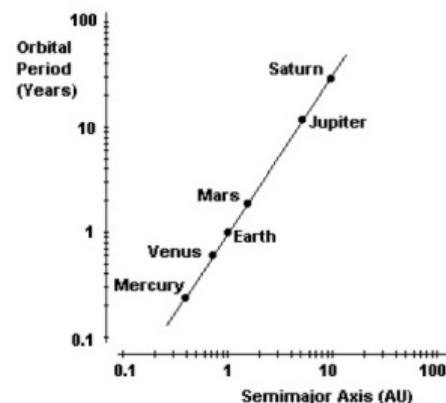


Datos y modelos

Data

Year	Date, Old Style			Longitude			Latitude			Mean Longitude			
	Day	Month	H	M	D	S	Sign	D	M	S	S	N	
I	1580	10			November	1	31	6	20	35	Gemini	1	-40
II	1582	28			December	3	16	55	30	00	Cancer	4	6
III	1585	30			January	19	14	11	36	10	Aries	4	32
IV	1587	6			March	7	23	25	43	0	Virgo	3	41
V	1589	14			April	6	23	4	23	0	Scorpio	1	12
VI	1591	8			June	7	43	26	43	0	Sagittarius	4	0
VII	1593	25			August	17	27	12	16	0	Pisces	6	2
VIII	1595	31			September	1	39	12	40	40	Taurus	0	8
IX	1597	13			October	15	2	25	0	00	Cancer	3	33
X	1600	18			January	14	2	8	38	0	LEO	4	30
XI	1602	20			February	14	13	12	27	0	Virgo	4	10
XII	1604	28			March	16	23	18	37	10	Libra	2	26
											N	6	27
											S	0	12

Model



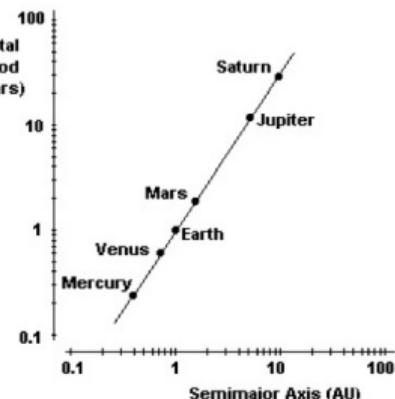
Aprendizaje automático

Data

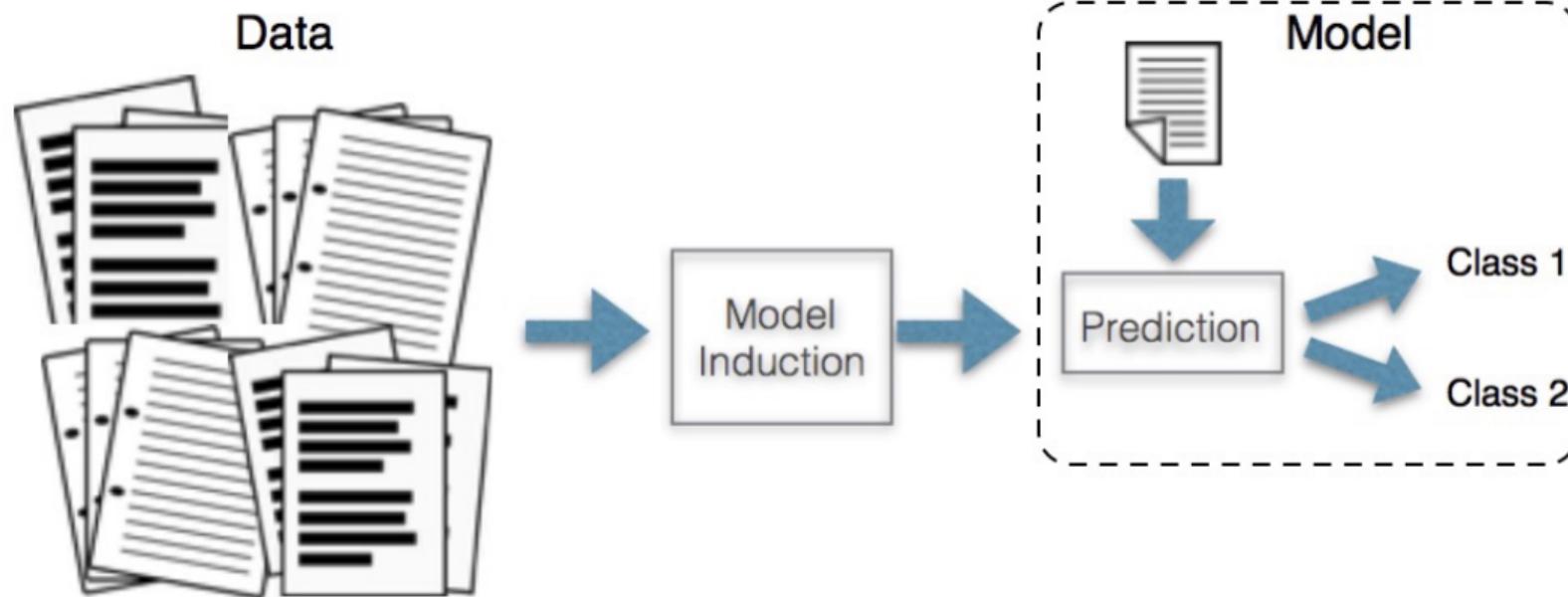
Year	Day	Month	Longitude					Latitude					Mean Longitude				
			H	M	D	M	S	Sign	D	M	S	S	D	M	S	S	D
I	1580	18	November	1	31	6	28	35	Gemini	1	40	N.	1	25	49	31	
II	1582	28	December	3	58	16	55	30	Cancer	4	6	N.	3	9	24	55	
III	1585	30	January	19	14	21	50	10	Leo	4	32	S.	4	29	5	7	
IV	1587	1	February	1	23	13	43	0	Virgo	3	41	N.	6	0	47	40	
V	1589	14	April	6	23	4	23	0	Scorpio	1	12	N.	7	14	18	26	
VI	1591	8	June	7	43	26	43	0	Sagittarius	4	0	S.	9	5	43	55	
VII	1593	25	August	17	27	12	16	0	Pisces	6	2	S.	11	9	49	31	
VIII	1595	31	October	0	39	17	31	40	Taurus	0	8	N.	1	9	55	4	
IX	1597	11	November	15	44	2	28	0	Cancer	3	33	N.	2	25	35	56	
X	1600	18	January	14	13	38	0	Libra	4	30	N.	4	4	38	50		
XI	1602	20	February	14	13	12	27	0	Virgo	4	10	N.	5	14	59	37	
XII	1604	28	March	16	23	18	37	10	Libra	2	26	N.	6	27	0	12	

Learning/
Model
Induction

Model



Aprendizaje automático con datos de texto



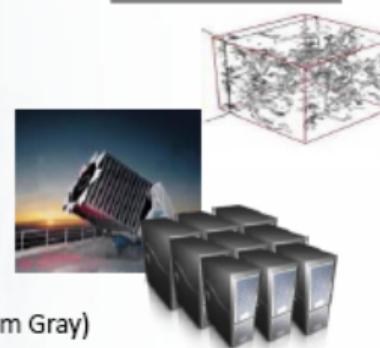
El cuarto paradigma

Emergence of a Fourth Research Paradigm

1. Thousand years ago – **Experimental Science**
 - Description of natural phenomena
 2. Last few hundred years – **Theoretical Science**
 - Newton's Laws, Maxwell's Equations...
 3. Last few decades – **Computational Science**
 - Simulation of complex phenomena
 4. Today – **Data-Intensive Science**
 - Scientists overwhelmed with data sets from many different sources
 - Data captured by instruments
 - Data generated by simulations
 - Data generated by sensor networks
- eScience is the set of tools and technologies to support data federation and collaboration
- For analysis and data mining
 - For data visualization and exploration
 - For scholarly communication and dissemination



$$\left(\frac{c}{a}\right)^2 = \frac{4\pi G\rho}{3} - K \frac{c^2}{a^2}$$



(With thanks to Jim Gray)

Machine Learning

Aprendizaje de Máquina

- Construcción y estudio de sistemas que pueden aprender de los datos
- **Problema principal:** encontrar patrones, relaciones y regularidades entre los datos, que permitan construir modelos *descriptivos* y *predictivos*.
- **Campos relacionados:**
 - Estadística
 - Reconocimiento de patrones y visión por computador
 - Minería de datos y descubrimiento de conocimiento
 - Análisis de datos

Breve historia

- Discriminante lineal de Fisher (Fisher, 1936)
- Modelo de Neurinal Artificial (McCulloch and Pitts, 1943)
- Perceptron (Rosenblatt, 1957) (Minsky & Papert, 1969)
- Aprendizaje correcto probablemente aproximado (Valiant, 1984)
- Perceptron multicapa y *Backpropagation* (Rumelhart et al., 1986)
- Árboles de decisión (Quinlan, 1987)
- Redes Bayesianas (Pearl, 1988)
- Maquinas de vectores de soporte (Cortes & Vapnik, 1995)
- Deep learning (Hinton et al., 2007)

Machine Learning en las noticias

Big Data

Google uses machine learning to fill in the blanks in your spreadsheet

Price	Year	Num. of Miles	Num. of Owners	Type
22000	2012	100000	2-Door	Car
12000	2010	40000	2-Door	Car
12000	2011	60000	2-Door	Car
10000	2010	70000	2-Door	Car
20000	2008	70000	4-Door	Truck

From online dating to driverless cars, machine learning is everywhere

Dr Michael Osborne from the University of Oxford answers our Q&A about the mysteries of a component of artificial intelligence

Why Facebook, Google, and the NSA Want Computers That Learn Like Humans

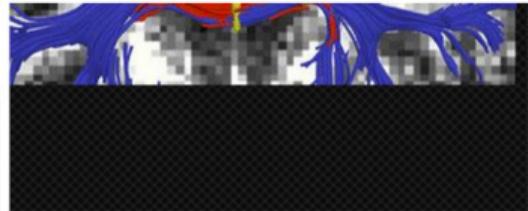
Deep learning could transform artificial intelligence. It could also get pretty creepy.

—By Dana Liebelson | September/October 2014 Issue

[Like](#) [Share](#) 442 [Tweet](#) 474 [Email](#) 70



Nicola Davis
river, Thursday 18 September 2014 07.00 BST
to comments (0)



Making sense of medical sensors

Computer scientists and electrical engineers are devising a useful new patterns in data produced by medical sensors.

Data analytics driving medical breakthroughs

Using big data to save lives

MORE LIK

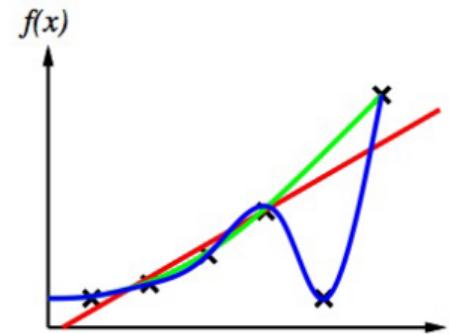
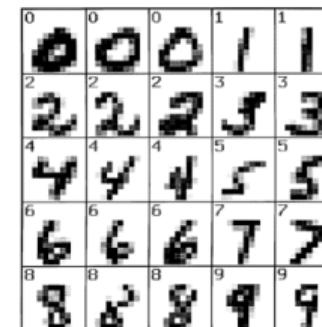
5 Business Analytics Tech Exploit Them

How to get a hot job in b

What's the big deal abou

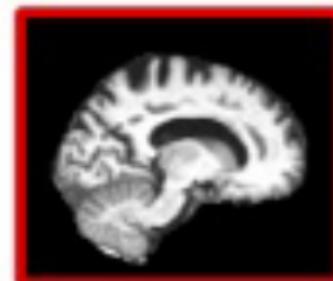
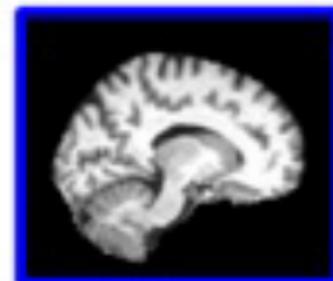
Aprendizaje supervisado

- **Problema fundamental:** encontrar una función que relacione un conjunto de entradas con un conjunto de salidas
- **Problemas típicos:**
 - Clasificación
 - Regresión



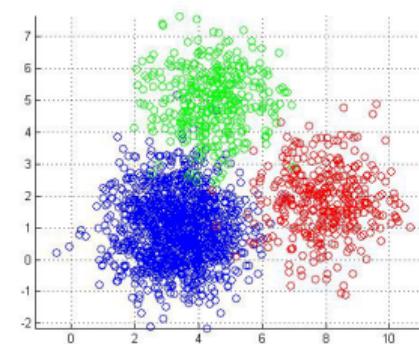
Aprendizaje supervisado

- **Problema fundamental:** encontrar una función que relacione un conjunto de entradas con un conjunto de salidas
- **Problemas típicos:**
 - Clasificación
 - Regresión



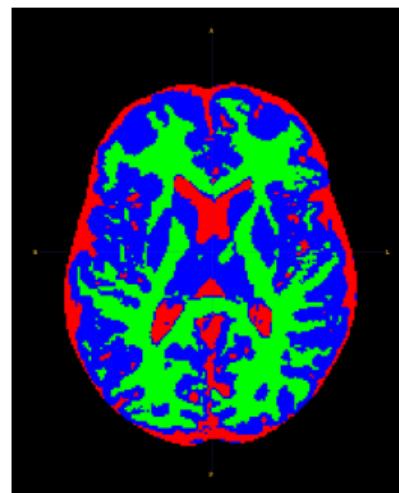
Aprendizaje no supervisado

- No hay etiquetas para las muestras de entrenamiento
- **Problema fundamental:** encontrar la estructura subyacente de un conjunto de datos de entrenamiento
- **Problemas típicos:** clustering, segmentación, reducción de la dimensionalidad, análisis de temas latentes
- Algunas muestras pueden tener etiquetas, en cuyo caso se denomina **aprendizaje semisupervisado**



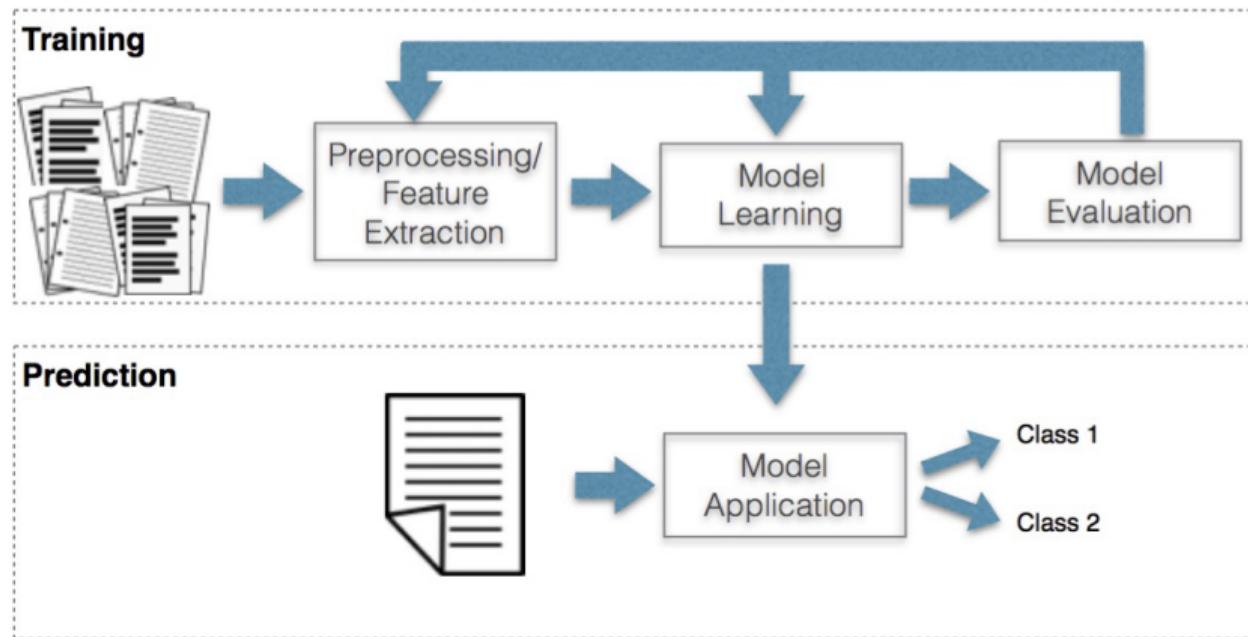
Aprendizaje no supervisado

- No hay etiquetas para las muestras de entrenamiento
- **Problema fundamental:** encontrar la estructura subyacente de un conjunto de datos de entrenamiento
- Problemas típicos: clustering, segmentación, reducción de la dimensionalidad, análisis de temas latentes
- Algunas muestras pueden tener etiquetas, en cuyo caso se denomina **aprendizaje semisupervisado**

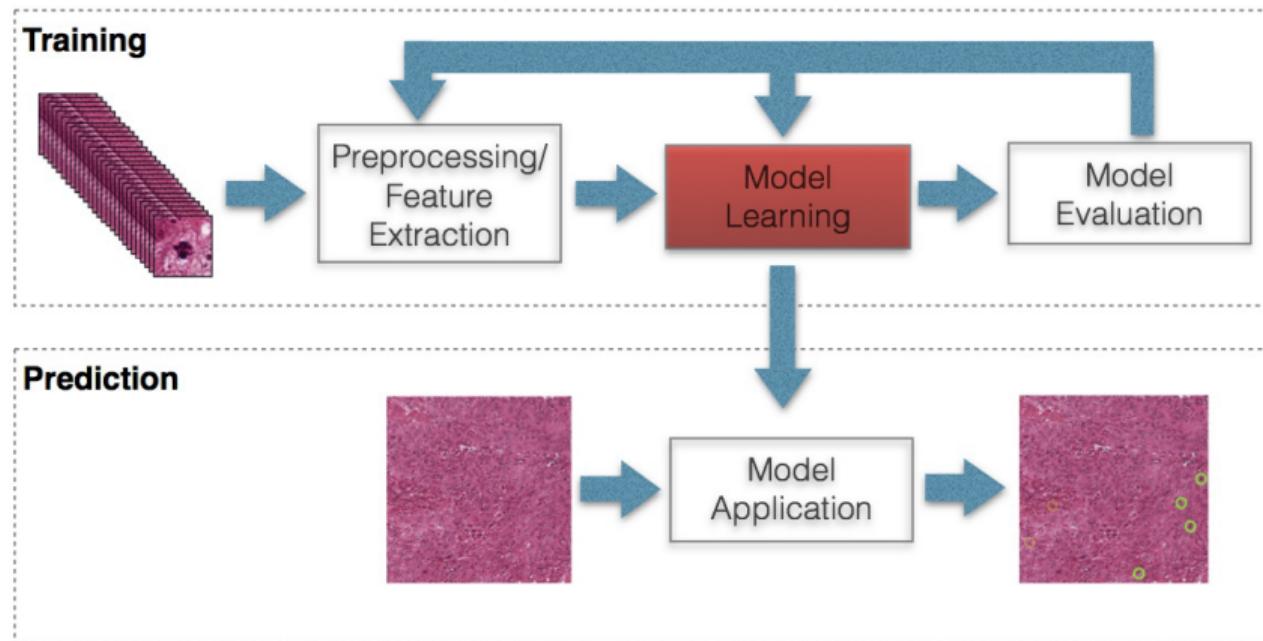


El proceso de aprendizaje automático

El proceso de aprendizaje automático



Aprendizaje de modelos



Inducción de modelos a partir de datos

- El aprendizaje es un problema *ill-posed* (más de una solución posible para el mismo problema particular. Las soluciones son sensibles a pequeños cambios en el problema)
- Es necesario hacer suposiciones adicionales acerca del tipo de patrones que queremos aprender
- **Espacio de hipótesis:** conjunto de patrones válidos que pueden ser aprendidos por el algoritmo de aprendizaje
- Occam's razor: "*En igualdad de condiciones, la solución más sencilla tiende a ser la mejor*".

Enfoques de aprendizaje

■ Probabilísticos:

- Modelos generativos: modelo $P(Y, X)$
- Modelos discriminativos: modelo $P(Y|X)$

■ Geométricos:

- **Manifold learning:** modela la geometría del espacio donde residen los datos
- **Max margin learning:** modela la separación entre las clases

■ Optimización:

- Minimización de energía/pérdida/riesgo

El aprendizaje como problema de optimización

■ Problema de Optimización general:

$$\min_{f \in H} L(f, D)$$

con H : espacio de Hipótesis, D : datos de entrenamiento, L : función de pérdida/error

■ Ejemplo de regresión logística:

■ Espacio de Hipótesis:

$$y(x) = P(C_+|x) = \sigma(w^T x)$$

■ Error de entropía cruzada:

$$E(w) = -\ln p(\mathbf{t}|w) = -\sum_{n=1}^l [t_n \ln y_n + (1-t_n) \ln(1-y_n)]$$

Métodos

■ Supervised generative:

- Naive Bayes
- Graphical models
- Markov random fields
- Hidden markov models

■ Supervised discriminative:

- Logistic regression
- Ridge regression
- Conditional random fields

■ Supervised geometrical

- Max margin classification (SVM)
- k-nearest neighbors

■ Non-supervised generative:

- Latent semantic analysis
- Latent Dirichlet allocation
- Gaussian mixtures

■ Unsupervised geometrical:

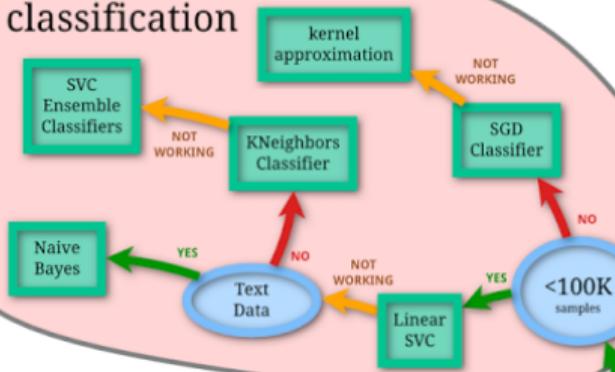
- k-means
- PCA
- Manifold learning

■ Otros

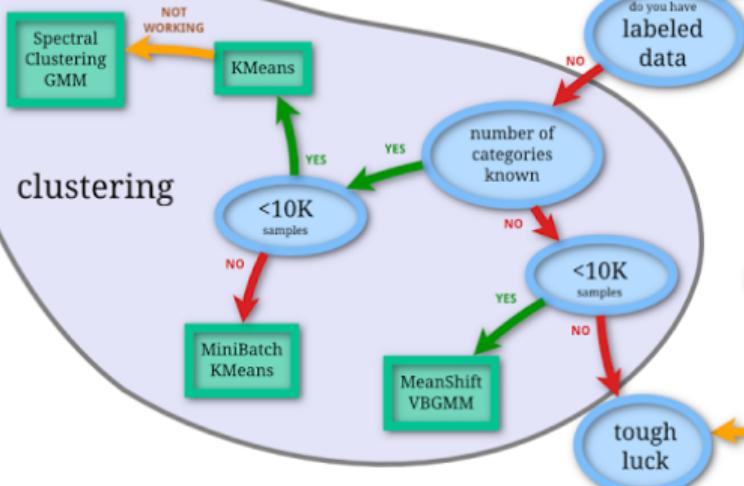
- Neural networks (Deep learning)
- Decision trees
- Association rules

scikit-learn algorithm cheat-sheet

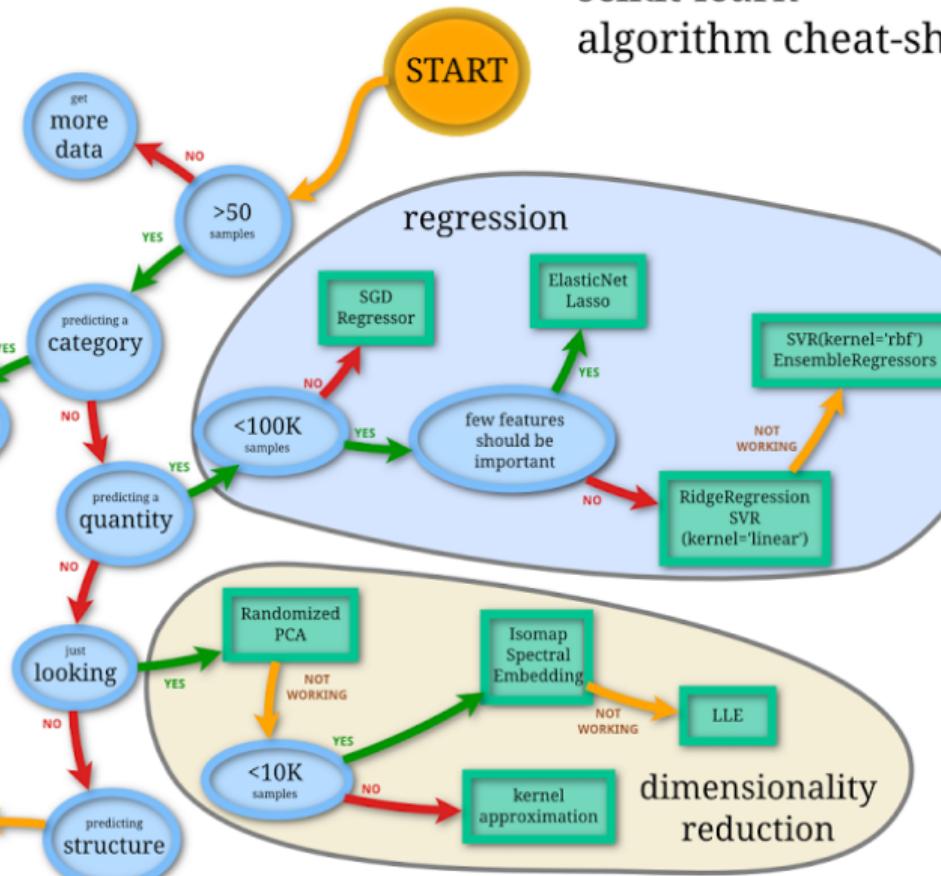
classification



clustering



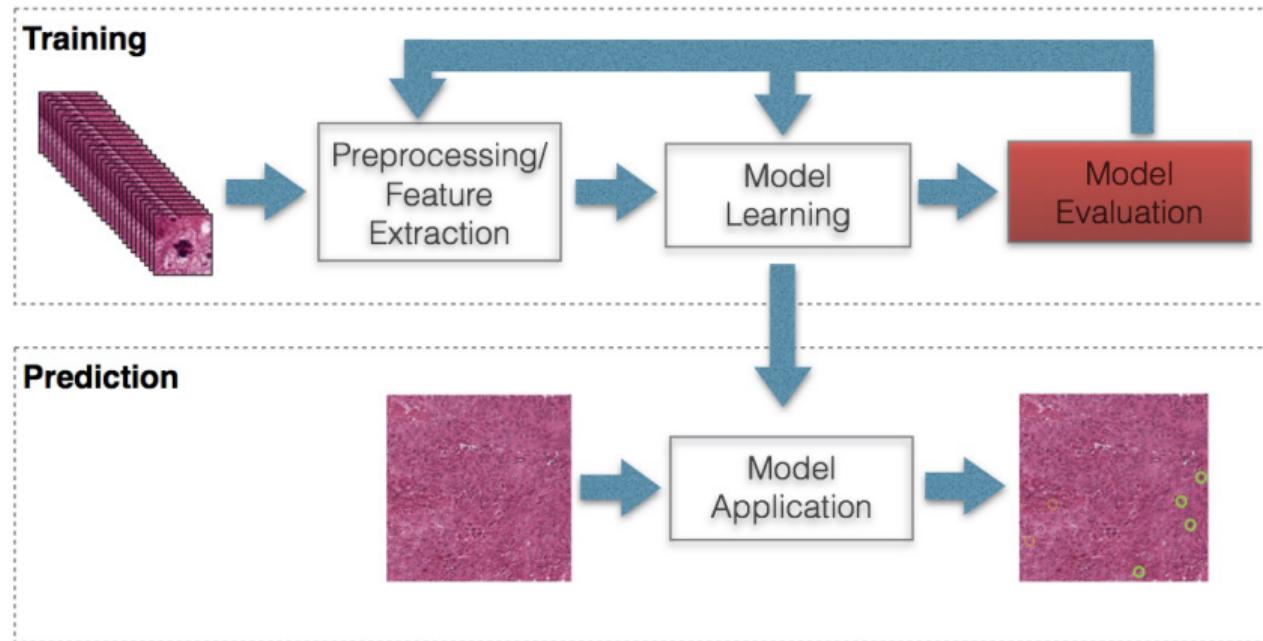
dimensionality reduction



Estrategias

- Optimización (no lineal, convexa, etc.)
- Gradiente Descendiente Estocástico (SGD)
- Métodos de *Kernel*
- Maximum likelihood estimation (Estimación de máxima verosimilitud)
- Maximum a posteriori estimation (Estimación a posteriori máxima)
- Estimación Bayesiana (variational learning, procesos gausianos)
- Expectation maximization (Esperanza-Maximización)
- Modelos de máxima entropía
- Muestreo (Markov Chain Monte Carlo, filtrado de partículas)

Evaluación



Error de entrenamiento vs error de generalización

- Error de entrenamiento:

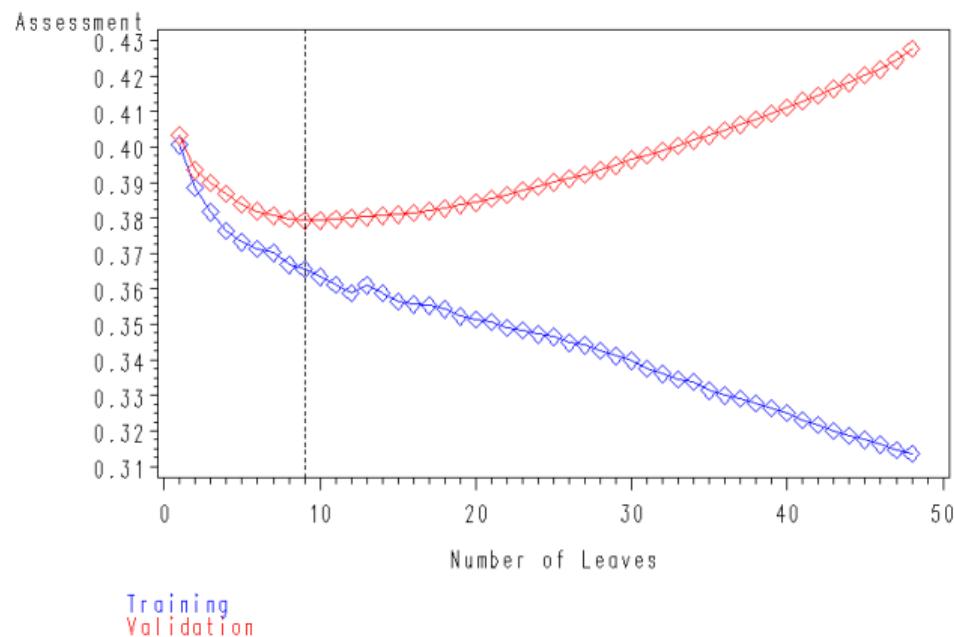
$$\sum_{i=1}^l L(f_w, S_i)$$

- Error de generalización:

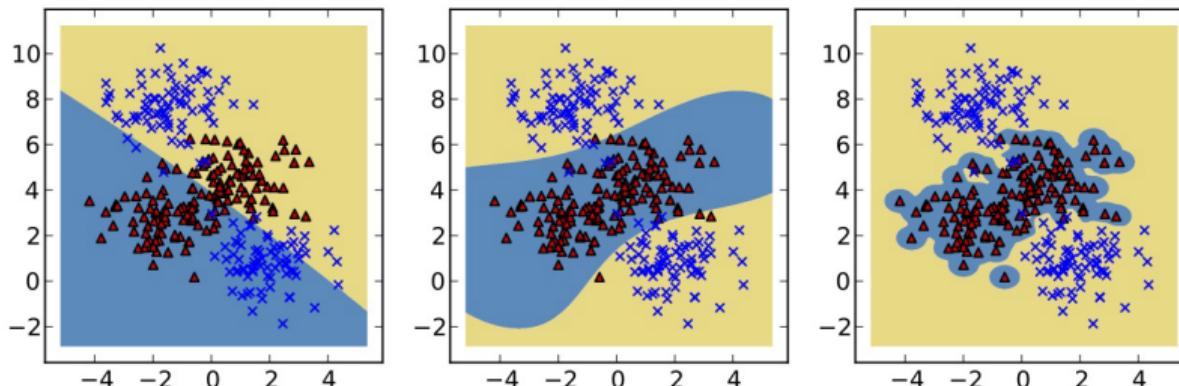
$$E[L(f_w, S)]$$

Validación cruzada

Average Square Error (Gini index)



Overfitting y underfitting (sobreajuste)

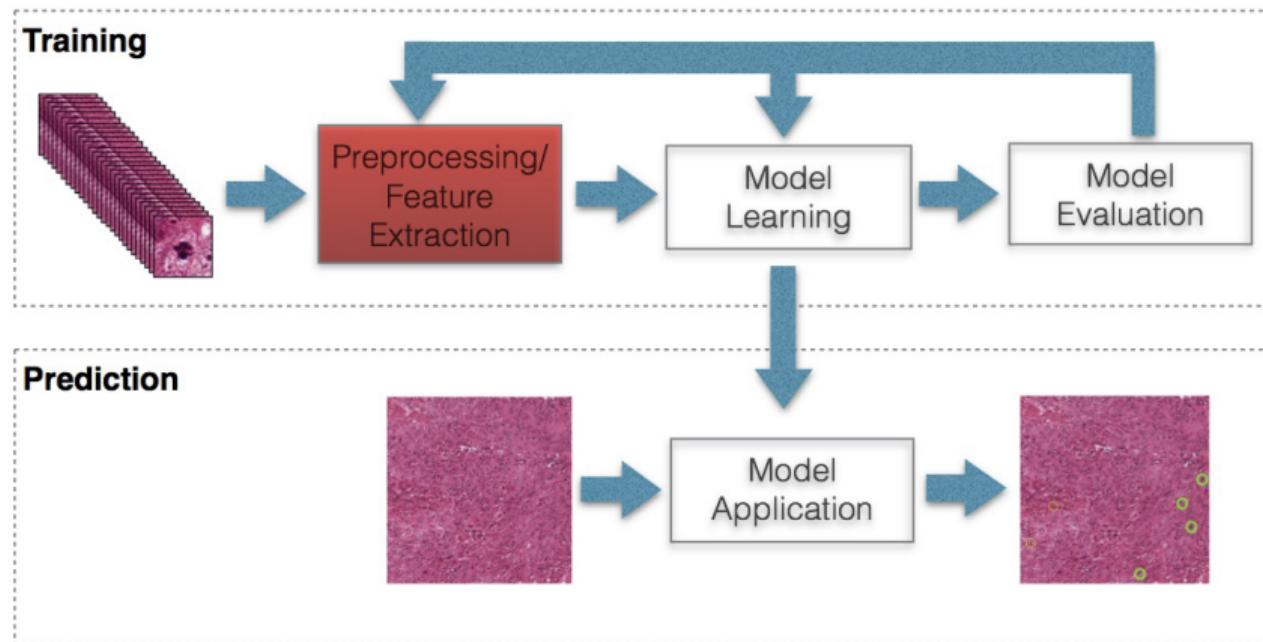


Regularización

$$\min_w -C \sum_{n=1}^{\ell} [t_n \ln y_n + (1-t_n) \ln(1-y_n)] + \|\mathbf{w}\|^2$$

- Controla la complejidad de un modelo aprendido
 - Por lo general, el término de regularización corresponde a una norma del vector parámetro (L_1 o L_2 los más comunes)
 - En algunos casos, equivale a la inclusión de un *prior* y encontrando una solución MAP

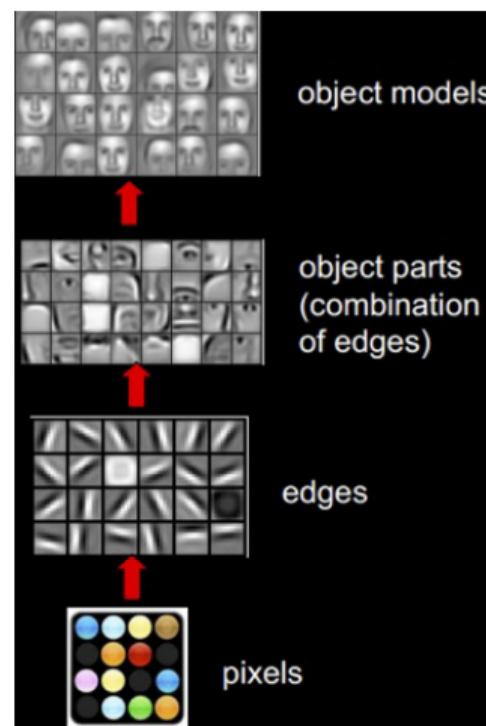
Extracción de características



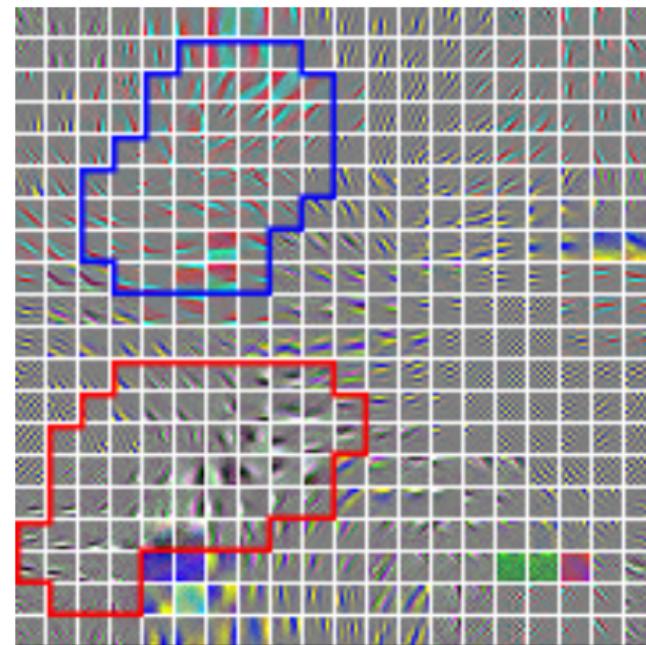
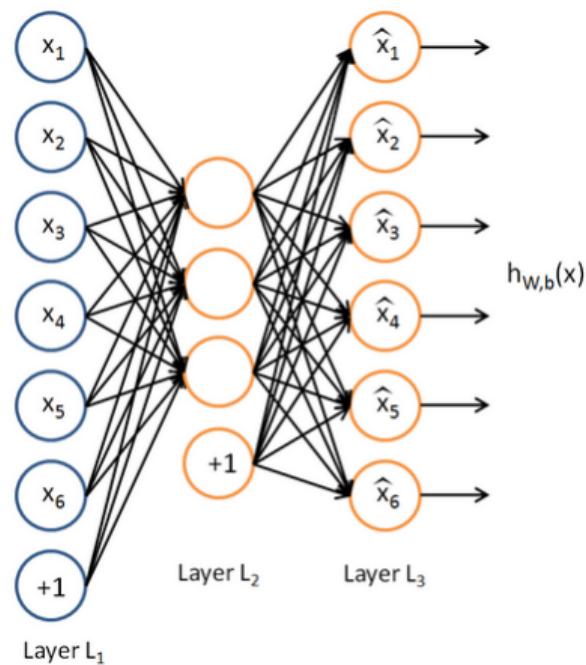
Características

- Las características representan nuestro conocimiento previo del problema
- Dependen del tipo de datos
- Existen características especiales para prácticamente cualquier tipo de datos (imágenes, vídeo, sonido, voz, texto, páginas web, etc.)
- **Ejemplo:** imágenes médicas:
 - Características estándar de visión por computador (color, forma, textura, bordes, local-global, etc.)
 - Características especializadas adaptadas al problema en cuestión
- **Nueva tendencia:** aprendizaje de características a partir de los datos

Aprendizaje de características



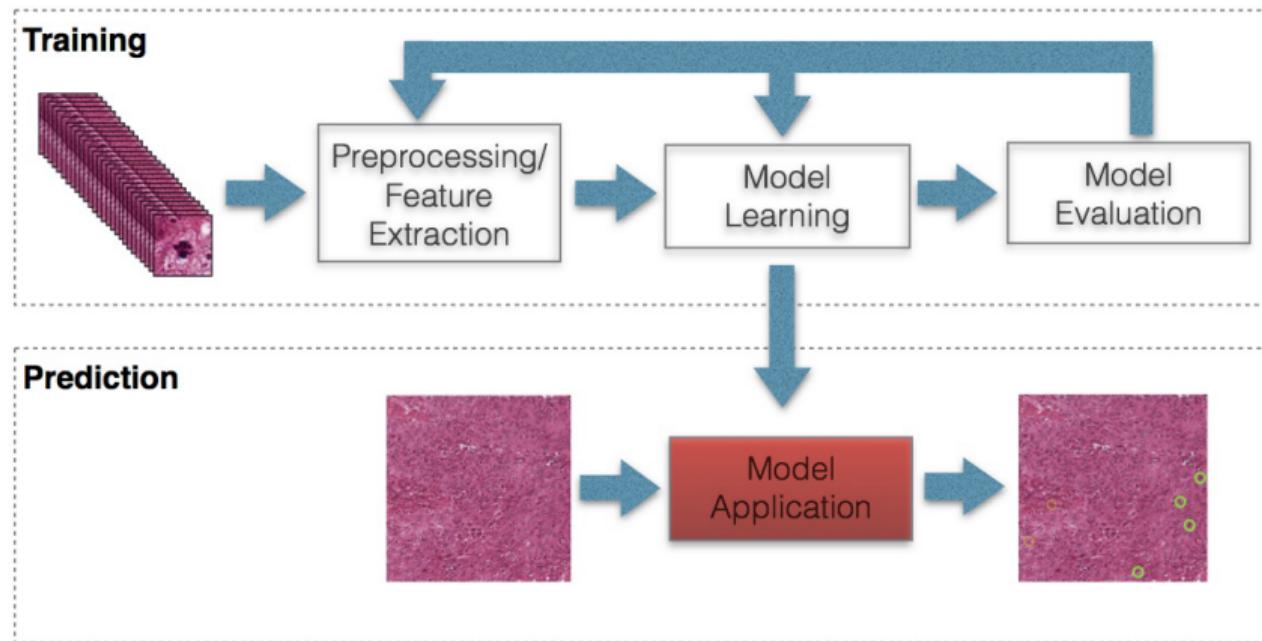
Aprendizaje de características no supervisado



Desafío AMIDA-MICCAI 2013

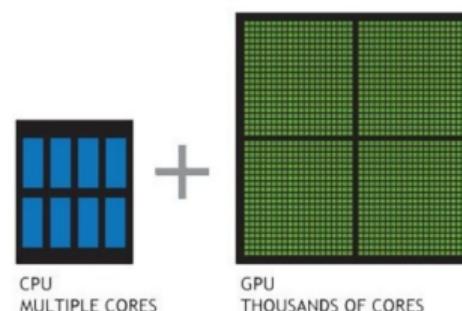
Team name	Precision	Recall	F ₁ -Score	
IDSIA	0.610	0.612	0.611	→ UFL using deep-leranin
DTU	0.427	0.555	0.483	→ Donut like histograms—SVM
SURREY	0.357	0.332	0.344	→ Shape, color, texture—SVM
ISIK	0.306	0.351	0.327	
PANASONIC	0.336	0.310	0.322	
CCIPD/MINDLAB	0.353	0.291	0.319	
WARWICK	0.171	0.552	0.261	
POLYTECH/UCLAN	0.186	0.263	0.218	
MINES	0.139	0.490	0.217	
SHEFFIELD/SURREY	0.119	0.107	0.113	
SEOUL	0.032	0.630	0.061	
NTUST	0.011	0.685	0.022	
UNI-JENA	0.007	0.077	0.013	
NIH	0.002	0.049	0.003	

Aplicación del modelo



Análisis de datos de alto rendimiento

- Aprendizaje automático a gran escala (Big-Data):
 - Gran número de muestras
 - Muestras grandes (*whole-slide images*, volúmenes de alta resolución 4D)
- Algoritmos de aprendizaje escalable (aprendizaje en línea)
- Arquitecturas de computación distribuida (Hadoop, Spark)
- Computación GPGPU y arquitecturas *multicore*
- Nueva tendencia: aprendizaje de características a partir de los datos



Muchas gracias por su atención

¿Preguntas?



Contacto: Marco Teran
webpage: marcoteran.github.io/