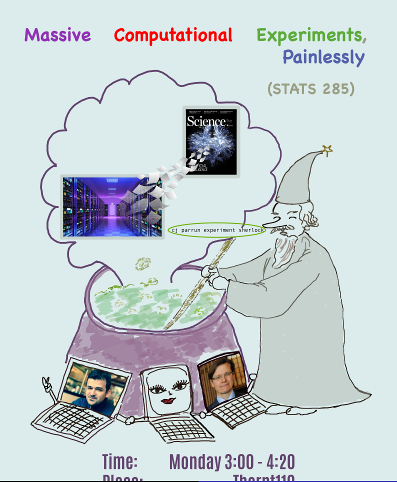
Lecture 10: Looking Back/Looking Ahead

D Donoho/ H Monajemi
Stats 285 Stanford

20171204

Stats 285 Fall 2017



Outline

Disclaimer

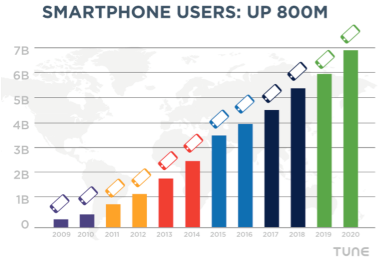
This document contains images obtained by routine Google Images searches. Some of these images may perhaps be copyright. They are included here for educational noncommercial purposes and are considered to be covered by the doctrine of **Fair Use***. In any event they are easily available from Google Images.*

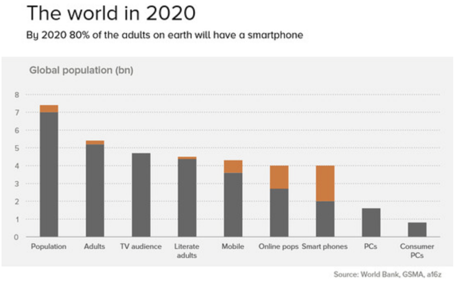
It's not feasible to give full scholarly credit to the creators of these images. We hope they can be satisfied with the positive role they are playing in the educational process.

The Mobile Revolution

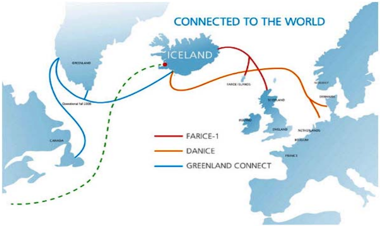



Smartphones are Spreading Everywhere





24/7 Deluge Spawns Global Computational Services



Cloud Paradigm

Cloud Paradigm:

- Billions of smart devices each drive queries to cloud servers
- Millions of business relying on cloud for all needs

Symbiosis of cloud and economy is *lasting* and *disruptive*.

Explosion of Computational Resources

Cloud Paradigm:

- Billions of smart devices each drive queries to cloud servers
- Millions of business relying on cloud for all needs

Symbiosis of cloud and economy is *lasting* and *disruptive*.

Cloud provides *any user* **same-day** delivery:

- Tens to hundreds of thousands of hours of CPU
- Pennies per CPU hour

Any user can consume *1 Million CPU hours* over a few days for a few $10K's.

Stack Paradigm

Stack Paradigm:

- Organizations combine software components from other providers in a `stack`
- Massive new capabilities emerge by *hybridizing components*

Examples:

- Uber
- Netflix relies on AWS
- Snap, Dropbox etc. small teams

Explosion of Convenience

Any user can deploy and control massive computational resources from a well-chosen stack of applications/libraries/services.

A Look Back, 2

- In Lecture 03, Eric Jonas showed how AWS Lambda creates new opportunities for research in computational science
- In Lecture 05, Percy Liang showed how Codalab+CodaWorksheets can run experiments and challenges on AWS/Azure/GCP
- In Lecture 07, Riccardo Murri showed how to make private clusters on AWS/Azure/GCP
- In Lecture 08, Andy Konwinski showed how to run large workflows painlessly on DataBricks(AWS)
- In Lecture 09, Hatef Monajemi told us that hybridizing ClusterJob+ElastiCluster can do pushbutton ML on AWS/Azure/GCP

A Look Back 3: Emergent Phenomena

The Rise of ...

- Prediction Challenges
- Software Frameworks
- Hyperparameter Search
- Workflows as Objects
- Equivalence of Efficiency, Reproducible, Painless computing

A Look Back 4: Lessons from Deep Learning

1. *Researchers who tweak more often, win more often!*
2. *If easier to implement tweaks and faster to evaluate them, more likely to win!*
3. Successful Research Environment
   - Easy to tweak models
   - Easy to score tweaks
   - Fast to score tweaks
4. Successful researchers perpetually motivated by *Game-ification*: tweaking, scoring, winning.
5. Easier to stay motivated when easier and more comfortable to play the game.
   - Elegant expression of tweaks
   - Rapid turn-around for scoring

A Look Back 5: Framework Wars

The real action is all in frameworks

1. Dream up, test, and publish better ...
   - Types of models
   - Types of tweaks
   - Properties for evaluation
2. Implement better *frameworks* ...
   - More elegant expression of models, tweaks
   - Distributed Learning across clusters
   - Smoother collection and analysis of results

AWS is Eating the world: Stock Market


TECH

TECH | MOBILE | SOCIAL MEDIA | ENTERPRISE | CYBERSECURITY | TECH GUIDE

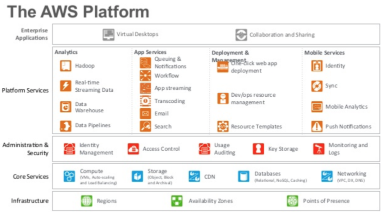
Amazon shares soar after massive earnings beat

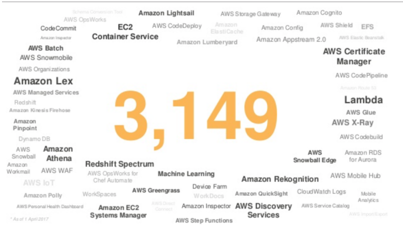
- Amazon reported its third quarter results Thursday after the bell.
- It was a huge beat across the board.
- Amazon shares jumped over 7 percent in after hours trading.

Eugene Kim | @eugenekim222
Published 3:24 PM ET Thu, 26 Oct 2017 | Updated 6:55 PM ET Thu, 26 Oct 2017

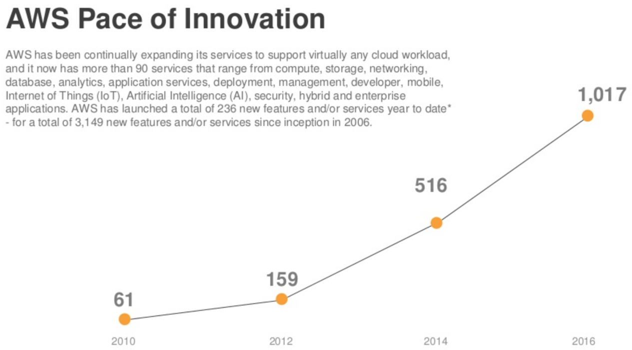
CNBC

AWS Services Are Ubiquitous





AWS Services are Proliferating



AWS Impact on Machine Learning, I

ImageNet dataset

- 14,197,122 labeled images
- 21,841 classes
- Labeling: more than a year of human effort via Amazon Mechanical Turk

AWS Impact on Machine Learning, II

Business Wire
A Berkshire Hathaway Company

HOME SERVICES NEWS EDUCATION ABOUT US

Search

AWS Announces Five New Machine Learning Services and the World's First Deep Learning-Enabled Video Camera for Developers

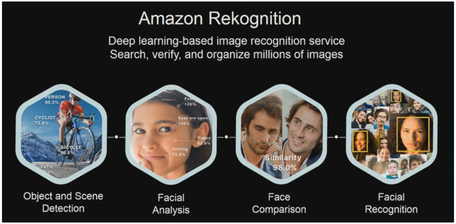
Amazon SageMaker makes it easy to build, train, and deploy machine learning models

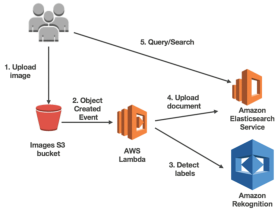
AWS DeepLens is the world's first deep learning-enabled wireless video camera built to give developers hands-on experience with machine learning

Amazon Transcribe, Amazon Translate, Amazon Comprehend, and Amazon Rekognition Video allow app developers to easily build applications that transcribe speech to text, translate text between languages, extract insights from text, and analyze videos

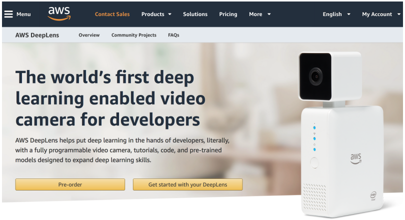
NFL, Intuit, Thomson Reuters, DigitalGlobe, Hotels.com, ZipRecruiter, Washington Post, Motorola

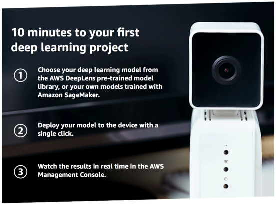
Impact on Machine Learning, II





Impact on Machine Learning, III



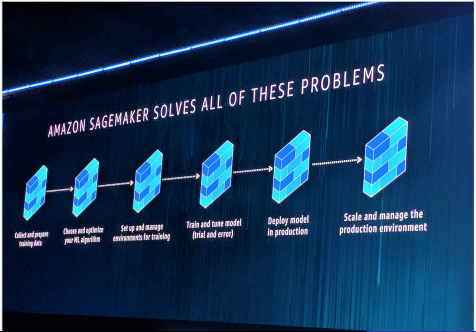


Impact on Machine Learning, III

AWS releases SageMaker to make it easier to build and deploy machine learning models

Posted 22 hours ago by Ron Miller (@ron_miller)

Next Story



Impact on Machine Learning, IV

Menu | aws | Products | Solutions | Pricing | Getting Started | Documentation | More | My Account | Create an AWS Accoun

Home | Category | Edition | Follow | Search Blogs

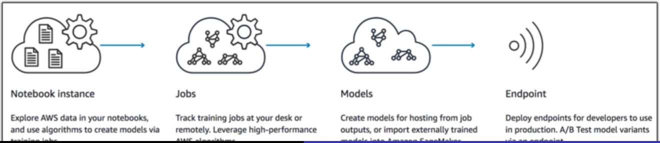
Amazon SageMaker – Accelerating Machine Learning

by Randall Hunt | on 29 NOV 2017 | in Artificial Intelligence*, AWS Re:Invent*, SageMaker | Permalink | Comments | Share

Machine Learning is a pivotal technology for many startups and enterprises. Despite decades of investment and improvements, the process of developing, training, and maintaining machine learning models has still been cumbersome and ad-hoc. The process of incorporating machine learning into an application often involves a team of experts tuning and tinkering for months with inconsistent setups. Businesses and developers want an end-to-end, development to production pipeline for machine learning.

Introducing Amazon SageMaker

Amazon SageMaker is a fully managed end-to-end machine learning service that enables data scientists, developers, and machine learning experts to quickly build, train, and host machine learning models at scale. This drastically accelerates all of your machine learning efforts and allows you to add machine learning to your production applications quickly.



AWS validates Stats 285 thesis

AWS perceives massive demand for

- Massive scale
- Convenience
- Hygiene
- Standardization of workflows

Future Science will ...

- View Science itself as data
- Test new methodology against historical corpus of science
- Measure success of **end-to-end pipelines**

Google: '50 Years of Data Science Donoho'

Two Examples below

- Cross-study performance of pipelines
- Cross-methodology performance of pipelines

JNCI Journal of the National Cancer Institute Advance Access published April 3, 2014
DOI:10.1093/jnci/dju049

ARTICLE

Comparative Meta-analysis of Prognostic Gene Signatures for Late-Stage Ovarian Cancer

Levi Waldron, Benjamin Haibe-Kains, Aedín C. Culhane, Markus Riester, Jie Ding, Xin Victoria Wang, Mahnaz Ahmadifar, Svitlana Tyekucheva, Christoph Bernau, Thomas Risch, Benjamin Frederick Ganzfried, Curtis Huttenhower, Michael Birrer, Giovanni Parmigiani

Manuscript received February 24, 2013; revised January 13, 2014; accepted January 29, 2014.

Correspondence to: Giovanni Parmigiani, PhD, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, 450 Brookline Ave, Boston, MA 02115 (e-mail: gp@jimmy.harvard.edu).

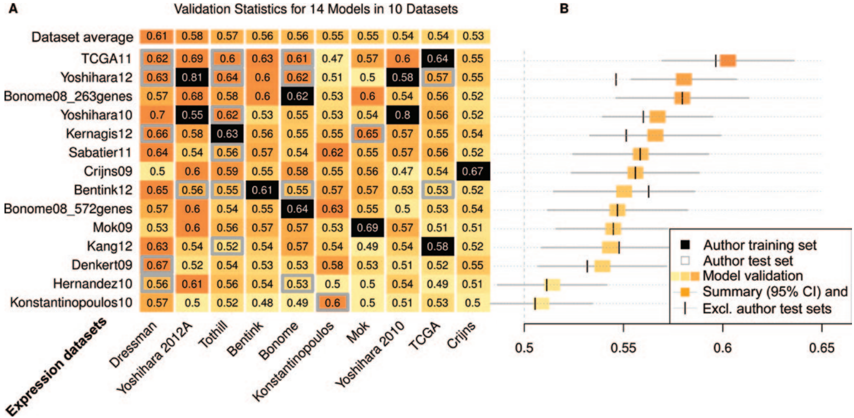
Background Ovarian cancer is the fifth most common cause of cancer deaths in women in the United States. Numerous gene signatures of patient prognosis have been proposed, but diverse data and methods make these difficult to compare or use in a clinically meaningful way. We sought to identify successful published prognostic gene signatures through systematic validation using public data.

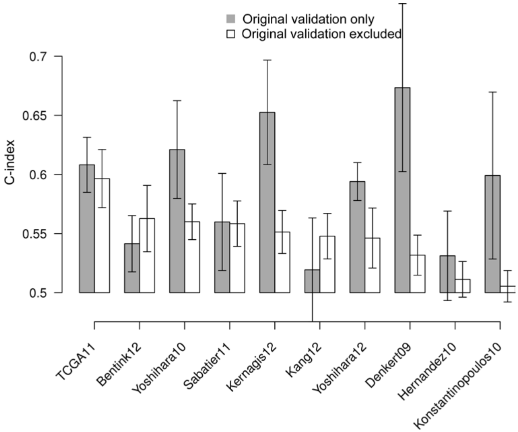
Methods A systematic review identified 14 prognostic models for late-stage ovarian cancer. For each, we evaluated its 1) reimplementation as described by the original study, 2) performance for prognosis of overall survival in independent data, and 3) performance compared with random gene signatures. We compared and ranked models by validation in 10 published datasets comprising 1251 primarily high-grade, late-stage serous ovarian cancer patients. All tests of statistical significance were two-sided.

Results Twelve published models had 95% confidence intervals of the C-index that did not include the null value of 0.5; eight outperformed 97.5% of signatures including the same number of randomly selected genes and trained on the same data. The four top-ranked models achieved overall validation C-indices of 0.56 to 0.60 and shared anti-correlation with expression of immune response pathways. Most models demonstrated lower accuracy in new datasets than in validation sets presented in their publication.

Table 1. Reproducibility of the 14 published models for prognosis of late-stage epithelial ovarian cancer selected for meta-analysis*

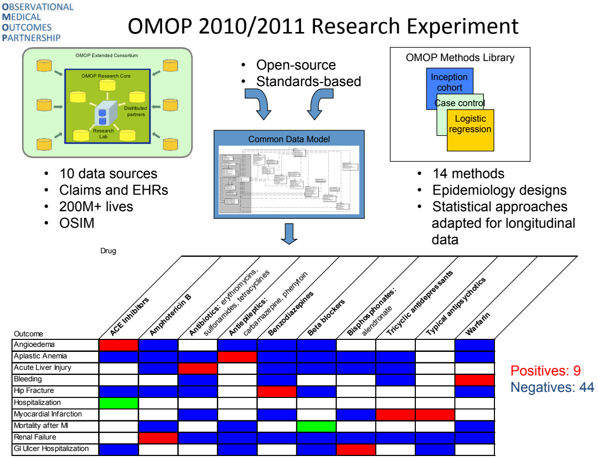
	Reproducibility†			
Model	**Model provided**	**Training data available**	**Validation data available**	**Verified implementation**
TCGA11 (12)	Yes	Yes	Yes	Yes
Denkert09 (13)	Yes	Yes	Yes	Yes
Bonome08_263genes (14)	Yes	Yes	Yes	Yes
Bonome08_572genes (14)	Yes	Yes	Yes	Yes
Mok09 (15)	No	Yes	Yes	Partially
Yoshihara12 (16)	Yes	—	Yes	Yes
Yoshihara10 (17)	Yes	—	Yes	Yes
Bentink12 (18)	Yes	—	Yes	Yes
Kang12 (19)	Yes	Yes	Yes	Partially
Crijns09 (20)	No	Yes	No	No
Kernagis12 (21)	Partially	Yes	Yes	Partially
Sabatier11 (22)	Partially	No	No	No
Konstantinopoulos10 (23)	Yes	—	Yes	Partially
Hernandez10 (24)	Partially	—	Yes	Partially


A
Validation Statistics for 14 Models in 10 Datasets
B
Dataset average 0.61 0.58 0.57 0.56 0.56 0.55 0.55 0.54 0.54 0.53
TCGA11 0.62 0.69 0.6 0.63 0.61 0.47 0.57 0.6 0.64 0.55
Yoshihara12 0.63 0.81 0.64 0.6 0.62 0.51 0.5 0.58 0.57 0.55
Bonome08_263genes 0.57 0.68 0.58 0.6 0.62 0.53 0.6 0.54 0.56 0.52
Yoshihara10 0.7 0.55 0.62 0.53 0.55 0.53 0.54 0.8 0.56 0.52
Kernagis12 0.66 0.58 0.63 0.56 0.55 0.55 0.65 0.57 0.55 0.54
Sabatier11 0.64 0.54 0.56 0.57 0.54 0.62 0.55 0.57 0.56 0.52
Crijns09 0.5 0.6 0.59 0.55 0.58 0.55 0.56 0.47 0.54 0.67
Bentink12 0.65 0.56 0.55 0.61 0.55 0.57 0.57 0.53 0.53 0.52
Bonome08_572genes 0.57 0.6 0.54 0.55 0.64 0.63 0.55 0.5 0.53 0.54
Mok09 0.53 0.6 0.56 0.57 0.57 0.53 0.69 0.57 0.51 0.51
Kang12 0.63 0.54 0.52 0.54 0.57 0.54 0.49 0.54 0.58 0.52
Denkert09 0.67 0.52 0.54 0.53 0.53 0.58 0.53 0.51 0.52 0.55
Hernandez10 0.56 0.61 0.56 0.54 0.53 0.5 0.5 0.54 0.49 0.51
Konstantinopoulos10 0.57 0.5 0.52 0.48 0.49 0.6 0.5 0.51 0.53 0.5
Expression datasets
Dressman
Yoshihara 2012A
Tothill
Bentink
Bonome
Konstantinopoulos
Mok
Yoshihara 2010
TCGA
Crijns
Author training set
Author test set
Model validation
Summary (95% CI) and
Excl. author test sets
0.5
0.55
0.6
0.65


Original validation only
Original validation excluded
C-index
0.7
0.65
0.6
0.55
0.5
TCGA11
Bentink12
Yoshihara10
Sabatier11
Kernagis12
Kang12
Yoshihara12
Denkert09
Hernandez10
Konstantinopoulos10

A Systematic Statistical Approach to Evaluating Evidence from Observational Studies

David Madigan,[1,2] Paul E. Stang,[2,3] Jesse A. Berlin,[4]
Martijn Schuemie,[2,3] J. Marc Overhage,[2,5]
Marc A. Suchard,[2,6,7,8] Bill Dumouchel,[2,9]
Abraham G. Hartzema,[2,10] and Patrick B. Ryan[2,3]


OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP
OMOP 2010/2011 Research Experiment
OMOP Extended Consortium
OMOP Research Core
Distributed partners
Research Lab
• 10 data sources
• Claims and EHRs
• 200M+ lives
• OSIM
• Open-source
• Standards-based
Common Data Model
OMOP Methods Library
Inception cohort
Case control
Logistic regression
• 14 methods
• Epidemiology designs
• Statistical approaches adapted for longitudinal data
Drug
ACE Inhibitors
Amphotericin B
Antibiotics: erythromycins, sulfonamides, tetracyclines
Antiepileptics: carbamazepine, phenytoin
Benzodiazepines
Beta blockers
Bisphosphonates: alendronate
Tricyclic antidepressants
Typical antipsychotics
Warfarin
Outcome
Angioedema
Aplastic Anemia
Acute Liver Injury
Bleeding
Hip Fracture
Hospitalization
Myocardial Infarction
Mortality after MI
Renal Failure
GI Ulcer Hospitalization
Positives: 9
Negatives: 44

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Ground truth for OMOP 2011/2012 experiments

isoniazid
fluticasone
indomethacin
clindamycin

	Positive controls	Negative controls	Total
Acute Liver Injury	81	37	118
Acute Myocardial Infarction	36	66	102
Acute Renal Failure	24	64	88
Upper Gastrointestinal Bleeding	24	67	91
Total	165	234	399

ibuprofen
loratadine
sertraline
pioglitazone

Criteria for positive controls:

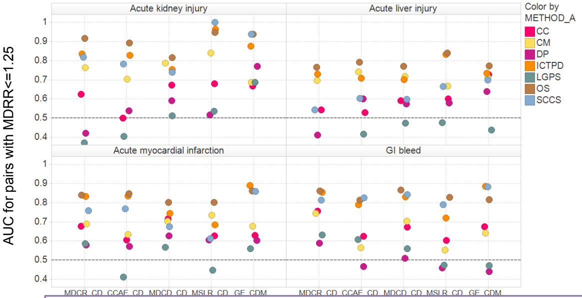
- Event listed in Boxed Warning or Warnings/Precautions section of active FDA structured product label
- Drug listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases"
- Literature review identified no powered studies with refuting evidence of effect

Criteria for negative controls:

- Event not listed anywhere in any section of active FDA structured product label
- Drug not listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases"
- Literature review identified no powered studies with evidence of potential positive association

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Performance across methods, by database



- All self-controlled designs (OS, ICTPD, SCCS) are consistently at or near the top of performance across all outcomes and sources
- Cohort and case-control designs have comparable performance, consistently lower than all self-controlled designs
- Substantial variability in performance across the optimal settings of each method

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Optimal methods (AUC) by outcome and data source

Data source	Acute kidney injury	Acute liver injury	Acute myocardial infarction	GI bleed
MDCR	**OS: 401002** (0.92)	**OS: 401002** (0.76)	**OS: 407002** (0.84)	**OS: 402002** (0.86)
CCAE	**OS: 404002** (0.89)	**OS: 403002** (0.79)	**OS: 408013** (0.85)	**SCCS: 1931010** (0.82)
MDCD	**OS: 408013** (0.82)	**OS: 409013** (0.77)	**OS: 407004** (0.80)	**OS: 401004** (0.87)
MSLR	**SCCS: 1939009** (1.00)	**OS: 406002** (0.84)	**OS: 403002** (0.80)	**OS: 403002** (0.83)
GE	**SCCS: 1949010** (0.94)	**OS: 409002** (0.77)	**ICTPD: 3016001** (0.89)	**ICTPD: 3034001** (0.89)

- Self-controlled designs are optimal across all outcomes and all sources, but the specific settings are different in each scenario
- AUC > 0.80 in all sources for acute kidney injury, acute MI, and GI bleed
- Acute liver injury has consistently lower predictive accuracy
- No evidence that any data source is consistently better or worse than others

31

Global Economy → Computing → Science