

Personalized Health with Gaussian Processes

Neil D. Lawrence

University of Sheffield

19th August 2015

What is Machine Learning?

data

- ▶ **data**: observations, could be actively or passively acquired (meta-data).

What is Machine Learning?

data +

- ▶ **data**: observations, could be actively or passively acquired (meta-data).

What is Machine Learning?

data + **model**

- ▶ **data**: observations, could be actively or passively acquired (meta-data).
- ▶ **model**: assumptions, based on previous experience (other data! transfer learning etc), or beliefs about the regularities of the universe. Inductive bias.

What is Machine Learning?

data + **model** =

- ▶ **data**: observations, could be actively or passively acquired (meta-data).
- ▶ **model**: assumptions, based on previous experience (other data! transfer learning etc), or beliefs about the regularities of the universe. Inductive bias.

What is Machine Learning?

$$\text{data} + \text{model} = \text{prediction}$$

- ▶ **data**: observations, could be actively or passively acquired (meta-data).
- ▶ **model**: assumptions, based on previous experience (other data! transfer learning etc), or beliefs about the regularities of the universe. Inductive bias.
- ▶ **prediction**: an action to be taken or a categorization or a quality score.

What's Changed (Changing) for Medical Data?

- ▶ Try Googling for: “patient data ”...



Image from [Wikimedia Commons](#)



Image from [Wikimedia Commons](#)



INF57

A brief history *of Registration*

For more information go to: www.direct.gov.uk/motoring

A brief history of registration

The early days

Prior to the appearance of the first railways in Britain, there was a brief development and interest in steam powered road going vehicles. In 1834, a Mr Hancock started a steam coach called the “Era”, carrying up to 14 passengers from Paddington to Regents Park and the City at 6d a head. And in the following year, a Mr Church built an omnibus capable of carrying 40 passengers for the London and Birmingham Steam Carriage Company.

However, the success of the railway movement drove all such traffic off the roads.

A **Parliamentary Commission of Enquiry in 1836** reported “strongly in favour of steam carriages on roads”, but subsequent Acts of Parliament tended to have a discouraging and restrictive effect. **The Locomotive Act 1861** limited the weight of steam engines to 12 tons and imposed a speed limit of 10 mph.

The Locomotive Act 1865 set a speed limit of 4 mph in the country and 2 mph in towns. The 1865 Act also provided for the famous “man with a red flag”. Walking 60 yards ahead of each vehicle, a man with a red flag or lantern enforced a walking pace, and warned horse riders and horse drawn traffic of the approach of a self propelled machine.

The Locomotive Amendment Act 1878 made the red flag optional under local regulations, and

[Crown Copyright Reserved.]



Ministry of Transport.

THE
HIGHWAY CODE

Issued by the Minister of Transport
with the authority of Parliament in
pursuance of Section 45 of the
Road Traffic Act, 1930.

LONDON :

PRINTED AND PUBLISHED BY HIS MAJESTY'S STATIONERY OFFICE
To be purchased directly from H.M. Stationery Office at the following addresses:
Admiral House, Kingsway, London, W.C.2; 120, George St., Edinburgh;
York Street, Manchester; 1, St. Andrew's Crescent, Cardiff;
15, Donegall Square West, Belfast;
or through any bookseller.

1931.

Price 1d. net.

55-166

What are the Issues?

- ▶ Who owns our data?
- ▶ Is it 'finders keepers'?
- ▶ Does ownership proliferate?
- ▶ What does data protection offer?
- ▶ Who has the right to share our data?
- ▶ Can we withdraw this right?

Moral Panics: Perhaps Rightly



The screenshot shows a web browser displaying the BBC News website. The page features a red header with the BBC logo and navigation links. The main article is titled "NHS Care.data information scheme 'mishandled'" and is written by Chris Vallance. It includes a photograph of a hand writing on a document and a sub-headline about a panel chair's advice. To the right, there are sections for "Top Stories" and "Features" with various news items.

18 April 2014 Last updated at 17:00

NHS Care.data information scheme 'mishandled'

By Chris Vallance
PM, BBC Radio 4



The chair of the panel set up to advise the NHS and ministers on the governance of patient information has told the BBC the Care.data programme was mishandled.

Under the scheme, GP records in England will be put on a database and combined with other data to improve care.

Top Stories


- Sarkozy placed under investigation
- Met 'deleted discrimination record'
- PM warns of antibiotic resistance
- Israel vows to find teens' killers
- Facebook faces UK probe over study

Features

- It's good to talk**
The fightback against text and email 'conversations'
- Hazardous waters**
The swimming lessons blighted by sewage and corpses
- Mutant insects**
The butterflies with wings made of Sim cards
- End of the road?**
Police suspicion squeezes Syria aid convoys
- Control issues**

Related Stories

- Care.data: How did it go so wrong?
- Giant NHS database

 **Listen to the Story**  + Playlist
Morning Edition 3 min 56 sec + Download
Transcript



Cyclists look at a Ferrari parked illegally and blocking the bicycle lane off a main road in Beijing, on April 7, 2011.

Frederic J. Brown/AFP/Getty images

Share

54 Comments

Before it became China's capital in 1949, Beijing was a fairly provincial little city of 2 million people.

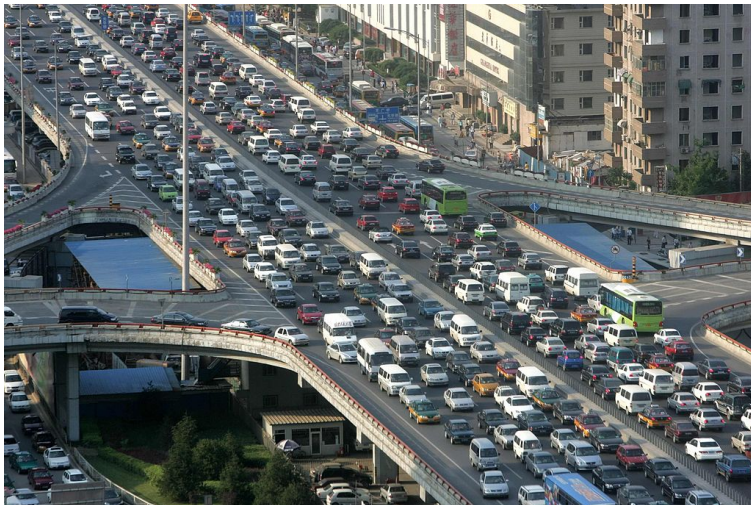


Image from [Wikimedia Commons](#)

What's Changed (Changing) for Medical Data?

- ▶ Genotyping.
- ▶ Epigenotyping.
- ▶ Transcriptome: detailed characterization of phenotype.
 - ▶ Stratification of patients.
- ▶ Massive unstructured data sources.

Open Data

- ▶ Automatic data curation: from curated data to curation of publicly available data.
- ▶ Open Data: <http://www.openstreetmap.org/?lat=53.38086&lon=-1.48545&zoom=17&layers=M>.

Open Data

- ▶ Automatic data curation: from curated data to curation of publicly available data.
- ▶ Open Data: <http://www.openstreetmap.org/?lat=53.38086&lon=-1.48545&zoom=17&layers=M>.



Open Data

- ▶ Automatic data curation: from curated data to curation of publicly available data.
- ▶ Open Data: <http://www.openstreetmap.org/?lat=53.38086&lon=-1.48545&zoom=17&layers=M>.



- ▶ Social network data, music information (Spotify), exercise.



Why Africa?

- ▶ Short circuit the process.
 - ▶ For UK—infrastructure paralysis.
 - ▶ For Africa—potential for distributed architectures.
 - ▶ User-centric models of data management.
- ▶ Store personal data on mobile phone within control of individual.

The image shows a browser window displaying the 'About us' page of citizenme.com. The browser's address bar shows the URL 'www.citizenme.com/us/'. The website's navigation menu includes 'Home', 'The app', 'About us', and 'Blog'. The main heading is 'About us' in large white text on a dark background. Below the heading, the text reads: 'citizenme was founded on the belief that people make the Internet extraordinary. Specifically, the 5 billion unique digital souls who contribute every day to the greatest utility ever invented. People like you.' Two columns of text follow, discussing digital citizenship and data transparency. The browser's taskbar at the bottom shows two open files: 'malaria.tex' and 'data_sources.tex', and a 'Show All' button.

citizenme

Home The app **About us** Blog

About us

citizenme was founded on the belief that people make the Internet extraordinary. Specifically, the 5 billion unique digital souls who contribute every day to the greatest utility ever invented. People like you.

As a digital citizen, you have an absolute right to benefit from your data contribution, to determine your online privacy, to control your digital identity. citizenme provides you with the tools to exercise these rights. The commercial future of the free Internet depends on the

certainty that your data is transparent, can be self-managed, and if you wish, transacted for your own benefit. Harvested for your data and left in the shadows? Never again. You make the Internet amazing. Value yourself.

malaria.tex data_sources.tex Show All

Outline

Diversity of Data

Massively Missing Data

Not the Scale it's the Diversity

The screenshot shows a web browser window with the URL `dataconomy.com/big-data-proving-to-be-a-real-challenge-for-data-scientists/`. The page features the Dataconomy logo and navigation menu at the top. The main article title is "Big Data Proving to Be A Real Challenge for Data Scientists", dated July 2, 2014, and written by Furhaad Shah. The article includes a large image of a person's silhouette looking at a complex data visualization of a city or network. Below the image are social media sharing icons for LinkedIn, Twitter, and Facebook. The article text discusses a survey by Paradigm4, highlighting that the challenge of big data is not just volume but the diversity of data types. A quote from Marilyn Matz, CEO of Paradigm4, is included. On the right side, there is a "Top Stories" section with two featured articles: "Predicting the World Cup with Big Data" and "Kreditech Raises \$40 Million at \$190 Million Valuation". A "Privacy & Cookies Policy" link is visible at the bottom right.

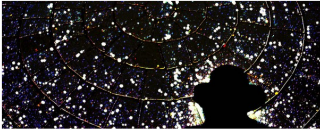
dataconomy.com/big-data-proving-to-be-a-real-challenge-for-data-scientists/

dataconomy NEWS EVENTS OPINION START UPS INDUSTRY RESOURCES ABOUT JOBS

BI & ANALYTICS MACHINE LEARNING DATA SCIENCE DATABASE TECHNOLOGY UNDERSTANDING BIG DATA

Big Data Proving to Be A Real Challenge for Data Scientists

July 2, 2014 Written by: Furhaad Shah Leave a reply



Category: Data Science, News, generalist


Tagged under: Big Data, Data Scientist, survey

In a recent survey conducted by [Paradigm4](#), a computational database company, it was revealed that big data was proving to be a challenge for data scientists – but not because of the amount, or volume, of data being produced, but rather the variety and diverse types of data these professionals have to handle.


"The increasing variety of data sources is forcing data scientists into shortcuts that leave data and money on the table," said Marilyn Matz, CEO of Paradigm4. "The focus on the volume of data hides the real challenge of data analytics today. Only diverse types of data will we be able to unlock the enormous potential of analytics."

Follow @DataconomyMedia

Top Stories



Predicting the World Cup with Big Data



Kreditech Raises \$40 Million at \$190 Million Valuation

Privacy & Cookies Policy

Scientists See Advances in Deep Learning, a Part of Artificial Intelligence

mes.com/2012/11/24/science/scientists-see-advances-in-deep-learning-a-part-of-artificial-intelligence.html

My Boosters LastPass - Dow... Add to TripIt Proverbi napol... IEEE Xplore - On... Google Maps Other Bookmarks

HOME PAGE TODAY'S PAPER VIDEO MOST POPULAR U.S. Edition

Subscribe: Digital / Home Delivery Log In Register Now Help

The New York Times **Science** Search All NYTimes.com

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION ARTS STYLE TRAVEL JOBS REAL ESTATE AUSTRALIA ENVIRONMENT SPACE & COSMOS

Scientists See Promise in Deep-Learning Programs



From [The New York Times](#)

A voice recognition program translated a speech given by Richard F. Rashid, Microsoft's top scientist, into Mandarin Chinese.

By **JOHN MARKOFF**
Published: November 23, 2012

Using an artificial intelligence technique inspired by theories about how the brain recognizes patterns, technology companies are reporting startling gains in fields as diverse as computer vision, speech recognition and the identification of promising new molecules for designing drugs.

The advances have led to widespread enthusiasm among researchers who design software to perform human

Log in to see what your friends are sharing on [Facebook](#) | [Log In With Facebook](#)
[nytimes.com](#) | [Privacy Policy](#) | [What's This?](#)

What's Popular Now

King Abdullah of Jordan Has Criticisms for All Concerned

7 Marines Killed in Nevada Training Exercise

MOST E-MAILED MOST VIEWED

1. **WELL**
Lost Sleep Can Lead to Weight Gain
2. **THIS LIFE**
The Stories That Bind Us
3. **WELL**
A New Approach to Hip Surgery
4. **Unwanted Electronic Gear Rising in Toxic Piles**
5. **DAVID BROOKS**
The Progressive Shift
6. **CONTINUING EDUCATION SPECIAL SECTION**
A Gray Jobs Market for All Ages
7. **Vatican's Bureaucracy Tests Even the Infallible**

FACEBOOK TWITTER GOOGLE+ SAVE E-MAIL SHARE PRINT

点击查看本文中文版

Connect With

Is "Deep Learning" a Rev... x Scientists See Advances x graphics - Is it possible to x

newyorker.com/online/blogs/newsdesk/2012/11/is-deep-learning-a-revolution-in-artificial-intelligence.html

My Boosters LastPass - Dow... Add to TripIt Proverbi napol... IEEE Xplore - On... Google Maps Other Bookmarks



THE NEW YORKER

SUBSCRIBE and Save up to 85%

- SUBSCRIBE
- RENEW
- ORDER A GIFT
- INTERNATIONAL ORDERS
- ONLINE ARCHIVE



SUBSCRIBE THIS WEEK'S ISSUE NEWS CULTURE POLITICS BOOKS BUSINESS CARTOONS HUMOR ARCHIVE

DOUBLE TAKE PHOTO BOOTH DAILY SHOUTS PAGE-TURNER DAILY COMMENT AMY DAVIDSON JOHN CASSIDY BOROWITZ RICHARD BRODY

THIS WEEK TOPICS | ONLINE ONLY

NEWS DESK

Reporting the latest on Washington and the world.



« How Susan Rice Sees the World | Main | Moral Machines »

NOVEMBER 29, 2012

IS "DEEP LEARNING" A REVOLUTION IN ARTIFICIAL INTELLIGENCE?

POSTED BY GARY MARCUS

f Share 877 Tweet 361 +1 1 PRINT + MORE COMMENTS

Can a new technique known as deep learning revolutionize artificial intelligence, as yesterday's front-page article at the New York Times suggests? There is good reason to be excited about deep learning, a sophisticated "machine learning" algorithm that far exceeds many of its predecessors in its abilities to recognize syllables and images. But there's also good reason to be skeptical. While the Times reports that "advances in an artificial intelligence technology that can recognize patterns



WELCOME SIGN IN | HELP | REGISTER

Search Web site Find

MOST POPULAR MOST E-MAILED THIS ISSUE

1. Andy Borowitz: Cheney Marks Tenth Anniversary of Pretending There Was Reason to Invade Iraq
2. Amy Davidson: Life After Steubenville
3. William Finnegan: Gina Rinehart, Australia's Mining Billionaire
4. Maria Bustillos: On Video Games and Storytelling: An Interview with Tom Bissell
5. Lena Dunham: Lifelong Canine Cravings

THE NEW YORKER DIGITAL

CLICK HERE

SUBSCRIBE TODAY!

THE NEW YORKER DIGITAL

f +

TABLET, MOBILE, AND MORE

Newsletter sign-up: Enter e-mail address Submit

Google To Expand Knowledge Graph Through Hire Of Geoffrey Hinton

Mar 14, 2013 • 8:23 am | (10)

by [Betsy Schwartz](#) | Filed Under [Google Search Engine](#)

If I had to place one search priority above all else, I'd say right now, Google's most ambitious project is the [knowledge graph](#). Yea, they are pushing Google+ big time, but the knowledge graph is a level above all of that technically.

Of course, Google has an outstanding team working on this project lead by one of the smartest people I've ever met Amit Singhal.

To take the knowledge graph to the next level, Google has hired/acquired Geoffrey Hinton and his team at DNNresearch. Geoffrey posted a note on his [Google+](#) page about it:



Last summer, I spent several months working with Google's Knowledge team in Mountain View, working with Jeff Dean and an incredible group of scientists and engineers who have a real shot at making spectacular progress in machine learning. Together with two of my recent graduate students, Ilya Sutskever and Alex Krizhevsky (who won the 2012 ImageNet competition), I am betting on Google's team to be the epicenter of future breakthroughs. That means we'll soon be joining Google to work with some of the smartest engineering minds to tackle some of the biggest challenges in computer science. I'll remain part-time at the University of Toronto, where I still have a lot of excellent graduate students, but at Google I will get to see what we can do with very large-scale computation.

I know we just scratched the surface of the knowledge graph and I am excited to see where it takes us in the future.

I am just glad I don't have to figure out how to get us there. I get to just sit and enjoy the ride.

[◀ PREV STORY](#) [NEXT STORY ▶](#)

49 10 16
[Tweet](#) [+1](#) [Like](#)
[SHARE](#) [Print](#)

SUBSCRIBE [t](#) [f](#) [g+](#) [r](#)

Enter Email Address [Subscribe Now](#)

[SUBSCRIBE OPTIONS >](#)

SEARCH BUZZ VIDEO [Subscribe](#)



[SUBSCRIBE >](#) [MORE VIDEOS >](#) [VIDEO DETAILS >](#)

ROUNDTABLE SPONSORS

BROWSE BY:

- [> Browse by Date](#)
- [> Find by Category](#)
- [> Discover by Author](#)
- [> Scan Most Recent](#)
- [> See Comments](#)
- [> View Tag Cloud](#)

SEM FORUM THREADS

[WebmasterWorld Forums](#)

Google Hires Geoffrey Hinton | Google Hires Brains that | Geoffrey Hinton - Google

www.wired.com/wiredenterprise/2013/03/google_hinton/

My Boosters | LastPass - Dow... | Add to TripIt | Proverbi napol... | IEEE Xplore - On... | Google Maps | Other Bookmarks

WIRED GEAR SCIENCE ENTERTAINMENT BUSINESS SECURITY DESIGN OPINION VIDEO INSIDER MAGAZINE SUBSCRIBE

ENTERPRISE | research | software | analytics

Google Hires Brains that Helped Supercharge Machine Learning

BY ROBERT MCMILLAN 03.13.13 6:30 AM

[Follow @bobmcmillan](#)

Like 270 | Tweet 216 | +1 114 | Share 34

FOLLOW WIRED ENTERPRISE

MOST RECENT WIRED POSTS

- Jawbong's Up Fitness Band Is Now Android-Compatible
- Review: Ecovacs Winbot, a Window-Cleaning Robot
- Sherlock, Professor X and Margaery Tyrell Team for Neil Gaiman Radio Play
- Video: Robo-Chopper Dives and Grabs Objects Like a Bird of Prey

Google Hires Geoffrey Hinton - Google+ | 1022889418997957626067/posts

My Boosters LastPass - Dow... Add to TripIt Proverbi napol... IEEE Xplore - On... Google Maps Other Bookmarks

Home Search Images Maps Play YouTube News Gmail Drive Calendar More -

Google+

Home Profile Explore Events Photos Communities **NEW** Post page Local Games Hangouts More

Geoffrey Hinton

1,234 have him in circles 23 in common

About Posts Photos Videos

Geoffrey Hinton 12 Mar 2013 - Public

Last summer, I spent several months working with Google's Knowledge team in Mountain View, working with Jeff Dean and an incredible group of scientists and engineers who have a real shot at making spectacular progress in machine learning. Together with two of my recent graduate students, Ilya Sutskever and Alex Krizhevsky (who won the 2012 ImageNet competition), I am betting on Google's team to be the epicenter of future breakthroughs. That means we'll soon be joining Google to work with some of the smartest engineering minds to tackle some of the biggest challenges in computer science. I'll remain part-time at the University of Toronto, where I still have a lot of excellent graduate students, but at Google I will get to see what we can do with very large-scale computation.

+1 415 167

64 comments

Reza Samahri 10 Mar 2013

Geoffrey Hinton congrats to you and your team from an old UofT eng grad. Wish I were young again to contribute to your endeavour.

Add a comment...

43 IN HIS CIRCLES

George Dahl Add

David Reichert Add

Nitish Srivastava Add

Jacqueline Ford Add

Aaron Hartzmann Add

Naveed Jaffy Add

23 IN COMMON WITH YOU

1,734 HAVE Chat

facebook's 'Deep Learnin... x

www.wired.com/wiredenterprise/2013/12/facebook-yann-lecun-qa/

Introducing Wa... LastPass - Dow... Getting Started My Boosters Add to Tri... Proverbi napol... IEEE Xplore - On... Other Bookmarks


ENTERPRISE | research | intelligence

Facebook's 'Deep Learning' Guru Reveals the Future of AI

BY CADE METZ 12:12:13 6:30 AM

Follow @cademetz

Share 452
Tweet 332
+1 115
Share 49
Print



Yann LeCun Photo: WREDU.com/Valerio

FOLLOW WIRED
ENTERPRISE

MOST RECENT WIRED POSTS

- Everything That Happened in 2013 in One Image
- Build a Bot
- Facebook Forces Video Ads on You Because Marketers Told It To
- Obama Taps Ex-Microsoft Exec to Fix Site That's Ruining His Presidency
- WIRED Space Photo of the Day: Vesta's Aella
- What Improves Every Party's Giant

Are you an expert in machine learning? Facebook is hiring

theconversation.com/are-you-an-expert-in-machine-learning-facebook-is-hiring-21438

Introducing Wa... LastPass - Dow... Getting Started My Boosters Add to Tri... Proverbi napol... IEEE Xplore - On... Other Bookmarks

Edition: AU | UK Newsletter

Become an author Sign up as a reader Sign in

THE CONVERSATION

Academic rigour. Journalistic fear.


Business + Economy Environment + Energy Health + Medicine Politics + Society **Science + Technology**

Follow Topics Nelson Mandela Hard Evidence Digital economy DNA sequencing H2O Explainer NHS Media

12 December 2013, 2:41pm GMT


Are you an expert in machine learning? Facebook is hiring

AUTHOR

 **Neil Lawrence**
Professor of Machine Learning and Computational Biology at University of Sheffield

DISCLOSURE STATEMENT


Neil Lawrence does not work for, consult to, own shares in or receive funding from any company or organisation that would benefit from the article, and has no relevant affiliations.



Do you know anything about machine learning? [View on YouTube](#)

"Move fast and break things." That is the Facebook motto plastered all over their California headquarters to remind engineers never to stop innovating. This week, the company

Sign in to Favourite
Post a Comment
Republish



AI 'cheating' scandal makes machine learning sound like a sport – it isn't

June 12, 2015 5:31am BST



Author

Neil Lawrence
Professor of Machine Learning at University of Sheffield

Disclosure statement

Neil Lawrence does not work for, consult, own shares in or receive funding from any company or organization that would benefit from this article, and has disclosed no relevant affiliations

THE CONVERSATION
Academic rigour, journalistic flair

Arts + Culture Business + Economy Education Environment + Energy Health + Medicine Politics + Society **Science + Technology**

Edition: United Kingdom - Events [Get newsletter](#) [f](#) [t](#) [r](#) Become an author Sign up as a reader Sign in

Search analysis, research, academics...

AI 'cheating' scandal makes machine learning sound like a sport – it isn't

https://theconversation.com/ai-cheating-scandal-makes-machine-learning-sound-like-a-sport-it-isn-t

Emmulated Intelligence

- ▶ There is a common thread to the applications that are 'falling over' in the face of deep learning.
- ▶ Massive data, massive compute 'perceptual' tasks.
- ▶ We are merely building algorithms that *emulate* human intelligence.
- ▶ This gives a misleading impression of achieving intelligence.

Outline

Diversity of Data

Massively Missing Data

Massive Missing Data

- ▶ If missing at random it can be marginalized.
- ▶ As data sets become very large (39 million in EMIS) data becomes extremely sparse.
- ▶ Imputation becomes impractical.

Imputation

- ▶ Expectation Maximization (EM) is gold standard imputation algorithm.
- ▶ Exact EM optimizes the log likelihood.
- ▶ Approximate EM optimizes a lower bound on log likelihood.
 - ▶ e.g. variational approximations (VIBES, Infer.net).
- ▶ Convergence is *guaranteed* to a local maxima in log likelihood.

Expectation Maximization

Require: An initial guess for missing data

Expectation Maximization

Require: An initial guess for missing data
repeat

Expectation Maximization

Require: An initial guess for missing data

repeat

 Update model parameters

(M-step)

Expectation Maximization

Require: An initial guess for missing data

repeat

 Update model parameters

(M-step)

 Update guess of missing data

(E-step)

Expectation Maximization

Require: An initial guess for missing data

repeat

 Update model parameters

(M-step)

 Update guess of missing data

(E-step)

until convergence

Imputation is Impractical

- ▶ In very sparse data imputation is impractical.
- ▶ EMIS: 39 million patients, thousands of tests.
- ▶ For most people, most tests are missing.
- ▶ M-step becomes confused by poor imputation.

Direct Marginalization is the Answer

- ▶ Perhaps we need joint distribution of two test outcomes,

$$p(y_1, y_2)$$

- ▶ Obtained through marginalizing over all missing data,

$$p(y_1, y_2) = \int p(y_1, y_2, y_3, \dots, y_p) dy_3, \dots, dy_p$$

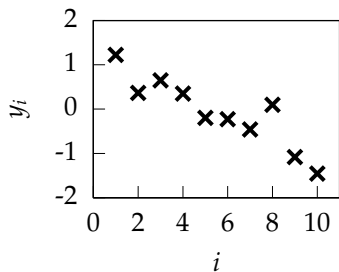
- ▶ Where y_3, \dots, y_p contains:
 1. all tests not applied to this patient
 2. all tests not yet invented!!

Magical Marginalization in Gaussians

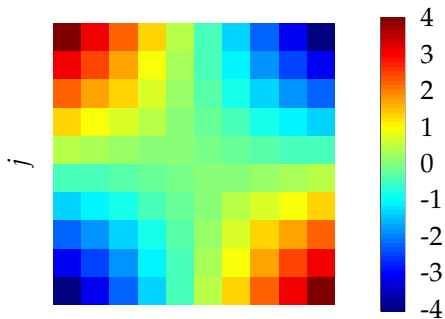
Multi-variate Gaussians

- ▶ Given 10 dimensional multivariate Gaussian, $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$.
- ▶ Generate a single correlated sample $\mathbf{y} = [y_1, y_2 \dots y_{10}]$.
- ▶ How do we find the marginal distribution of y_1, y_2 ?

Gaussian Marginalization Property



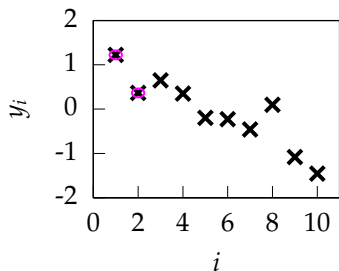
(a) A 10 dimensional sample



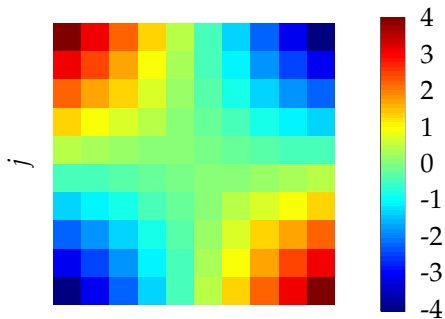
(b) colormap showing covariance between dimensions.

Figure: A sample from a 10 dimensional correlated Gaussian distribution.

Gaussian Marginalization Property



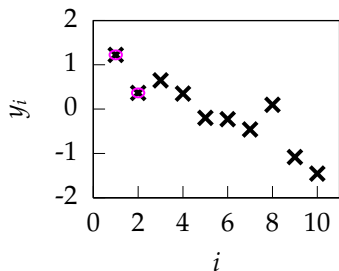
(a) A 10 dimensional sample



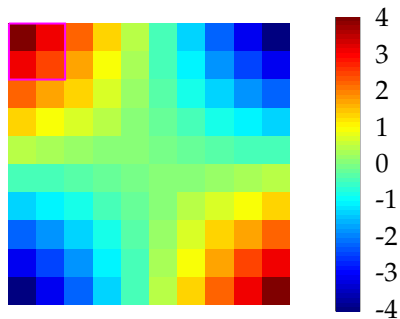
(b) colormap showing covariance between dimensions.

Figure: A sample from a 10 dimensional correlated Gaussian distribution.

Gaussian Marginalization Property



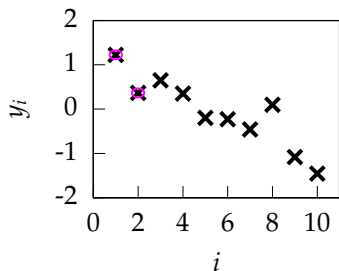
(a) A 10 dimensional sample



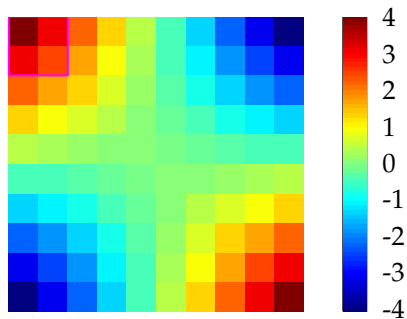
(b) colormap showing covariance between dimensions.

Figure: A sample from a 10 dimensional correlated Gaussian distribution.

Gaussian Marginalization Property



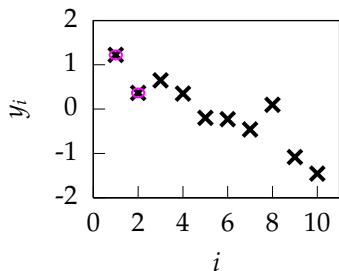
(a) A 10 dimensional sample



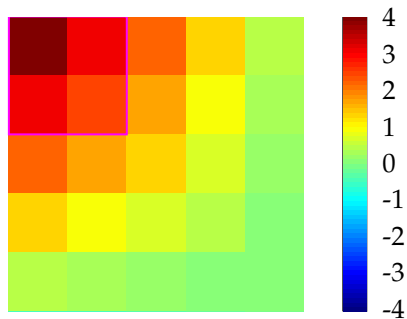
(b) colormap showing covariance between dimensions.

Figure: A sample from a 10 dimensional correlated Gaussian distribution.

Gaussian Marginalization Property



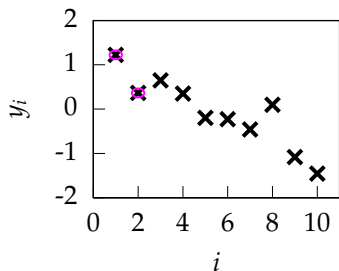
(a) A 10 dimensional sample



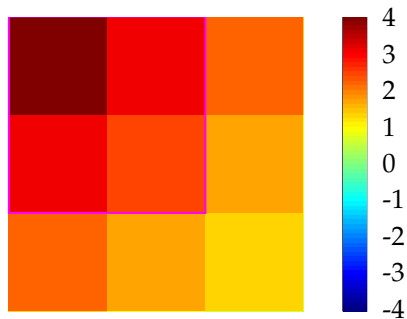
(b) colormap showing covariance between dimensions.

Figure: A sample from a 10 dimensional correlated Gaussian distribution.

Gaussian Marginalization Property



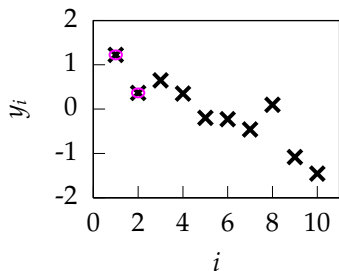
(a) A 10 dimensional sample



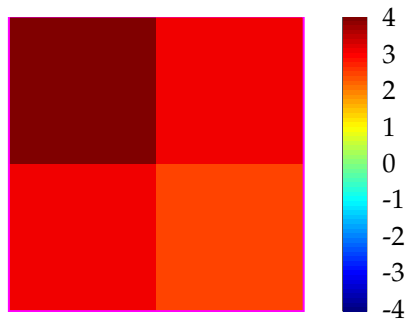
(b) colormap showing covariance between dimensions.

Figure: A sample from a 10 dimensional correlated Gaussian distribution.

Gaussian Marginalization Property



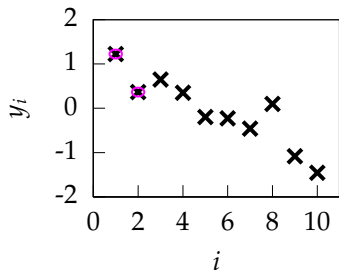
(a) A 10 dimensional sample



(b) colormap showing covariance between dimensions.

Figure: A sample from a 10 dimensional correlated Gaussian distribution.

Gaussian Marginalization Property



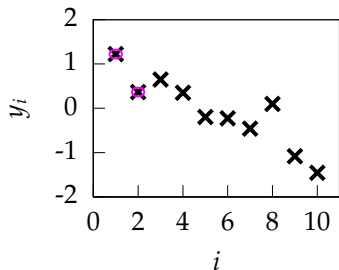
(a) A 10 dimensional sample



(b) covariance between y_1 and y_2 .

Figure: A sample from a 10 dimensional correlated Gaussian distribution.

Gaussian Marginalization Property



(a) A 10 dimensional sample



(b) correlation between y_1 and y_2 .

Figure: A sample from a 10 dimensional correlated Gaussian distribution.

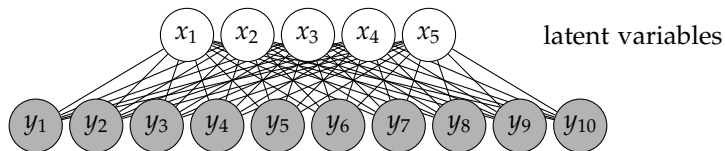
Avoid Imputation: Marginalize Directly



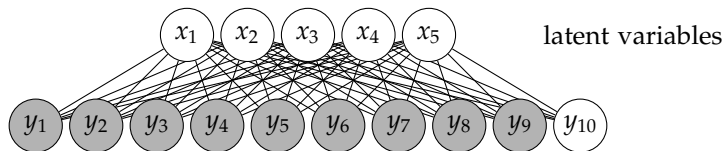
- ▶ Our approach: Avoid Imputation, Marginalize Directly.
- ▶ Explored in context of Collaborative Filtering.
- ▶ Similar challenges:
 - ▶ many users (patients),
 - ▶ many items (tests),
 - ▶ sparse data
- ▶ Implicitly marginalizes over all future tests too.

Work with Raquel Urtasun (Lawrence and Urtasun, 2009) and ongoing work with Max Zwiefsele and Nicolás Fusi.

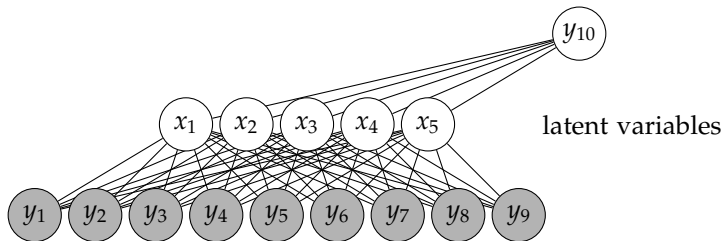
Marginalization in Bipartite Undirected Graph



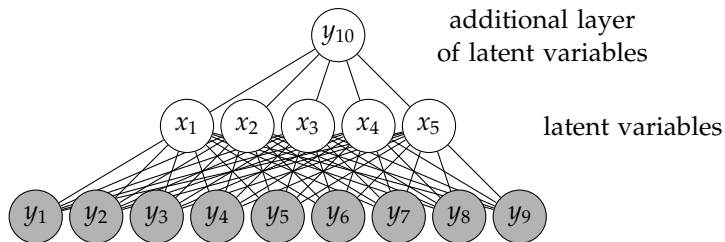
Marginalization in Bipartite Undirected Graph



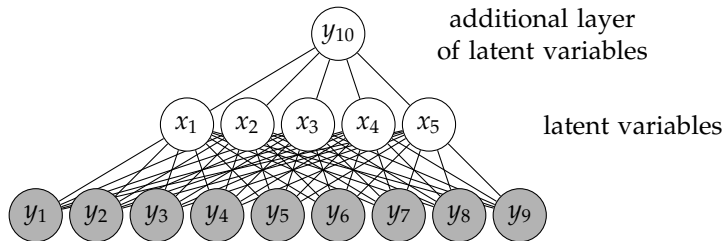
Marginalization in Bipartite Undirected Graph



Marginalization in Bipartite Undirected Graph



Marginalization in Bipartite Undirected Graph



For *massive missing data*, how many additional latent variables?

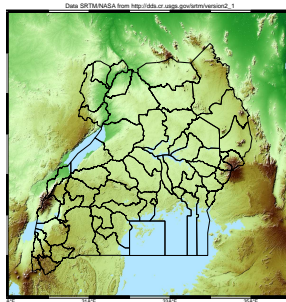
Methods that Interrelate Covariates

- ▶ Need Class of models that interrelates data, but allows for variable p .
- ▶ Common assumption: high dimensional data lies on low dimensional manifold.
- ▶ Want to retain the marginalization property of Gaussians but deal with non-Gaussian data!

Example: Prediction of Malaria Incidence in Uganda

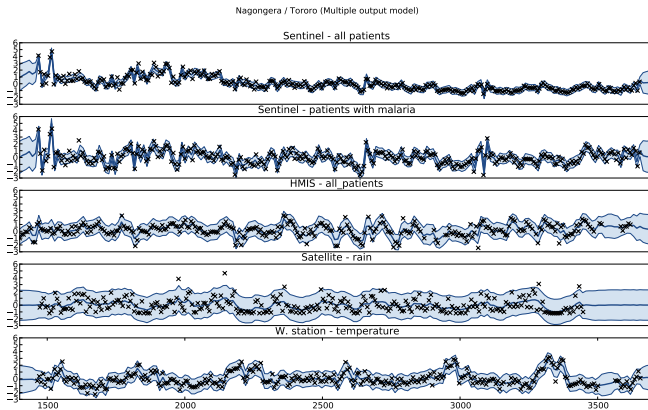
- ▶ Work with John Quinn and Martin Mubaganzi (Makerere University, Uganda)
- ▶ See <http://air.ug/research.html>.

Malaria Prediction in Uganda



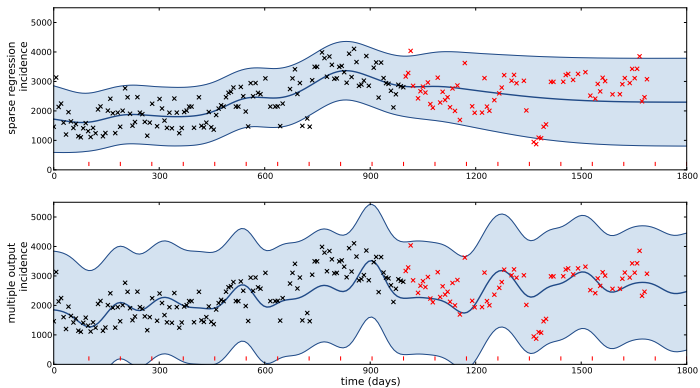
(Andrade-Pacheco et al., 2014; Mubangizi et al., 2014)

Malaria Prediction in Uganda



Malaria Prediction in Uganda

Mubende



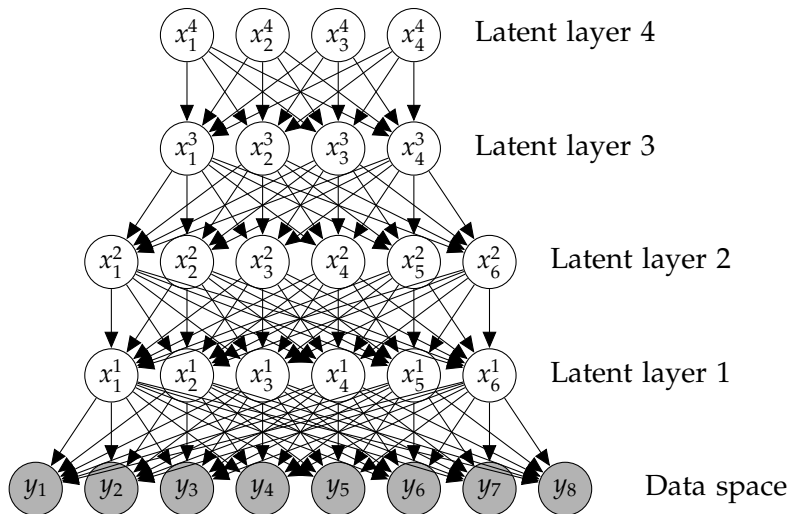
GP School at Makerere



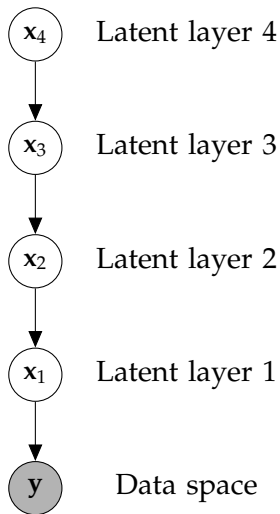
Early Warning Systems

Early Warning Systems

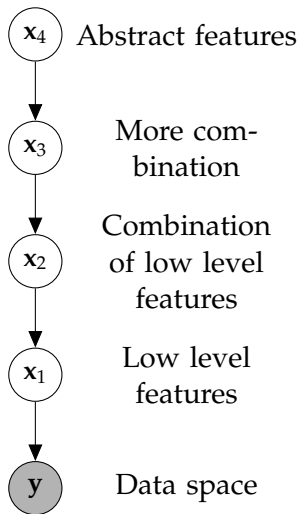
Deep Models



Deep Models



Deep Models



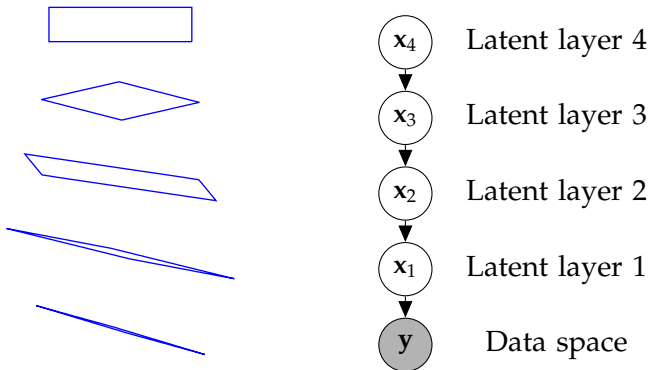
Deep Gaussian Processes



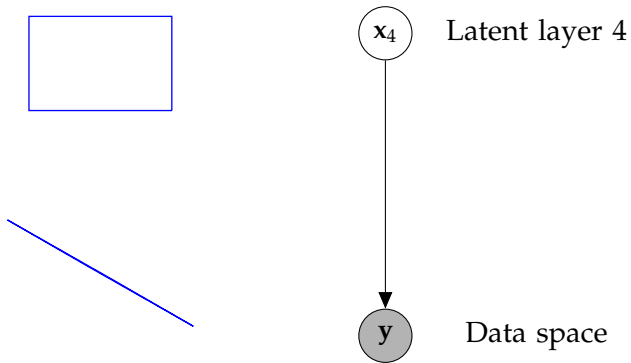
Damianou and Lawrence (2013)

- ▶ Deep architectures allow abstraction of features (Bengio, 2009; Hinton and Osindero, 2006; Salakhutdinov and Murray, 2008).
- ▶ We use variational approach to stack GP models.

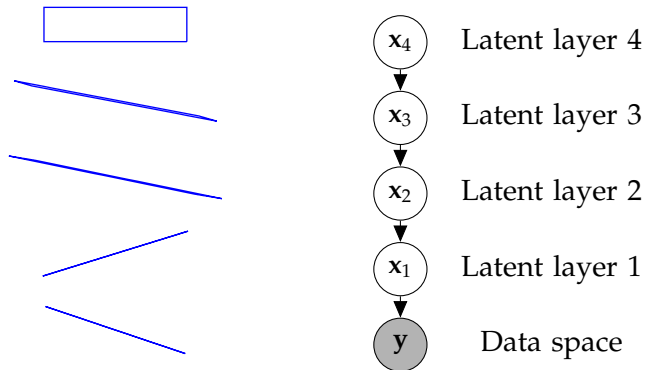
Stacked PCA



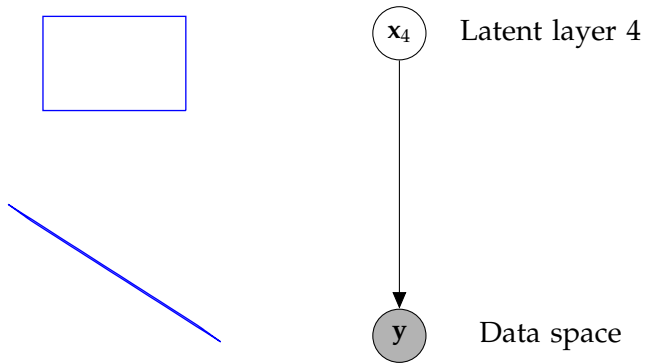
Stacked PCA



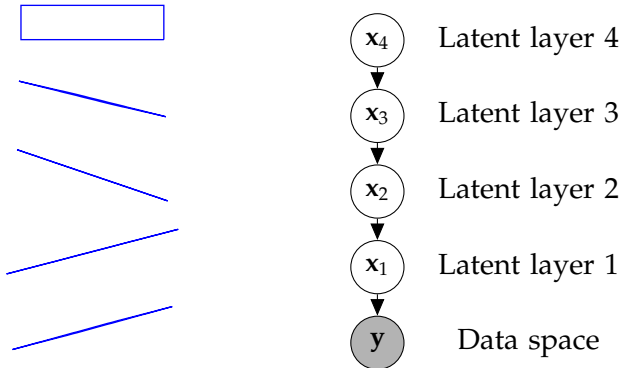
Stacked PCA



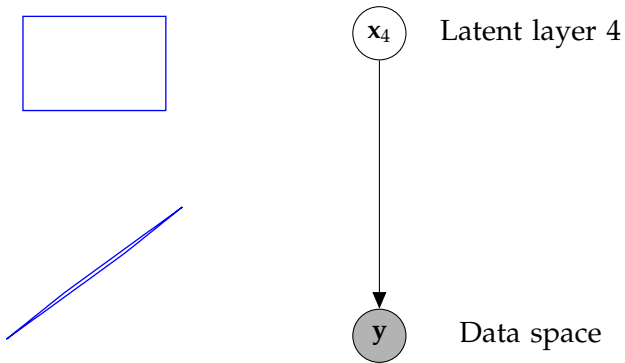
Stacked PCA



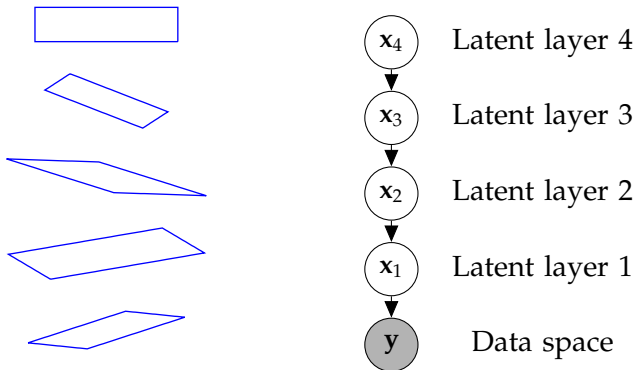
Stacked PCA



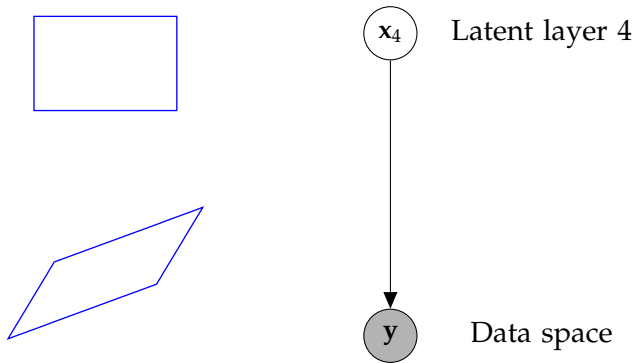
Stacked PCA



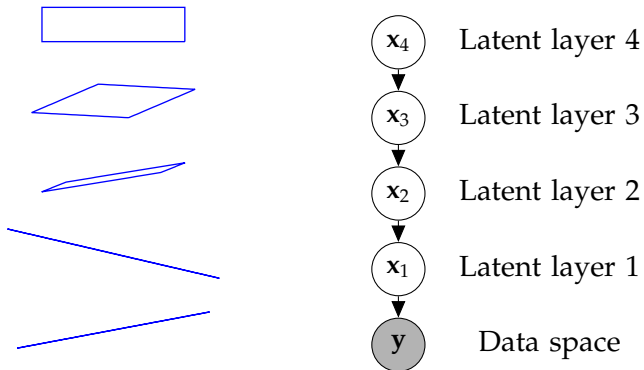
Stacked PCA



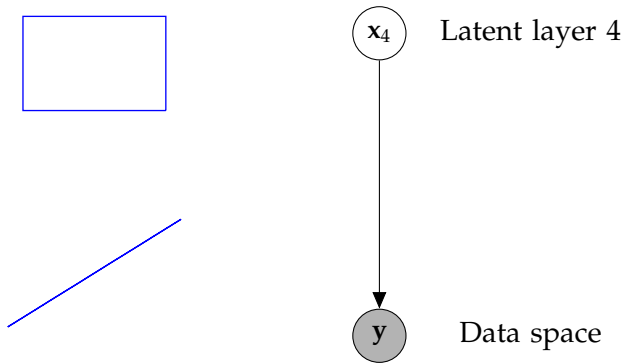
Stacked PCA



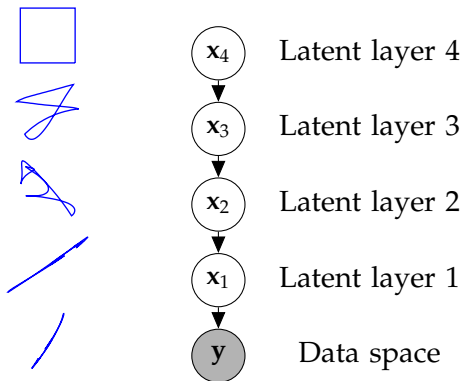
Stacked PCA



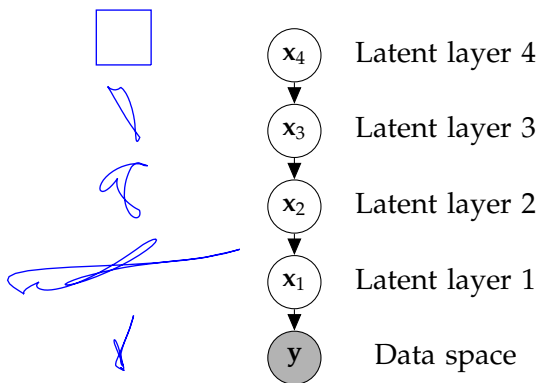
Stacked PCA



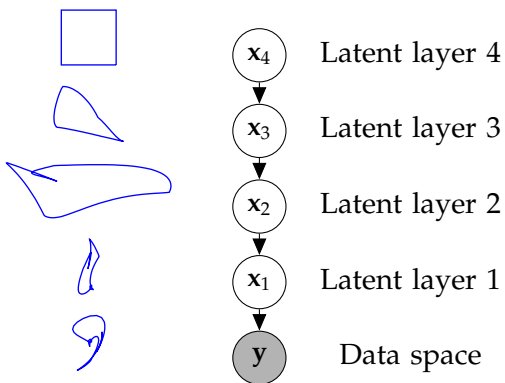
Stacked GPs



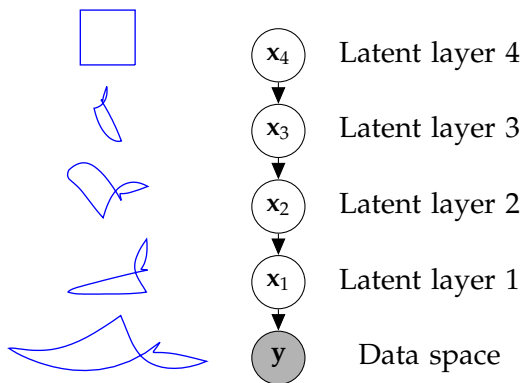
Stacked GPs



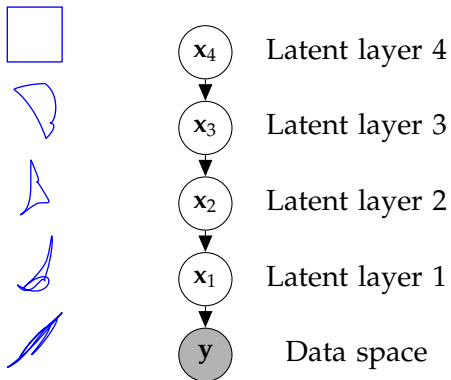
Stacked GPs



Stacked GPs



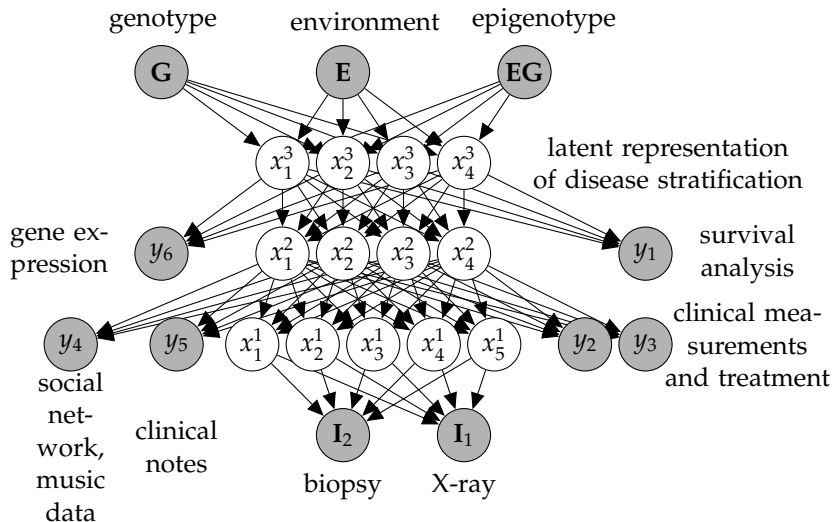
Stacked GPs



What Can Academics Do that Google Can't?

- ▶ Google's resources give them access to volumes of data (or Facebook, or Microsoft, or Amazon).
- ▶ Is there anything for Universities to contribute?
- ▶ Assimilation of multiple views of the patient: each perhaps from a different patient.
- ▶ This may be done by small companies (with support of Universities).
- ▶ A Facebook app for your personalised health.
- ▶ These methodologies are part of that picture.

Deep Health



The Patient Experience: Bedside Manner

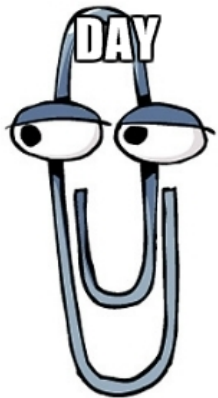
- ▶ A good bedside manner is a key part of the patient experience.
- ▶ How can information be delivered to patients?
- ▶ Public health significantly changed: tailored health advice.

Bedside Manner



Steen: Doctor and His Patient, Image from [Wikimedia Commons](#).

**YOU'RE SMOKING 20 A
DAY**



**WOULD YOU LIKE ME TO ESTIMATE YOUR YEAR
OF DEATH?**

memegenerator.net

Interventions?



Poster shown under US fair use. Copyright DreamWorks

Guided Behaviour

The screenshot shows the Amazon.co.uk homepage with the following elements:

- Navigation:** Search bar, "Shop by Department" menu, and account options like "Hello, Neil Your Account", "Try Prime", "Basket", and "Wish List".
- Product Categories:** Music Apps, LOVEFILM, Kindle, Cloud Drive, Appstore for Android, and Audible Audiobooks.
- Kindle Family Promotion:** "All-New Kindle Family" featuring Kindle Fire HDX, Kindle Fire HD, and Kindle.
- Amazon Prime:** "Free One-Day Delivery" with "No minimum order value" and "Try Prime Free for 30 Days".
- Right-Side Promotions:** "BritishRedCross" donation link, "Christmas is coming" banner, "Lightning Deals" banner, and "New 2013 Advent Calendars".
- More Items to Consider:** A section titled "Customers who viewed this also viewed" featuring books like "Learning and Inference in Neural Networks", "Systems Biology: Simulation of...", "Mathematical Modeling in Systems Biology", and "Introduction to Bioinformatics".
- Best Sellers:** A section for "Beauty - Women" featuring "Best Selling Fragrances for Women Updated hourly" with products by Sarah Jessica Parker.
- Footer:** A row of image thumbnails labeled "Minority_Report...jpg", "pl8vk.jpg", "pl8tg.jpg", and "pl8sk.jpg", followed by a "Show all downloads..." link.

Facebook Knew I was Gay

The image shows a screenshot of a web browser displaying a BuzzFeed FWD article. The browser's address bar shows the URL www.buzzfeed.com/katieheaney/facebook-knew-i-was-gay-before-my-family-did. The page features the BuzzFeed FWD logo and a navigation bar with categories like NEWS, ENTERTAINMENT, LIFE, VIDEO, and MORE. The article title is "Facebook Knew I Was Gay Before My Family Did" by Katie Heaney, posted on March 19, 2013. The article text states: "A well-placed ad led one Facebook user to wonder if the social media site was reading his text messages. In truth, he was probably outed by an algorithm." Below the article is a screenshot of a Facebook advertisement for Rick Clemons, "The Coming Out Coach," which includes a rainbow flag and the text "COMING OUT? NEED HELP? Coming Out Support: individuals, parents, spouses, kids - ALL WELCOME! CLICK, LIKE, SHARE." To the right of the article are social media sharing options and a "Connect with BuzzFeed Tech" section. Below the article is a "Suggested Page" for Rick Clemons and a "Get Our Weekly Tech Email" sign-up form. At the bottom of the browser window, there are several download thumbnails labeled "Minority_Repor... .jpg", "pl8vk.jpg", "pl8tg.jpg", and "pl8sk.jpg", along with a "Show all downloads..." link.

Facebook Knew I Was Gay

www.buzzfeed.com/katieheaney/facebook-knew-i-was-gay-before-my-family-did

Introducing Wa... LastPass - Dow... Getting Started My Boosters Add to Tri... Proverbi napol... IEEE Xplore - On... Other Bookmarks

BuzzFeed FWD LOL win omg cute trashy kawaii WTF

NEWS ENTERTAINMENT LIFE VIDEO MORE **NEW!** Travel!

Like Follow

Facebook Knew I Was Gay Before My Family Did

A well-placed ad led one Facebook user to wonder if the social media site was reading his text messages. In truth, he was probably outed by an algorithm.

posted on March 19, 2013 at 9:03pm EDT

Katie Heaney
BuzzFeed Staff

Follow @Ktheaney

Share Like 283 Tweet Email Pin it

Connect with BuzzFeed Tech

Get Our Weekly Tech Email

your email address

Sign Up!

Rick Clemons, The Coming Out Coach
Suggested Page

COMING OUT? NEED HELP?
Coming Out Support: individuals, parents, spouses, kids - ALL WELCOME! CLICK, LIKE, SHARE.

3,977 people like this. - Sponsored

A screenshot from Matt's News Feed.

The Defining Breakthrough in Next-Gen Graphics: Floating Garbage

19 Things You Probably Didn't Know About Super Mario Bros.

Minority_Repor... .jpg pl8vk.jpg pl8tg.jpg pl8sk.jpg Show all downloads...

Target told my Dad I was Pregnant

The screenshot shows a web browser window with the URL consumerist.com/2012/02/17/target-figures-out-teen-girl-is-pregnant-before-her-father-does-sends-helpful-coupons/. The page features the Consumerist logo with a silhouette of a man in a hat, a 'DONATE' button, and a 'SUBSCRIBE' button. The article is categorized under 'TARGET' and is dated February 17, 2012, by Mary Beth Quirk. The main text discusses how Target's data analysis allowed them to identify pregnant shoppers and send them coupons. A quote from the article states: "As [his] computers crawled through the data, he was able to identify about 25 products that, when analyzed together, allowed him to assign each shopper a 'pregnancy prediction' score. More important, he could also estimate her due date to within a small window, so Target could send coupons timed to very specific stages of her pregnancy." The right sidebar contains a 'BITE BACK: SUBMIT A TIP' button, a search bar, a 'DONT PANIC!' section about a new commenting system, and a 'POPULAR POSTS' section with links to various articles like 'This is What An Uncooked McRib Looks Like' and 'HOA Fines Homeowner \$5K For Planting Trees That Are Currently Too Small'.

Target Figures Out Teen Girl Is Pregnant Before Her Father Does, Sends Helpful Coupons

By Mary Beth Quirk February 17, 2012

We didn't really believe it when we were told in 7th grade that math could unlock the secrets of the universe, but after reading about the coupon-wielding power of a Target statistician, which resulted in a mighty surprise for one father of a teenage girl, we might be converts. Doesn't make math any better though.

The *New York Times* (via *Forbes*) had some time chatting with Target's statistician royale, Andrew, before he was told to zip his lips by the company. He discussed how retailers figure out how to sort out your purchases — from what you need, what you will use a coupon for and your personal preferences. Oh yeah, and they can decode if you're pregnant even before you buy diapers.

In Target's case, it all comes down to your Guest ID number tied your credit card, name, and other info, which saves all kinds of data about what you buy. Statistician Andrew mined that data and saw patterns in it, for example — women on baby registries buy larger amounts of unscented lotion around the beginning of their second trimester. Bam! Send'em some coupons for other baby items. More Andrew magic!

As [his] computers crawled through the data, he was able to identify about 25 products that, when analyzed together, allowed him to assign each shopper a "pregnancy prediction" score. More important, he could also estimate her due date to within a small window, so Target could send coupons timed to very specific stages of her pregnancy.

BITE BACK: SUBMIT A TIP

Search this website... **Search**

DONT PANIC! BETA
We are currently testing a new commenting system. Want to help? Request an invite.
LOGIN **REQUEST AN INVITE**

POPULAR POSTS

- This is What An Uncooked McRib Looks Like**
- HOA Fines Homeowner \$5K For Planting Trees That Are Currently Too Small**
- Funny How The Promo Photo Lil' Kim Is Using For Her Single Looks A Lot Like Artist's Zombie Work**
- Insider: Everyone At Best Buy Dreads 6 P.M. Thanksgiving Day Opening**

Minority_Repor... .jpg pl8vk.jpg pl8tg.jpg pl8sk.jpg Show all downloads...

The screenshot shows a web browser window displaying the Patient.co.uk website. The browser's address bar shows the URL <https://myhealth.patient.co.uk/actions/myaction/>. The website has a green header with the Patient.co.uk logo and navigation links. The main content area is divided into several sections:

- Header:** Includes the Patient.co.uk logo, a search bar, and navigation tabs for Home, Wellbeing, Health Information, Medicines, Professional Reference, Forums, Directory, and Patient Access.
- MyHealth Section:** A central dashboard for the user, Neil, showing their current and target Patient Q Score (Q77), a list of next steps for health improvement, and a section for next goals.
- Right Sidebar:** Contains a 'Why not challenge your friends?' section with social sharing options (Facebook and Twitter) and a 'Discuss your score in the forums' section with a 'Visit forums' button.
- Footer:** A download bar at the bottom of the browser window showing several image files (e.g., Steen_Doctor_a_...jpg, Minority_Repor...jpg, pl8vk.jpg, pl8tg.jpg) and a 'Show all downloads...' link.

Track your health status
Welcome back Neil
You have the following actions that require your attention.
[Set your health goals](#)

Your Patient Q Score [View details](#)
Current **Q77** → Target **Q77**
[Update Patient Q Score](#)
Your First Q Score **Q77** 0 places since joining

Next steps
Step 1: [Improve your diet](#)
Step 2: [Stop smoking](#)
Step 3: [Maintain your healthy cholesterol profile](#)
Step 4: [Maintain your healthy stress levels](#)
Step 5: [Maintain your healthy blood pressure](#)
Step 6: [Maintain your healthy exercise levels](#)
Step 7: [Maintain your healthy alcohol intake](#)
Step 8: [Maintain your healthy body weight](#)

Next goals
[Add some goals](#)

Why not challenge your friends? Share your score now!
[Share](#) [Tweet](#)

Discuss your score in the forums
[Visit forums](#)

How to Handle this?

- ▶ A *potential* answer.
 - ▶ Give the patients control of their own data.
 - ▶ Make patients the gatekeeper of what can be cross-linked.

Summary

- ▶ Intention is to deploy probabilistic machine learning for assimilating a wide range of data types in personalized health:
 - ▶ Social networking, text (clinical notes), survival times, medical imaging, phenotype, genotype, mobile phone records, music tastes, Tesco club card
- ▶ Requires population scale models with millions of features.
- ▶ May be necessary for early detection of dementia or other diseases with high noise to signal.
- ▶ Major issues in privacy and interfacing with the patient.
- ▶ But: the revolution *is* coming. We need to steer it.

References I

- R. Andrade-Pacheco, M. Mubangizi, J. Quinn, and N. D. Lawrence. Consistent mapping of government malaria records across a changing territory delimitation. *Malaria Journal*, 13, 2014. [\[DOI\]](#).
- Y. Bengio. Learning Deep Architectures for AI. *Found. Trends Mach. Learn.*, 2(1):1–127, Jan. 2009. ISSN 1935-8237. [\[DOI\]](#).
- A. Damianou and N. D. Lawrence. Deep Gaussian processes. In C. Carvalho and P. Ravikumar, editors, *Proceedings of the Sixteenth International Workshop on Artificial Intelligence and Statistics*, volume 31, AZ, USA, 2013. JMLR W&CP 31. [\[PDF\]](#).
- G. E. Hinton and S. Osindero. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:2006, 2006.
- N. D. Lawrence and R. Urtasun. Non-linear matrix factorization with Gaussian processes. In L. Bottou and M. Littman, editors, *Proceedings of the International Conference in Machine Learning*, volume 26, San Francisco, CA, 2009. Morgan Kauffman. [\[PDF\]](#).

References II

- M. Mubangizi, R. Andrade-Pacheco, M. T. Smith, J. Quinn, and N. D. Lawrence. Malaria surveillance with multiple data sources using gaussian process models. In *1st International Conference on the Use of Mobile ICT in Africa*, 9–10 Dec 2014. [\[PDF\]](#).
- R. Salakhutdinov and I. Murray. On the quantitative analysis of deep belief networks. In S. Roweis and A. McCallum, editors, *Proceedings of the International Conference in Machine Learning*, volume 25, pages 872–879. Omnipress, 2008.