CURSOS DE VERANO 2014

APROXIMACIÓN PRÁCTICA A LA CIENCIA DE DATOS Y BIG DATA: HERRAMIENTAS KNIME, R, HADOOP Y MAHOUT

Experiencia práctica en Ciencia de Datos: La Competición de KAGGLE como plataforma para la adquisición de experiencia profesional

José Antonio Guerrero Durán

Contenidos

- Kaggle: plataforma de referencia mundial para análisis predictivo de datos
- Estado del arte en análisis predictivo: Paradigmas de la Estadística Multivariable vs el Aprendizaje Automático
- Experiencias en competiciones Kaggle

Origen de Kaggle

kaggle

- 2006: Algoritmo de recomendación de películas
  - 100 M de votos
  - 480 K clientes
  - 17.7 K películas
- Mejora del 10% en la predicción -> 1.000.000 $
- 2009: 40.000 equipos participantes

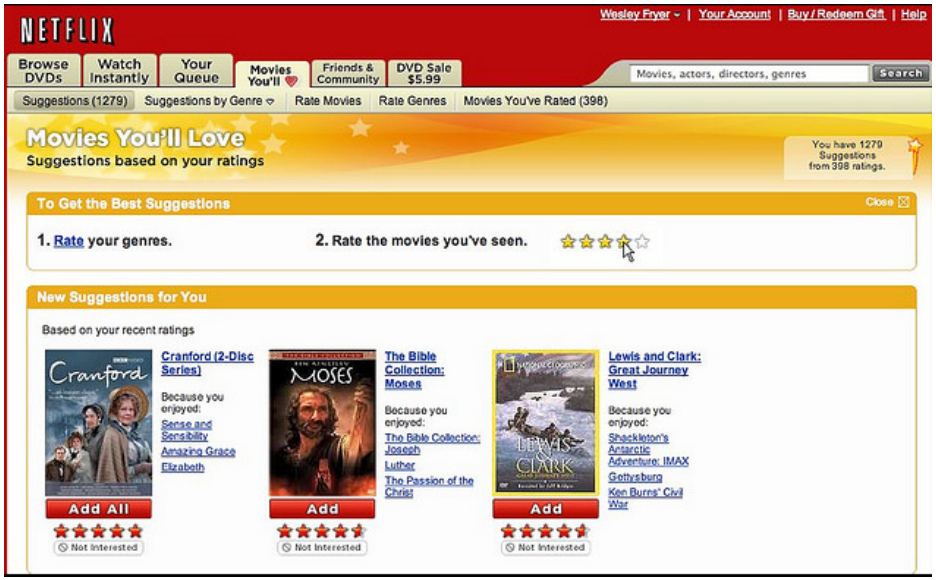
Origen de Kaggle

		Best Score	% Improvement
1	The Ensemble	0.8553	10.10
2	BellKor's Pragmatic Chaos	0.8554	10.09
	Grand Prize Barrier	**0.8563**	
3	Grand Prize Team	0.8571	9.91
4	Opera Solutions and Vandelay United	0.8573	9.89
5	Vandelay Industries !	0.8579	9.83
6	Pragmatic Theory	0.8582	9.80
7	BellKor in BigChaos	0.8590	9.71
8	Dace_	0.8603	9.58
9	Opera Solutions	0.8611	9.49
10	BellKor	0.8612	9.48

Origen de Kaggle

Origen de Kaggle

Origen de Kaggle



- Ensamblado
- SVD matrices dispersas
- Producción científica
- La totalidad de sistemas de recomendación actuales incorporan mejoras obtenidas de esta competición

Origen de Kaggle

commendo

Basics | Research | References | Company

Home | Contact | Sitemap | Imprint

Home > References > Awards

+ A -

References

- Awards
- Customers
- Public relations

recommendo The Engine

Customized Systems

Basic Information

Achievements

- Expedia Challenge: Personalize Expedia Hotel Searches - ICDM 2013
  Kaggle Competition (11/2013)
- KDD Cup 2012: 2nd place (Track 2) - together with Opera Solutions
  ACM Sig Conference on Knowledge Discovery and Data Mining (08/2012)
- KDD Cup 2011: 2nd place (Track 1) & 3rd place (Track 2)
  ACM Sig Conference on Knowledge Discovery and Data Mining (08/2011)
- KDD Cup 2010: 3rd place.
  ACM Sig Conference on Knowledge Discovery and Data Mining (06/2010)
- INNOward 2010.
  The State Economic Chamber Styria, Austria (02/2010)
- Austrian Federal State Award 'Consulting/Knowledge based services' .
  Republic of Austria, Federal Ministry of Economy (11/2009)
- Western Styrian Business Award - Category Innovation
  Wirtschaftsbund Steiermark - Voitsberg, Austria (10/2009)
- ebiz egovernment Award Styria: 2nd place.

- **Georg Preßler, Michael Schrotter, Michael Jahrer, Andreas Töscher**

Origen de Kaggle

"We're making data science into a sport" ... Kaggle founder Anthony Goldbloom.

- **2010: We're making data science into a sport**

Kaggle, referencia para el análisis predictivo

- 400 retos en un formato similar a Netflix:
  - Problemas de predicción o de optimización.
  - Variable respuesta
  - Métrica para evaluar el error cometido en las estimaciones.
- Training set: conjunto de datos con variable respuesta.
- Test set: conjunto de datos para validar los modelos
- Clasificación pública: sobre una parte del test set
- Clasificación privada: otra parte del test set no utilizada hasta la finalización

Kaggle, referencia para el análisis predictivo

- **200.000 miembros registrados**
- **400 retos (250 son prácticas de cursos de instituciones académicas)**
  - **Patrocinados por empresas para resolver un problema real**
  - **Organizados por instituciones académicas (congresos, investigación)**
  - **Promocionados por Kaggle (conocimiento)**

Kaggle, referencia para el análisis predictivo

- **Puntos en función del número de participantes y de la posición alcanzada.**

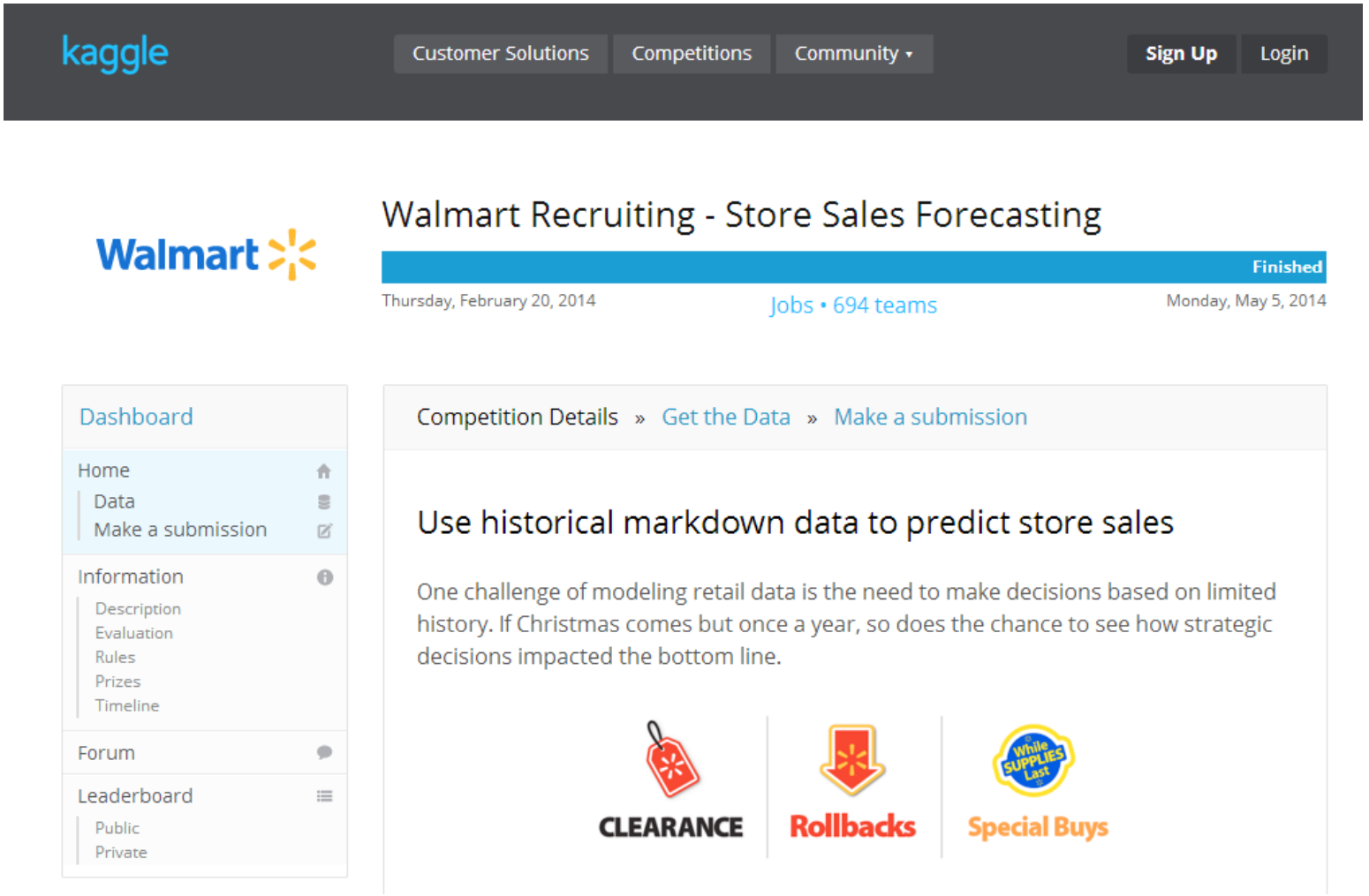
- **Los puntos se van acumulando mediante una fórmula que tiene en cuenta el tiempo.**

$$\text{Competition Points} = \frac{100000}{\#\,\text{Team Members}} (\text{Team Rank})^{-0.75} \log_{10}(\#\,\text{Teams}) \frac{2\,\text{years} - \text{time elapsed since deadline}}{2\,\text{years}}$$

Kaggle, referencia para el análisis predictivo

- Ofertas de trabajo: 700 en dos años
- Algunas ligadas a ranking global o de alguna competición
- Retos específicos para seleccionar personal:
  - WalMart
  - Facebook

Kaggle, referencia para el análisis predictivo



Kaggle, referencia para el análisis predictivo

kaggle | Customer Solutions | Competitions | Community ▾ | Sign Up | Login

Facebook Recruiting III - Keyword Extraction

Finished

Friday, August 30, 2013 — Jobs • 367 teams — Friday, December 20, 2013

Dashboard

- Home
  - Data
  - Make a submission
- Information
  - Description
  - Evaluation
  - Rules
  - Prizes
- Forum
- Leaderboard
  - Public
  - Private

Leaderboard

Competition Details » Get the Data » Make a submission

Identify keywords and tags from millions of text questions

Looking for a data science position at Facebook? After two successful prior Kaggle competitions, Facebook continues their mission to identify the best data scientists and software engineers that Kaggle has to offer. In this third installment, they seek candidates who have experience text mining large amounts of data.

Kaggle, referencia para el análisis predictivo

kaggle in Class

Stanford University + kaggle

Sign Up Login

Predicting unemployment in the Great Recession

Finished

Monday, September 23, 2013 — Knowledge • 84 teams — Friday, December 6, 2013

Dashboard

- Home
  - Data
  - Make a submission
- Information
  - Description
  - Evaluation
  - Rules
  - Prizes
- Forum
- Leaderboard
  - Public
  - Private

Leaderboard

1. GobbleGobble
2. Bryan
3. Jim Monteleone
4. DMFK
5. mudkips
6. Steven Balough

Competition Details » Get the Data » Make a submission

This competition is private-entry. You can view but not participate.

Stats 202 Prediction Challenge.

Who can participate?

This challenge is restricted to students enrolled in Stats 202 at Stanford University in the fall of 2013.

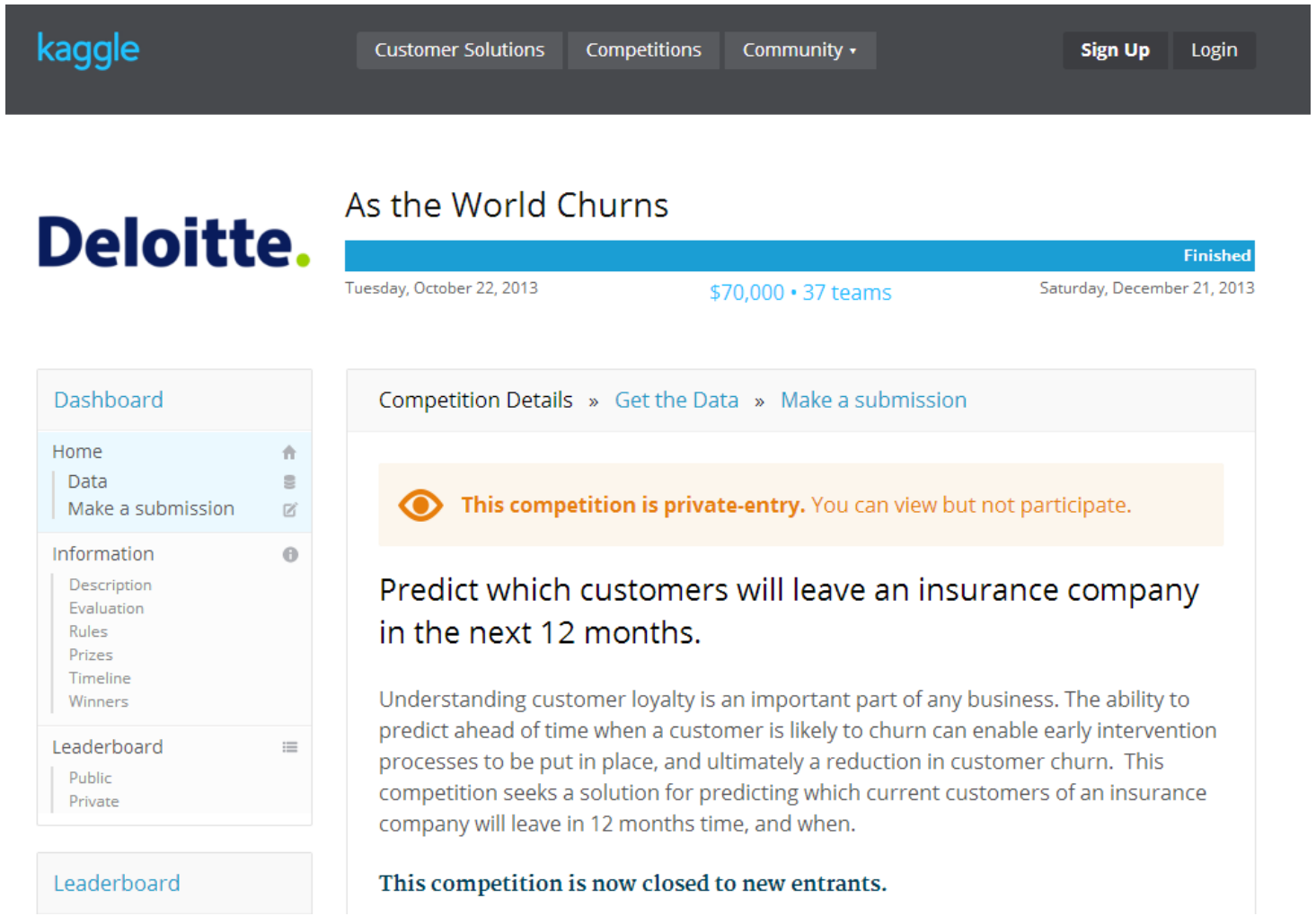
The data

We will use the National Longitudinal Study of Youth (NLSY79). This landmark study published by the National Bureau of Labor Statistics has followed a cohort of baby boomers since 1979, recording various data on employment, education, poverty, family, attitudes, etc.

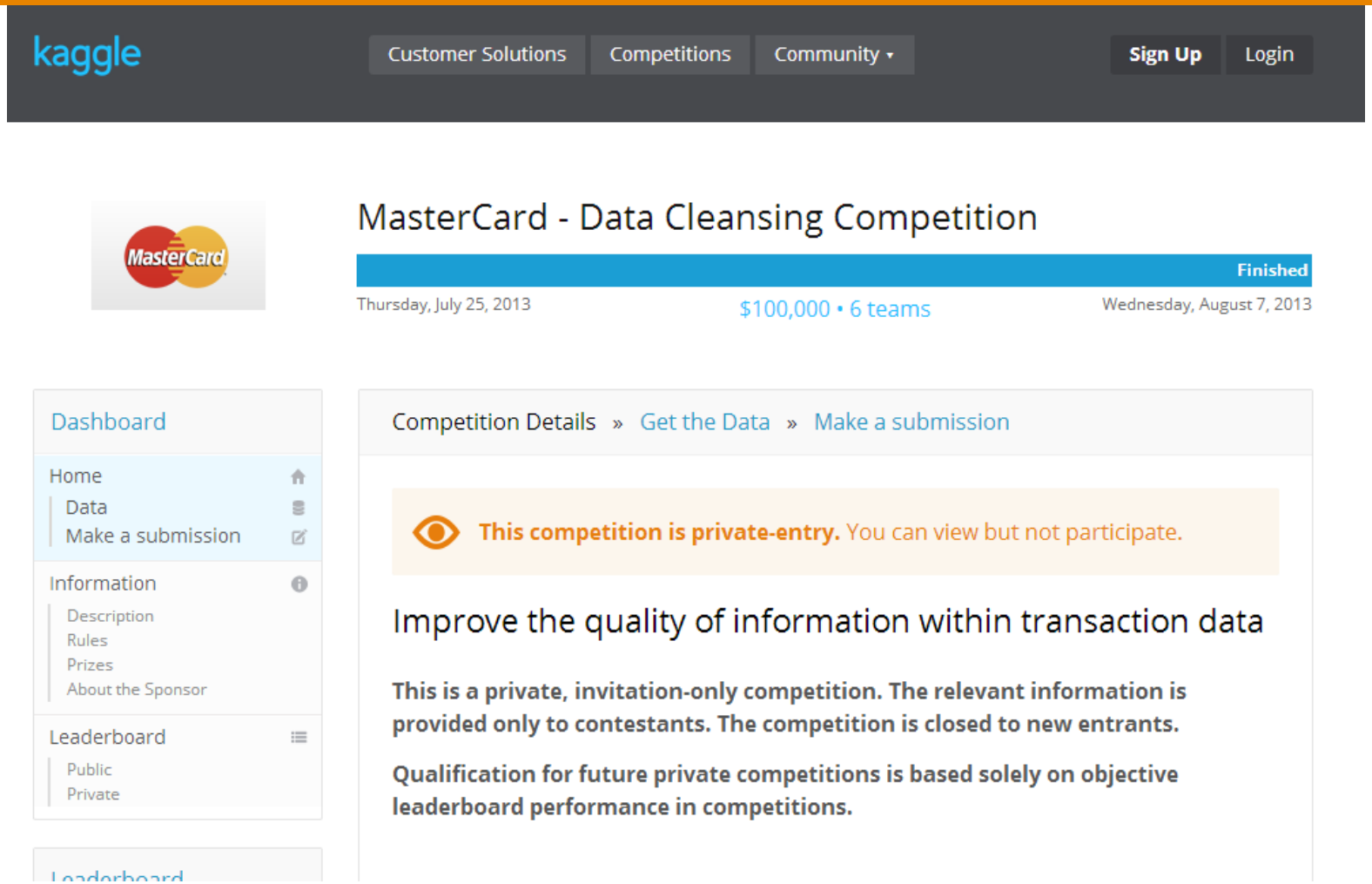
What is your goal?

Based on interviews spanning from 1979 to 2006, you will be tasked with predicting the number of weeks that each person was unemployed during 2010, at the peak of the Great Recession.

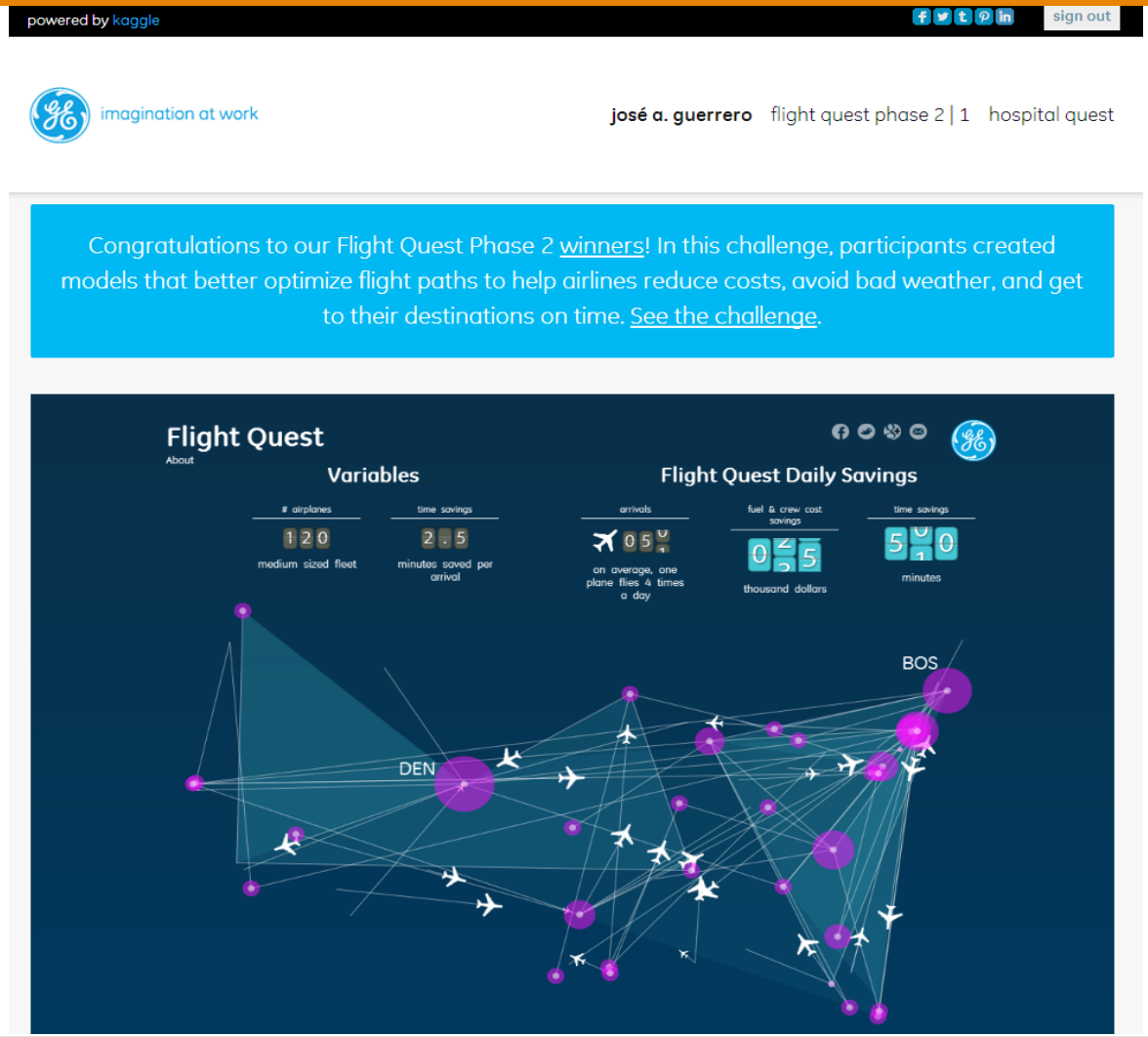
Kaggle, referencia para el análisis predictivo



Kaggle, referencia para el análisis predictivo



Kaggle, referencia para el análisis predictivo



Kaggle, referencia para el análisis predictivo

kaggle | Customer Solutions | Competitions | Community ▾ | Sign Up | Login

Allstate. You're in good hands.

Allstate Purchase Prediction Challenge

Finished

Tuesday, February 18, 2014 — $50,000 • 1,571 teams — Monday, May 19, 2014

- Dashboard
- Home
  - Data
  - Make a submission
- Information
  - Description
  - Evaluation
  - Rules
  - Prizes
  - Timeline
- Forum
- Leaderboard
  - Public
  - Private

Competition Details » Get the Data » Make a submission

Predict a purchased policy based on transaction history

Kaggle, referencia para el análisis predictivo

kaggle

Customer Solutions | Competitions | Community ▾ | Sign Up | Login

Genentech
A Member of the Roche Group

Flu Forecasting

Finished

Thursday, December 19, 2013 — $125,000 • 50 teams — Monday, March 3, 2014

Dashboard

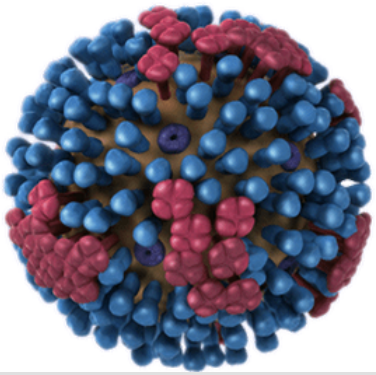
- Home
  - Data
  - Make a submission
- Information
  - Description
  - Evaluation
  - Rules
  - Prizes
  - About Genentech
  - Enter the Competition
  - Timeline
  - Winners
- Leaderboard
  - Public
  - Private

Leaderboard

Competition Details » Get the Data » Make a submission

This competition is private-entry. You can view but not participate.

Predict when, where and how strong the flu will be



Kaggle, referencia para el análisis predictivo

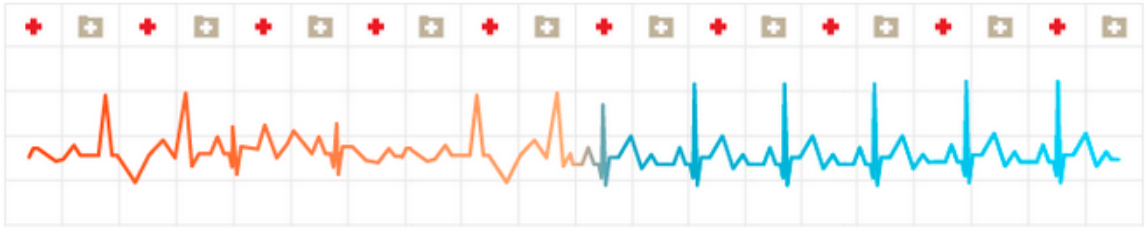
HERITAGE PROVIDER NETWORK HEALTH PRIZE

Sign Up | In the News | Judging Panel | Visit HPN

Dashboard

- Home
- Data
- Information
  - Description
  - Evaluation
  - Rules
  - Dos and Don'ts
  - FAQ
  - Milestone Winners
  - Timeline
- Forum
- Leaderboard
  - Public
  - Private

Leaderboard



Improve Healthcare, Win $3,000,000.

Identify patients who will be admitted to a hospital within the next year using historical claims data. (Enter by 06:59:59 UTC Oct 4 2012)

Please note: This competition is over! The leaderboard now displays the final results.

Kaggle, referencia para el análisis predictivo

Kaggle Rankings

Sorted by **Rank** (Beta)

1st	2nd	3rd	4th	5th
685,783 pts	584,479 pts	569,575 pts	541,648 pts	447,273 pts
				
José A. Guerrero	**Leustagos**	**BreakfastPirate**	**Owen**	**Naokazu Mizuta**
20 competitions Spain	28 competitions Belo Horizonte Brazil	18 competitions Indianapolis United States	16 competitions NYC United States	28 competitions Tokyo Japan

1st	2nd	3rd	4th	5th
821,194 pts	539,227 pts	498,581 pts	496,081 pts	457,483 pts
				
Owen	**BreakfastPirate**	**Leustagos**	**David Thaler**	**José A. Guerrero**
26 competitions NYC United States	25 competitions Indianapolis United States	36 competitions Belo Horizonte Brazil	14 competitions Seattle United States	28 competitions Spain

Estado del Arte

- ¿Cómo son los datos facilitados en este tipo de problemas y que tamaño tienen?
- ¿Qué herramientas son las que se usan más frecuentemente?
- ¿Cuál es el perfil de las personas que se dedican a resolverlos?
- ¿Qué técnicas son las más utilizadas?

Estado del Arte: Datos

- 10M – 250 M Registros
- 100 – 10000 variables
- 1 – 20 Gb
- Microarrays: Variables:Registros -> 10:1
- Raw data vs blind data

Estado del Arte: Datos

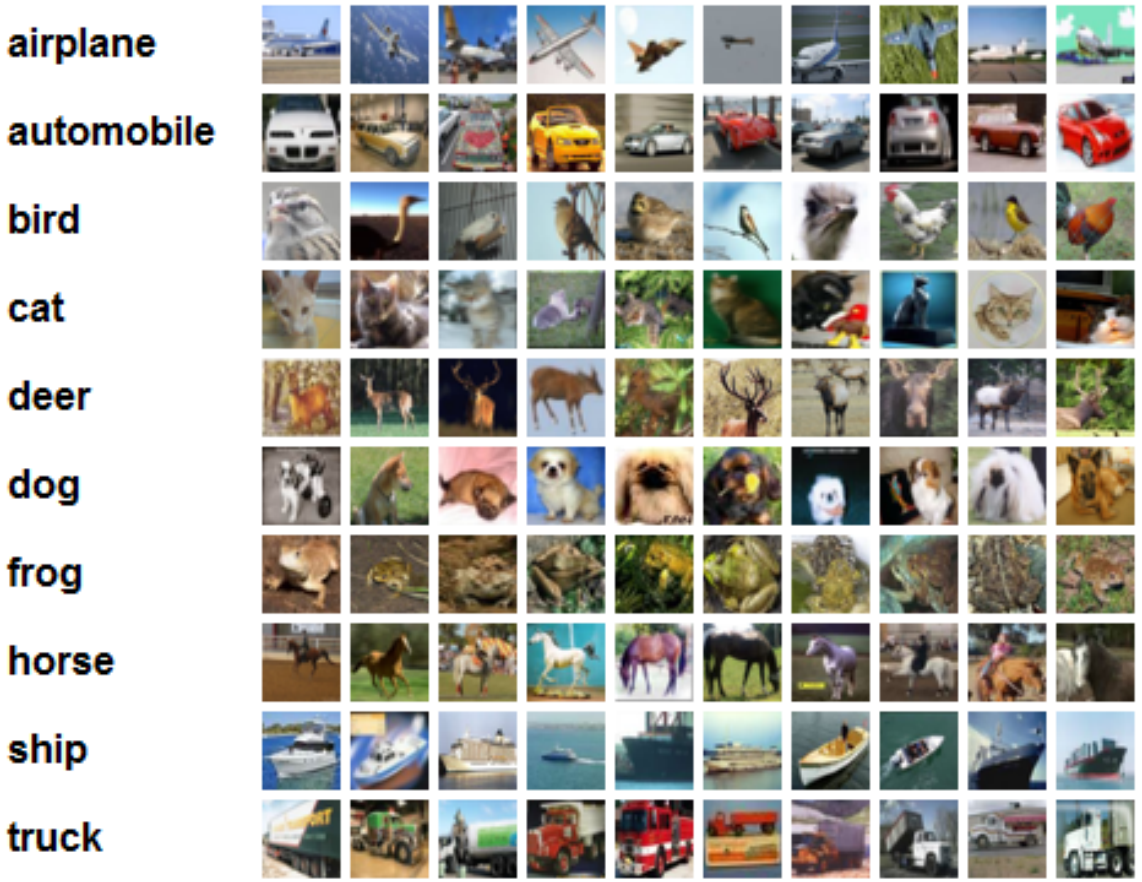
- Información texto:
- Corpus (construido o externo)
- Tf-idf (Term frequency – Inverse document frequency)
- N-gramas: palabras próximas aportan información contextual
- Correctores gramaticales, ortográficos, traductores, análisis de sentimiento
- Aproximaciones bayesianas
- Cadenas de Markov

Estado del Arte: Datos

- Web 1T 5-gram Version 1 (Google 2006)
- Construido a partir de 95.000 millones de frases con más de 1.000.000.000.000 de palabras.
  - 13.5 M 1-gramas
  - 314 M bi-gramas
  - 977 M tri-gramas
  - 1.300 M 4-gramas
  - 1.174 M 5-gramas

Estado del Arte: Datos

- **CIFAR10** 60.000 imágenes en 10 categorías

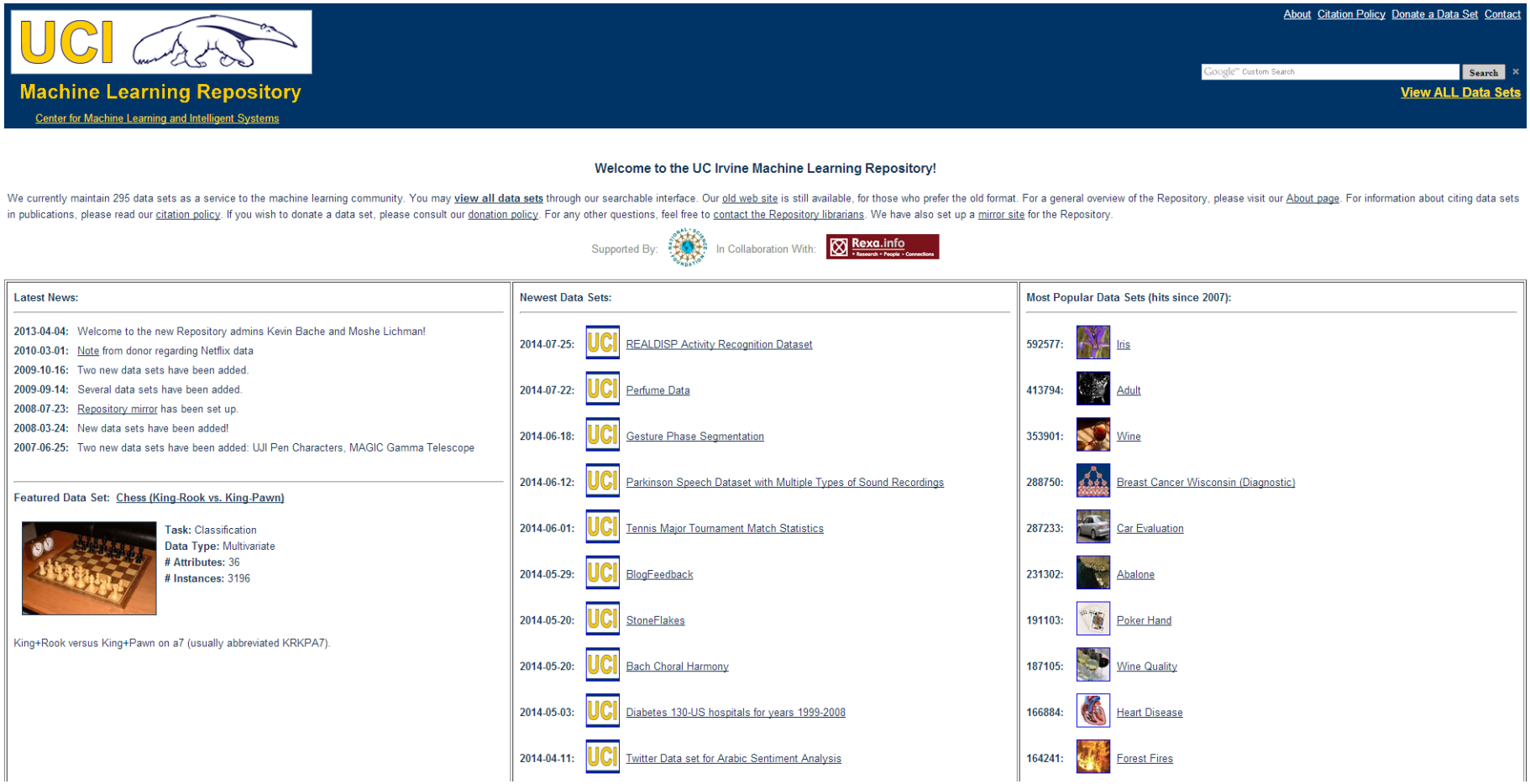


Estado del Arte: Datos

- **MNIST (Mixed Nat. Institute of Standards and Technology) 70.000 dígitos**

Estado del Arte: Datos

UCI Machine Learning Repository



UCI

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

About Citation Policy Donate a Data Set Contact

Google™ Custom Search Search

View ALL Data Sets

Welcome to the UC Irvine Machine Learning Repository!

We currently maintain 295 data sets as a service to the machine learning community. You may **view all data sets** through our searchable interface. Our old web site is still available, for those who prefer the old format. For a general overview of the Repository, please visit our About page. For information about citing data sets in publications, please read our citation policy. If you wish to donate a data set, please consult our donation policy. For any other questions, feel free to contact the Repository librarians. We have also set up a mirror site for the Repository.

Supported By: In Collaboration With: Rexa.info

Latest News:

2013-04-04: Welcome to the new Repository admins Kevin Bache and Moshe Lichman!
2010-03-01: Note from donor regarding Netflix data
2009-10-16: Two new data sets have been added.
2009-09-14: Several data sets have been added.
2008-07-23: Repository mirror has been set up.
2008-03-24: New data sets have been added!
2007-06-25: Two new data sets have been added: UJI Pen Characters, MAGIC Gamma Telescope

Featured Data Set: Chess (King-Rook vs. King-Pawn)

Task: Classification
Data Type: Multivariate
 # Attributes: 36
 # Instances: 3196

King+Rook versus King+Pawn on a7 (usually abbreviated KRKPA7).

Newest Data Sets:

2014-07-25: REALDISP Activity Recognition Dataset
2014-07-22: Perfume Data
2014-06-18: Gesture Phase Segmentation
2014-06-12: Parkinson Speech Dataset with Multiple Types of Sound Recordings
2014-06-01: Tennis Major Tournament Match Statistics
2014-05-29: BlogFeedback
2014-05-20: StoneFlakes
2014-05-20: Bach Choral Harmony
2014-05-03: Diabetes 130-US hospitals for years 1999-2008
2014-04-11: Twitter Data set for Arabic Sentiment Analysis

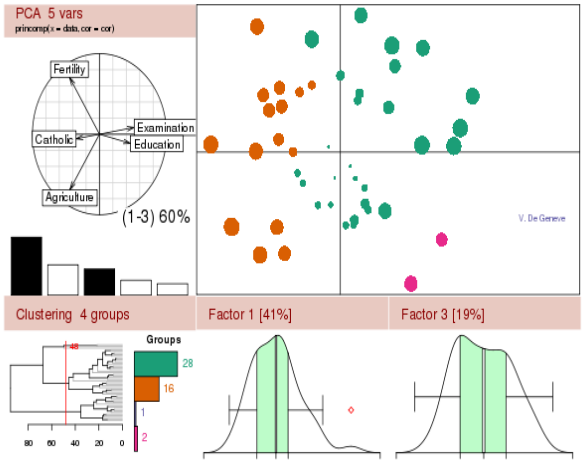
Most Popular Data Sets (hits since 2007):

592577: Iris
413794: Adult
353901: Wine
288750: Breast Cancer Wisconsin (Diagnostic)
287233: Car Evaluation
231302: Abalone
191103: Poker Hand
187105: Wine Quality
166884: Heart Disease
164241: Forest Fires

Estado del Arte: Herramientas

www.r-project.org

The R Project for Statistical Computing



About R
What is R?
Contributors
Screenshots
What's new?

Download, Packages
CRAN

R Project
Foundation
Members & Donors
Mailing Lists
Bug Tracking
Developer Page
Conferences
Search

Documentation
Manuals
FAQs
The R Journal
Wiki
Books
Certification
Other

Misc
Bioconductor
Related Projects
User Groups
Links

Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To download R,
- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you

News:

- **R version 3.1.0** (Spring Dance) has been released on 2014-04-10.
- **R version 3.0.3** (Warm Puppy) has been released on 2014-03-06.
- The R Journal Vol.5/2 is available.
- useR! 2013, took place at the University of Castilla-La Mancha, Albacete, Spain, July 10-12 2013.
- **R version 2.15.3** (Security Blanket) has been released on 2013-03-01.

This server is hosted by the Institute for Statistics and Mathematics of WU (Wirtschaftsuniversität Wien).

Estado del Arte: Herramientas

- **R : Lenguaje interpretado de propósito general**
- **Más de 5.500 paquetes disponibles**
- **Facilidad de uso interactivo**
- **Puntos débiles:**
  - **Gestión de la memoria**
  - **Paralelización**
- **Paquetes clave:**
  - **data.table**
  - **Bigmemory**
  - **foreach**

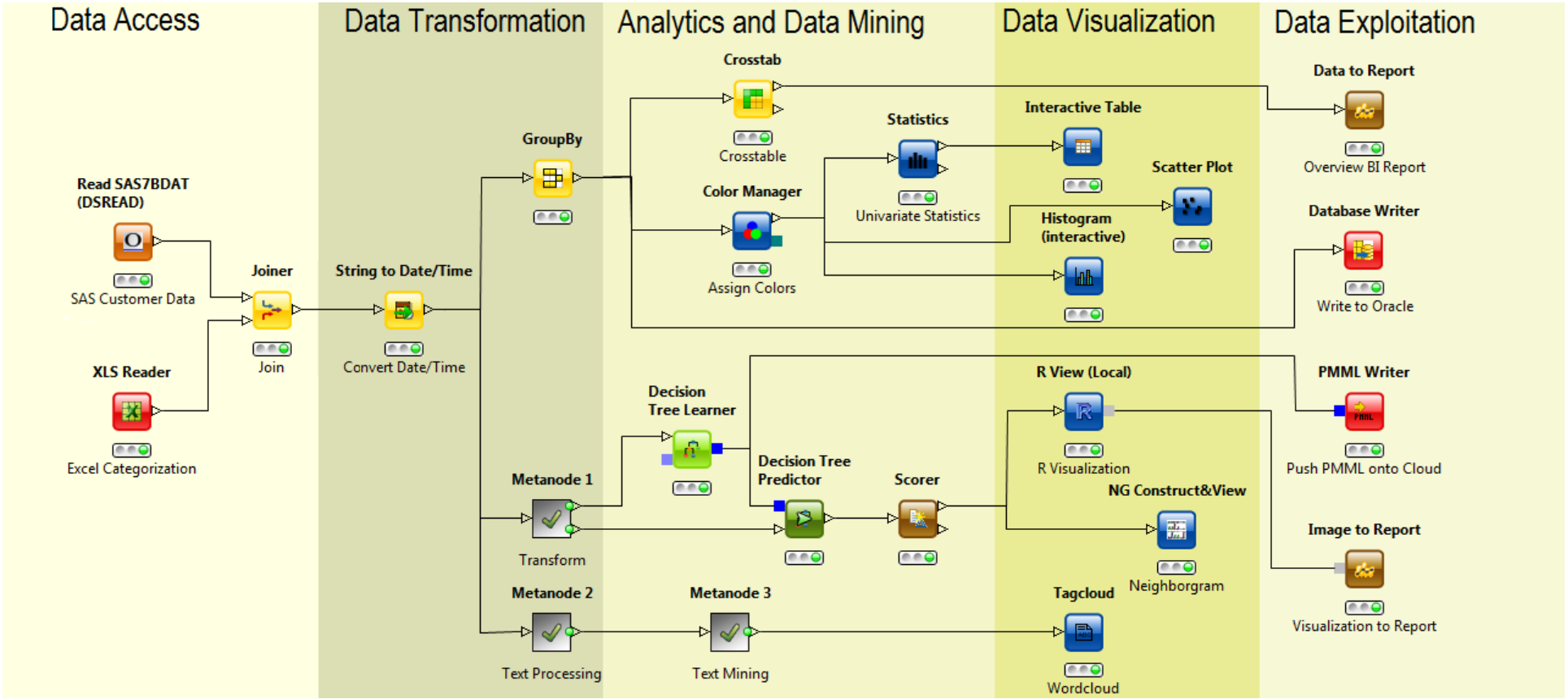
Estado del Arte: Herramientas

Estado del Arte: Herramientas

- Python: Lenguaje compilado
- Orientado a objetos
- Eficiente gestión de memoria
- Paralelización
- Menos paquetes disponibles
- Mejor para textos e imágenes que R
- Referencia para NN (pylearn2)

Estado del Arte: Herramientas

- www.knime.org (Konstanz Information Miner)



Estado del Arte: Herramientas

Estado del Arte: Herramientas

Elmer Fudd
Vorpal Rabbit

YAHOO!

Microsoft

(Fast & Incremental Learning)

John Langford

Estado del Arte: Herramientas

Sofia – ML (Fast & incremental Learning)

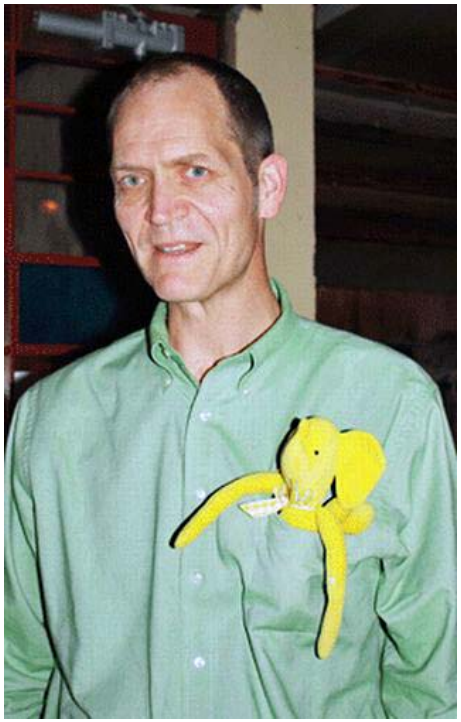
Google™

"This implementation is especially well suited to large, sparse learning problems, and can be used to learn linear SVM models on hundreds of thousands or **millions of data points using a only fraction of a second of CPU time for training on a normal laptop**".

David Sculley

Estado del Arte: Herramientas

Google™

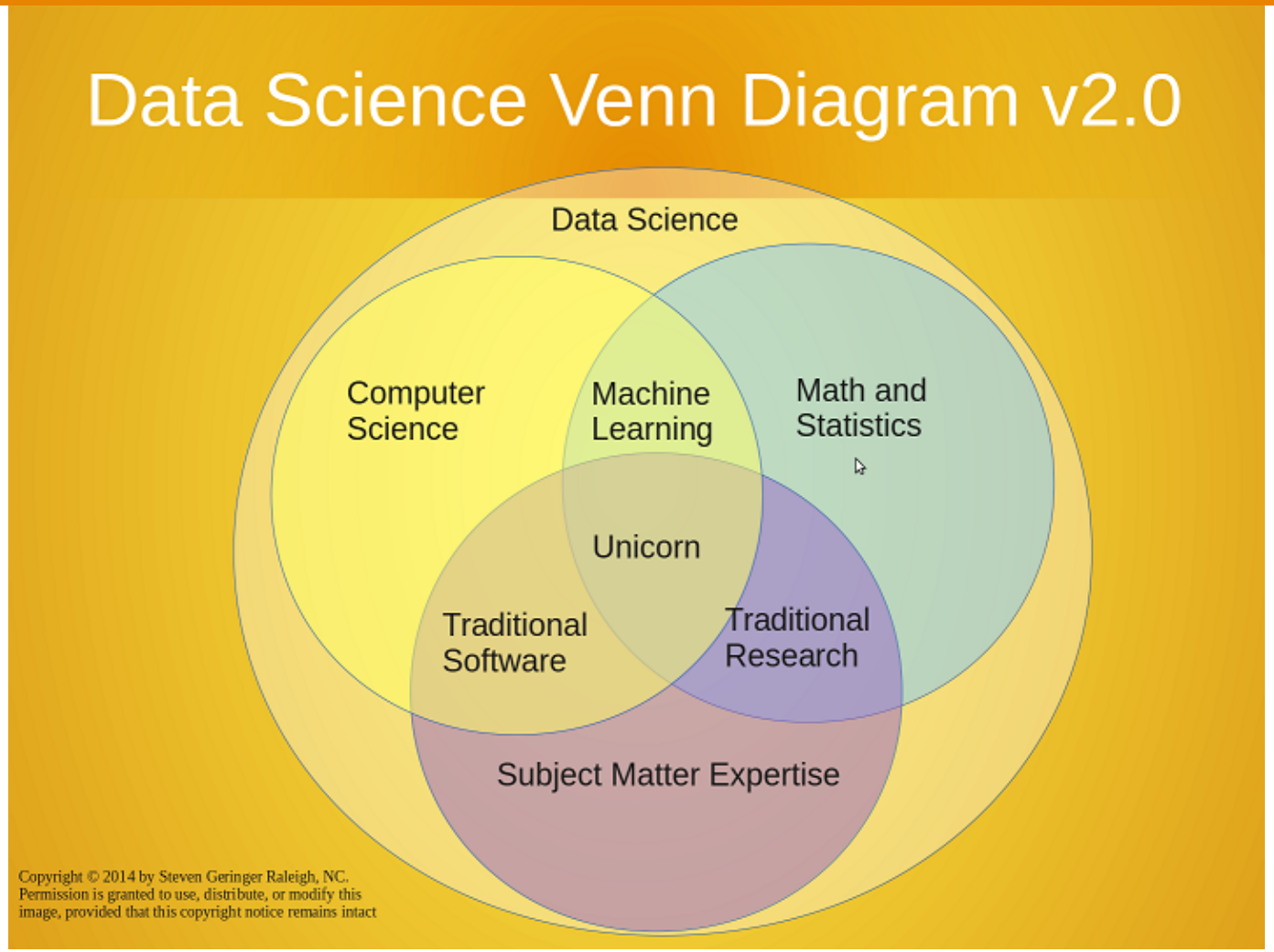


ML para Big Data

Aplicaciones distribuidas

Doug Cutting

Estado del Arte: Profesionales



Estado del Arte: Estadística Multivariable - Machine Learning

Estado del Arte: Estadística Multivariable

- **Hipótesis**
  - **Normalidad**
  - **No correlación de errores**
  - **Homocedasticidad**
  - **No colinealidad**

X , Y

- **Bondad del ajuste:**
  - **Grados de libertad**
  - **Descomposición de la varianza**
  - **Estimaciones puntuales y por IC de errores y coeficientes**
  - **Contraste de hipótesis**

Estado del Arte: Estadística Multivariable

- **DEBILIDADES**
  - **Asumir hipótesis sobre la distribución de los datos**
  - **Mal manejo de la colinealidad**
  - **Convergencia y estabilidad de las soluciones**
  - **La limitación en la forma funcional del modelo**
  - **Alta sensibilidad a observaciones extremas**
  - **Mal manejo de observaciones desconocidas**
  - **Problemas de escalabilidad**
  - **Mal manejo variables >> casos**

Estado del Arte: Estadística Multivariable

- **FORTALEZAS**
  - **Reproducibles**
  - **Rápidos de ajustar**
  - **Modelos interpretables (expresión analítica)**
  - **Importancia relativa de variables**
  - **Inferencia (bondad de ajuste, coeficientes)**

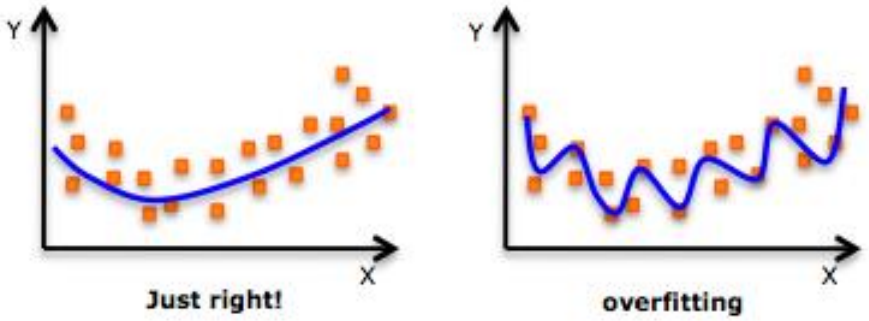
Estado del Arte: Machine Learning

- **Arthur Samuel (1959):**
  - **"Field of study that gives computers the ability to learn without being explicitly programmed“**
- **Tom M. Mitchell (1997):**
  - **"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E".**

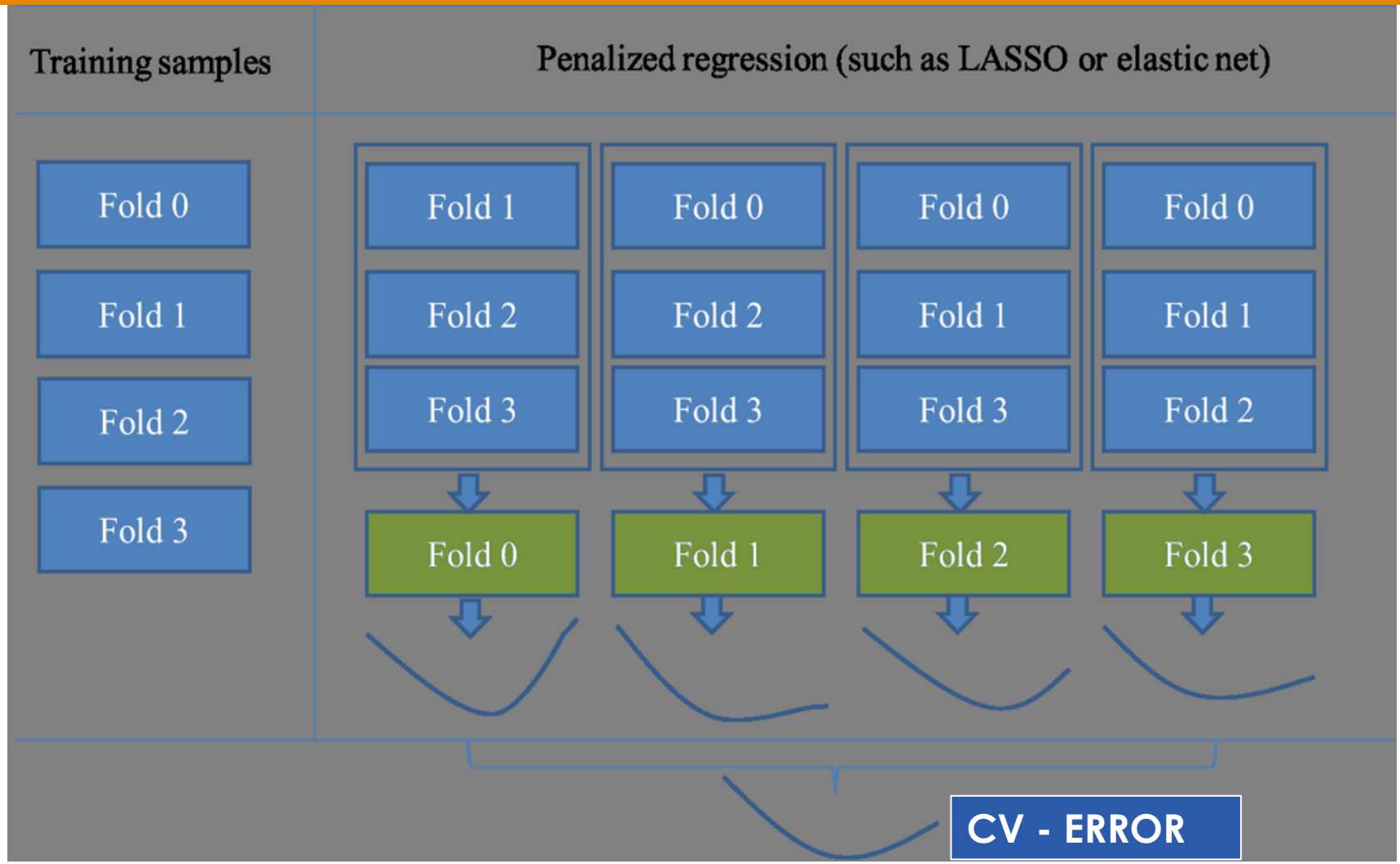
Estado del Arte: Sobreajuste

Estado del Arte: Sobreajuste

- El modelo ajusta el error aleatorio en vez de la verdadera relación entre los predictores y la respuesta
- Modelos complejos (muchos parámetros en relación con observaciones)
- Pocos grados de libertad
- Cross-validación, regularización



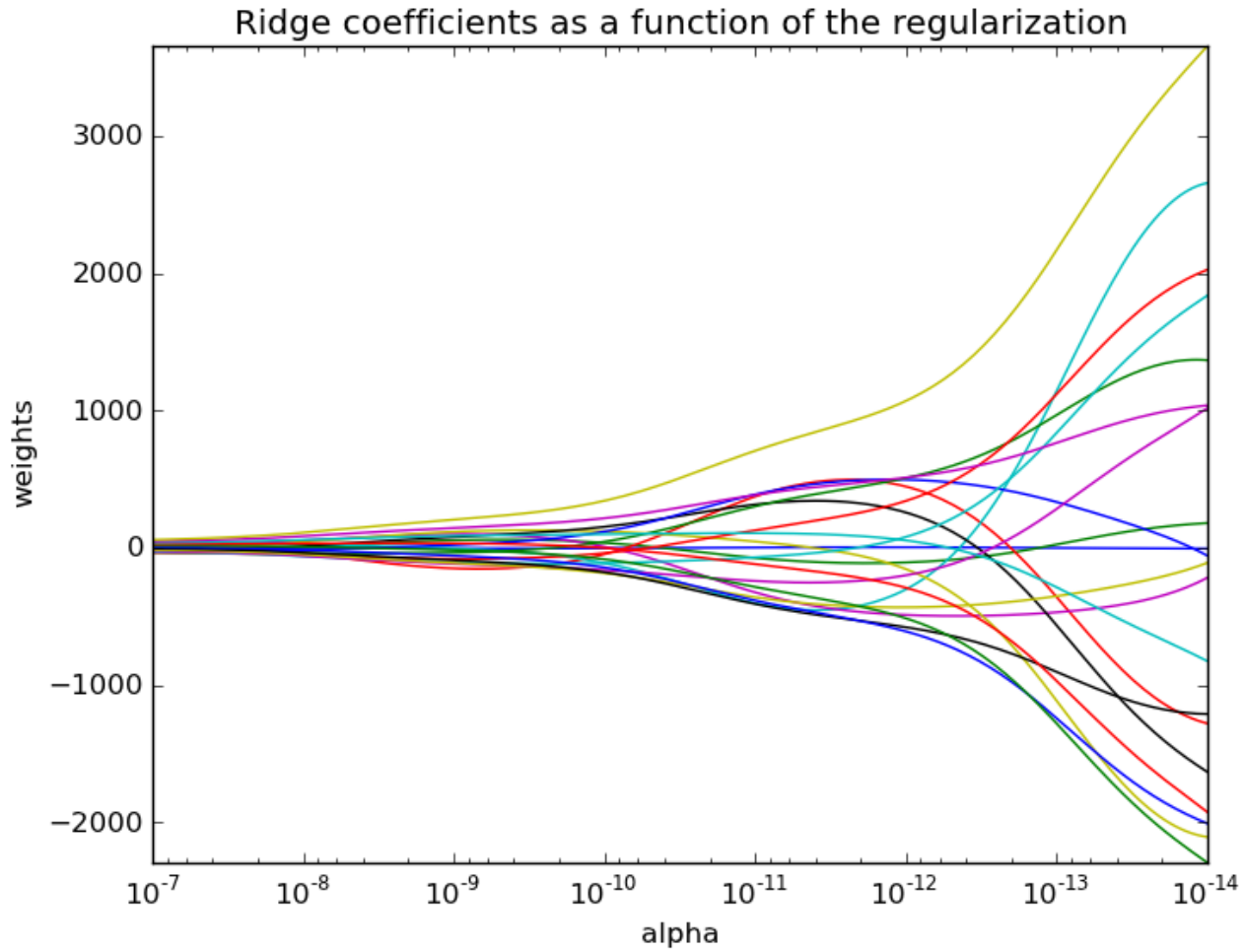
Estado del Arte: Validación cruzada



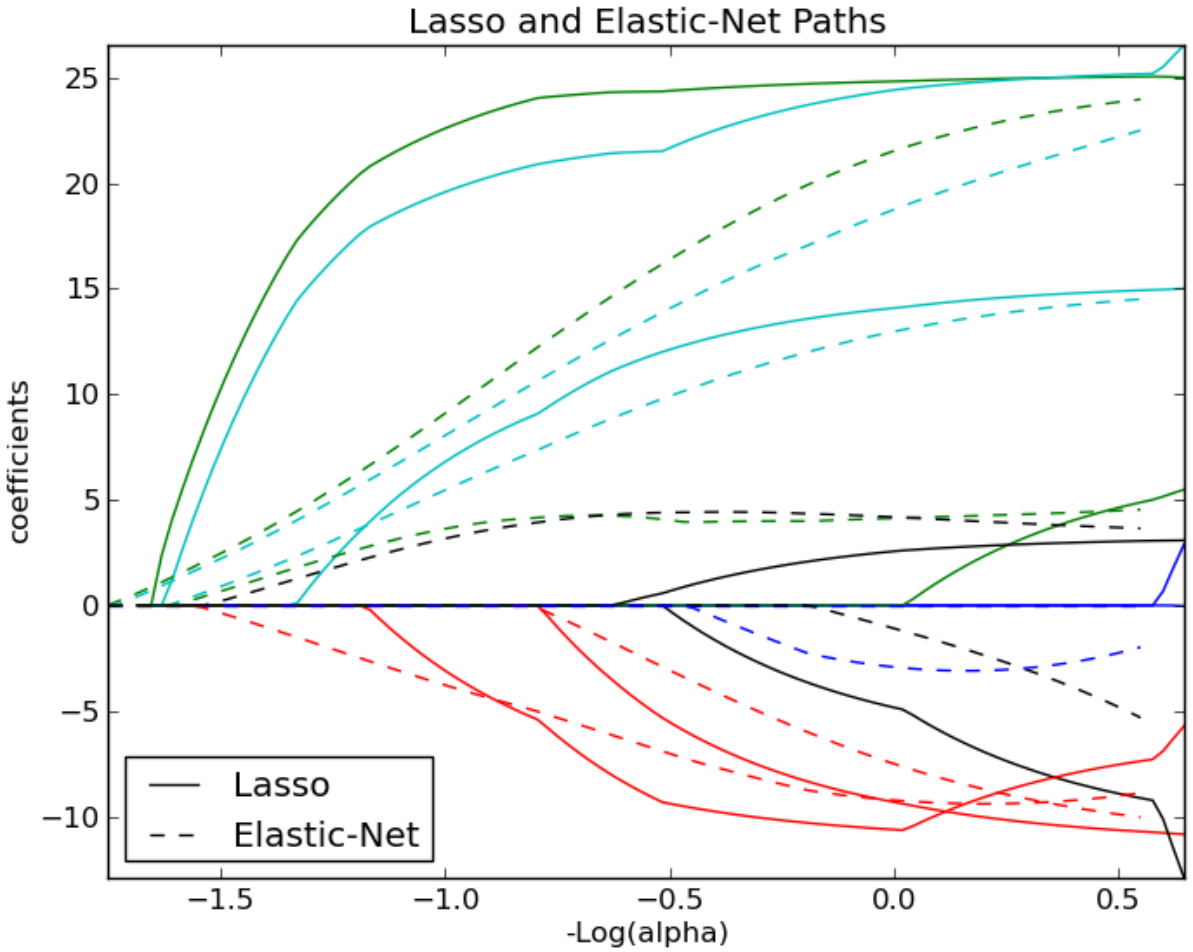
Estado del Arte: Regularización

- Introducción en el modelo restricciones para evitar el sobreajuste
- Penalizaciones a la complejidad del modelo
- Regresión lineal:
  - Lambda, factor de penalización sobre los coeficientes:
  - L1 (Lasso) (1996) R. Tibshirani.
  - L2 (Ridge) (1970) A. Tikhonov, A.E. Hoerl

Estado del Arte: Regularización



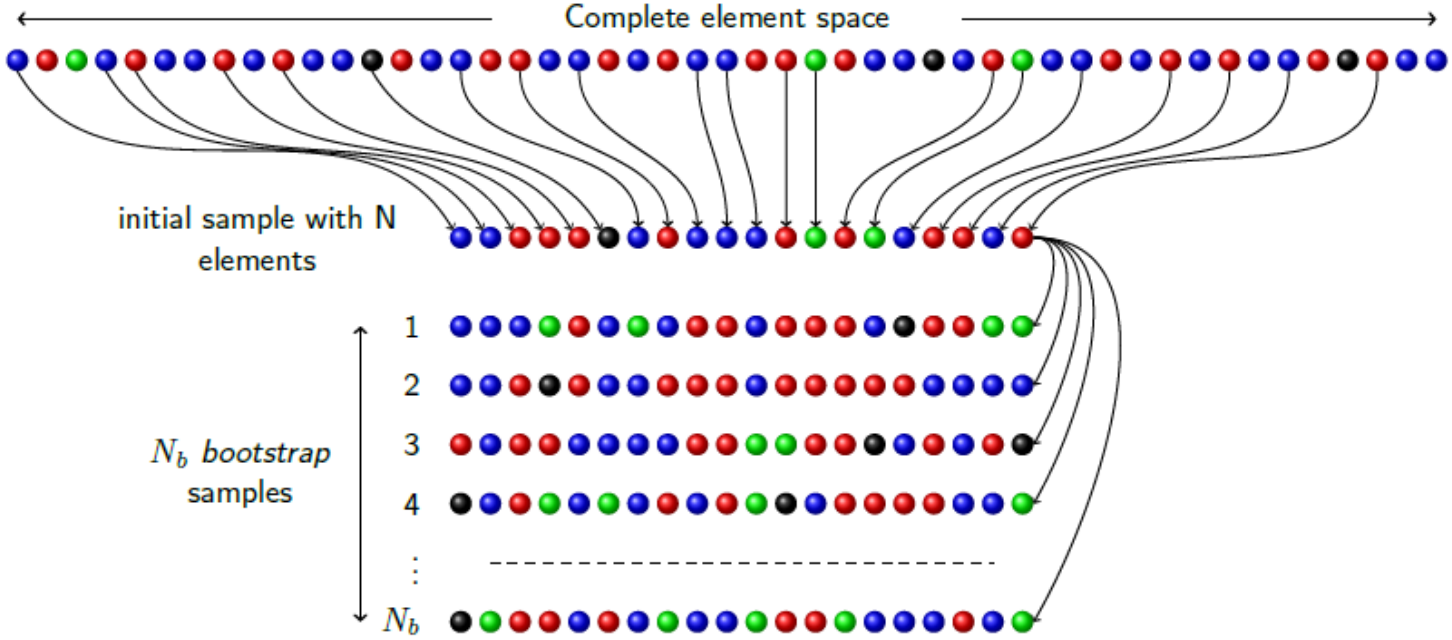
Estado del Arte: Regularización



Estado del Arte: Reducción dimensionalidad

- PCA (Análisis Componentes Principales)
- Stepwise
- Regularización Lasso
- Métodos de permutación
- Random Forest stepwise
- Muestreo de variables y ensamblado
- K-means

Estado del Arte: Bootstrapping (remuestreo)

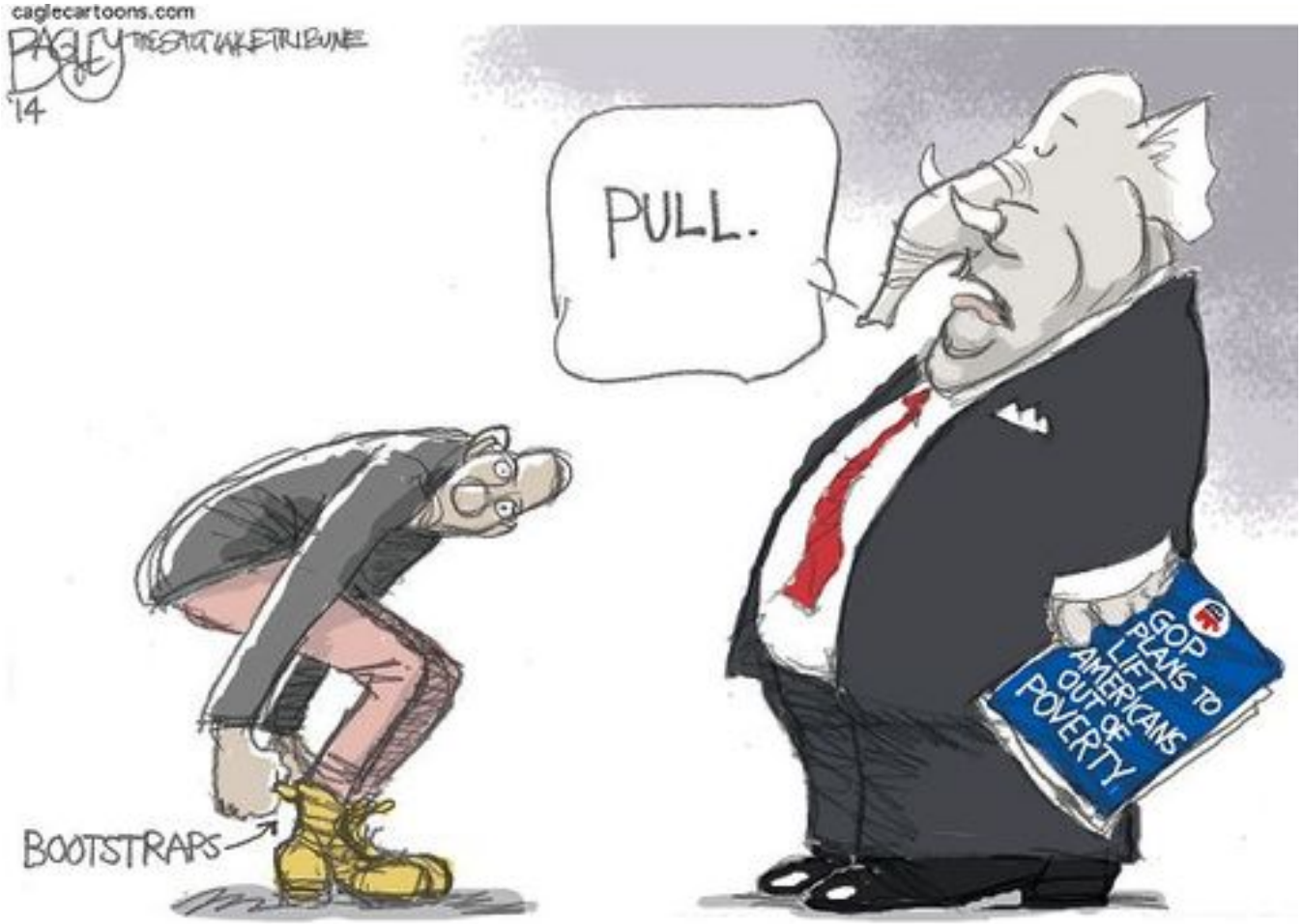


Theorem (B. Efron, Ann. Statist. 1979)

When N tend to infinity, the distribution of average values computed from bootstrap samples is equal to the distribution of average values obtained from ALL samples with N elements which can be constructed from the complete space. Thus the width of the distribution gives an evaluation of the sample quality.

Estado del Arte: Bootstrapping (remuestreo)

'To lift himself up by his bootstraps'



Estado del Arte: Técnicas Machine Learning

- **Según conocimiento que se tiene de la variable respuesta:**
- **Supervisados:**
  - **Conjunto de datos etiquetados (clasificación) o medidos (regresión) para el training**
- **No supervisados:**
  - **No necesario las clases a las que pertenecen las observaciones**
  - **Tampoco cuántas clases hay en total**
- **Semisupervisados:**
  - **Elevado coste de la clasificación manual del training set**
  - **Label propagation (asignación iterativa de etiquetas)**
  - **Manifolds (hipótesis sobre dimensionalidad espacio observaciones)**

Estado del Arte: Random Forest (2001)

Adele Cutler

Leo Breiman

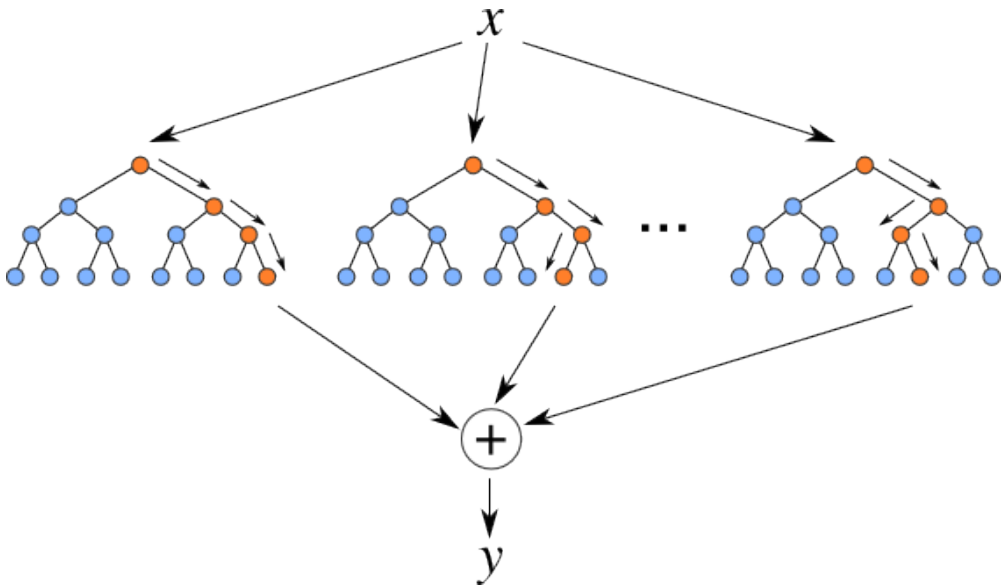
Phil Cutler

Estado del Arte: Random Forest (2001)

- Ensamblado de árboles construidos con bootstrapping
- Regresión o clasificación
- En cada nodo se selecciona un subconjunto de variables y para ellas se determina el mejor punto de corte
- No causan sobreajuste
- Detectan interacciones muy fractalizadas
- Pueden utilizarse para imputación de valores perdidos
- Puede inducirse una distancia entre casos

Estado del Arte: Random Forest (2001)

- OOB error es generalizable
- Importancia de variables por permutación
- Categorías son tratadas como numéricas
- Pueden agregarse
- Buen manejo outliers



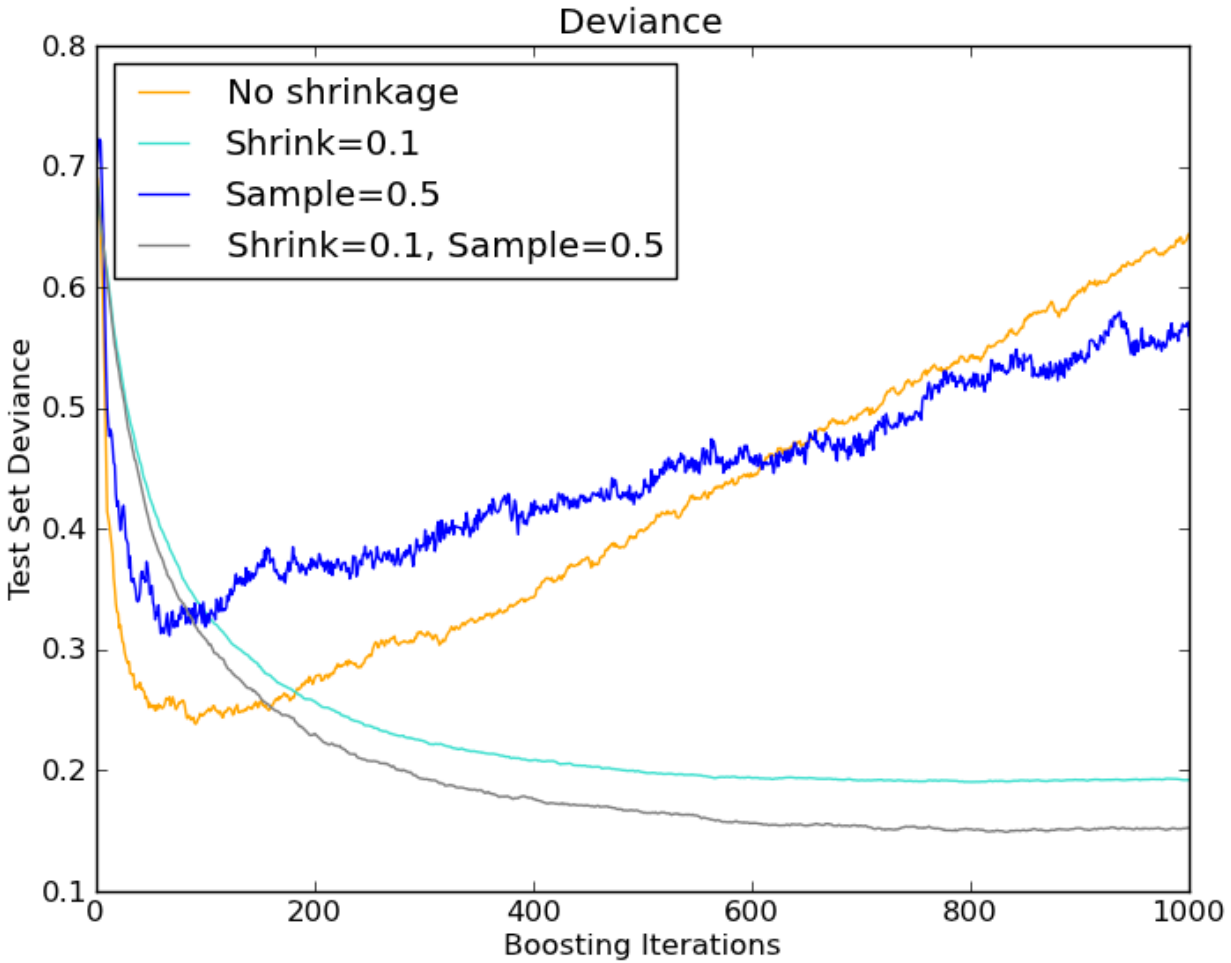
Estado del Arte: Gradient Boosting (1999)

- Jerome Friedman
- Modelo ensamblado de otros modelos de una forma iterativa
- Para cualquier función error diferenciable
- Se inicia el modelo con una constante
- Se calculan los pseudoresiduos: menos el gradiente de la función error en las estimaciones actuales.
- Se ajusta un modelo a los pseudoresiduos
- Se actualizan las estimaciones en la dirección de dicho modelo eligiendo un paso que minimice el error.

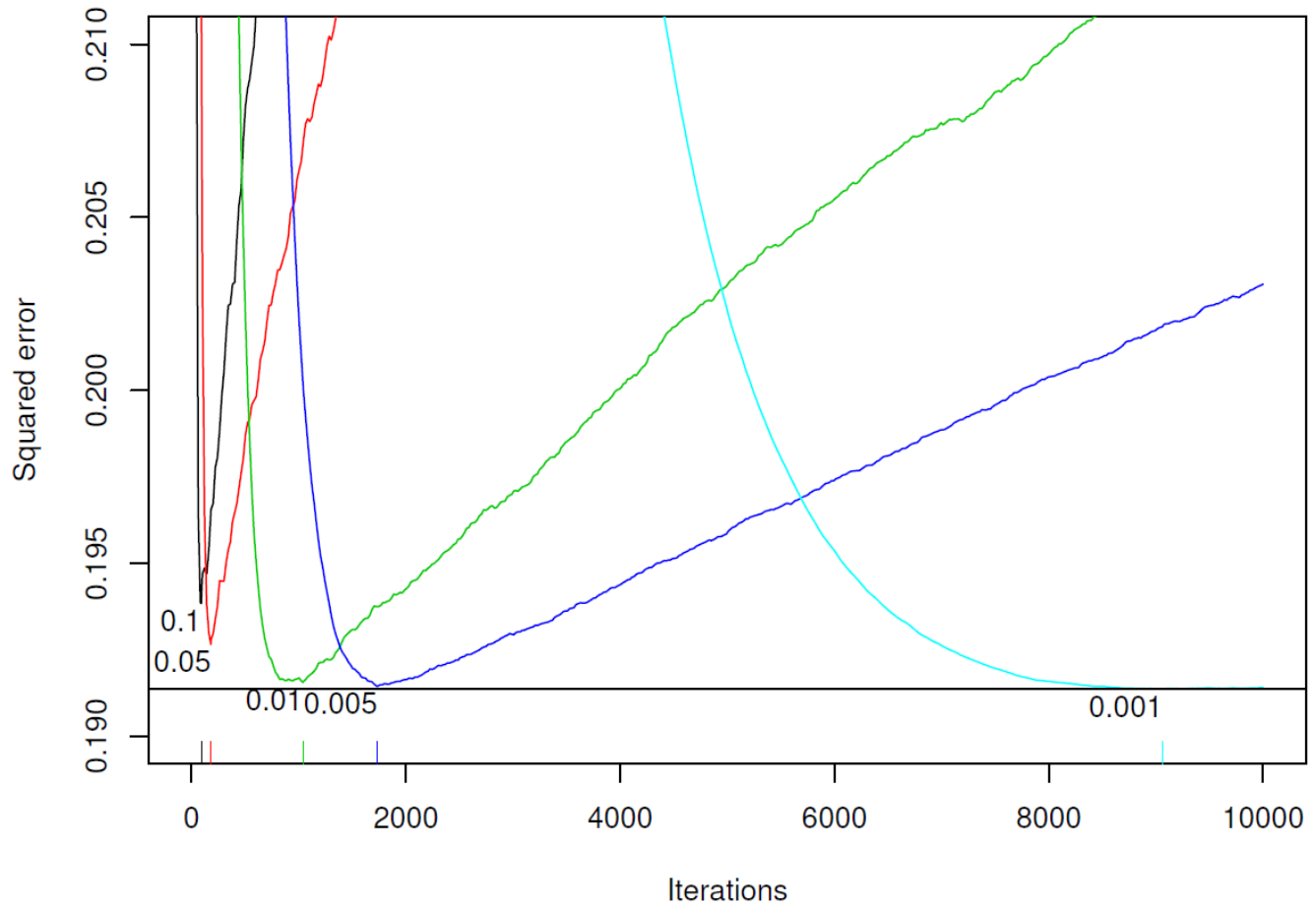
Estado del Arte: Gradient Boosting

- **Regularizaciones:**
  - **Interaction – depth (Número de niveles del árbol)**
  - **Mínimo número de observaciones en nodos finales**
  - **Shrinkage: constante que limita cada paso en la dirección del gradiente**
  - **Bag fraction: Fracción de submuestreo.**
  - **Criterio de parada por cross-validación**

Estado del Arte: Gradient Boosting (gbm)

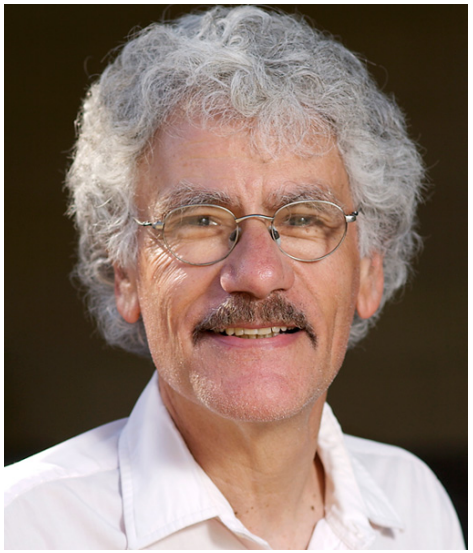


Estado del Arte: Gradient Boosting (gbm)



Estado del Arte: Gradient Boosting

STANFORD UNIVERSITY



Jerome Friedman

Trevor Hastie

Robert Tibshirani

Estado del Arte: Gradient Boosting

OFFICE OF JUSTICE PROGRAMS

ABOUT NIJ | WHAT'S NEW | ALL TOPIC

NATIONAL INSTITUTE OF JUSTICE

Research • Development • Evaluation

Advanced

NIJ HOME PAGE | FUNDING & AWARDS | EVENTS | PUBLICATIONS & MULTIMEDIA | TRAINING

NIJ Home Page > About NIJ > Director of NIJ

THE NIJ DIRECTOR

- ABOUT ACTING DIRECTOR DR. GREGORY K. RIDGEWAY
- SPEECHES AND REMARKS

About the NIJ Director

The Director is appointed by the President to lead the National Institute of Justice and establish the agency's objectives, guided by the needs of the field and the priorities of the U.S. Department of Justice.

Upon the departure of Director John H. Laub, Deputy Director Greg Ridgeway became Acting Director. (Read Dr. Laub's farewell message.)

On this page you can follow NIJ Directors through their speeches and commentary.

Posted February 12, 2013

Statement on NIJ's Role in the National Dialogue on Gun Violence

Our entire country is talking about gun violence. The recent spate of mass gun

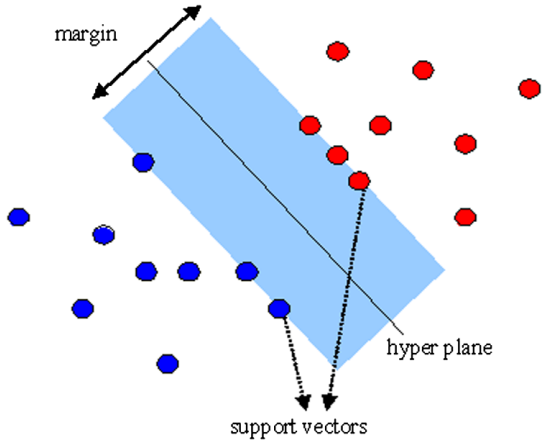
Estado del Arte: Support Vector Machine (SVM)

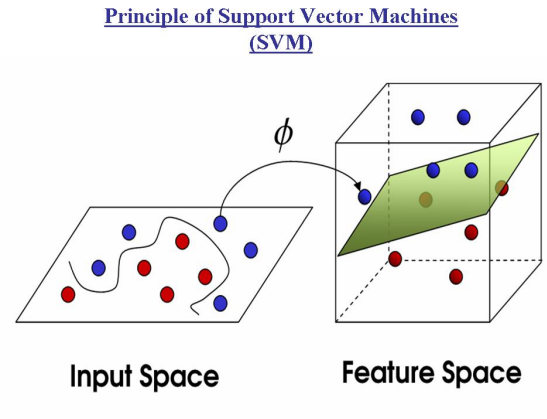
Corinna Cortes

“For 2,000 years, we believed logic was the only instrument for solving intellectual problems. Now, our analysis of machine learning is showing us that to address truly complex problems, we need images, poetry, and metaphors as well”

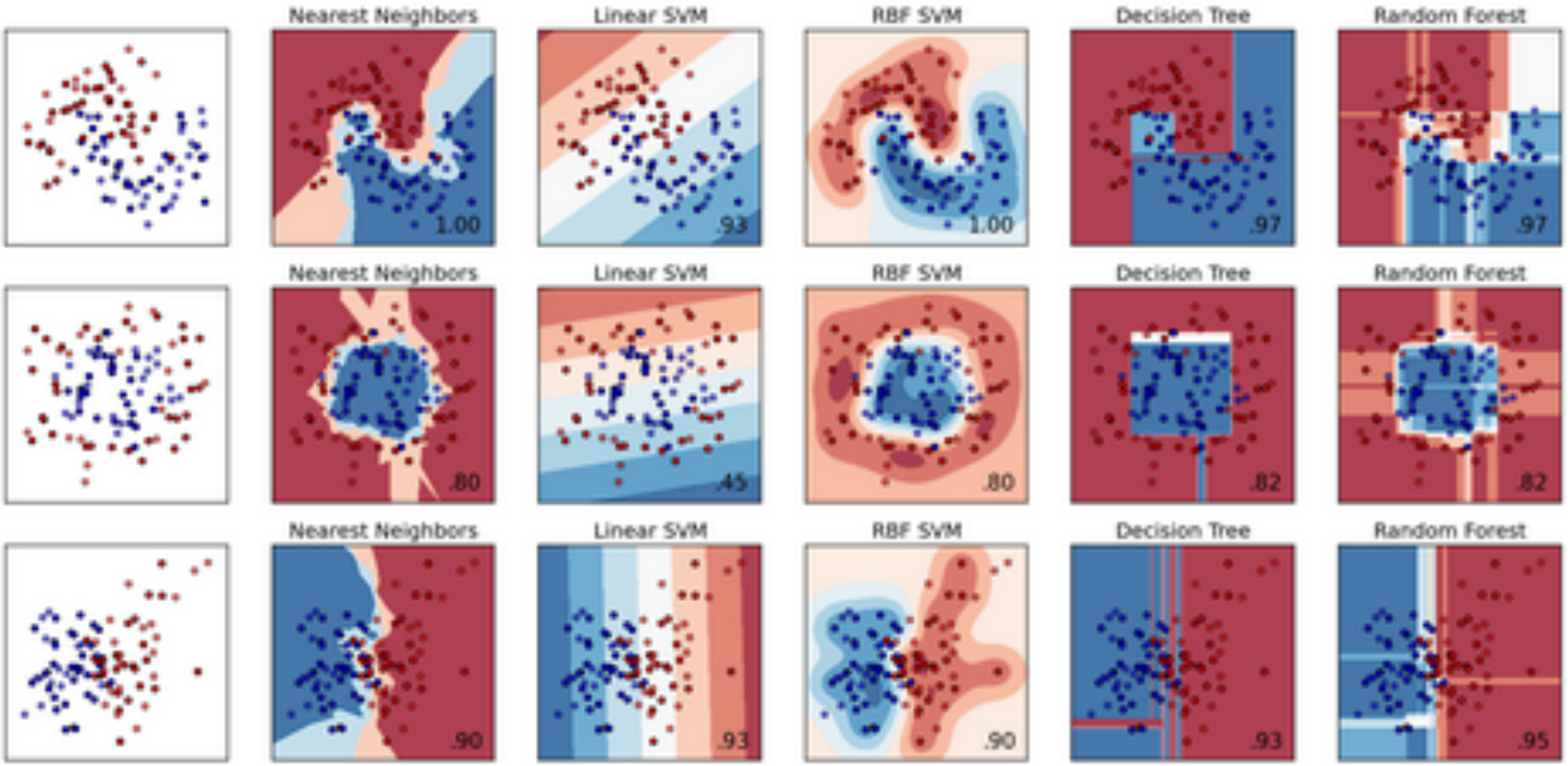
Estado del Arte: Support Vector Machine (SVM)

- Hiperplanos con máximo margen
- C: penalización para los casos mal clasificados (Soft margin)
- Mapeo de casos a una dimensión mayor para facilitar la separación: Kernels
- Kernels: lineal, gauss, polinomial
- Extensiones para regresión, clasificación multinomial, ranking





Estado del Arte: Support Vector Machine (SVM)



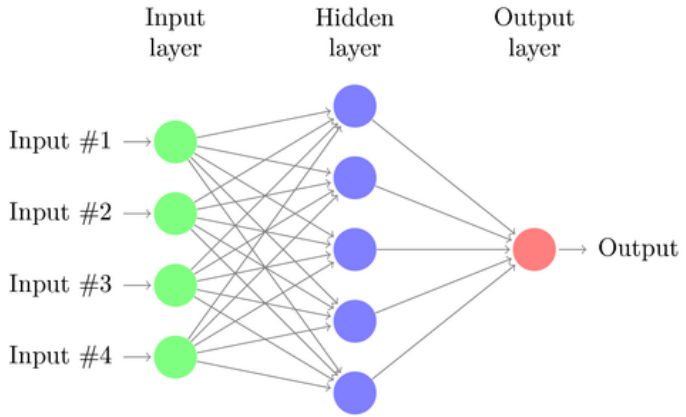
Estado del Arte: Support Vector Machine (SVM)

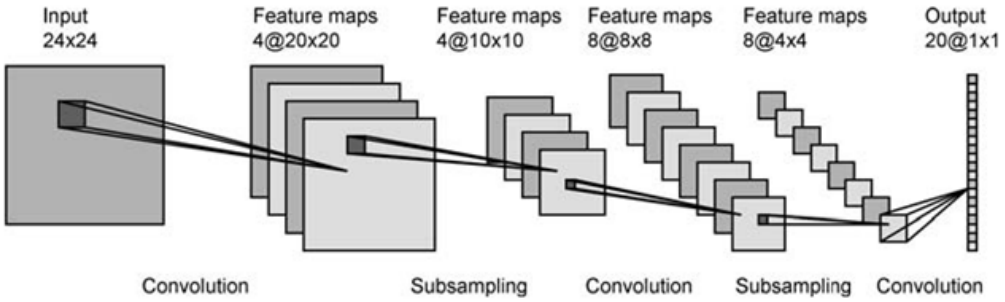
Chih-Jen Lin
LibSVM
LibLinear

Thorsten Joachim
SVMlight

Estado del Arte: Redes neuronales (NN)

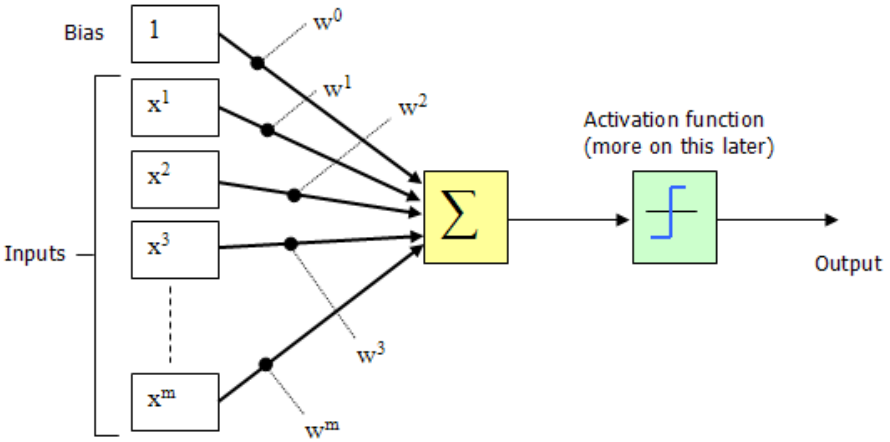
- **Optimización multietapa (Back propagation)**
- **Teorema de aproximación universal: Una capa es suficiente para aproximar cualquier función continua en un conjunto compacto bajo ciertas condiciones de la función de activación.**
- **MLP: Multilayer Perceptron**
- **CNN: Convolution Neural Network**

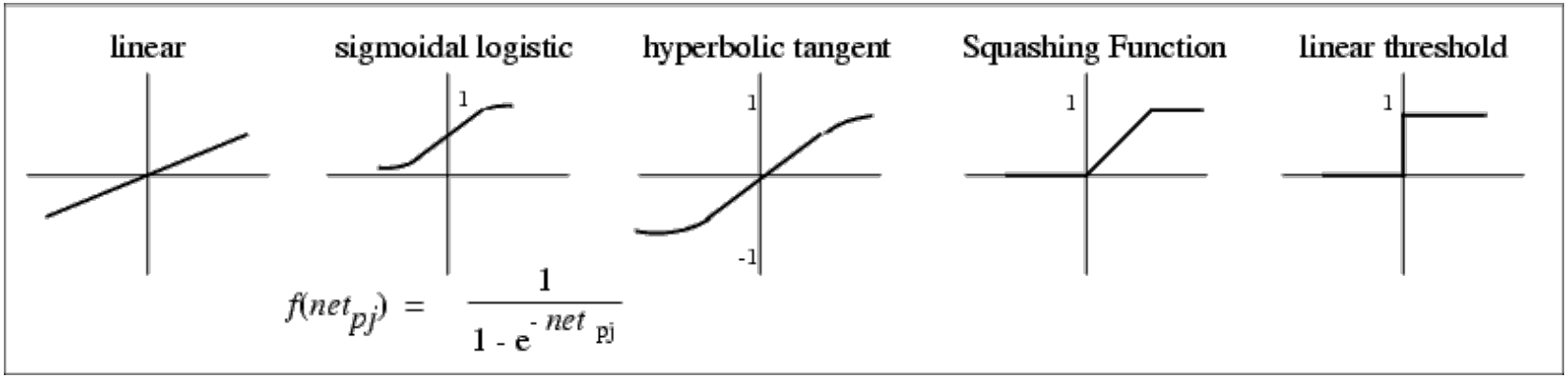




Estado del Arte: Redes neuronales (NN)

- **Funciones de activación y de transferencia**





Estado del Arte: Redes neuronales (NN)

UNIVERSITY OF TORONTO

Université de Montréal

Geoffrey Hinton

Ian Goodfellow (pylearn2)

Yoshua Bengio

Experiencias

kaggle

Customer Solutions | Competitions | Community ▾ | Sign Up | Login

José A. Guerrero

Data Scientist

http://www.kressala.com/

MASTER ?

Highest	Current
1st	3rd /176,495

579,416.8 points
Joined 3 years ago

Profile | Results | Forum

1ST	1ST	2ND	3RD	5TH	5TH	6TH	9TH	
1st/532	1st/146	2nd/50	3rd/1353	5th/132	5th/12	6th/337	9th/362	25 Competitions

Experiencias

kaggle | Customer Solutions | Competitions | Community ▾ | José A. Guerrero | Logout

Traveling Santa Problem

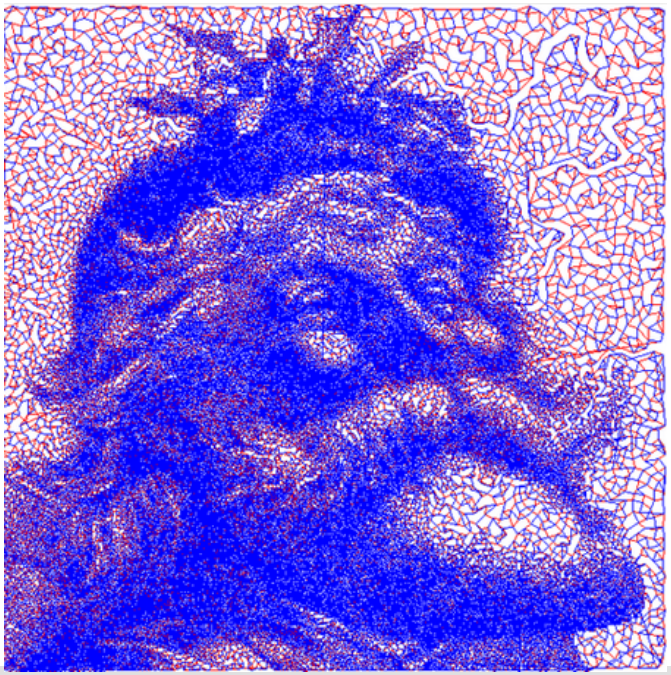
Finished

Friday, December 14, 2012 — $3,000 • 355 teams — Saturday, January 19, 2013

150.000 chimeneas

Leaderboard

1. Farmers peeing further
2. wleite
3. Rudolph
4. Olexandr Topchylo
5. TravelingSasquatch
6. kc
7. Leustagos
8. Helsgaun
9. Titericz & Santooma
10. 4259018372



Experiencias

kaggle | Customer Solutions | Competitions | Community ▾ | José A. Guerrero | Logout

Personalize Expedia Hotel Searches - ICDM 2013

Finished

Tuesday, September 3, 2013 — $25,000 • 337 teams — Monday, November 4, 2013

Dashboard

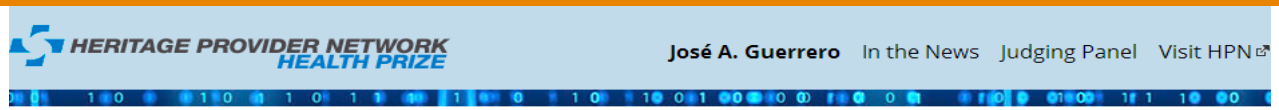
Leaderboard - Personalize Expedia Hotel Searches - ICDM 2013

This competition has completed. This leaderboard reflects the final standings.

See someone using multiple accounts? Let us know.

#	Δ1w	Team Name * in the money	Score	Entries	Last Submission UTC (Best – Last Submission)
1	-	commendo - part of Opera Solutions * • Michael Jahrer • AndreasToescher	0.54075	55	Mon, 04 Nov 2013 22:41:11
2	-	Owen *	0.53984	35	Mon, 04 Nov 2013 22:35:38
3	↑4	Jun Wang@UniGe *	0.53839	55	Mon, 04 Nov 2013 23:12:44 (-47.1h)
4	↓1	idle_speculation *	0.53366	46	Mon, 04 Nov 2013 14:20:46 (-14.6h)
5	↑1	binghsu & MLRush & BrickMover	0.53102	101	Mon, 04 Nov 2013 22:13:23
6	↓2	**J.A. Guerrero**	**0.53069**	**48**	**Mon, 04 Nov 2013 22:16:26**

Experiencias



Dashboard

Private Leaderboard - Heritage Health Prize

This competition has completed. This leaderboard reflects the final standings.

See someone using multiple accounts?
Let us know.

#	Δ1w	Team Name * in the money	Score	Entries	Last Submission UTC (Best – Last Submission)
1	-	POWERDOT *	0.461197	671	Thu, 04 Apr 2013 05:12:00 (-12.3d)
2	↑60	EXL Analytics	0.462247	555	Thu, 04 Apr 2013 00:06:09 (-3.4d)
3	↑15	**J.A. Guerrero**	**0.462417**	**173**	**Thu, 04 Apr 2013 06:03:09**
4	↑10	PANDA	0.462644	63	Thu, 04 Apr 2013 01:44:15 (-23.9h)
5	↓3	CombinedPower	0.463052	286	Thu, 04 Apr 2013 06:15:57 (-173.5d)
6	↓2	Xing Zhao	0.463125	161	Thu, 04 Apr 2013 00:13:08 (-5d)
7	↓4	Ambrosia	0.463133	134	Mon, 10 Dec 2012 20:34:10 (-2.5d)
8	↓3	Areté Associates	0.463526	112	Sun, 31 Mar 2013 20:54:52 (-226.8d)
9	↓1	Analytics Inside	0.464170	162	Thu, 04 Apr 2013 00:00:31
10	↓4	Old Dogs With New Tricks	0.464452	370	Wed, 03 Apr 2013 20:04:56 (-4.7d)
11	↓4	luscus	0.464457	52	Fri, 15 Mar 2013 05:00:27 (-12.2d)
12	↓3	J Analysis	0.464542	165	Thu, 24 Jan 2013 01:22:13 (-37.9d)
13	↓3	Dolphin	0.464684	555	Thu, 04 Apr 2013 05:37:08 (-22.4d)
14	↓3	Hopkins Biostat	0.464803	444	Thu, 04 Apr 2013 03:06:25 (-7.6d)

Experiencias

kaggle

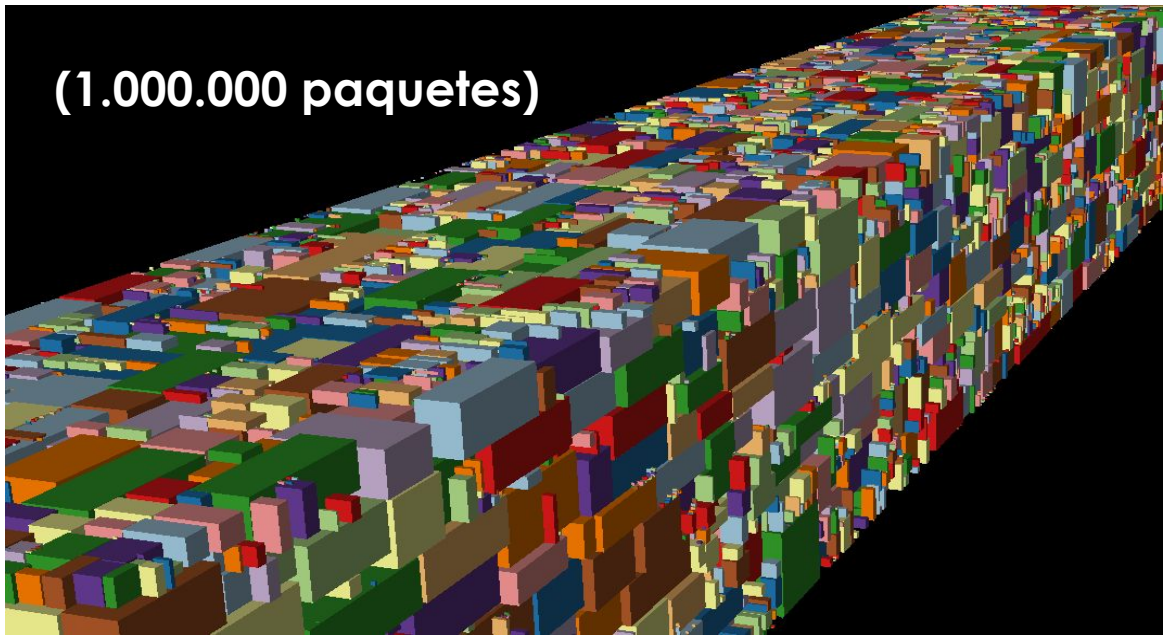
Customer Solutions | Competitions | Community ▾ | José A. Guerrero | Logout

MathWorks

Packing Santa's Sleigh

Finished

Monday, December 2, 2013 — $10,000 • 362 teams — Sunday, January 26, 2014



Experiencias

Experiencias

kaggle | Customer Solutions | Competitions | Community ▾ | José A. Guerrero | Logout

CHALEARN Gesture Challenge 2

Finished

Tuesday, May 8, 2012 — $10,000 • 31 teams — Tuesday, September 11, 2012

Dashboard

Leaderboard - CHALEARN Gesture Challenge 2

This competition has completed. This leaderboard reflects the final standings.

See someone using multiple accounts? Let us know.

#	Δ1w	Team Name * in the money	Score	Entries	Last Submission UTC (Best – Last Submission)
1	↑20	alfnie *	0.07099	7	Sat, 08 Sep 2012 05:09:02
2	↑8	Joewan *	0.14476	58	Mon, 10 Sep 2012 02:48:33
3	-	Wayne Zhang *	0.16084	16	Sun, 09 Sep 2012 10:27:54
4	↑16	Manavender	0.19246	12	Sun, 09 Sep 2012 00:37:42
5	↑21	HIT_CS	0.20078	11	Tue, 11 Sep 2012 14:09:02 (-16.4h)
6	↑1	vigilant	0.22352	9	Mon, 10 Sep 2012 00:21:25
7	new	Mikalai Drabovich	0.31115	12	Tue, 11 Sep 2012 17:55:45
		Principal Motion benchmark	0.31725		
8	↓7	JNB Concepts	0.90682	41	Tue, 11 Sep 2012 23:01:11 (-95d)
		Winner round 1	1.00000		

Cursos y documentación

Especializaciones

www.coursera.org

Universidad Johns Hopkins

Data Science

Siguiente sesión:
may. 5º 2014

Cursos

Universidad Johns Hopkins

Aprendizaje automático en la práctica

con Jeff Leek, Brian Caffo & Roger D. Peng

jun. 2º 2014
4 weeks de duración
Especializaciones

Universidad Stanford

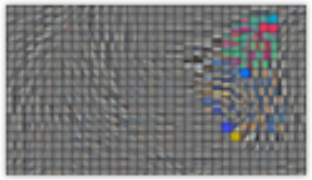
Aprendizaje automático

con Andrew Ng

jun. 16º 2014
10 weeks de duración

Cursos y documentación

www.coursera.org

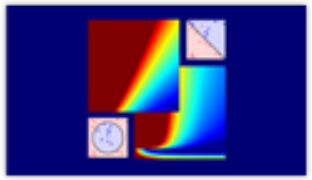


Universidad de Toronto
Las redes neuronales y el aprendizaje automático
con Geoffrey Hinton

oct. 1º 2012
8 weeks de duración

Universidad Stanford
Modelos en grafo de probabilidades
con Daphne Koller

abr. 8º 2013
11 weeks de duración



Universidad Nacional de Taiwán
機器學習基石 (Fundamentos del aprendizaje automático)
con Hsuan-Tien Lin

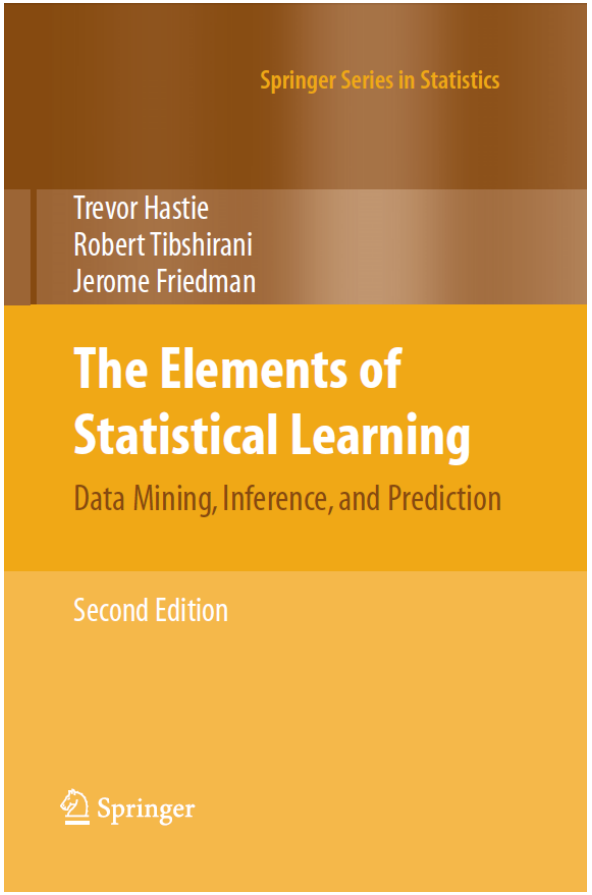
nov. 26º 2013
8 weeks de duración

Universidad de Washington
Aprendizaje automático
con Pedro Domingos

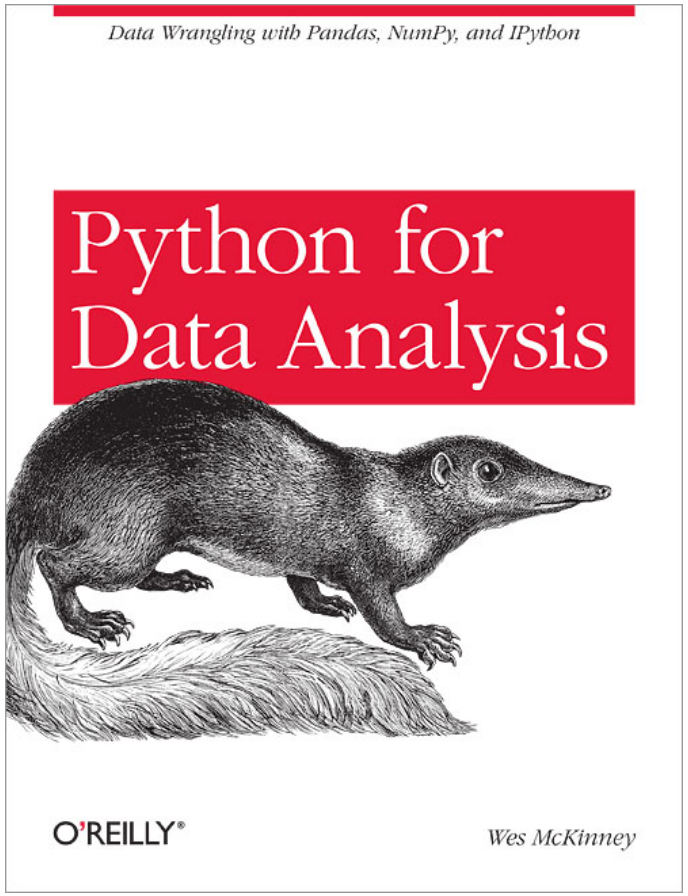
No hay sesiones abiertas.

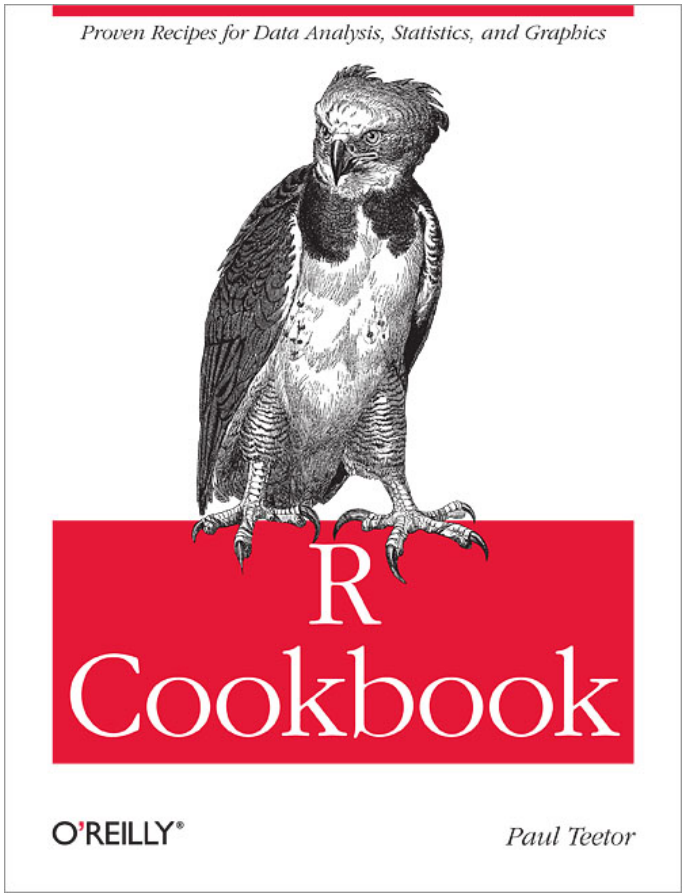
Cursos y documentación

http://statweb.stanford.edu/~tibs/ElemStatLearn/download.html

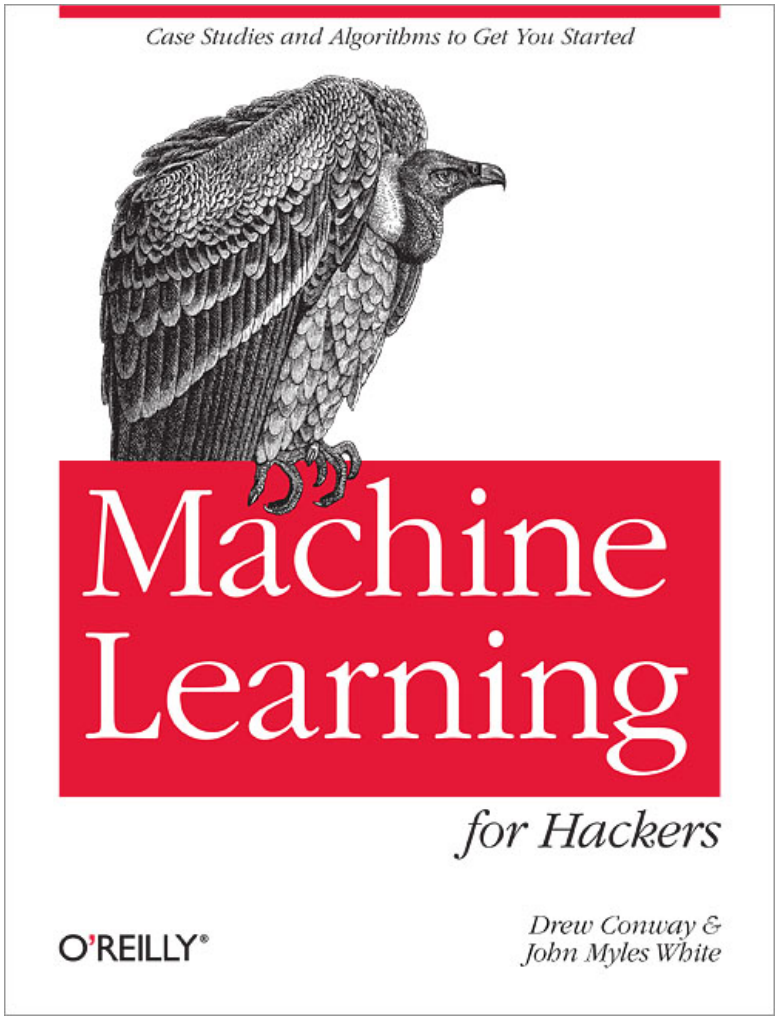


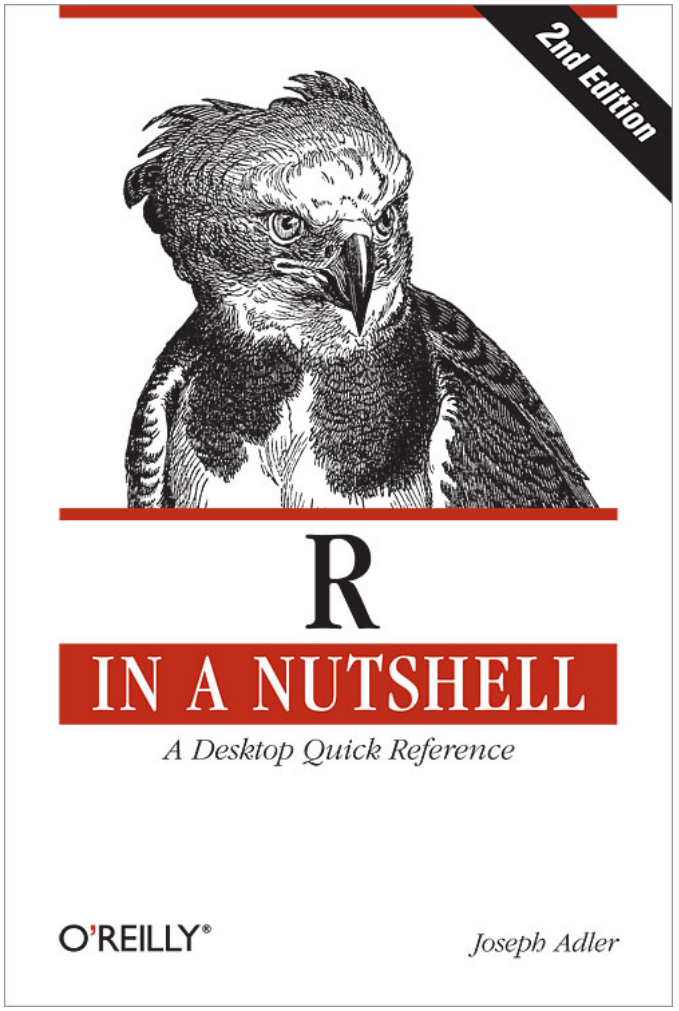
Cursos y documentación





Cursos y documentación





Contacto

José Antonio Guerrero

jaguerrerod@ono.com