

机器学习与数据挖掘

潘 炎

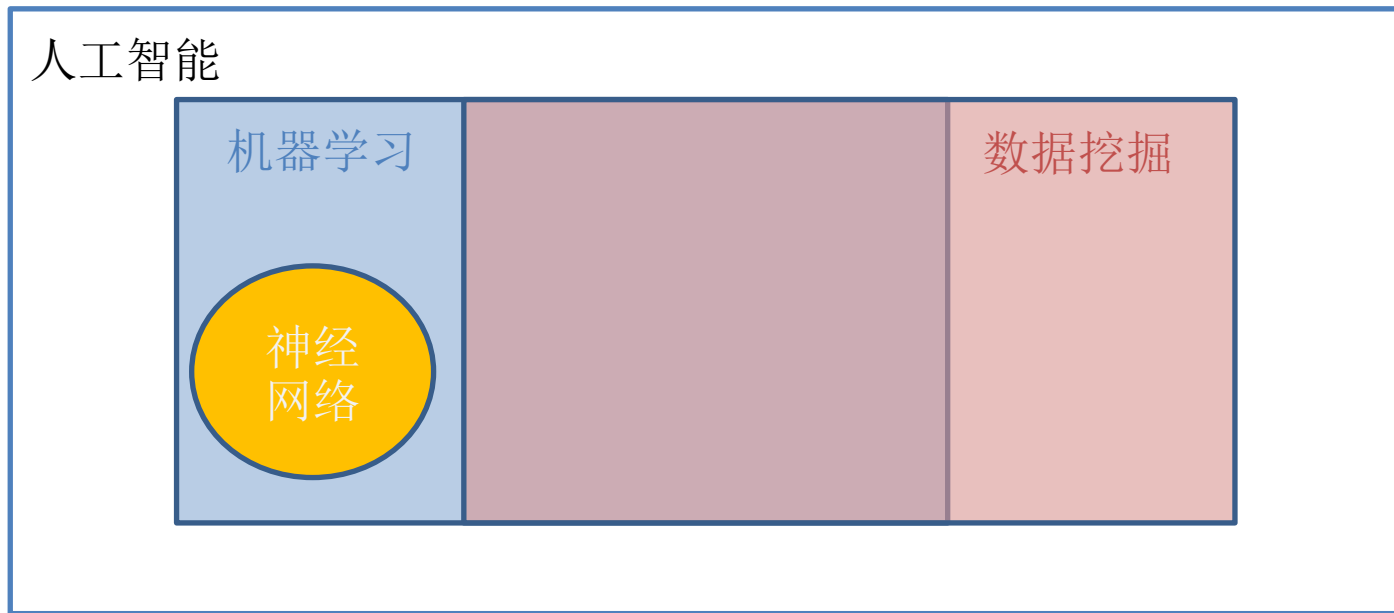
中山大学数据科学与计算机学院

panyan5@mail.sysu.edu.cn

The speaker

- 研究兴趣：机器学习
 - Deep learning for information retrieval
 - Learning to hash （图片搜索）
 - Ranking and recommendation （广告排序，个性化推荐）
 - Machine learning for quantitative trading

人工智能、机器学习、数据挖掘、神经网络



About this course

- **Practical** machine learning and data mining techniques

- 基本概念，核心工具

- 学习系统的框架，优化方法，分类，回归，聚类，关联规则，链接分析

- 主流技术，典型实际问题的解决方法

- 深度神经网络，Boosting/随机森林+决策树
 - 并行化机器学习算法（大数据下的机器学习）
 - 应用：搜索引擎、推荐系统、图像搜索，量化交易等
 - Kaggle, PASCAL VOC, 腾讯社交广告预测比赛，百度快速人像识别比赛经验分享

Machine Learning Part

- Optimization methods
 - Gradient descent
 - Stochastic gradient descent
 - Coordinate descent
- Regression
 - Linear regression
- Classification
 - Logistic Regression (Parallelization)
 - Boosting + decision trees, Random Forest (Parallelization)
 - Deep Neural Networks
- Clustering
 - K-means
- Recommendation Systems
 - matrix factorization / field-aware factorization machines

Data Mining Part

- Association Rules
- Similarity Search
 - Learning to hash
- Link Analysis
 - Page Rank

Useful Links

- Stanford machine learning course
 - <http://www.ml-class.org/>
 - <http://cs231n.stanford.edu/>
 - <http://cs229.stanford.edu/materials.html>
- Stanford data mining course
 - <http://infolab.stanford.edu/~ullman/mmds.html>
- Machine learning competitions
 - <http://www.kaggle.com>

Textbooks

- There is no need to buy a textbook.
- Recommended online textbooks:
 - The Elements of Statistical Learning
<http://infolab.stanford.edu/~ullman/mmds.html>
 - Pattern Recognition and Machine Learning
 - Mining of massive datasets
<http://infolab.stanford.edu/~ullman/mmds.html>

Other Books



机器学习实战 2013-06
哈林顿 (Peter Harrington)、李锐



机器学习 2016-01
周志华



凸优化(英文) 2013-10
鲍迪 (Stephen Boyd)

About the exam

- 3次个人作业（Kaggle比赛）
 - 线性回归
 - 非线性分类（GBDT或随机森林）
 - 非线性分类（深度学习）
- 1个实验报告

软件的现状

- 过去十多年，我们开发的软件系统主要是收集数据
 - ERP系统=单据电子化
 - 邮件系统=信件电子化
 - 企业运作产生大量的数据
 - 互联网的网页/图片/视频数据
 - 互联网/移动应用中的用户行为数据
 - 金融行情/交易数据，各种经济数据，上市公司财务数据
 - 监控摄像头（小区，交通，商铺）或其他传感器产生的数据
 - 医疗数据
 -

软件的未来

- 从收集数据到分析数据
- 市场上需要“聪明”的软件
 - www.kaggle.com
 - 能否给我推荐我喜欢的电影/衣服/新闻/游戏/朋友？
 - 汽车/无人机能否自动驾驶？工业机器人能否代替工人？
 - 机器能否理解图片内容、视频内容、自然语言？
 - 机器能否代替人工决策？
 - ...

Data is money?

- Lots of data is being collected and warehoused
 - Web data, e-commerce
 - purchases at supermarket/grocery stores
 - Bank/Credit Card transactions
- Computers have become cheaper and more powerful
- Competitive Pressure is Strong
 - Provide better, customized services for an edge (e.g. in Customer Relationship Management)





HERITAGE PROVIDER NETWORK HEALTH PRIZE

- 在美国，每年有超过7100万病人入院。据06年的一项调查表明，每年有超过300亿美金花在不必要的入院治疗上
- 我们能否预测出高风险的可能要入院治疗的人群，并且保证他们得到适当而必要的治疗？同时，预测出可能的“骗保”行为。
- HPN（美国的医疗保障机构）在kaggle上赞助了一个奖金为300万美金的比赛，希望参赛者开发一个算法，通过分析病人使用医疗保险的历史数据来预测和避免不必要的入院治疗（骗保险）
- 具体地，参赛者需要设计一个算法来预测每个（买保险的）病人在下一年会住院多少天。有了这个预测信息，医院可以为急诊病人提供合适的治疗计划和安排，从而可以减少不必要的入院自疗，降低成本。

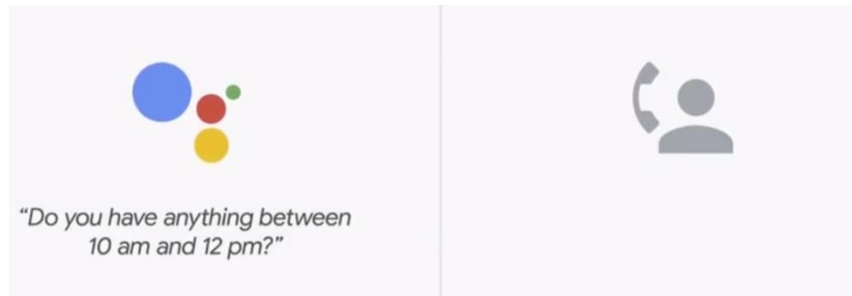
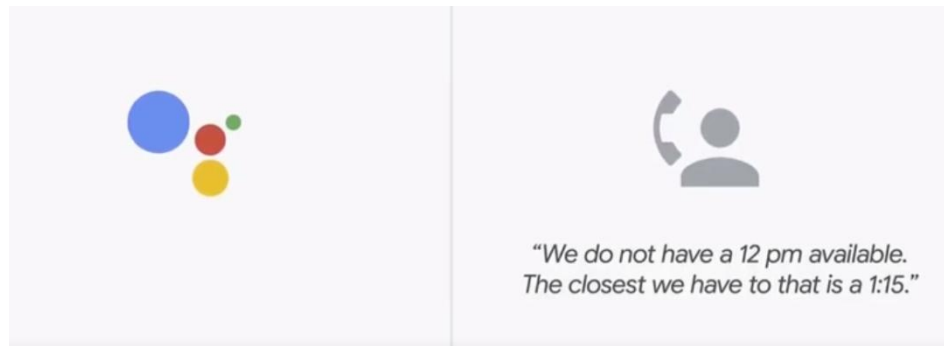
微软的“同声翻译”

- 微软研究院2013年发布了一段很有意思的视频，视频中，微软首席研究官Rick Rashid演示了一项关于语音识别和机器翻译方面的重大进展：计算机首先识别和理解他用英语说的话，同时（延时大概3-5秒）将这些英语立即翻译成普通话，并用自己的语音语调说出，听起来像自己说的一样。**该技术在2014年成功运用于skype上。**



Google的语音机器人助手

- 像真人一样帮您打电话订酒店，流畅地进行复杂的对话。
- 相关技术：基于深度神经网络的语音识别、自然语言理解、语音合成、对话生成等等。



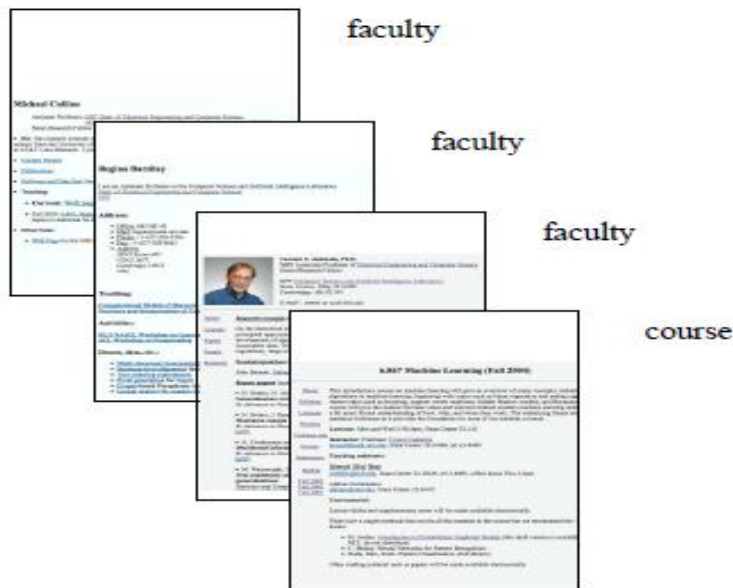
AlphaGo击败人类顶尖棋手

- 2016年3月，Google旗下的Deepmind公司开发的程序AlphaGo以4：1击败韩国顶尖棋手李世石。
- AlphaGo关键技术：蒙特卡罗树搜索（MCTS），深度卷积神经网络。
 - 走棋网络（Policy Network），给定当前局面，预测/采样下一步的走棋。
 - 快速走子（Fast rollout），目标和1一样，但在适当牺牲走棋质量的条件下，速度要比1快1000倍。
 - 估值网络（Value Network），给定当前局面，估计是白胜还是黑胜。
 - 蒙特卡罗树搜索（Monte Carlo Tree Search, MCTS），把以上这三个部分连起来，形成一个完整的系统。
- 升级版AlphaGo：利用强化学习，通过自我对弈进行学习，不再需要利用棋谱。



例子：网页分类

- Example problem: text/document classification



财经
科技
娱乐
军事
社会
体育
汽车

.....

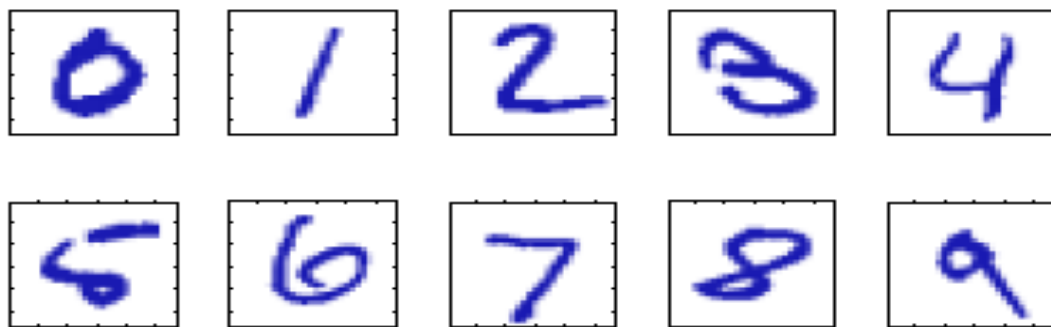
- a few labeled training documents (webpages)
- goal to label yet unseen documents

例子：垃圾邮件过滤

		From / To	Date Sent	Thread
2949		← Support	10/02/09 7:45 +0...	Message from eBay com au
2950		← Ken Johnston	10/02/09 14:12 +...	Fool them once, fool them twice, fool...
2951		→ christfried.web...	10/02/09 3:14 -0...	Assistance, Petersen
2952		→ Air Sep	9/02/09 4:53 -0800	Un negocio de por vida 1000% Renta...
2953		→ Oslia John	10/02/09 17:33 +...	Now Contact my secretary ask him fo...
2954		← Air Sep	9/02/09 0:38 -0800	Un negocio de por vida 1000% Renta...
2955		→ Air Sep	9/02/09 10:12 -0...	Un negocio de por vida 1000% Renta...
2956		→ MISS MERCY...	29/01/09 23:13 -...	Urgent Attention(YOUR FILE HAVE...
2957		→ PEPSI BOTTL...	25/07/08 11:23 -...	OEP00934/UK
2958		→ JOSEPH POON	11/02/09 12:04 +...	MY PROPOSAL!!!
2959		→ MADAMERL...	11/02/09 13:41 +...	LOOKING FOR A TRUSTWORTHY...
2960		→ REBECA RO...	11/02/09 18:48 +...	Dear sir/madam:
2961		→ REBECA RO...	11/02/09 18:48 +...	Dear sir/madam:
2962		→ Elinor Shannon	11/02/09 22:37 +...	I shall look forward to hearing from you
2963		→ Air Sep	10/02/09 14:57 +...	Un negocio de por vida 1000% Renta...
2964		→ Foreign Payme...	1/02/09 16:13 +0...	Goodday,
2965		→ JANET KUEN	12/02/09 16:11 +...	Dear sir/madam:
2966		→ Abubakar Mar...	10/02/09 19:04 +...	OUR DEAR FRIEND
2967		→ JAMES ROBE...	12/02/09 23:12 -...	From James Roberts
2968		→ Bases de Email...	13/02/09 10:50 -...	Nuevas Bases de Datos de Mexico
2969		→ Barrister Willi...	15/02/09 1:23 +0...	WILL AND TESTAMENT
2970		→ Isolde	15/02/09 9:45 -0...	A Valentine's Day Ecard Special Deli...
2971		→ NTI eNews	15/02/09 12:25 -...	Super Sweet Deals From NTIus.com

- Given some examples what the user defined as junk mail.
- From these examples, learn to identify new incoming junk mail.

例子：手写识别



- Given handwritten ZIP codes on letters, money amounts on cheques etc.
- From these examples, learn to recognise the correct digit written by hand.

例子：人脸识别

- Example problem: face recognition



Training data: a collection of images and labels (names)




Evaluation criterion: correct labeling of new images

例子：搜索引擎结果排序

- Given a number of web sites which match some search phrase:
Learn which pages most of the users are looking for.


Web [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more ▾](#) [Sign in](#)

 [Advanced Search](#) [Preferences](#)

Web Results 1 - 10 of about 5,870,000 for [Introduction to Statistical Machine Learning](#). (0.22 seconds)

[Introduction to Statistical Machine Learning](#)
15 May 2008 ... The other speakers will detail or built upon this introduction. **Statistical machine learning** is concerned with the development of algorithms ...
[videolectures.net/mlss08au_hutter_ismil/](#) - 75k - [Cached](#) - [Similar pages](#)

[Statistical Machine Learning \(SML\) Group, NICTA](#)
This course provides a broad but thorough introduction to the methods and practice of **statistical machine learning**. Topics covered will include Bayesian ...
[sml.nicta.com.au/ismil08.html](#) - 24k - [Cached](#) - [Similar pages](#)

 [An Introduction to Statistical Machine Learning - Introduction -](#)
File Format: PDF/Adobe Acrobat - [View as HTML](#)
Statistical Machine Learning. - Introduction -. Samy Bengio. [bengio@idiap.ch](#). Dalle Molle Institute for Perceptual Artificial Intelligence (IDIAP) ...
[bengio.abracadoudou.com/lectures/old/tex_intro.pdf](#) - [Similar pages](#)

例子：机器翻译

简体中文翻译

[2010年上海世博会中国](#)

作者：世界文化盛会。“2010年上海世博会，”在Decemer 2004年推出，是世博会唯一官方杂志2010年上海中国。更多>>...

[en.expo2010.cn/](#) - 44k - [网页快照](#)

[亭子的Participants Participation Expo中国2010年上海](#)

城市狂想曲：对上海世博会的网站预览 ▪ 世界博览会中国2010年上海推广Enblem拖车 ▪ 世界博览会中国2010年上海...

[en.expo2010.cn/participation/pop/moren.htm](#) - 23k - [网页快照](#)

[2010年世博会- 维基百科，自由的百科全书](#)

上海世博会，也是最大的世界博览会，网址为5.28永远...胡锦涛主席发起成立了上海世博会开幕。...

[en.wikipedia.org/wiki/Expo_2010](#) - 129k - [网页快照](#)

[上海世博会](#)

世界博览会是人类灵感和思想库。自1851年时，万国工业大展览在伦敦举行，...

[www.expo2010.cn/expo/expoenglish/oe/index.html](#) - 25k - [网页快照](#)

[2010年世博会澳大利亚国家馆- 首页](#)

澳大利亚庆祝2010年上海世博会与今天的国庆...西部的2010年上海世博会澳大利亚发射3日举行...

[www.australianpavilion.com/](#) - 68k - [网页快照](#)

英语原文- [隐藏英语结果](#)

[Expo 2010 Shanghai China](#)

A Grand Gathering of the World Cultures. "Expo 2010 Shanghai," launched in Decemer 2004, is the sole official magazine of Expo 2010 Shanghai China. more>> ...

[en.expo2010.cn/](#) - 44k - [Cached](#)

[Pavilions of Participants Participation Expo 2010 Shanghai China](#)

City Rhapsody: A Preview of the Site of Expo Shanghai ▪ World Exposition 2010 Shanghai China Promo Enblem Trailer ▪ World Exposition 2010 Shanghai China ...

[en.expo2010.cn/participation/pop/moren.htm](#) - 23k - [Cached](#)

[Expo 2010 - Wikipedia, the free encyclopedia](#)

The Shanghai World Expo is also the largest World's Fair site ever at 5.28 ... President Hu Jintao inaugurated the opening of the Shanghai World Expo. ...

[en.wikipedia.org/wiki/Expo_2010](#) - 129k - [Cached](#)

[Shanghai World Expo](#)

World Expositions are galleries of human inspirations and thoughts. Since 1851 when the Great Exhibition of Industries of All Nations was held in London, ...

[www.expo2010.cn/expo/expoenglish/oe/index.html](#) - 25k - [Cached](#)

[Australian Pavilion EXPO 2010 - Home](#)

Australia celebrated its national day at the Shanghai World Expo 2010 today with ... The Western Australian launch of Shanghai World Expo 2010 was held on 3 ...

[www.australianpavilion.com/](#) - 68k - [Cached](#)

例子：广告排序



新闻 网页 贴吧 知道 音乐 图片 视频 地图 文库 更多»

平板电脑

百度一下

推荐：[用手机随时随地上百度](#)

JD 平板电脑京东便宜!正品行货!

平板电脑正品行货底价促销!货到付款,全国联保,满39元免运费!全场全网底价,货到付款,配送极速,京东给您不一样的购物新体验!

[www.jd.com](#) 2014-02 - 推广

全新平板电脑三星Galaxy Note8.0 无尽可能随心创写

三星平板电脑Galaxy Note8.0全新N5100,N5110,四核1.6GHz处理器,全新Sp en和S日历,8英寸WXGA高清屏,多屏多任务,无尽可能 随心创写的平板电脑尽在三星Galaxy.

[www.samsung.com](#) 2014-02 - 推广

选购平板电脑首选英特尔芯平板 性能更强劲

台电 P89 mini Galaxy Tab3 10 蓝魔 i12 戴尔 venue 7
蓝魔 i9 联想 Miix2 华硕 ME372 W4 英特尔芯平: [更多](#)

[www.intel.cn](#) 2014-02 - 推广

苏宁易购去<苏宁易购> 100%正品行货!全场包邮!

平板电脑 <苏宁易购>好选择!正品保证,全国联保,100城市半日达,一站式7*24时服务!苏宁易购品牌网购商城,数千国内外知名品牌专属网络销售,支持货到付款!

[www.suning.com](#) 2014-02 - 推广

相关平板电脑



三星平板
电脑



安卓平板
电脑



pad



爱国者平
板电脑

相关电子产品



苹果笔记
本



平板电脑
手机



平板电视



联想笔记
本

相关品牌

例子：相似图片搜索

Query



Top



Bottom

例子：相机中的人脸检测



几乎所有数码相机的芯片里都实现了基于Adaboost算法的人脸检测

更多例子

- www.kaggle.com上有很多由大公司发起，来源于真实企业需求的比赛。
 - QE: 航班到达时间预测
 - Facebook: 社交网络分析
 - Yandex: 搜索引擎的个性化排序
 - 腾讯: 微博推荐, 微博好友推荐 (KDDCup 2012)
 -

机器学习人才的机会

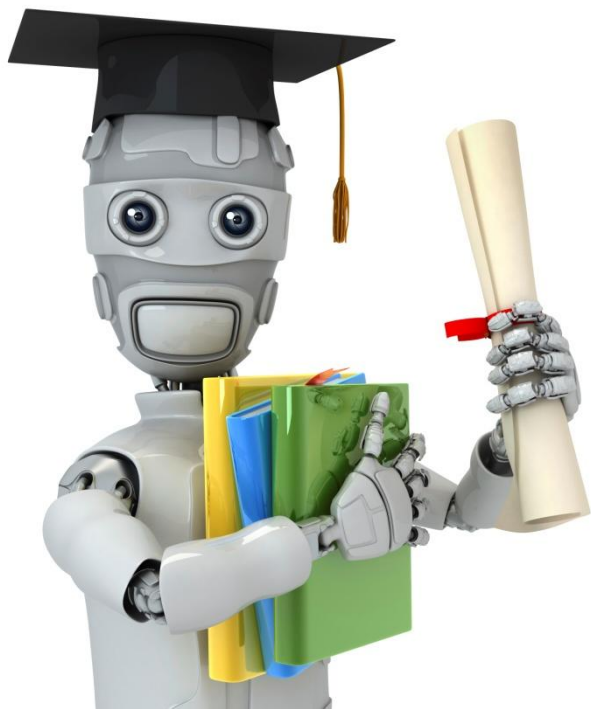
- 2013年，Google以惊人的4700万美金收购多伦多大学的一个初创公司。
 - 这个公司只成立了几个月，一无产品二无专利，只有一个将近80岁的老教授和他的两个刚毕业的博士生
 - 这个老教授就是Geoffrey Hinton
- 2013年，Facebook宣布成立人工智能实验室，招揽纽约大学的Yann Lecun来做“实验室主任”，还专门为了他在纽约建了个办公室。
- 2013年，百度成立深度学习研究院（IDL）。高薪挖来Andrew Ng.，余凯，张潼等人，多个基于深度学习的产品（广告点击率预估，识图等）上线。
- 2014年，Google以4亿美元收购Deepmind公司。
- 2014年，阿里巴巴在硅谷成立机器学习研究院。
- 百度，阿里巴巴，腾讯等开始招聘机器学习/数据挖掘工程师。
- 大量机器学习/数据挖掘的创业公司出现。

Discussion: Machine Learning in Industry

- 大规模数据+人工特征+线性模型为主
 - 高质量的人工特征（特征工程）+并行化训练+高吞吐量预测
 - 中等规模的数据，可用非线性模型（如Boosting+决策树，随机森林）
- 深度学习
 - 在一些难以获得高质量人工特征的应用中获得巨大成功
 - 图像/视频/语音/自然语言理解，广告点击率预估

Discussion: Machine Learning in Industry

- 关键技术
 - 并行化（分布式）学习算法，内存/训练时间限制
 - 典型场景：给你1000台机器，10000GB数据，要求你的算法在1小时内完成训练
 - 高吞吐量预测
 - 典型场景：每个CPU每秒钟要完成800-1000个预测
 - 模型稳定性
 - 增量训练
- 机器学习人才需求
 - Developer：负责实现并行化、高吞吐量的算法；特征工程
 - Researcher：设计新的机器学习算法，提升算法的效果（如准确率）



Machine Learning

Introduction

Welcome

Machine Learning

- Grew out of work in AI
- New capability for computers

Examples:

- Database mining
 - Large datasets from growth of automation/web.
 - E.g., Web click data, medical records, biology, engineering
- Applications can't program by hand.
 - E.g., Autonomous helicopter, handwriting recognition, most of Natural Language Processing (NLP), Computer Vision.

Machine Learning

- Grew out of work in AI
- New capability for computers

Examples:

- Database mining

Large datasets from growth of automation/web.

E.g., Web click data, medical records, biology, engineering

- Applications can't program by hand.

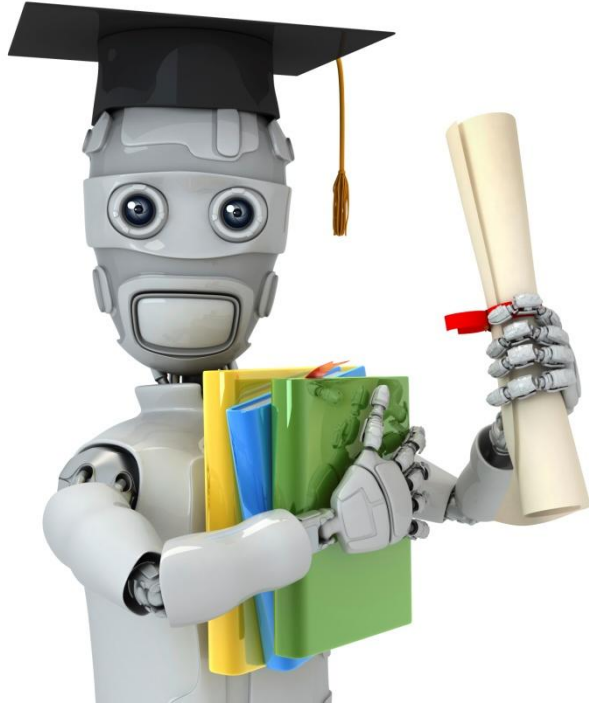
E.g., Autonomous helicopter, handwriting recognition, most of Natural Language Processing (NLP), Computer Vision.

- Self-customizing programs

E.g., Amazon, Netflix product recommendations

- Understanding human learning (brain, real AI).





Machine Learning

Introduction

What is machine learning




Machine Learning definition

- Tom Mitchell (1998) Well-posed Learning Problem: A computer 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 .

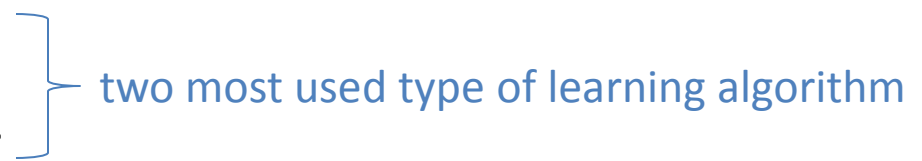
“A computer program is said to *learn* from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.”

e.g. predicting results of football matches

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

- ☐ Classifying emails as spam or not spam.  task T
- ☐ Watching you label emails as spam or not spam.  experience E
- ☐ The number (or fraction) of emails correctly classified as spam/not spam.  measure P
- ☐ None of the above—this is not a machine learning problem.

Machine learning algorithms:

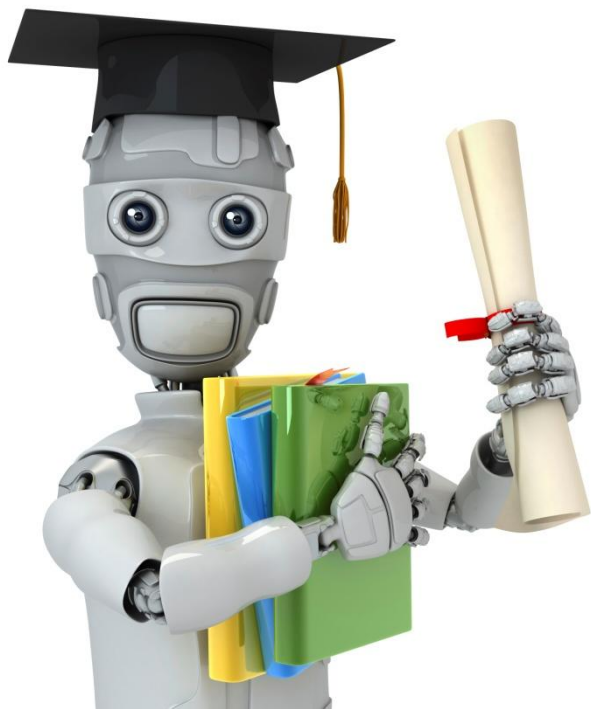
- Supervised learning
 - Unsupervised learning
- 
- two most used type of learning algorithm

Others: Reinforcement learning, recommender systems.

Also talk about: Practical advice for applying learning algorithms.

data mining vs. machine learning

- 数据挖掘：“数据模型”的发现过程，即从数据中发现有用的“规律”
 - 有人认为机器学习和数据挖掘是同义词
 - 机器学习擅长的典型场景是人们对数据中的寻找目标（规律）几乎一无所知。例如Netflix电影推荐
 - 但是，当挖掘的目标能够比较清楚地描述时，机器学习的方法并不成功。例如“在web上定位人们的简历”，基于规则的方法优于机器学习方法。

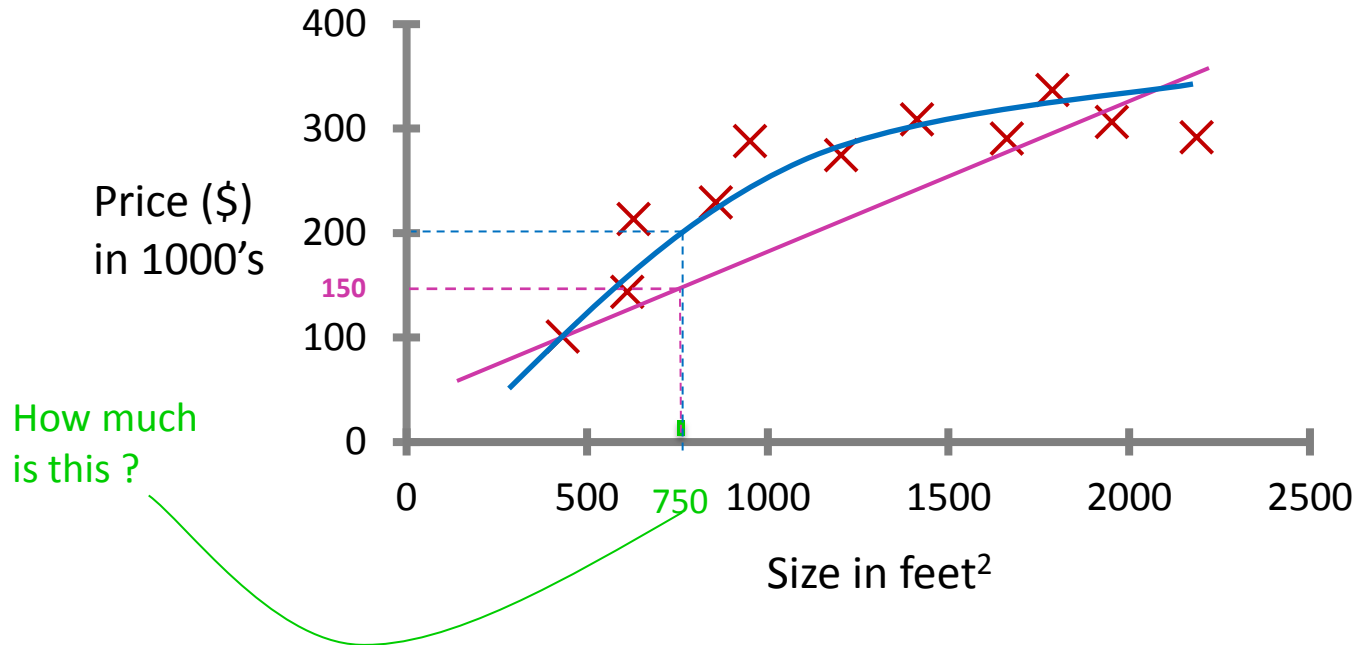


Machine Learning

Introduction

Supervised Learning

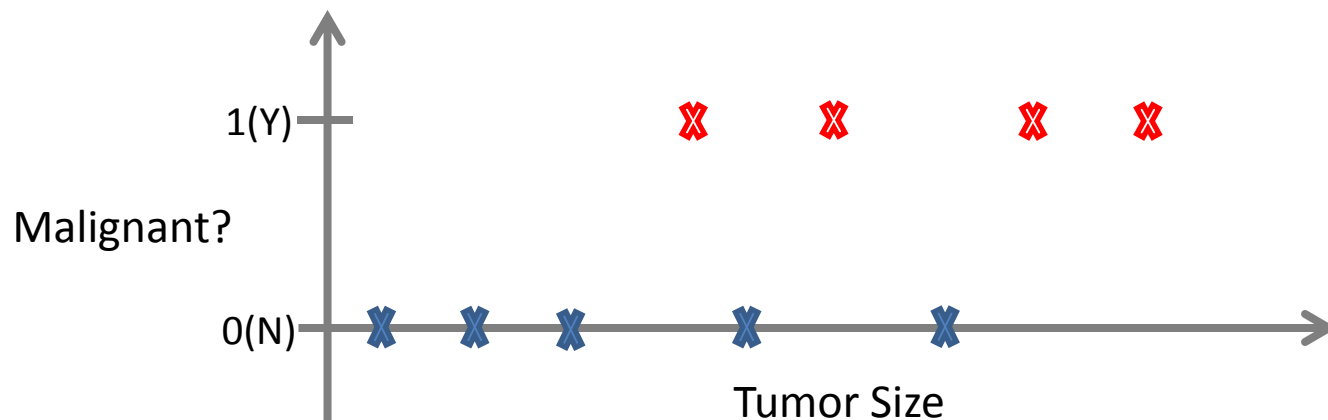
Housing price prediction.



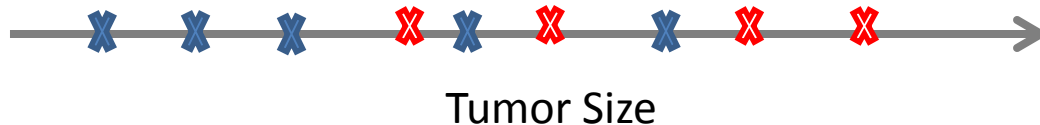
Supervised Learning
“right answers” given

Regression: Predict continuous
valued output (price)

Breast cancer (malignant, benign)



Another way to plot these data. Use one feature, the tumor size



Classification

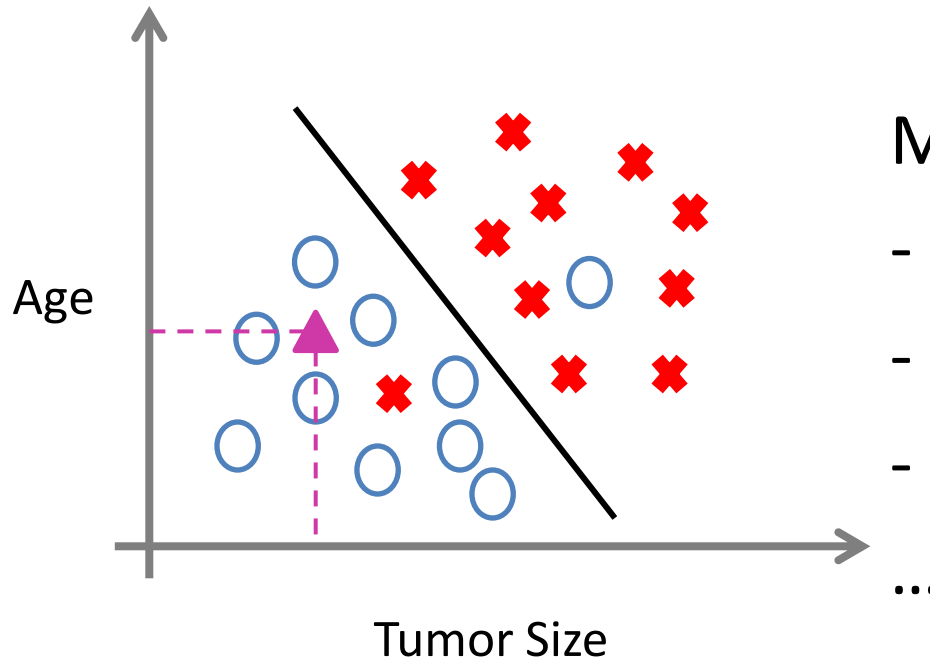
Discrete valued output (0 or 1)

maybe			
0	1	2	
↓	↓	↓	
benign	Type 1 cancer	Type 2 cancer	

○ :benign

✖ : malignant

▲ : dose it malignant or benign?



More features

- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

...

Use two features, age and tumor size.

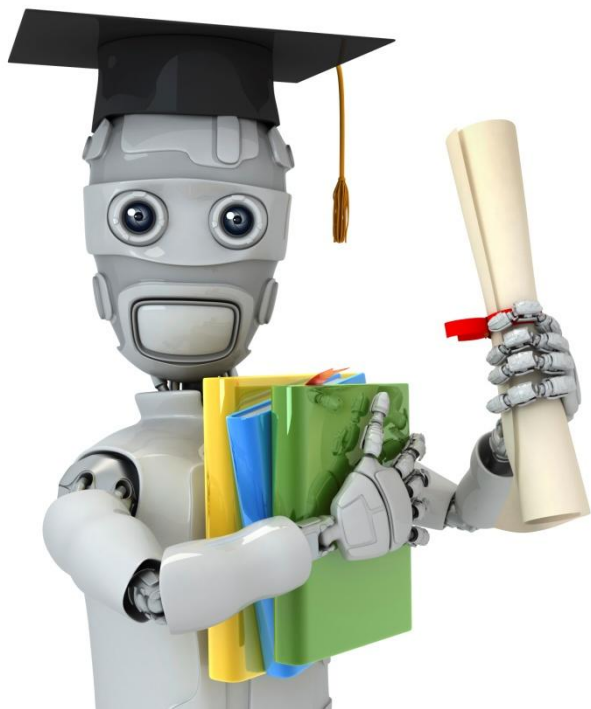
You're running a company, and you want to develop learning algorithms to address each of two problems.

Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.

Problem 2: You'd like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised. (例如：盗号检测，信用卡被盗刷)

Should you treat these as classification or as regression problems?

- ☐ Treat both as classification problems.
- ☐ Treat problem 1 as a classification problem, problem 2 as a regression problem.
- ☒ Treat problem 1 as a regression problem, problem 2 as a classification problem.
- ☐ Treat both as regression problems.

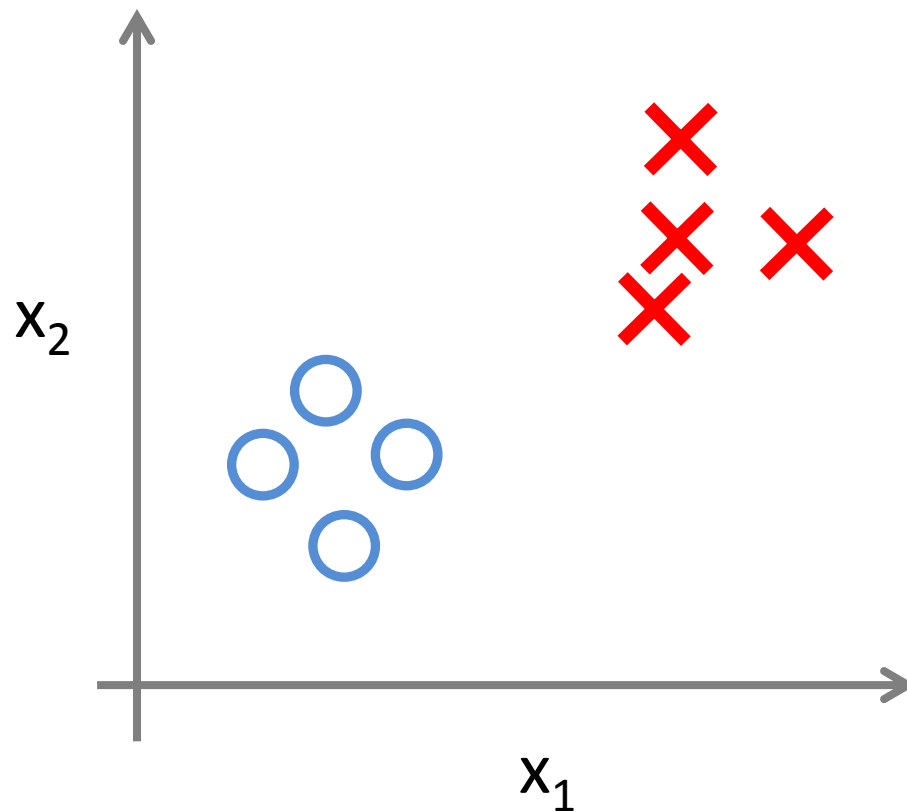


Machine Learning

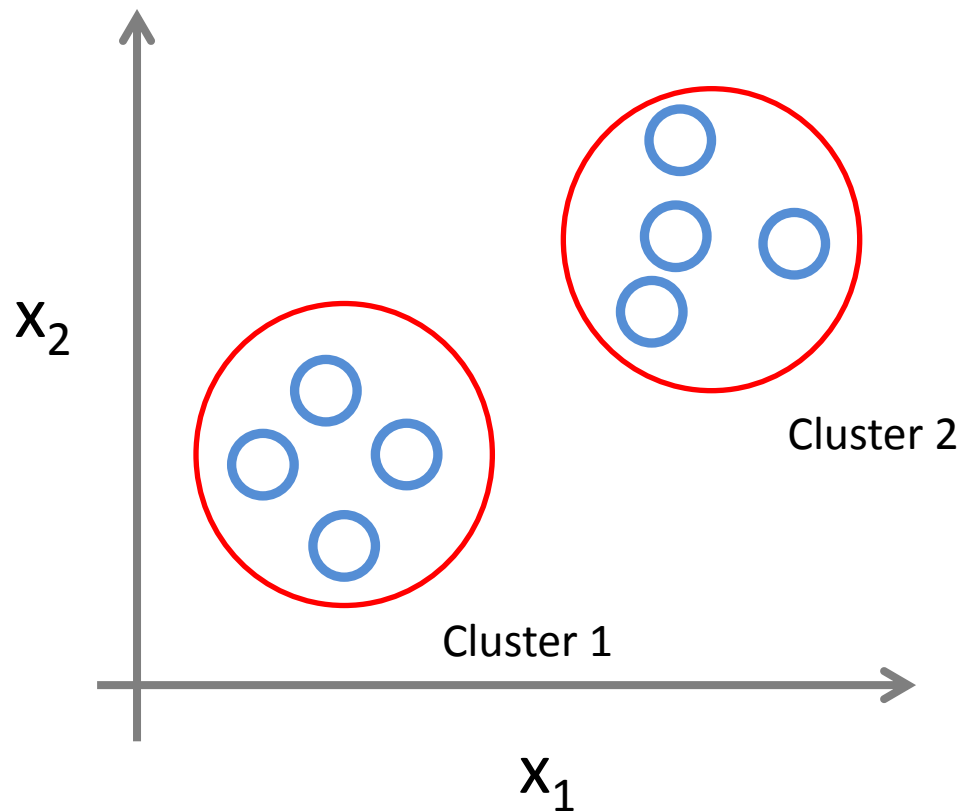
Introduction

Unsupervised Learning

Supervised Learning



Unsupervised Learning



Google News
news.google.com

Web Images Videos Maps News Shopping Gmail more ▾ andrewyantakng@gmail.com | Web History | Settings ▾ | Sign out

Google news Search News Search the Web
Advanced news search

U.S. edition ▾ Add a section »

Top Stories

Deepwater Horizon
Fed meeting
Foreign exchange market
Lindsay Lohan
IBM
Tom Brady
Toronto International Film Festival
Paris Hilton
Iran
Hurricane Igor

Starred ☆

San Francisco Bay Area
World
U.S.
Business
Sci/Tech
More Top Stories
Spotlight
Health
Sports
Entertainment


All news


Headlines
Images


Top Stories


Christine O'Donnell »
White House official denies Tea Party-focused ad campaign
CNN International - Ed Henry - 1 hour ago
Democratic sources say the White House is not considering an ad campaign tying Republicans to the Tea Party. Washington (CNN) -- A top White House official sharply denied a report that claims President Obama's political advisers are weighing a national ...
Tea Party is misplacing the blame, former President Bill Clinton claims
New York Daily News
GOP tea party backer defends Christine O'Donnell The Associated Press
Atlanta Journal Constitution - Politics Daily - MyFox Washington DC - Salon all 726 news articles »

US Stocks Climb After Recession Called Over, Homebuilders Gain
MarketWatch - Kristina Peterson - 16 minutes ago
NEW YORK (MarketWatch) -- US stocks climbed Monday, gaining speed after a key nonprofit organization officially called the recession over, giving investors a boost of confidence in the gradual economic recovery.
Longest recession since 1930s ended in June 2009, group says
Los Angeles Times
Downturn Was Longest in Decades, Panel Confirms New York Times
Wall Street Journal - AFP - CNN - USA Today all 276 news articles »

Deepwater Horizon »
BP Oil Well, Site of National Catastrophe, Dies at One
Vanity Fair - Juli Weiner - 22 minutes ago
The BP oil well, site of the Deepwater Horizon explosion that led to the worst oil spill in US history, died today at one year old.
Video: Blown-out BP Well Finally Killed in Gulf  The Associated Press
Weiss Doubts BP Would End Operations in Gulf of Mexico: Video Bloomberg
CNN International - Wall Street Journal (blog) - The Guardian - New York Times
all 2,292 news articles »


CNN Interna...


MyFox Phila...


Reuters

Recent

Recession officially ended in June 2009 ☆ ▾
CNMoney - Chris Isidore - 39 minutes ago

Hurricane Igor lashes Bermuda ☆ ▾
USA Today - Gerry Broome - 5 minutes ago

'Explain what you want from us.' reads front-page editorial ☆ ▾
msnbc.com - Olivia Torres - 10 minutes ago

Crisis response: Pakistan floods

San Francisco Bay Area - Edit

Clorox »
Bay Biz Buzz: Clorox close to selling STP, Armor All ☆ ▾
San Jose Mercury News - 48 minutes ago - all 24 articles »

Google's official beekeeper keeps the company buzzing with excitement ☆ ▾
San Jose Mercury News - Bruce Newman - 1 hour ago

Jon Sylvia »
Martinez man still unconscious as police investigate weekend shooting ☆ ▾
San Jose Mercury News - Robert Salonga - 48 minutes ago - all 6 articles »

Spotlight

Sarkozy rages at EU 'humiliation' ☆ ▾
Financial Times - Peggy Hollinger - Sep 16, 2010

THE WALL STREET JOURNAL.

Log In • Register For Free • Subscribe Now, Get 2 Weeks Free

<< PREVIOUS

THE SOURCE

NEXT >>

Financial
Services

Transport

Leisure

Insurance

Oil & Gas

Sport

Caught on the
Web

Betting

Technology

SEPTEMBER 20, 2010, 12:44 PM GMT

BP Kills Macondo, But Its Legacy Lives On

Search The Source

SEARCH

Article

Comments (2)

THE SOURCE HOME PAGE >



Email



Print



Permalink



Like

2



+ More



Text

About The Source

Follow Us:



The Source is WSJ.com Europe's home for rapid-fire analysis of the day's big business and finance stories. It is edited by Lauren Mills, based in London.



Most Recent

Articles

Comments

1. Who Needs Plaza II Anyway
2. Will Banks Be Forced to Split Retail And Banking Arms?
3. Timing of Ratings Award Intriguing
4. BP Kills Macondo, But Its Legacy Lives On
5. We Already Need a Sequel to Basel III

By James Herron

BP confirmed late Sunday that the Macondo well that leaked almost five million barrels of oil into the Gulf of Mexico has been permanently sealed, but the well will continue to affect BP and the wider oil industry for many years.

The most immediate worry for BP and its shareholders is how the authorities will apportion blame for the spill. BP's own investigation spread responsibility across



Associated Press

Fire boat response crews battled the blazing remnants of the off shore oil rig Deepwater Horizon on April 21, 2010.

Allen: Well is dead, but m... x

edition.cnn.com/2010/US/09/20/gulf.oil.disaster/

☆ ↗

EDITION: INTERNATIONAL | U.S. | MÉXICO | ARABIC

Set edition preference

CNN

Sign up | Log in

SEARCH


POWERED BY Google

Home | Video | World | U.S. | Africa | Asia | Europe | Latin America | Middle East | Business | World Sport | Entertainment | Tech | Travel | iReport

Allen: Well is dead, but much Gulf Coast work remains

By the CNN Wire Staff
September 20, 2010 -- Updated 1317 GMT (2117 HKT)

Click to play ▶



What next for Gulf oil spill?

Recommend

Be the first of your friends to recommend this.

Mixx | Twitter | Share | Email | Save | Print

Most Popular »

Today's five most popular stories

- White House official denies Tea Party-focused ad campaign
- Swedish far-right leader: Success due to immigration backlash
- Lady Gaga to lead rally against military's 'don't ask, don't tell'
- Why Facebook is blue -- six facts about Mark Zuckerberg
- China, Japan relations sour in dispute over fishing captain

More

STORY HIGHLIGHTS

(CNN) -- The ruptured Macondo well, a mile under the Gulf of Mexico off the Louisiana coast, has been pronounced dead.

guardian.co.uk

Search

Business ▾

Search

[News](#) | [Sport](#) | [Comment](#) | [Culture](#) | [Business](#) | [Money](#) | [Life & style](#) | [Travel](#) | [Environment](#) | [TV](#) | [Video](#) | [Community](#) | [Blogs](#) | [Jobs](#)

Business > BP

BP oil spill cost hits nearly \$10bn

BP has set up a \$20bn compensation fund after the Deepwater Horizon disaster, which has so far paid out 19,000 claims totalling more than \$240m

Julia Kollewe

guardian.co.uk, Monday 20 September 2010 08.33 BST

[Article history](#)

BP's costs for the Deepwater Horizon disaster have hit \$10bn. Photograph: Ho/Reuters



Tweet this (68)

Comments (10)



A larger | smaller

Business

BP

Environment

BP oil spill - Oil - Oil spills

World news

United States

More news

More on this story



US government declares well 'dead'

BP oil spill news on Twitter

Latest news on the BP oil spill in the Gulf of Mexico



BP_America: Find fact sheets, open house info and e-mail alert sign-ups at @EPAgov's oil spill community page: <http://bit.ly/diuHoH>
about 3 weeks ago



guardianeco: UN report on Nigeria oil spills relies too much on data from Shell | Nnimmo Bassey <http://bit.ly/dBd7Ru>
about 3 weeks, 5 days ago



BP_America: Newly discovered microbe thriving from consumption of #oil in the #Gulf of Mexico: <http://bit.ly/9kQYwa>
about 3 weeks, 5 days ago

- Archive page: read more BP oil spill tweets
- Follow BP oil spill news with our Twitter list
- Full BP oil spill coverage from guardian.co.uk

On Business

Most viewed

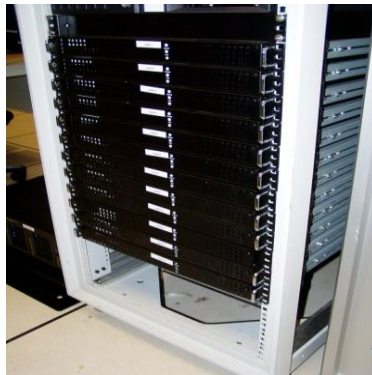
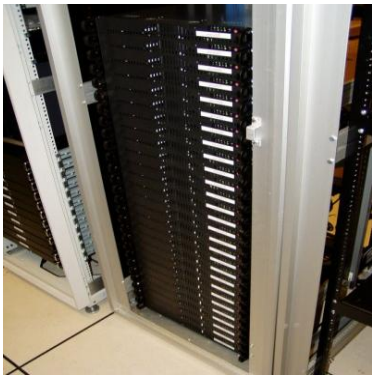
Zeitgeist

Latest

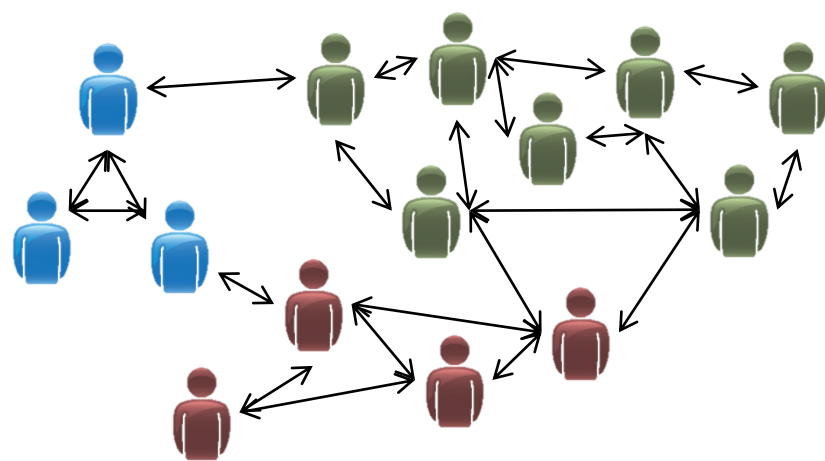
Last 24 hours



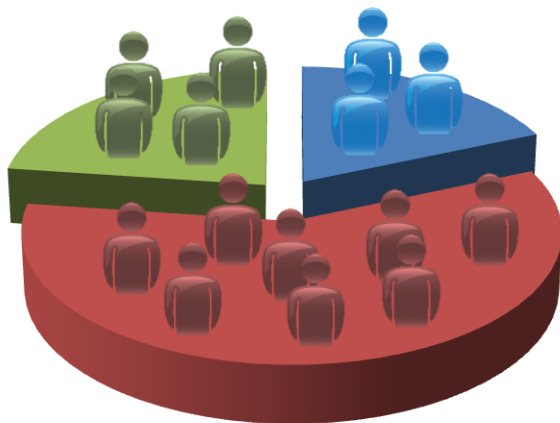
1. Britain keeps AAA credit rating



Organize computing clusters



Social network analysis



Market segmentation



Astronomical data analysis

Of the following examples, which would you address using an unsupervised learning algorithm? (Check all that apply.)

- ☐ Given email labeled as spam/not spam, learn a spam filter.
- ☒ Given a set of news articles found on the web, group them into set of articles about the same story.
- ☒ Given a database of customer data, automatically discover market segments and group customers into different market segments.
- ☐ Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.