Personalized Health with Gaussian Processes

Neil D. Lawrence

Universidad Nacional de Colombia
Sede Manizales

19th February 2014

Outline

Health

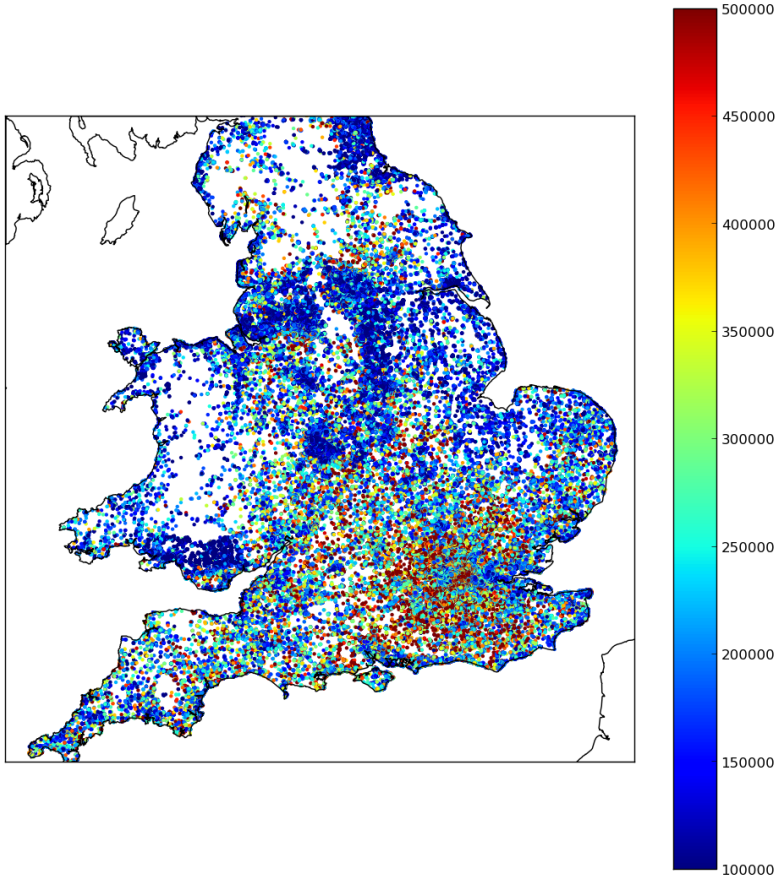
Data Heterogenity

Deep Learning

Conclusions

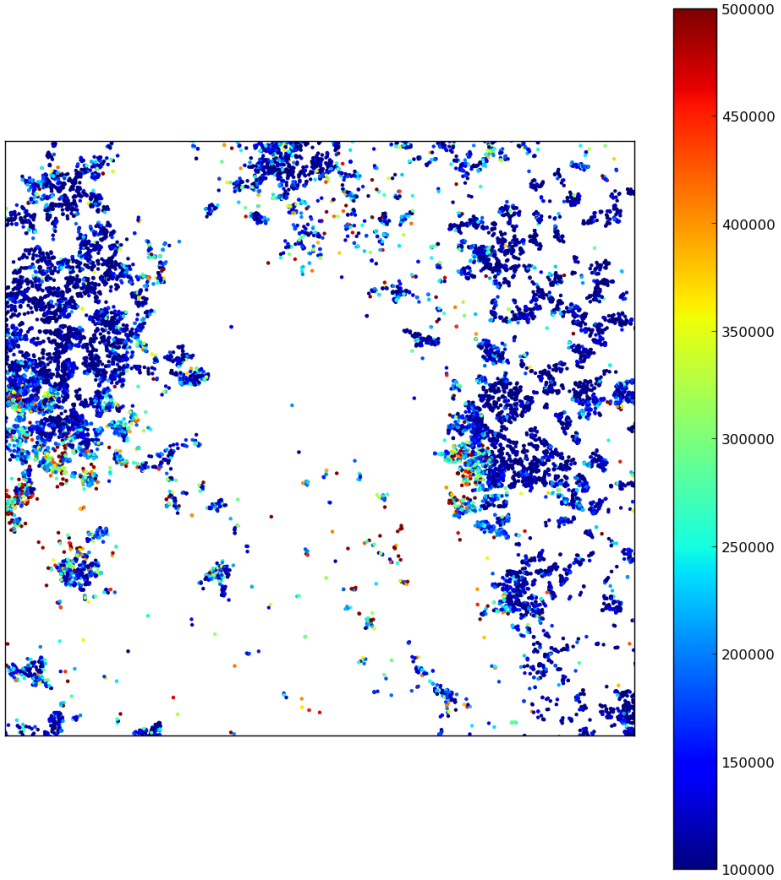
What's Changed (Changing) for Medical Data?

Modern data availability



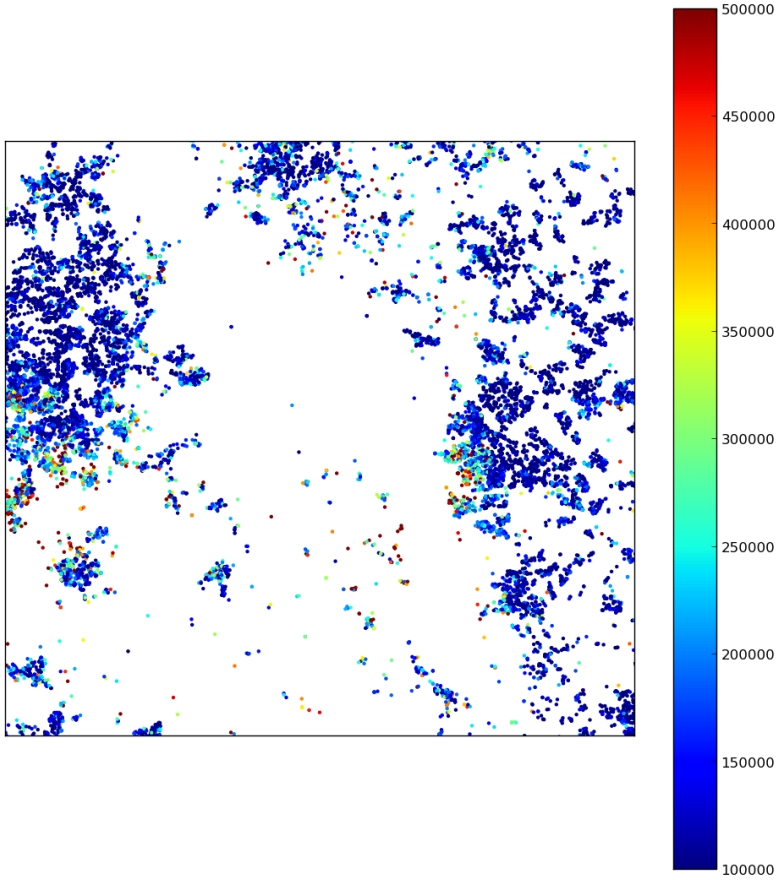
What's Changed (Changing) for Medical Data?

Proxy for index of deprivation?



What's Changed (Changing) for Medical Data?

Actually index of deprivation is a proxy for this ...



Hensman et al. (2013)

Gaussian Processes for Big Data

James Hensman*
Dept. Computer Science
The University of Sheffield
Sheffield, UK

Nicolò Fusi*
Dept. Computer Science
The University of Sheffield
Sheffield, UK

Neil D. Lawrence*
Dept. Computer Science
The University of Sheffield
Sheffield, UK

Abstract

We introduce stochastic variational inference for Gaussian process models. This enables the application of Gaussian process (GP) models to data sets containing millions of data points. We show how GPs can be variationally decomposed to depend on a set

Even to accommodate these data sets, various approximate techniques are required. One approach is to partition the data set into separate groups [e.g. Snelson and Ghahramani, 2007, Urtasun and Darrell, 2008]. An alternative is to build a low rank approximation to the covariance matrix based around 'inducing variables' [see e.g. Csató and Opper, 2002, Seeger et al., 2003, Quiñonero Candela and Rasmussen, 2005, Tit-

Hensman et al. (2013)

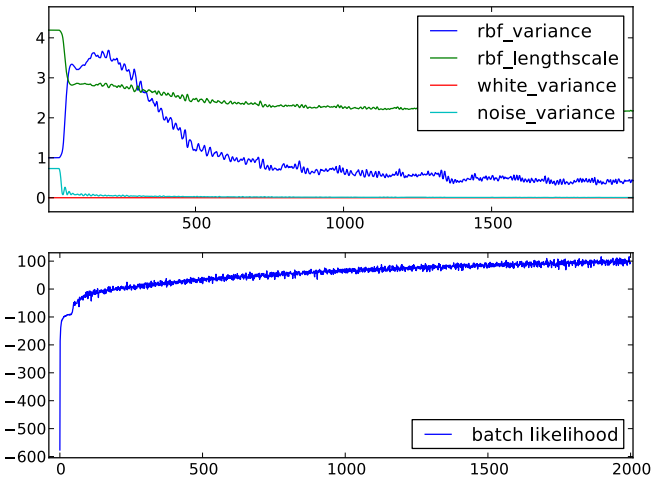


Figure 4: Convergence of the SVIGP algorithm on the two dimensional toy data

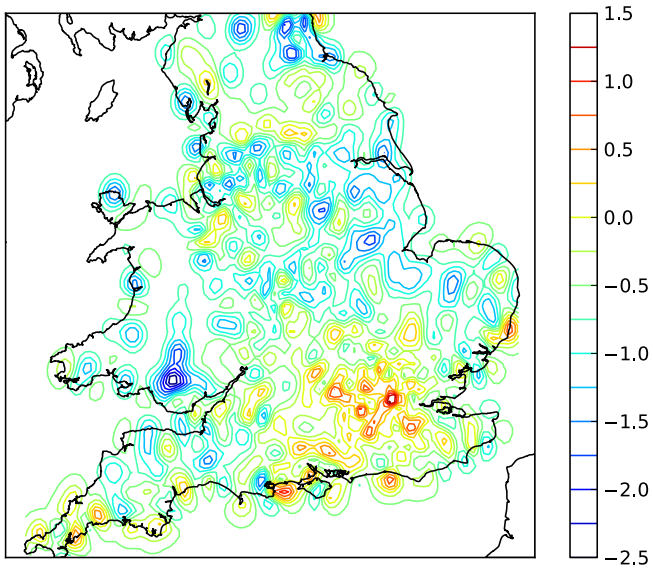


Figure 5: Variability of apartment price (logarithmically!) throughout England and Wales.

`land-registry-monthly-price-paid-data/`, which covers England and Wales, and filtered for apartments. This resulted in a data set with 75,000 entries,

ted a GP with the same covariance function as our

What's Changed (Changing) for Medical Data?

- Try Googling for: "patient data "...

Image from Wikimedia Commons

Image from Wikimedia Commons

Driver and Vehicle Licensing Agency

INF57

A brief history *of Registration*

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

A brief history of registration

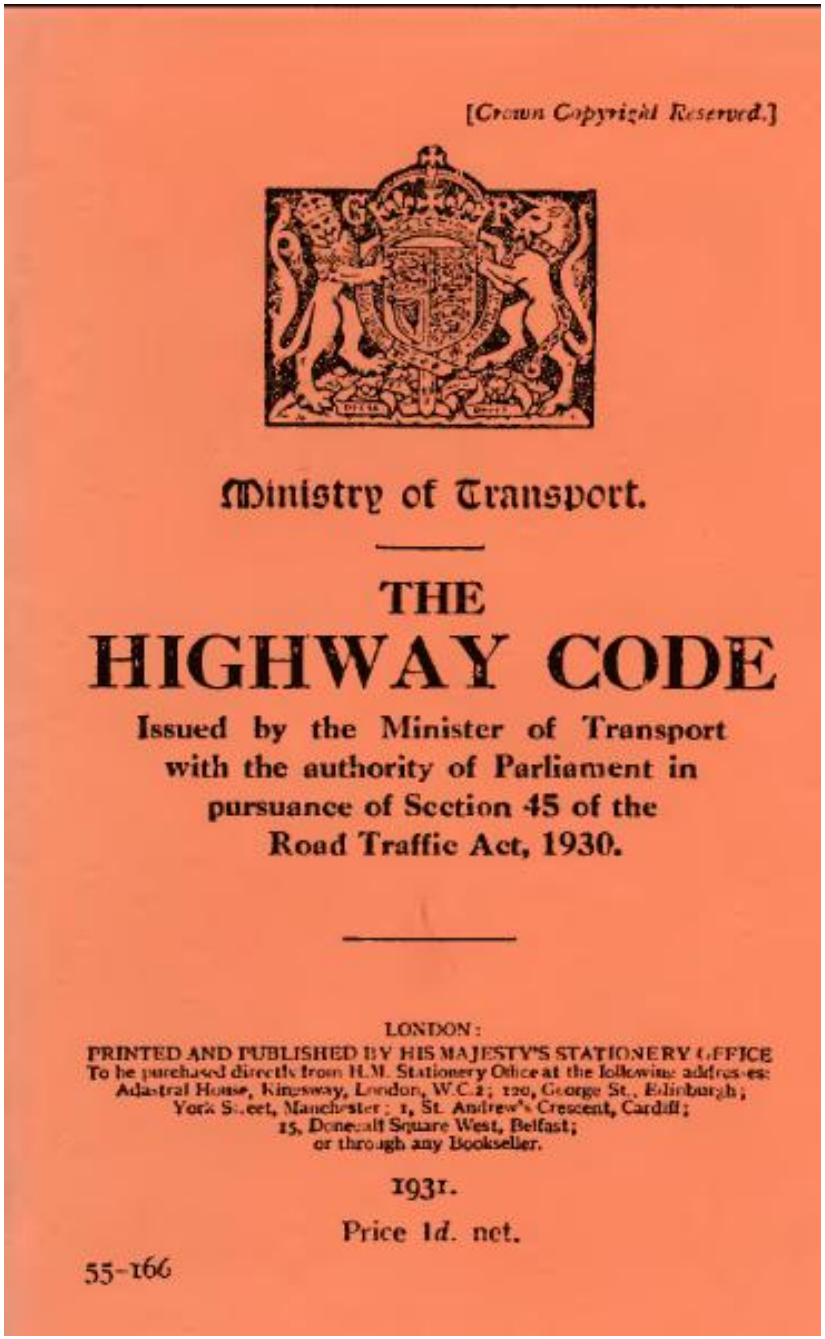
The early days

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:
Adastral 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

Image from Wikimedia Commons

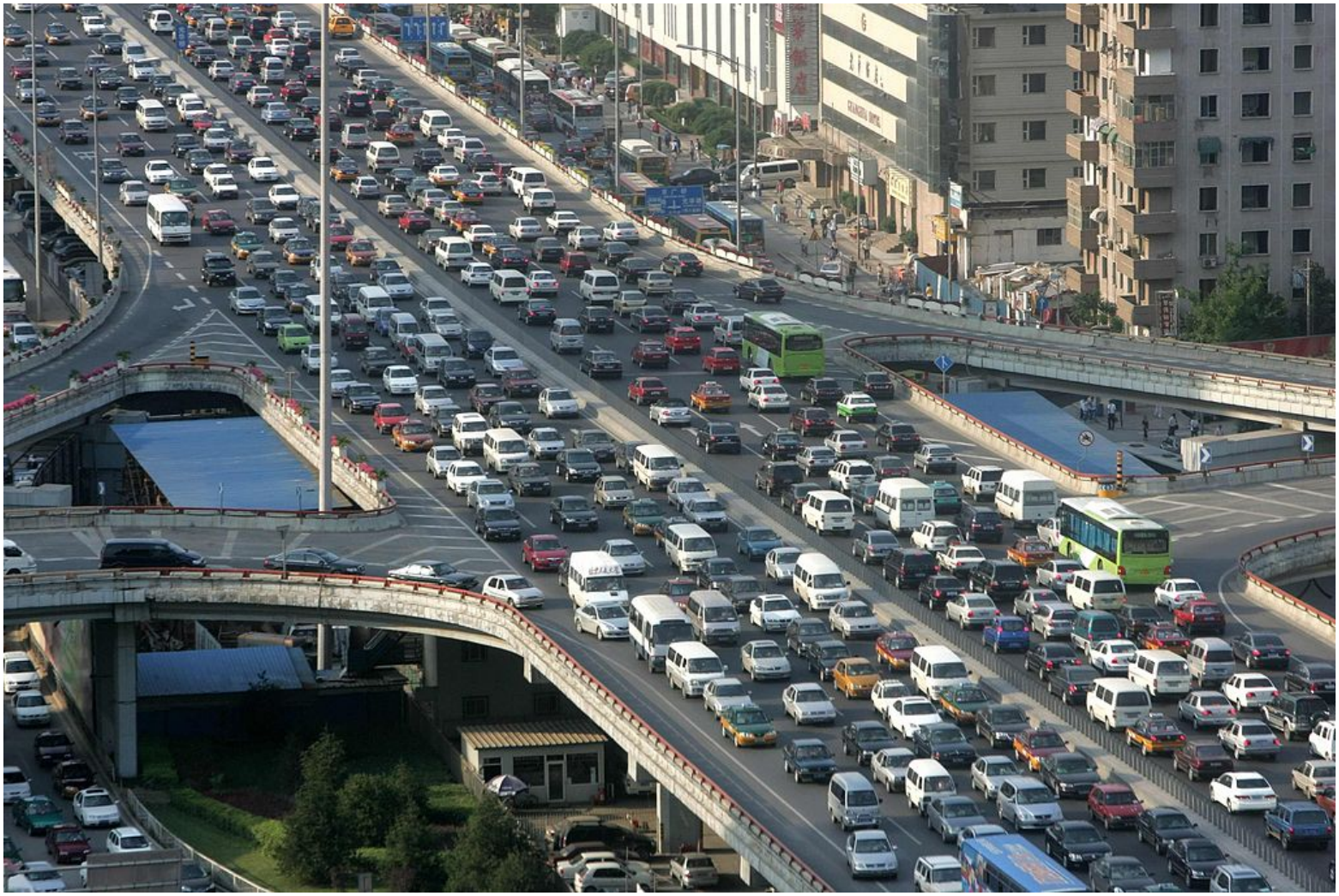


Image from Wikimedia Commons

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.

UK Government Stipulation on Data Availability

Patients will view their NH

www.telegraph.co.uk/health/healthnews/9673802/Patients-will-view-their-NHS-records-online-in-three-years.html

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

Privacy and cookies | Log in | Register | Subscribe

The Telegraph

Search - enhanced by Google

Thursday 14 November 2013

Home News World Sport Finance Comment Culture Travel Life Women Fashion Luxury Tech Dating Offers Jobs

Women | Men | Motoring | Health | Property | Gardening | Food | History | Relationships | Expat | Puzzles | Announcements | Shop

Health News | Health Advice | Diet and Fitness | Wellbeing | Expat Health | Pets Health

HOME » HEALTH » HEALTH NEWS

Patients will view their NHS records online in three years

NHS patients will be given online access to their health records in the next three years under plans to be announced by the Government today.

The move for online health records comes despite the decision by Andrew Lansley, the previous health secretary, to cancel a massive NHS national database Photo: ALAMY

By Robert Winnett, Political Editor
7:00AM GMT 13 Nov 2012

Print this article
Share 143
Facebook 76
26 Comments

Telegraph Politics

Bupa Care Services »

Promotions »

Health Most Viewed

TODAY | PAST WEEK | PAST MONTH

1. Millions more told to take statins
2. 111 line 'doing more bad than good'
3. Children's heart rates on the rise
4. NHS faces ruin and it will take brave decisions to save it
5. 'Only one region can cope with baby boom'

More from the web

Steen_Doctor_a....jpg | Minority_Repor....jpg | pl8vk.jpg | pl8tg.jpg | Show all downloads...

Patient Online: Roadmap

Patient Online: the route

www.rcgp.org.uk/news/2013/march/patient-online-launch-with-secretary-of-state.aspx

Home | About us | News | Publications | Support our work | Rooms for hire | Shop | Jobs | My RCGP

Royal College of General Practitioners

Search GO

Shop | Empty

e-Learning | Revalidation ePortfolio | Trainee ePortfolio

Revalidation | CPD | Exams | Policy | Clinical | Membership | RCGP near you

- RCGP in the news
- About the RCGP Press team
- News archive

Home › News › Patient Online launch with Secretary of State

Patient Online launch with Secretary of State

Publication date: 06 March 2013

Patient Online: the route to electronic access

(Left-right) Dr Peter Short, GP; Chris Ghush, Head of RCGP CIRC; Dr Imran Rafi, Chair of CIRC, who led the programme; Jeremy Hunt, Secretary of State for Health; Dr Clare Gerada, Chair of RCGP; and Dr Arvind Madan, GP

Patient Online: the route to electronic access

New guidance to support GP practices in providing online access for

Find courses & events

Enter keyword(s)

Topic

Region

From Date

To Date

Advanced search

Find

RSS News Feed

Steen_Doctor_a....jpg | Minority_Repor....jpg | pl8vk.jpg | pl8tg.jpg | Show all downloads...

PULSE Report

RCGP rules out full online

oday.co.uk/your-practice/practice-topics/it/rcgp-rules-out-full-online-access-to-gp-records-for-most-patients/20002168.article#.UoTHqXW9jAQ

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

This site is intended for health professionals only

Pulse Learning | Pulse Events | Pulse Live

PULSE

At the heart of general practice since 1960

Mountain medicine

How to combine a love for expeditions with a career as a GP

Search Pulse

Enter your search term | Search

Sign In | Forgotten password

Thursday 14 November 2013

Home | News | Views | Clinical | Your Practice | Commissioning | CPD on Pulse Learning | About Pulse | Jobs | Sponsored Information |

Articles | News | Practice topics | Video | Comment | Resources | Dilemmas | Working Life | Special reports | Finance diary |

Home > Your Practice > Practice topics > IT

RCGP rules out full online access to GP records for most patients

6 March 2013 | By Madlen Davies

Print | Email | Like 0 | Tweet 0 | g+1 0 | Comments (16) | Save

GPs should not be forced to give patients retrospective access to information in their medical records as it would pile work on practices and risk destabilise the doctor-patient relationship, says the RCGP in its road-map for Government plans for online records access by 2015.

The Department of Health-commissioned report says that online access to records should be 'prospective' by default and that practices should be able to set an 'access from' date for all records.

The college recommends practices assess whether access to information entered prior to this date should be allowed for patients with complex diseases and only on a 'case-by-case' basis.

The report, *Patient Online: The Road Map*, also warns of the 'unintended consequence' of an increase in queries from patients when allowing access to patient records online. It also recommends practices should be able to specify which patients were able to see their test results before the GP had reviewed them.

The GPC said the report was a blow to the Department of Health's plans to give all patients online access to their full patient record by 2015.

According to the report, while 75% of practices have the capability to provide electronic access to medical records, less than 1% of practices- 63 in total- have

RELATED ARTICLES

- GPs given slice of £240m Government funding for online access
  04 September 2013

PULSE Daily

Click here to manage your newsletter subscription for PulseToday and Pulse Learning

MOST POPULAR | MOST COMMENTED

- BMA calls for practices to work together to extend GP access in blueprint on future of general practice
- GP takes 'unlawful' decision to opt patients out of care.data programme
- Offer more same-day phone consultations for urgent patients to reduce burden on 'creaking' A&E, GPs told
- Monitor calls for greater competition between GPs and walk-in centres
- Husband appeals for safe return of missing GP

Steen_Doctor_a....jpg | Minority_Repor....jpg | pl8vk.jpg | pl8tg.jpg | Show all downloads...

EMIS Patient Access

How do I view my medical

https://patient.uservoice.com/knowledgebase/articles/214226-how-do-i-view-my-medical-record-

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

Patient.co.uk

How do I view my medical record?

← Patient Access - Medical Record

New and returning users may **sign in**

If your practice offers this service there is a link called View your medical record in the Medical Record section of the home page after you have signed in. This area of the site requires an extra sign in process so you will have to request details from your practice to gain access.

Appointments Book an appointment

Date	Time	Clinician	Location	Action
You have no appointments booked				

Medical Record

View your medical record

This link will open in another window and you will need to sign in there to view your record. Use your Access user details and security word. When finished remember to sign out and close the window.

Repeat Prescriptions Make a request See your repeat prescriptions See requests detail

Watch a video on how to view your medical record

Patient Access - Medical Record

How do I view my medical record?

Why can't I view my medical record?

Why do I have to sign in again to view my medical record?

Search

Give feedback

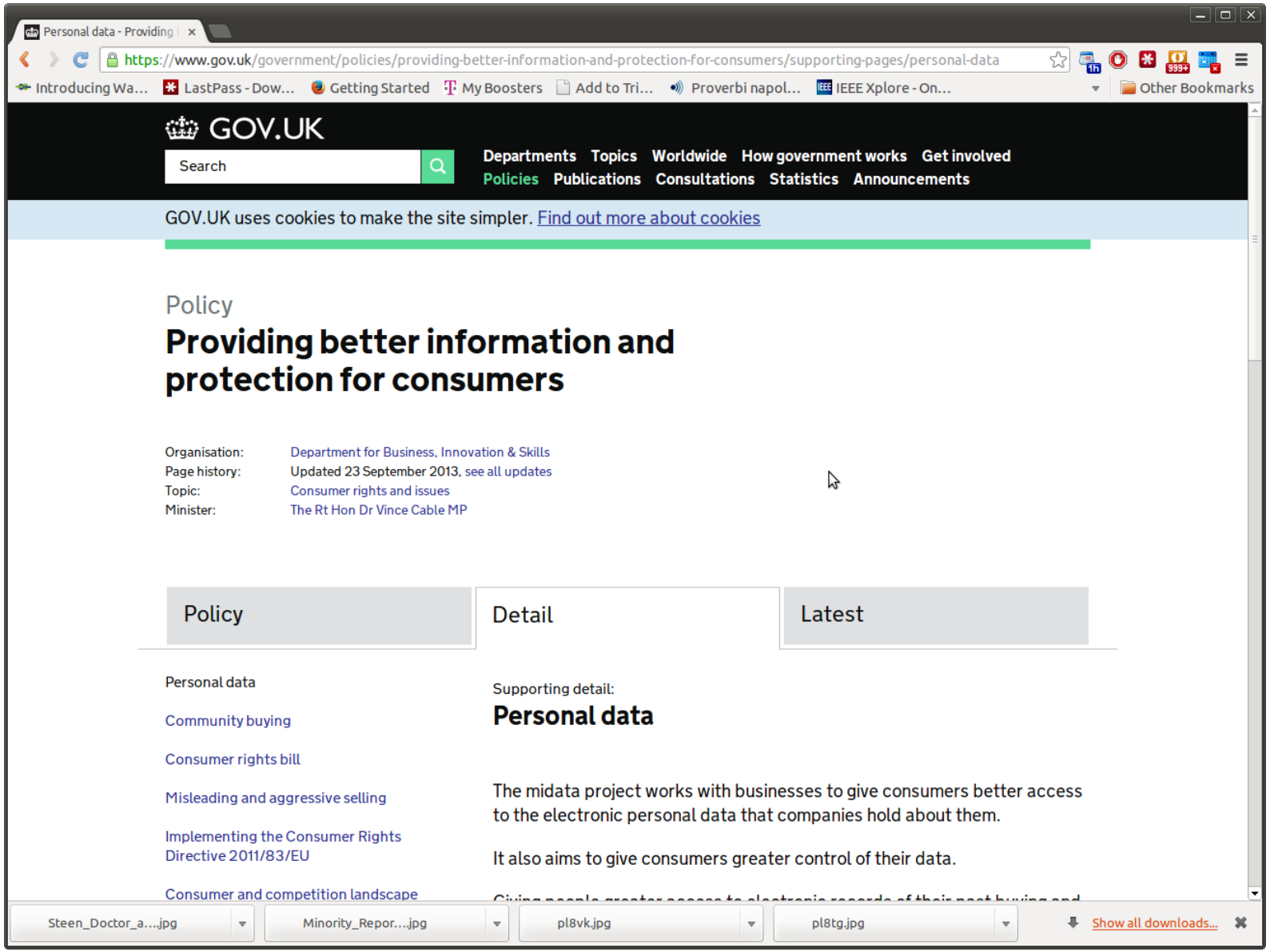
Knowledge Base

Patient Access - Getting Started 5
Patient Access - Registering 12
Patient Access - Signing In and User Details 13
Patient Access - Appointments 15
Patient Access - Prescriptions 13
Patient Access - Message Service 4
Patient Access - Medical Record 3
Patient Access - Reporting Issues 2
Patient Access - Videos 1
MyHealth 1
Advertising 4
All articles

Patient.co.uk

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

midata project



Outline

Health

Data Heterogenity

Deep Learning

Conclusions

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.

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,

$$p(y_1, y_2) = \int p(y_1, y_2, y_3, \ldots, y_p) \mathrm{d}y_3, \ldots \mathrm{d}y_p$$

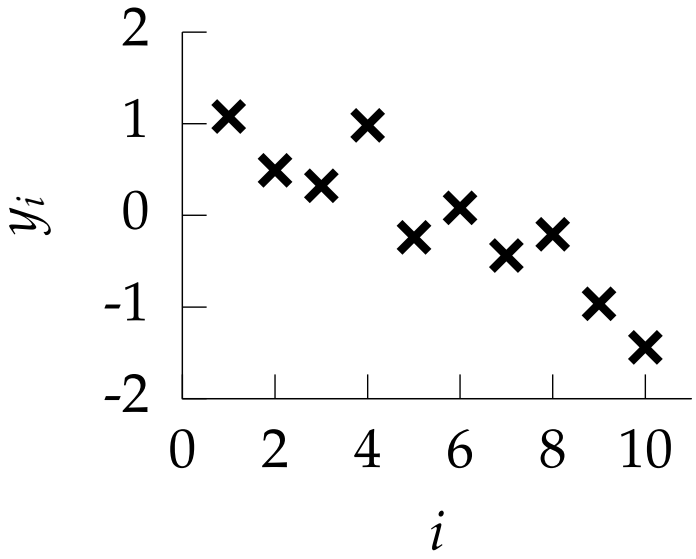
- Where $y_3, \ldots, y_p$ contains:
  1. all tests not applied to this patient
  2. all tests not yet invented!!

Magical Marginalization in Gaussians

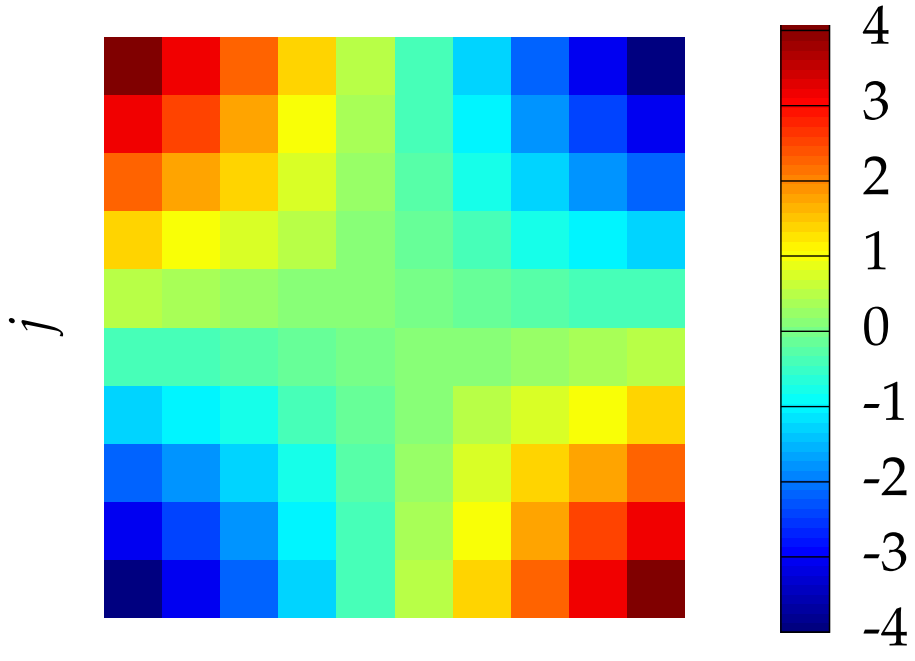
Multi-variate Gaussians

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

Gaussian Marginalization Property



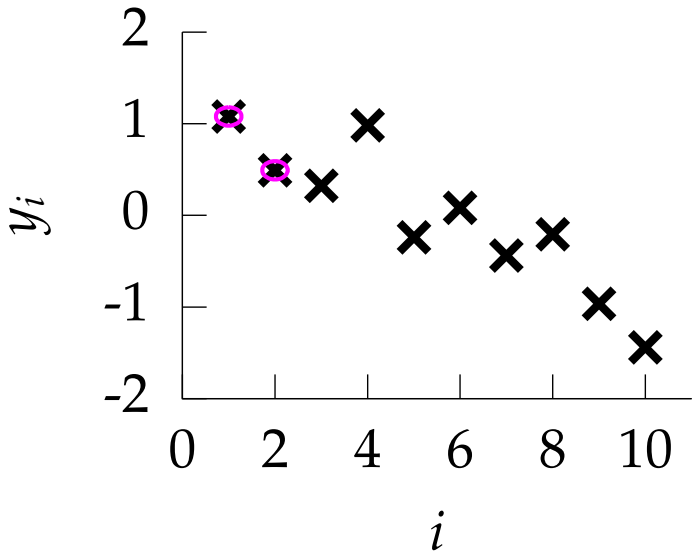
(a) A 10 dimensional sample



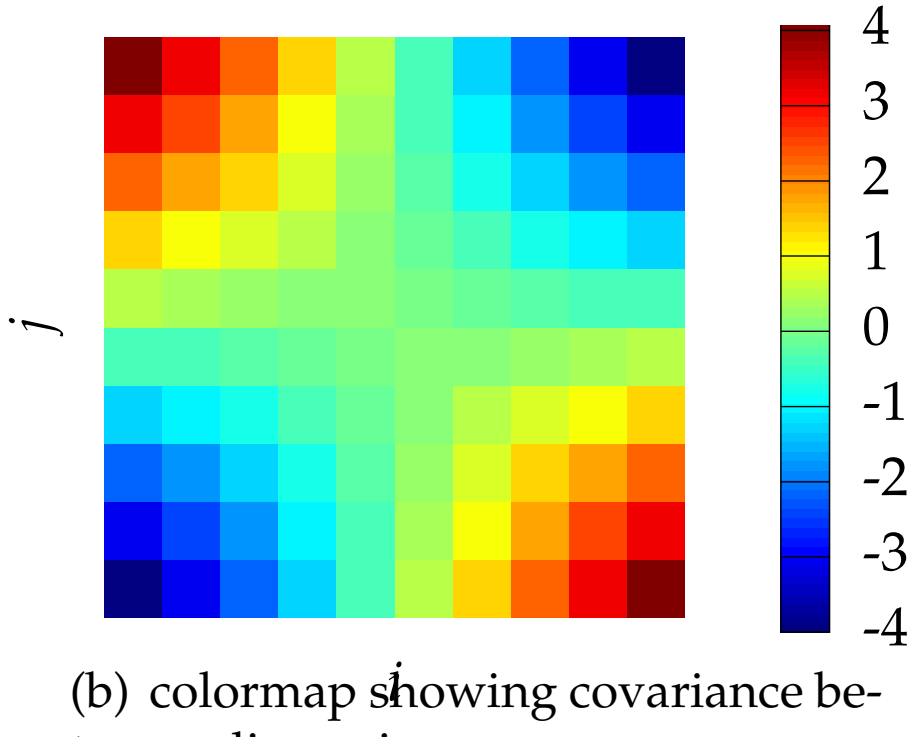
(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property



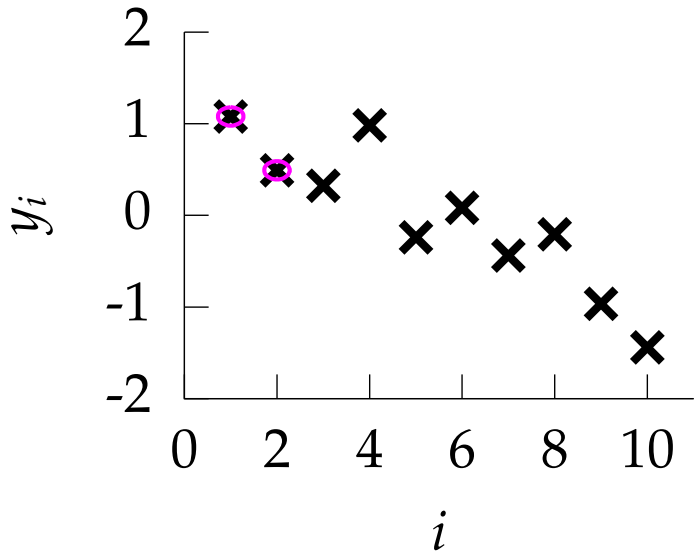
(a) A 10 dimensional sample



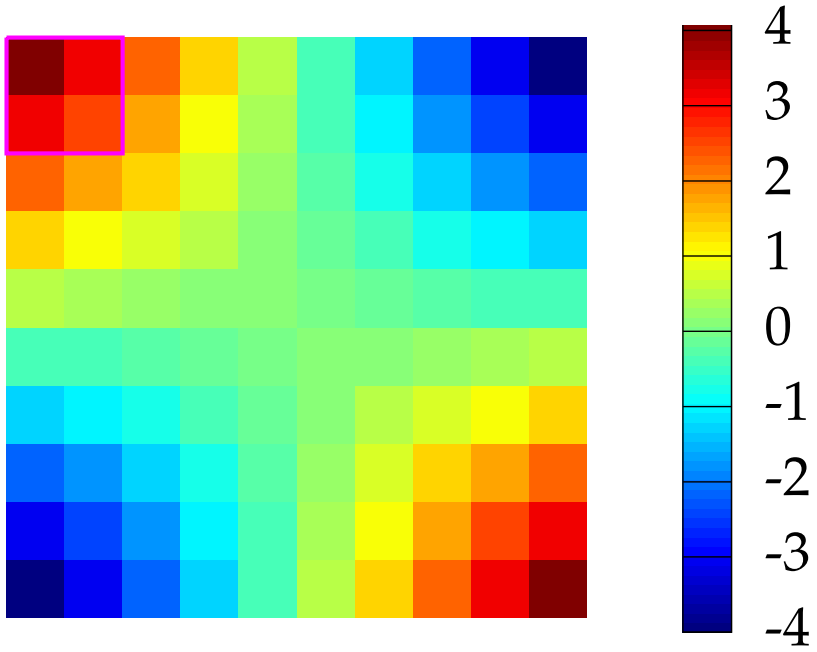
(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property



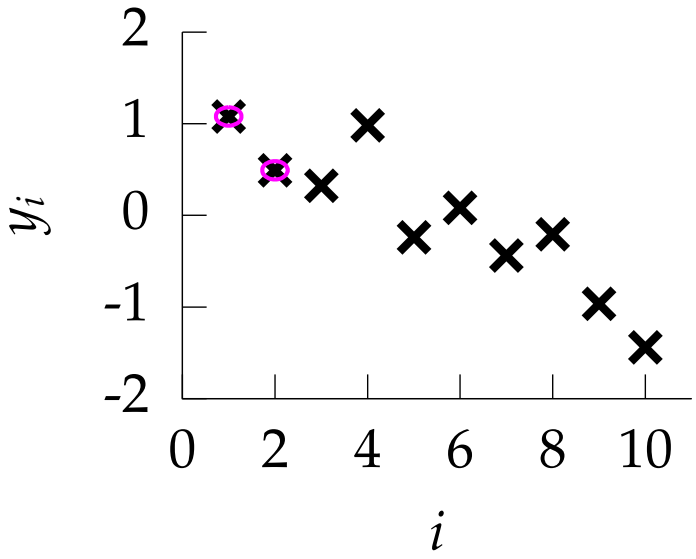
(a) A 10 dimensional sample



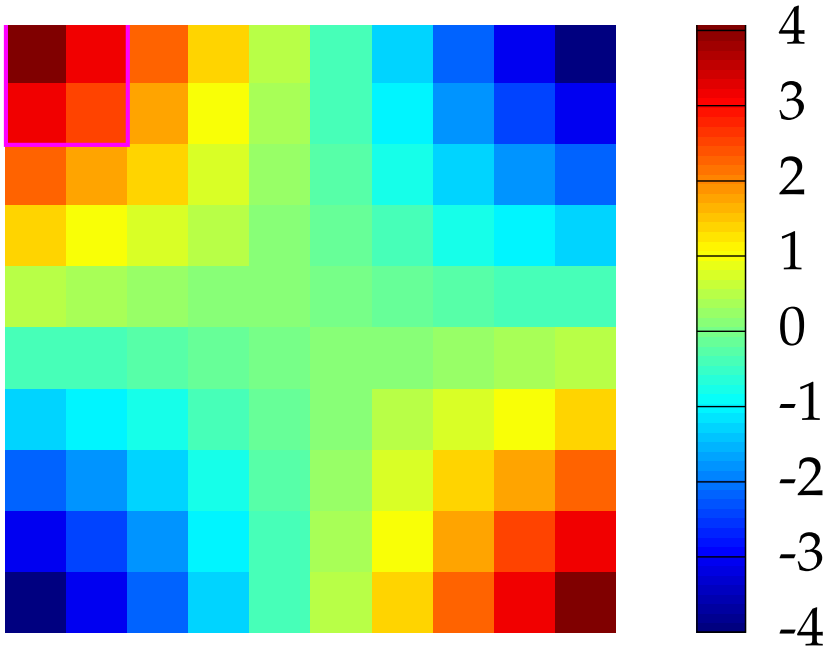
(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property



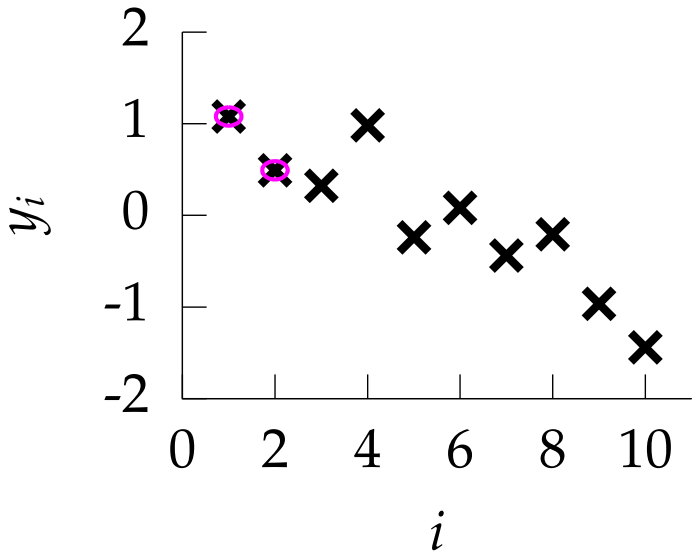
(a) A 10 dimensional sample



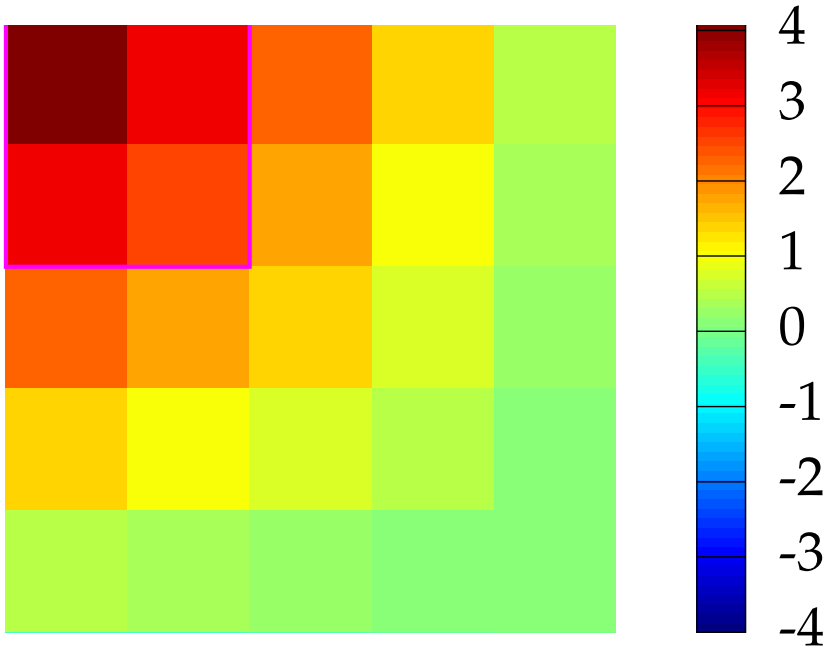
(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property



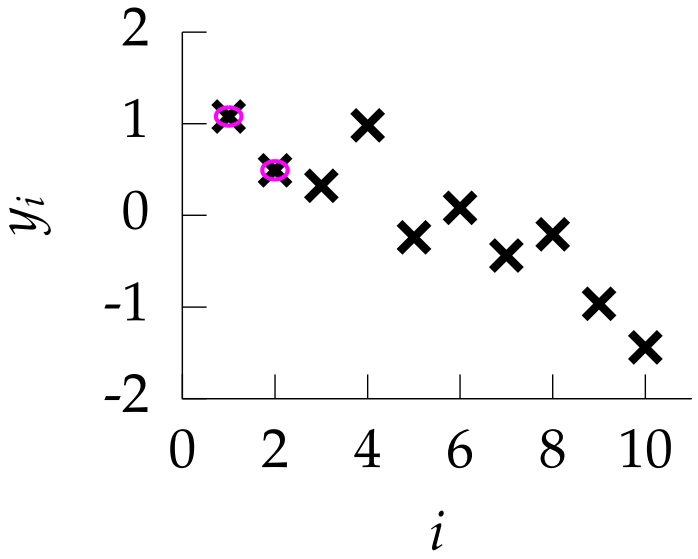
(a) A 10 dimensional sample



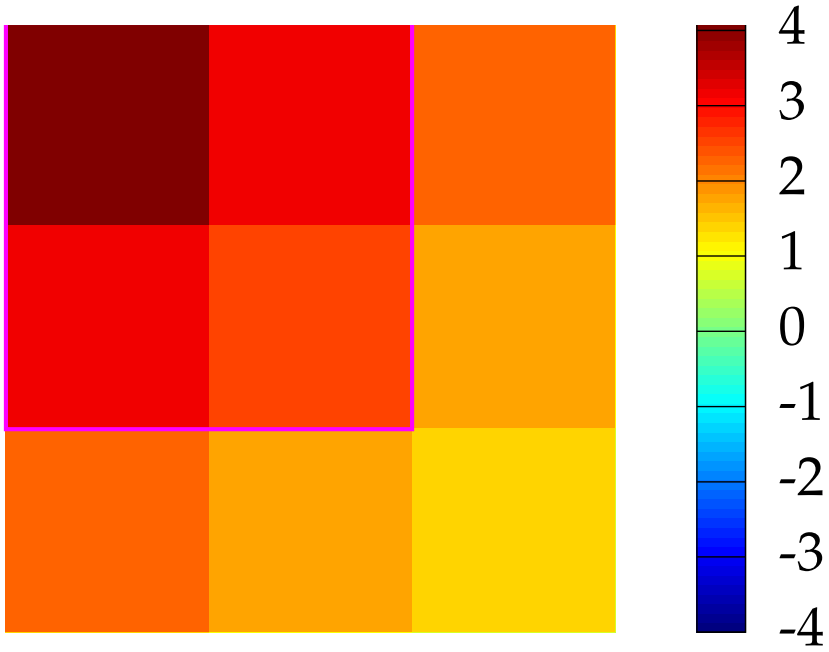
(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property



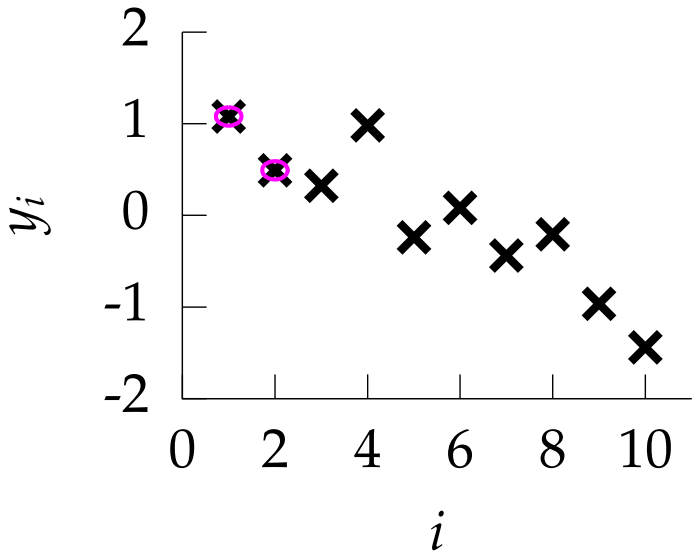
(a) A 10 dimensional sample



(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property



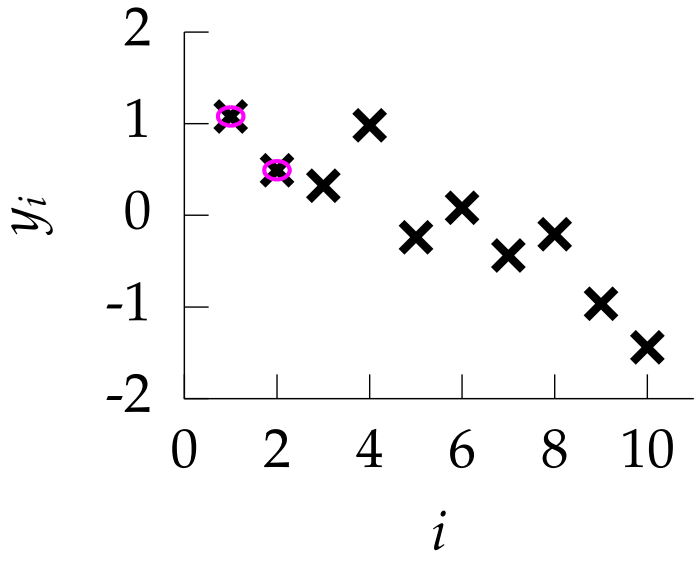
(a) A 10 dimensional sample

$$\begin{bmatrix} 1 & 0.96793 \\ 0.96793 & 1 \end{bmatrix}$$

(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property



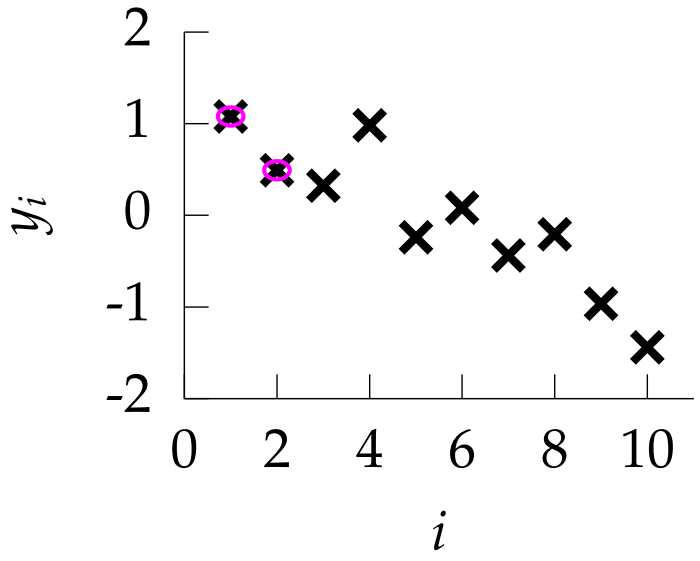
(a) A 10 dimensional sample

$$\begin{bmatrix} 4.1 & 3.1111 \\ 3.1111 & 2.5198 \end{bmatrix}$$

(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

$$\begin{bmatrix} 1 & 0.96793 \\ 0.96793 & 1 \end{bmatrix}$$

(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 recent submission with Nicoló Fusi.

Methods that Interrelate Covariates

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

Linear Dimensionality Reduction

Linear Latent Variable Model

- Represent data, $\mathbf{Y}$, with a lower dimensional set of latent variables $\mathbf{X}$.
- Assume a linear relationship of the form

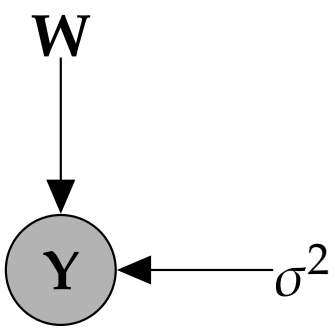
$$\mathbf{y}_{i,:} = \mathbf{W}\mathbf{x}_{i,:} + \boldsymbol{\epsilon}_{i,:},$$

where

$$\boldsymbol{\epsilon}_{i,:} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}\right).$$

Linear Latent Variable Model II

Probabilistic PCA Max. Likelihood Soln (Tipping and Bishop, 1999)



$$p\left(\mathbf{Y}|\mathbf{W}\right) = \prod_{i=1}^{n} \mathcal{N}\left(\mathbf{y}_{i,:}|\mathbf{0}, \mathbf{W}\mathbf{W}^{\top} + \sigma^2\mathbf{I}\right)$$

Linear Latent Variable Model II

Probabilistic PCA Max. Likelihood Soln (Tipping and Bishop, 1999)

$$p\left(\mathbf{Y}|\mathbf{W}\right) = \prod_{i=1}^{n} \mathcal{N}\left(\mathbf{y}_{i,:}|\mathbf{0}, \mathbf{C}\right), \quad \mathbf{C} = \mathbf{W}\mathbf{W}^{\top} + \sigma^2\mathbf{I}$$

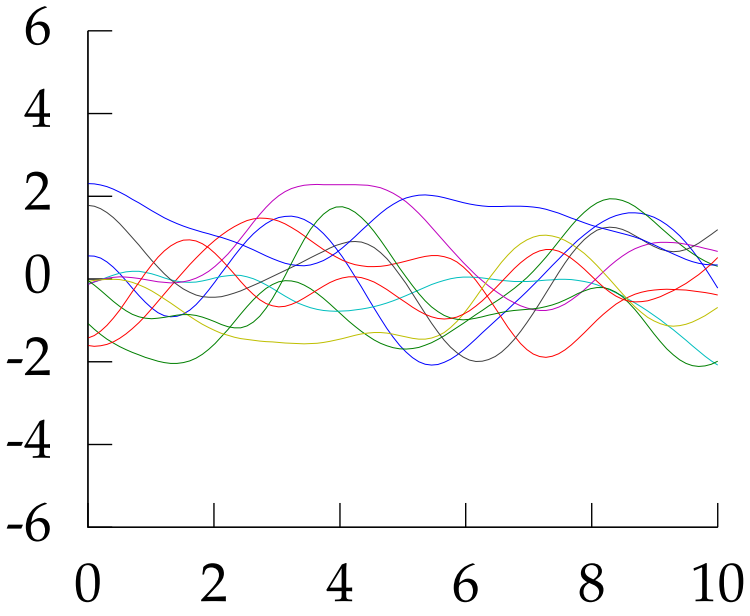
$$\log p\left(\mathbf{Y}|\mathbf{W}\right) = -\frac{n}{2}\log|\mathbf{C}| - \frac{1}{2}\mathrm{tr}\left(\mathbf{C}^{-1}\mathbf{Y}^{\top}\mathbf{Y}\right) + \text{const.}$$

If $\mathbf{U}_q$ are first q principal eigenvectors of $n^{-1}\mathbf{Y}^{\top}\mathbf{Y}$ and the corresponding eigenvalues are $\mathbf{\Lambda}_q$,

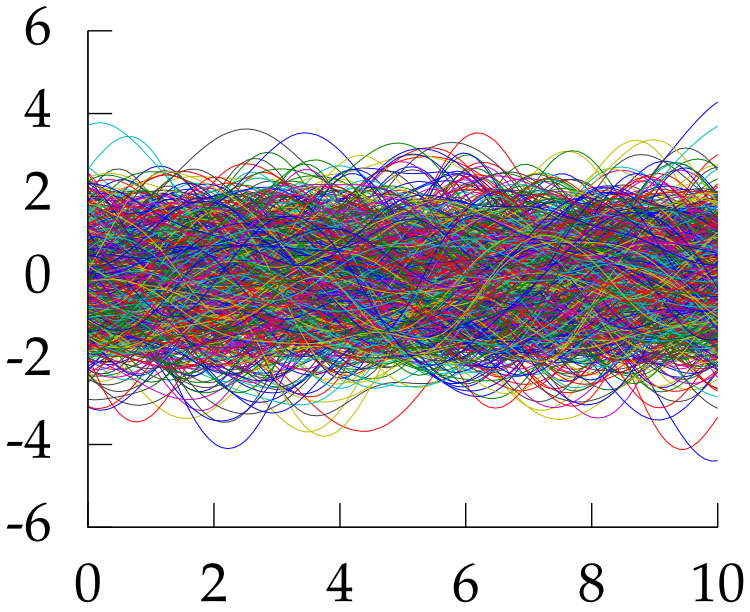
$$\mathbf{W} = \mathbf{U}_q\mathbf{L}\mathbf{R}^{\top}, \quad \mathbf{L} = \left(\mathbf{\Lambda}_q - \sigma^2\mathbf{I}\right)^{\frac{1}{2}}$$

where $\mathbf{R}$ is an arbitrary rotation matrix.

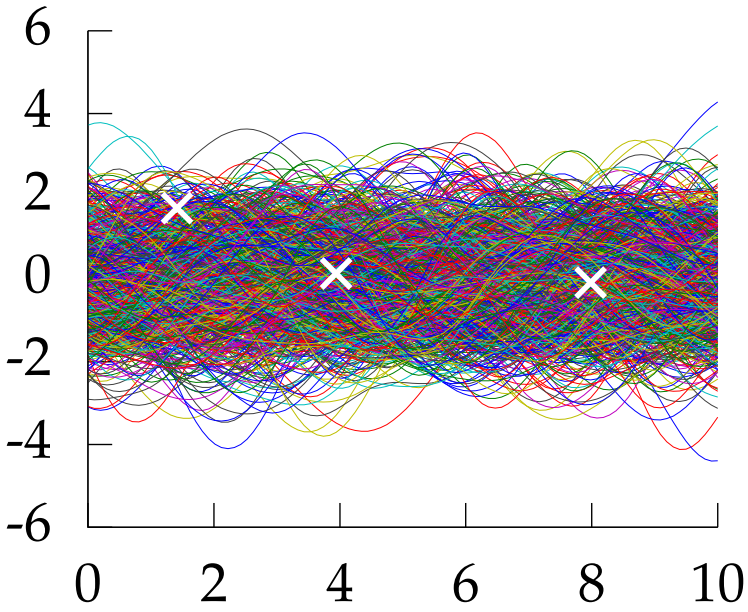
Gaussian Processes



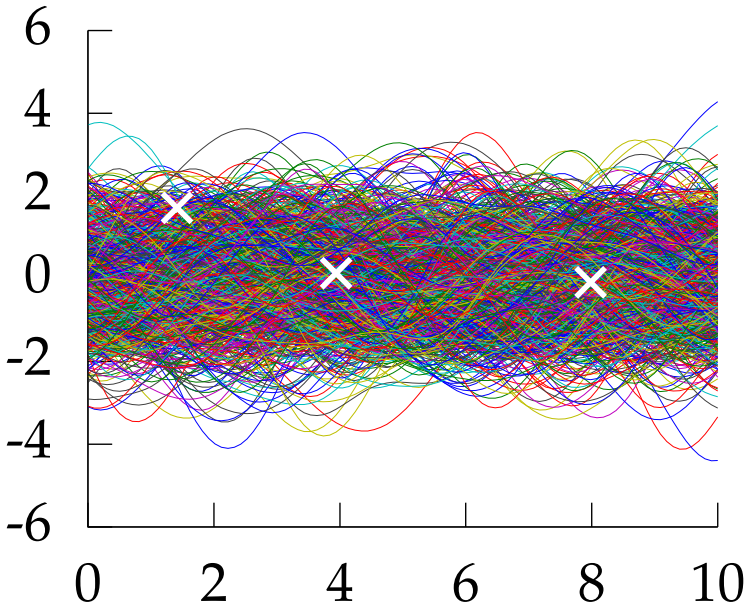
Gaussian Processes

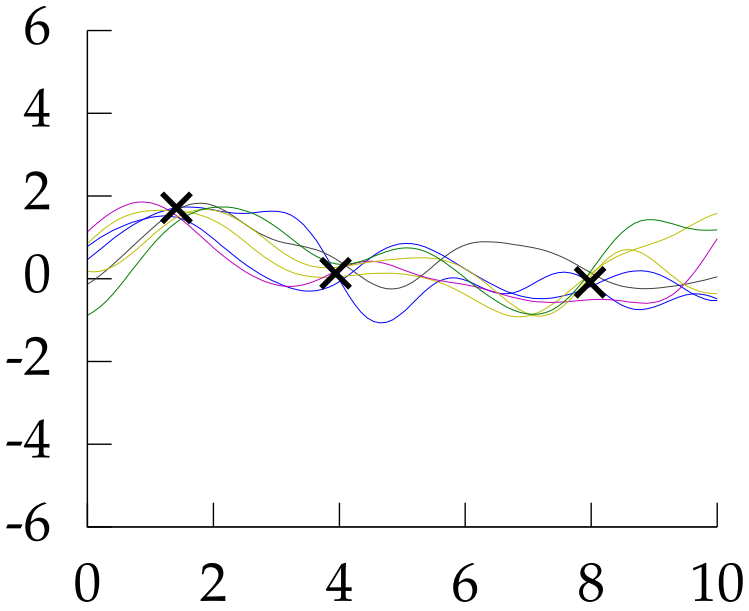


Gaussian Processes



Gaussian Processes





Dealing with Non Gaussian Data

- Marginalization property of Gaussians very attractive.
- How to incorporate non-Gaussian data?
  - Data which isn't missing at random.
  - Binary data.
  - Ordinal categorical data.
  - Poisson counts.
  - Outliers.

Project Back into Gaussian

- Combine non-Gaussian likelihood with Gaussian prior.
- Either:
  - Project back to Gaussian posterior that is nearest in KL sense.
  - Expectation propagation.
- Or:
  - Fit a locally valid Gaussian approximation.
  - Laplace Approximation.

Ongoing work with Ricardo Andrade Pacheco (EP) and Alan Saul (Laplace) also James Hensman.

Gaussian Noise

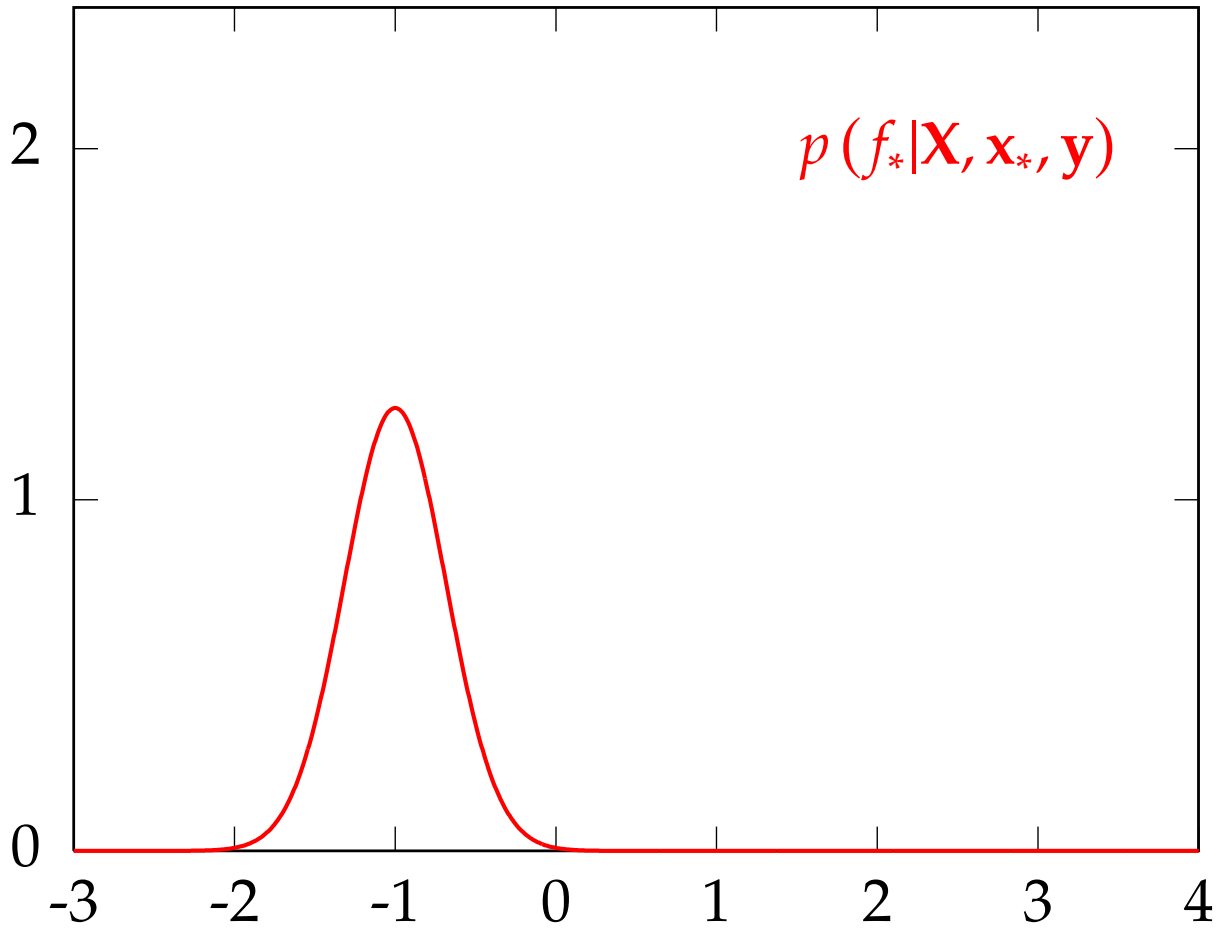


Figure : Inclusion of a data point with Gaussian noise.

Gaussian Noise

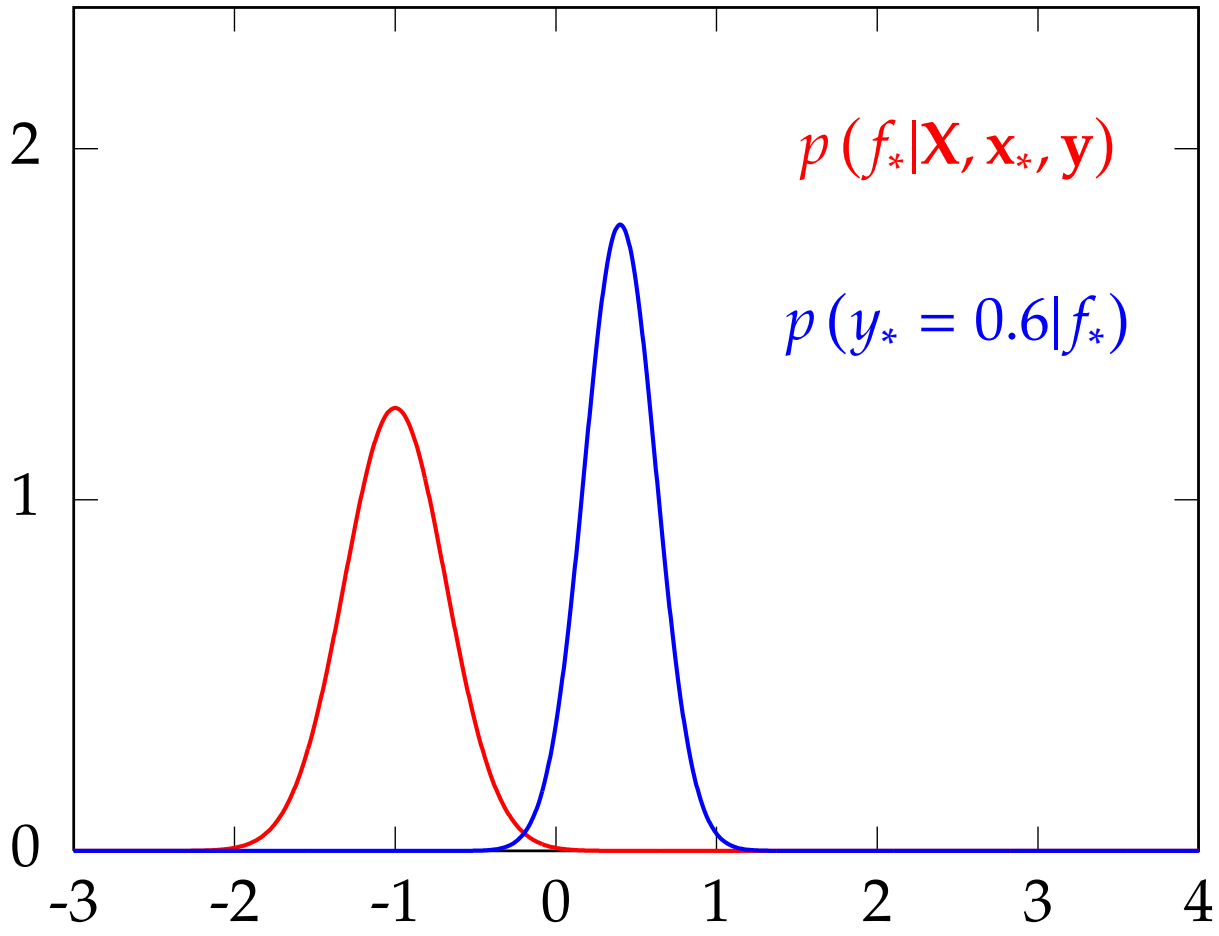


Figure : Inclusion of a data point with Gaussian noise.

Gaussian Noise

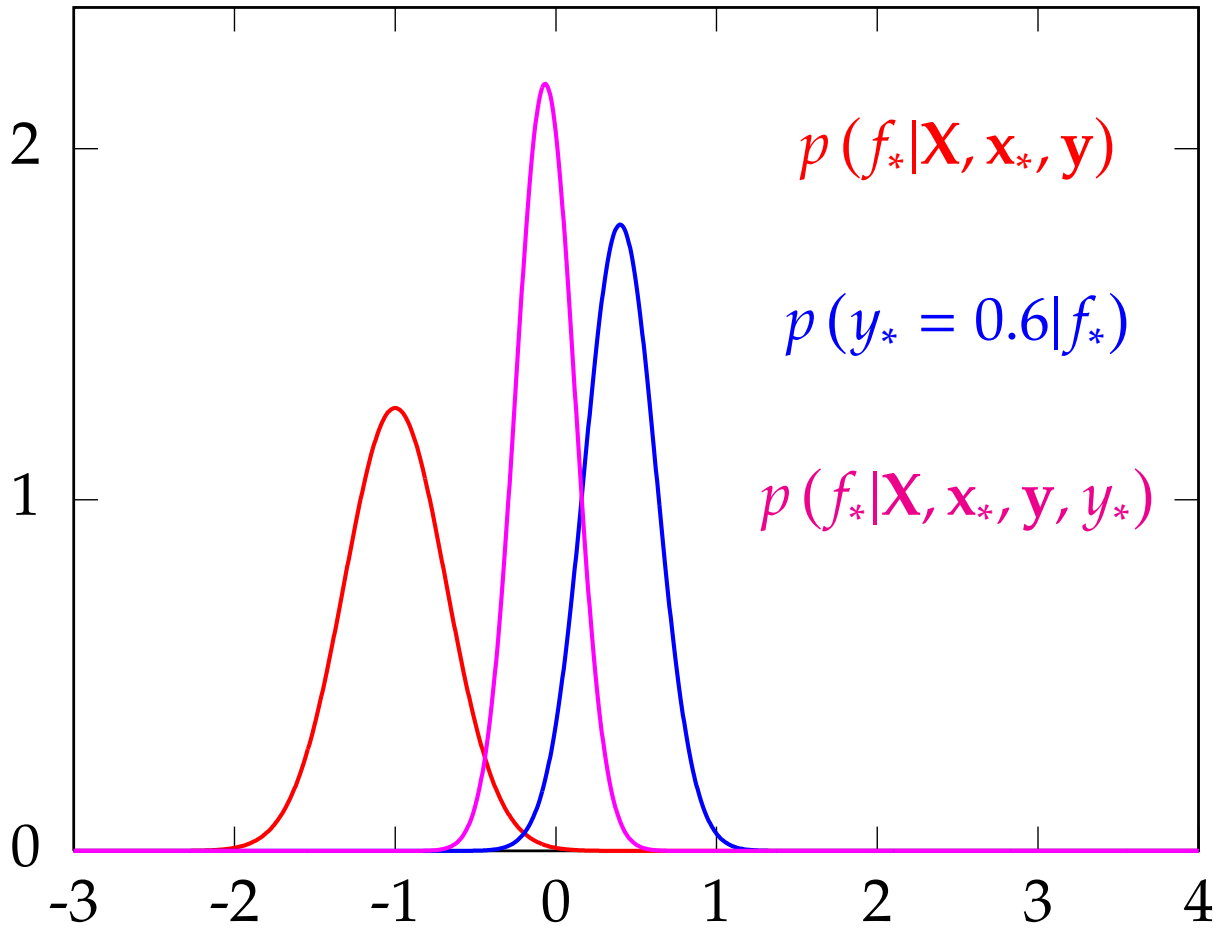


Figure : Inclusion of a data point with Gaussian noise.

Classification Noise Model

Probit Noise Model

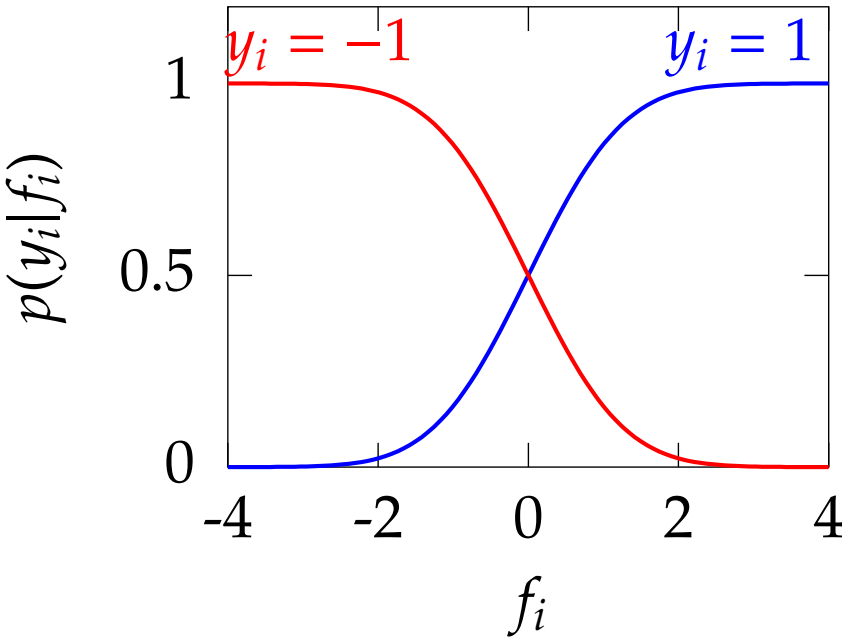


Figure : The probit model (classification). The plot shows $p(y_i|f_i)$ for different values of y_i. For $y_i = 1$ we have $p(y_i|f_i) = \phi(f_i) = \int_{-\infty}^{f_i} \mathcal{N}(z|0,1)\,\mathrm{d}z$.

Classification

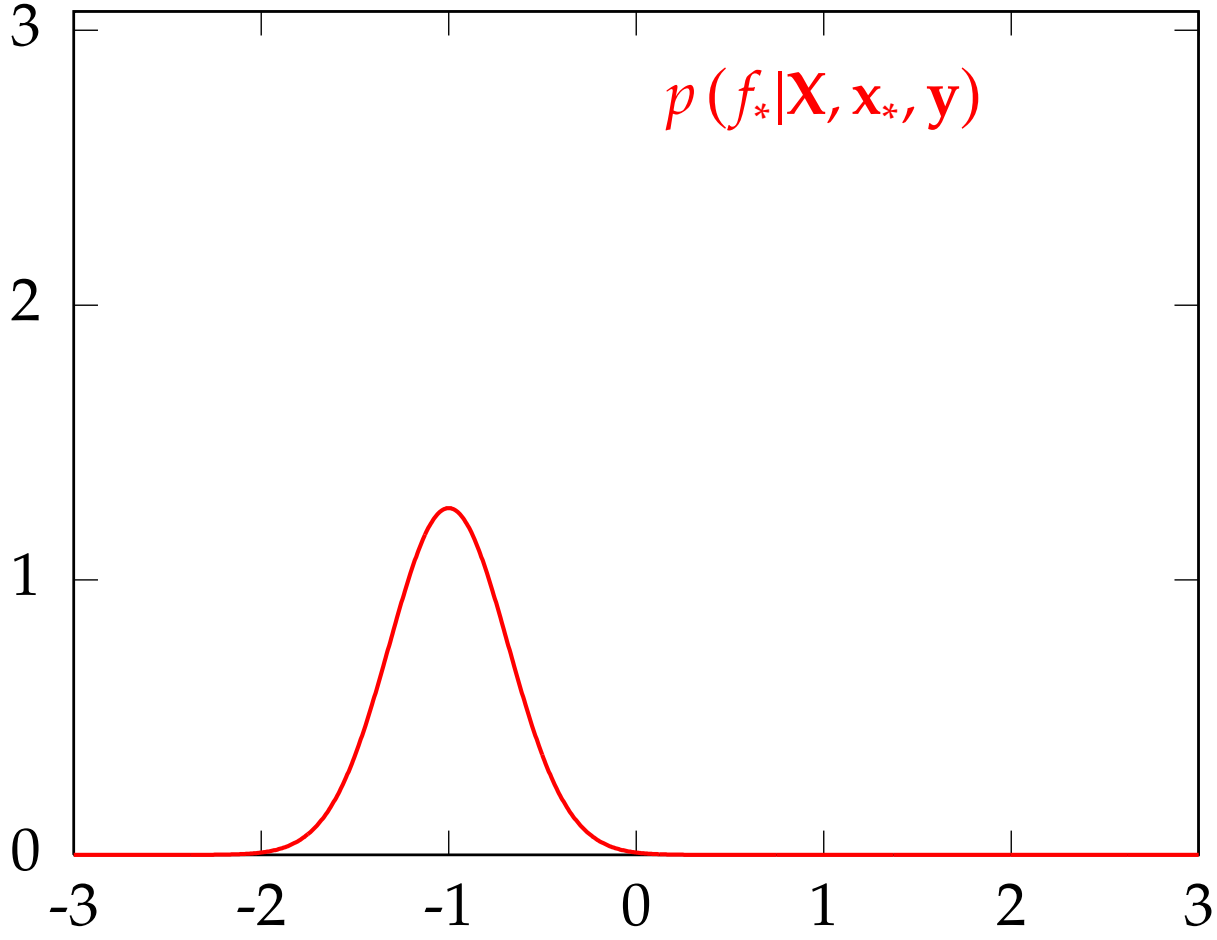


Figure : An EP style update with a classification noise model.

Classification

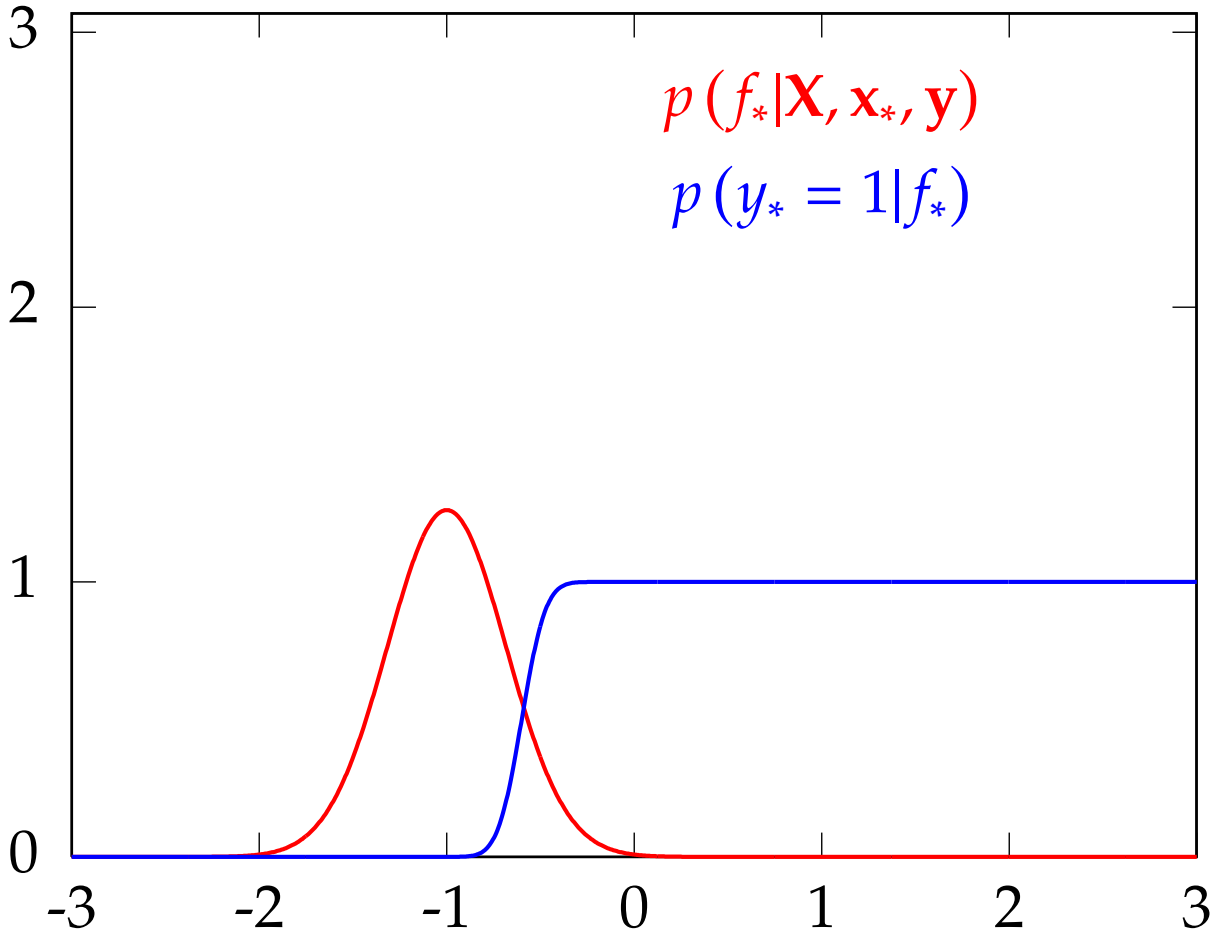


Figure : An EP style update with a classification noise model.

Classification

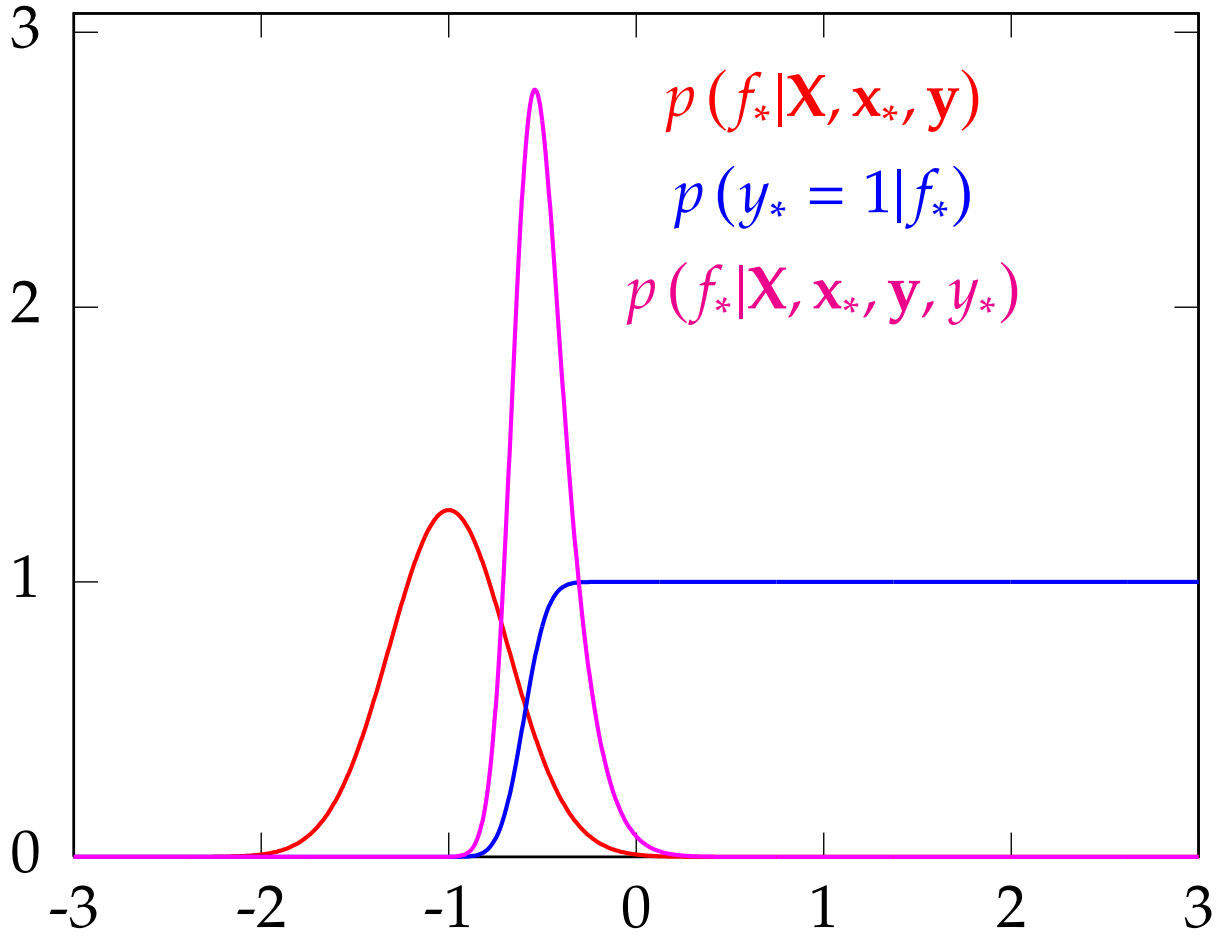


Figure : An EP style update with a classification noise model.

Classification

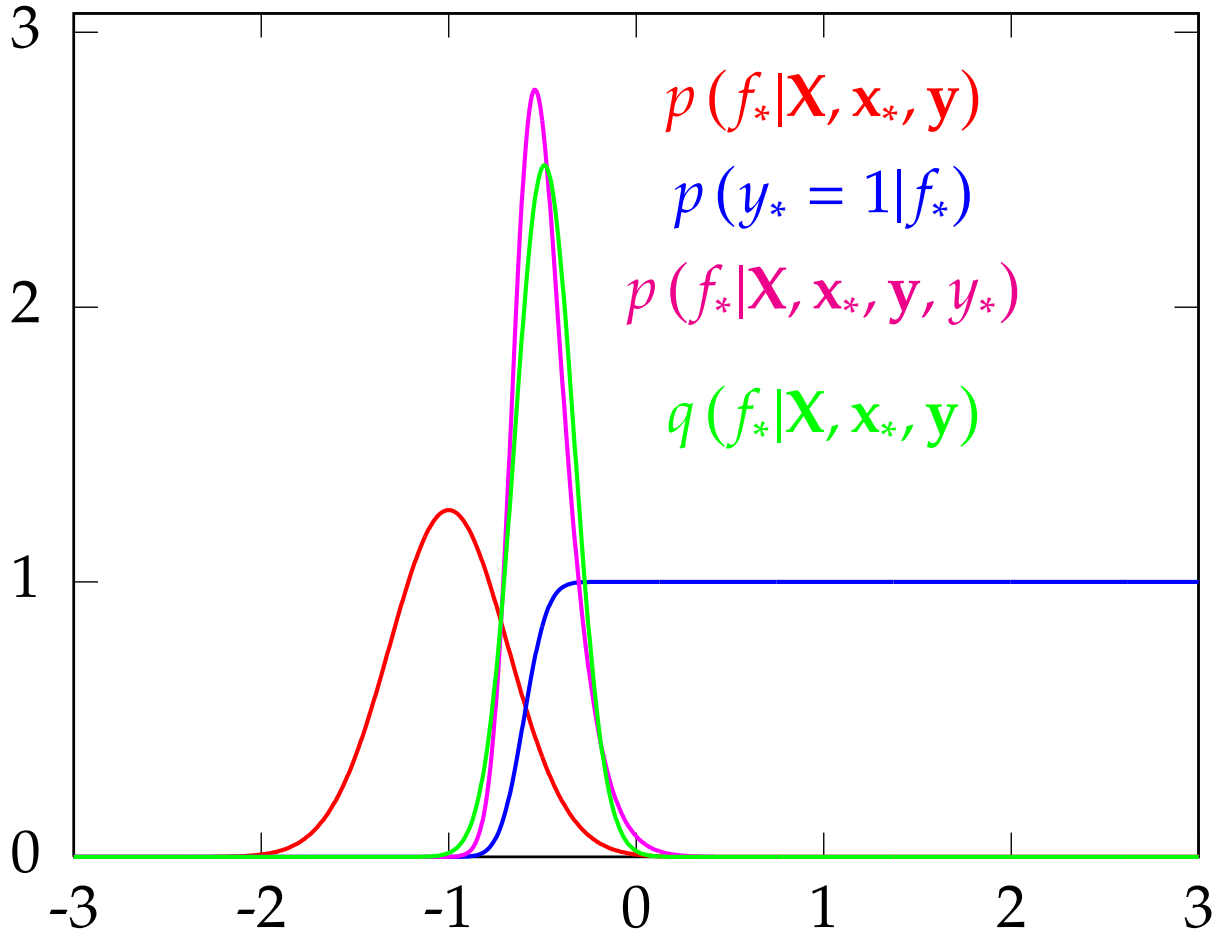


Figure : An EP style update with a classification noise model.

Ordinal Noise Model

Ordered Categories

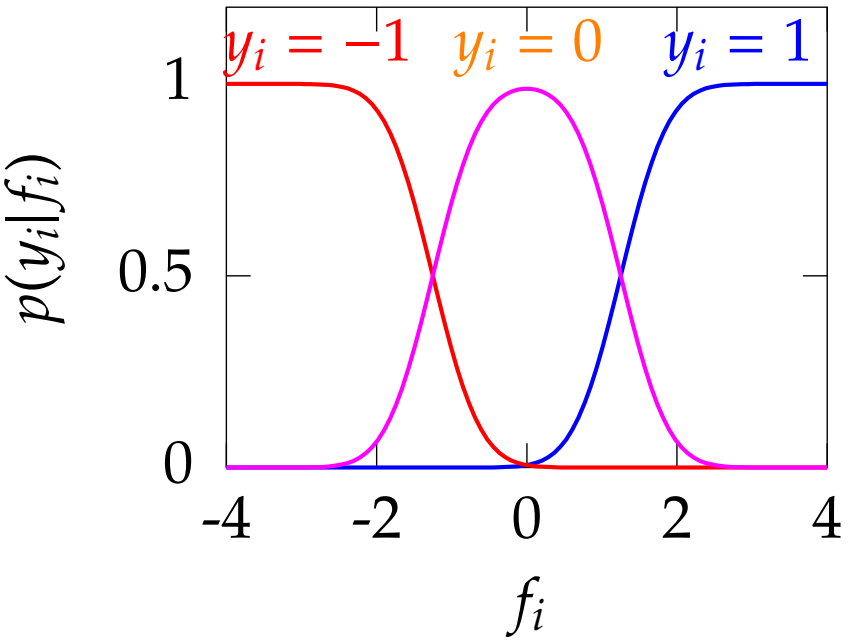


Figure : The ordered categorical noise model (ordinal regression). The plot shows $p(y_i|f_i)$ for different values of y_i. Here we have assumed three categories.

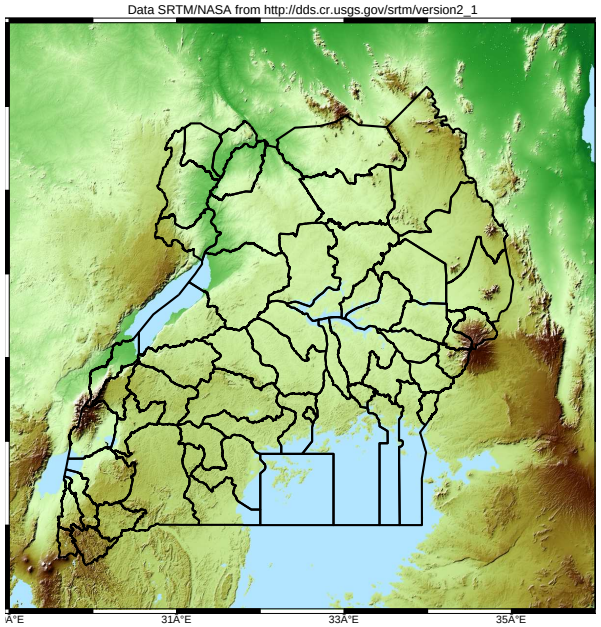
Other Challenges

- Spatial Data (workshop in November with Peter Diggle, work with Ricardo Andrade Pacheco and John Quinn's group).
- Survival Data (work with Alan Saul and Aki Vehtari's group and HeRC).
- Image Data (work with Teo de Campos, Violet Snell and imminent arrival of Zhenwen Dai)
- Text Data (planned collaboration with Trevor Cohn)

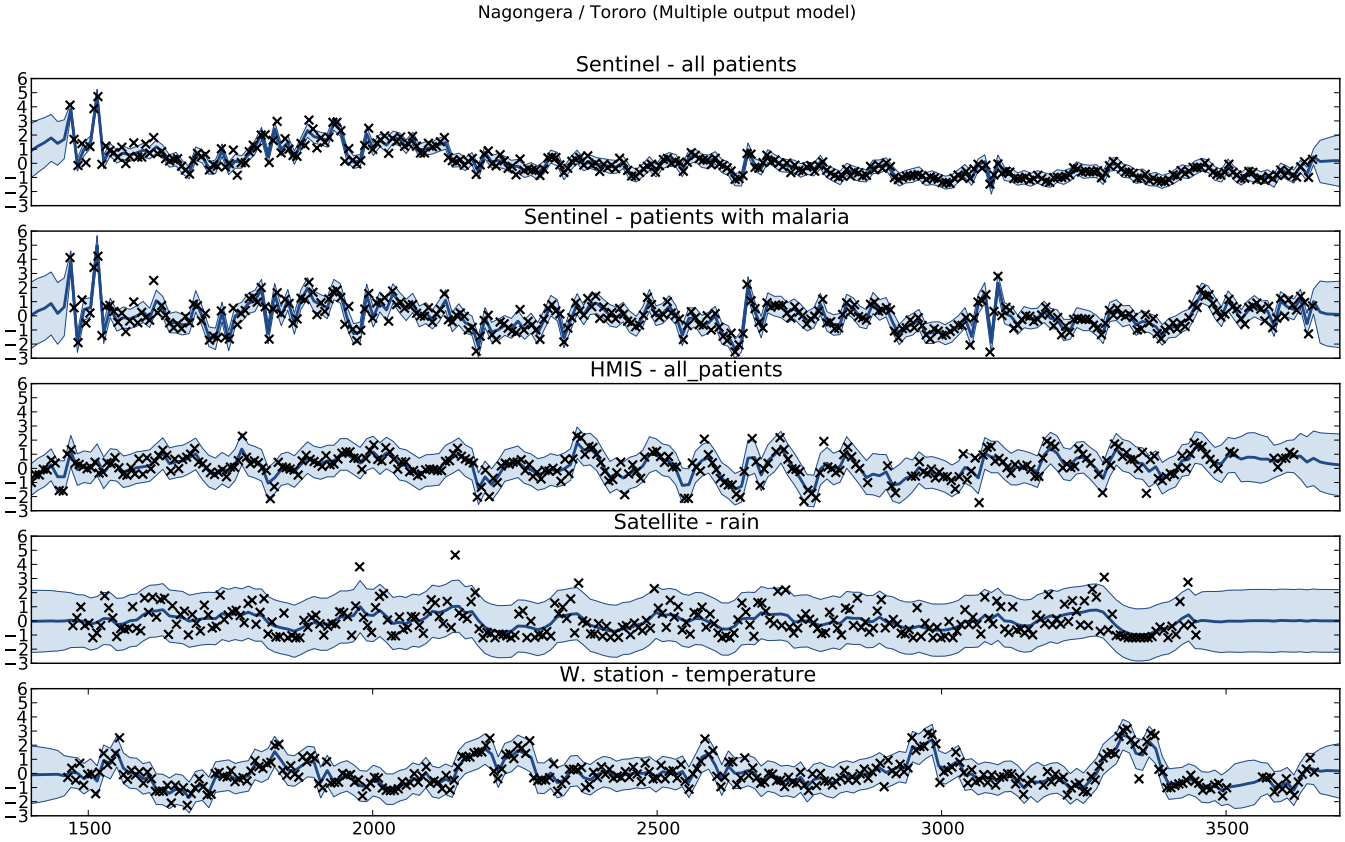
Example: Prediction of Malaria Incidence in Uganda

- Work with John Quinn and Martin Mubaganzi (Makerere University, Uganda)
- See http://cit.mak.ac.ug/cs/aigroup/.

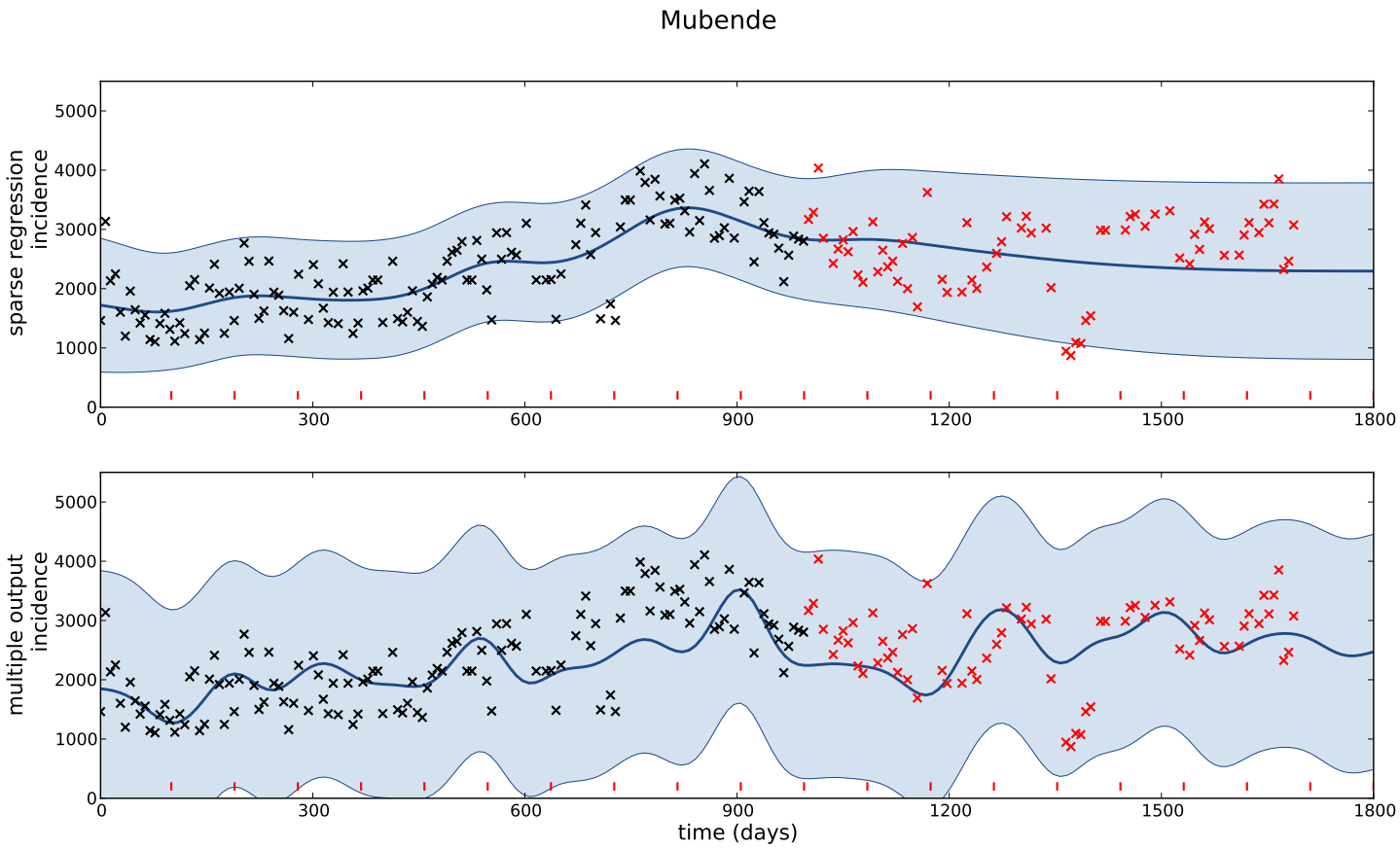
Malaria Prediction in Uganda



Malaria Prediction in Uganda



Malaria Prediction in Uganda



Visit to Uganda

Outline

Health

Data Heterogenity

Deep Learning

Conclusions

Scientists See Promise in Deep-Learning Programs

Hao Zhang/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

NEWS DESK

Reporting the latest on Washington and the world.

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

NOVEMBER 25, 2012

IS "DEEP LEARNING" A REVOLUTION IN ARTIFICIAL INTELLIGENCE?

POSTED BY GARY MARCUS

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

Google To Expand Knowledge Graph Through Hire Of Geoffrey Hinton

Mar 14, 2013 • 8:23 am | (10)
by Barry 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.

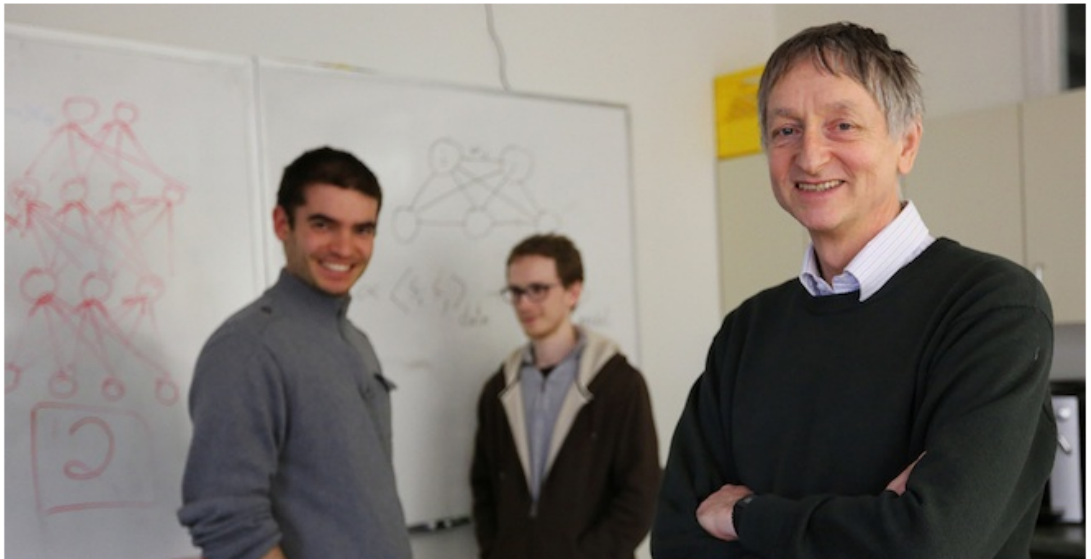
To take the knowledge graph to the next level, Google has hired/acquired Geoffrey Hinton and his team at DNNresearch. Geoffrey posted a note on his Google+ page about it:

> 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.

Google Hires Brains that Helped Supercharge Machine Learning

BY ROBERT MCMILLAN 03.13.13 6:30 AM



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

https://plus.google.com/102889418997957626067/posts

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

+Neil Search Images Maps Play YouTube News Gmail Drive Calendar More

Google+ Neil Lawrence 0 + Share

Home
Profile
Explore
Events
Photos
Communities
NEW Find people
Local
Games
Hangouts
More

Geoffrey Hinton

1,734 have him in circles

ML People
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 +418 167

64 comments

Reza Samahin 15 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 Hertzmann Add
Navdeep Jaitly Add

23 IN COMMON WITH YOU

1,734 HAVE

Chat

Facebook's 'Deep Learnin

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 | uncategorized

FOLLOW WIRED ENTERPRISE

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

BY CADE METZ 12.12.13 6:30 AM

Follow @cademetz

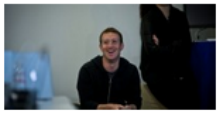
Share 452
Tweet 332
+1 125
Share 65
Pin it

Yann LeCun. *Photo: WIRED/Josh Valcarcel*

MOST RECENT WIRED POSTS

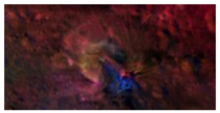
Everything That Happened in 2013 in One Image

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 Aelia

What Improves Every Party? Giant

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 this article, and has no relevant affiliations.

Do you know anything about machine learning? newyork

"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

direction for further research.

11.1. HAVE WE THROWN THE BABY OUT WITH THE BATH WATER?

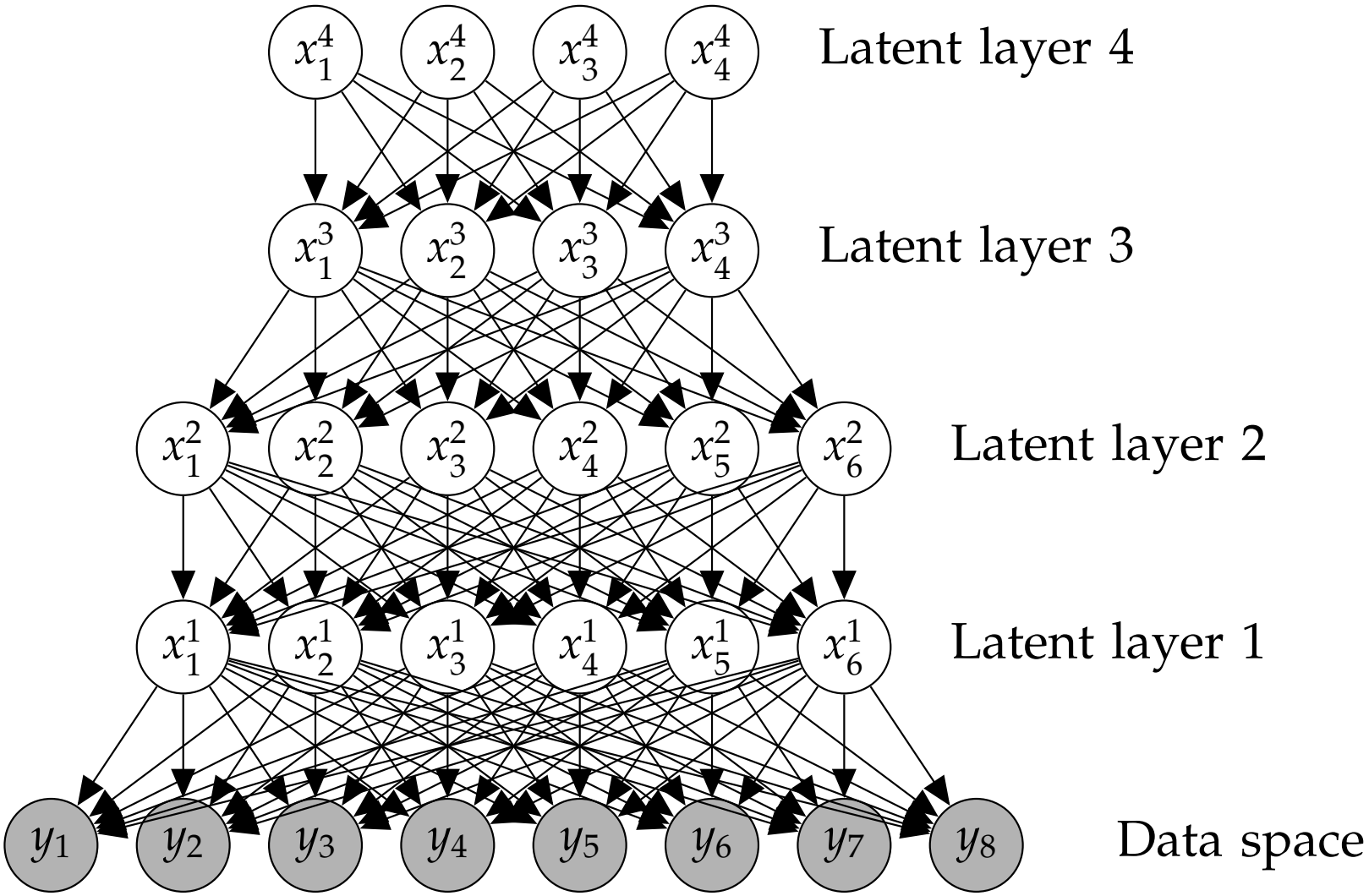
According to the hype of 1987, neural networks were meant to be intelligent models which discovered features and patterns in data. Gaussian processes in contrast are simply smoothing devices. How can Gaussian processes possibly replace neural networks? What is going on?

I think what the work of Williams and Rasmussen (1996) shows is that many real–world data modelling problems are perfectly well solved by sensible smoothing methods. The most interesting problems, the task of feature discovery for example, are not ones which Gaussian processes will solve. But maybe multilayer perceptrons can't solve them either. On the other hand, it may be that the limit of an infinite number of hidden units, to which Gaussian processes correspond, was a bad limit to take; maybe we should backtrack, or modify the prior on neural network parameters, so as to create new models more interesting than Gaussian processes. Evidence that this infinite limit has lost something compared with finite neural networks comes from the observation that in a finite neural network with more than one output, there are non–trivial correlations between the outputs (since they share inputs from common hidden units); but in the limit of an infinite number of hidden units, these correlations vanish. Radford Neal has suggested the use of non–Gaussian priors in networks with multiple hidden layers. Or perhaps a completely fresh start is needed, approaching the problem of machine learning from a paradigm different from the supervised feedforward mapping.

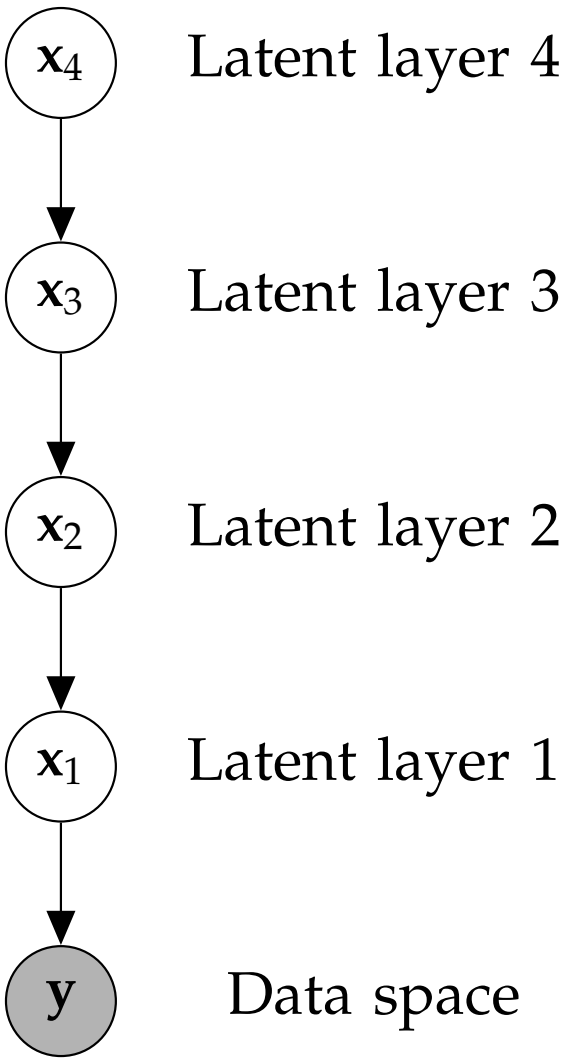
Structure of Priors

MacKay: NIPS Tutorial 1997 "Have we thrown out the baby with the bathwater?" (Published as MacKay, 1998) Also noted by (Wilson et al., 2012)

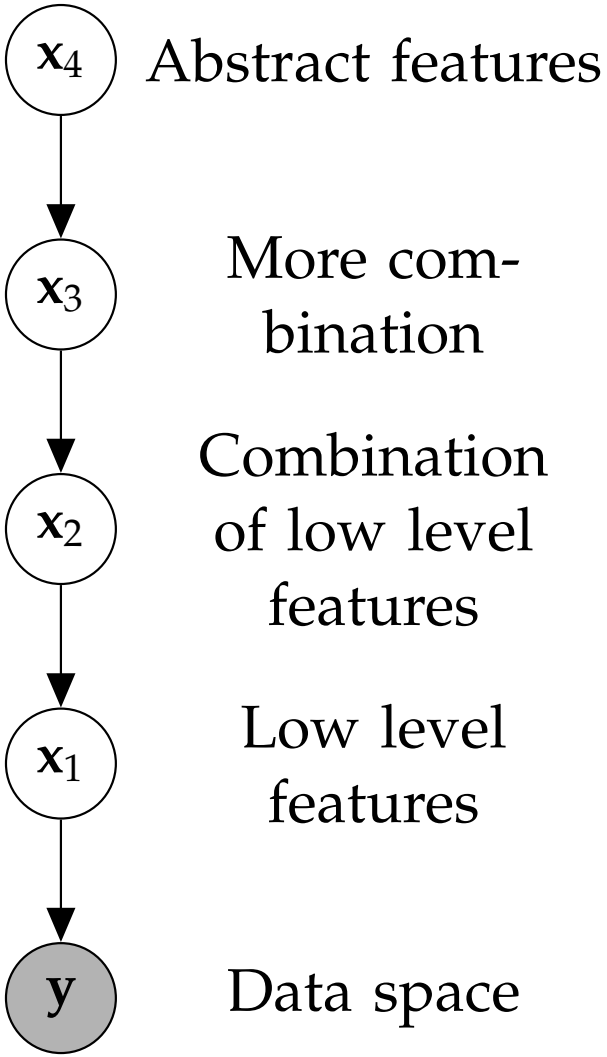
Deep Models


x_1^4 x_2^4 x_3^4 x_4^4
Latent layer 4
x_1^3 x_2^3 x_3^3 x_4^3
Latent layer 3
x_1^2 x_2^2 x_3^2 x_4^2 x_5^2 x_6^2
Latent layer 2
x_1^1 x_2^1 x_3^1 x_4^1 x_5^1 x_6^1
Latent layer 1
y_1 y_2 y_3 y_4 y_5 y_6 y_7 y_8
Data space

Deep Models


$\mathbf{x}_4$
Latent layer 4
$\mathbf{x}_3$
Latent layer 3
$\mathbf{x}_2$
Latent layer 2
$\mathbf{x}_1$
Latent layer 1
$\mathbf{y}$
Data space

Deep Models

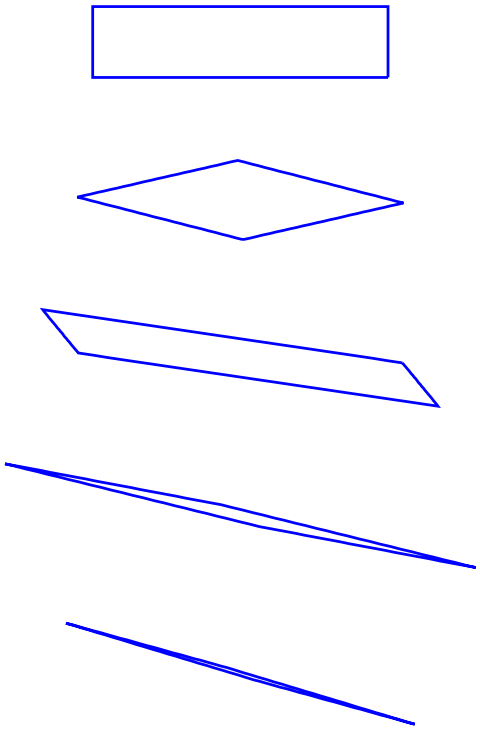

$\mathbf{x}_4$
Abstract features
$\mathbf{x}_3$
More combination
$\mathbf{x}_2$
Combination of low level features
$\mathbf{x}_1$
Low level features
$\mathbf{y}$
Data space

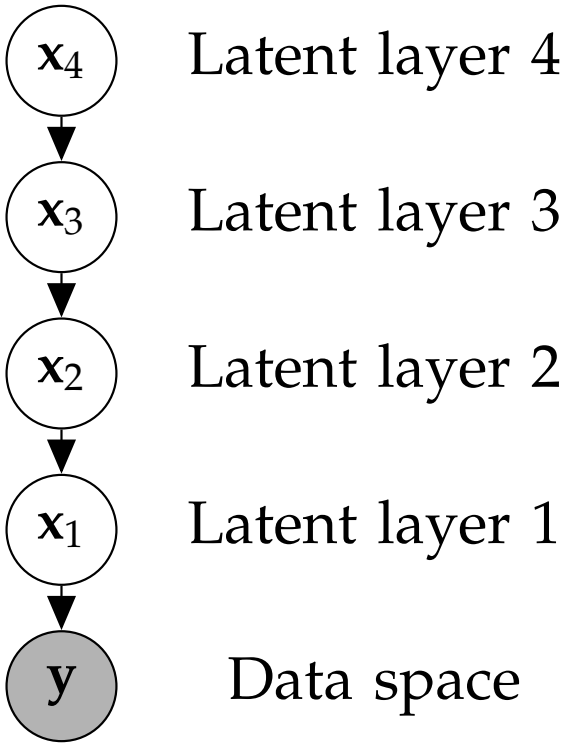
Deep Gaussian Processes

Damianou and Lawrence (2013)

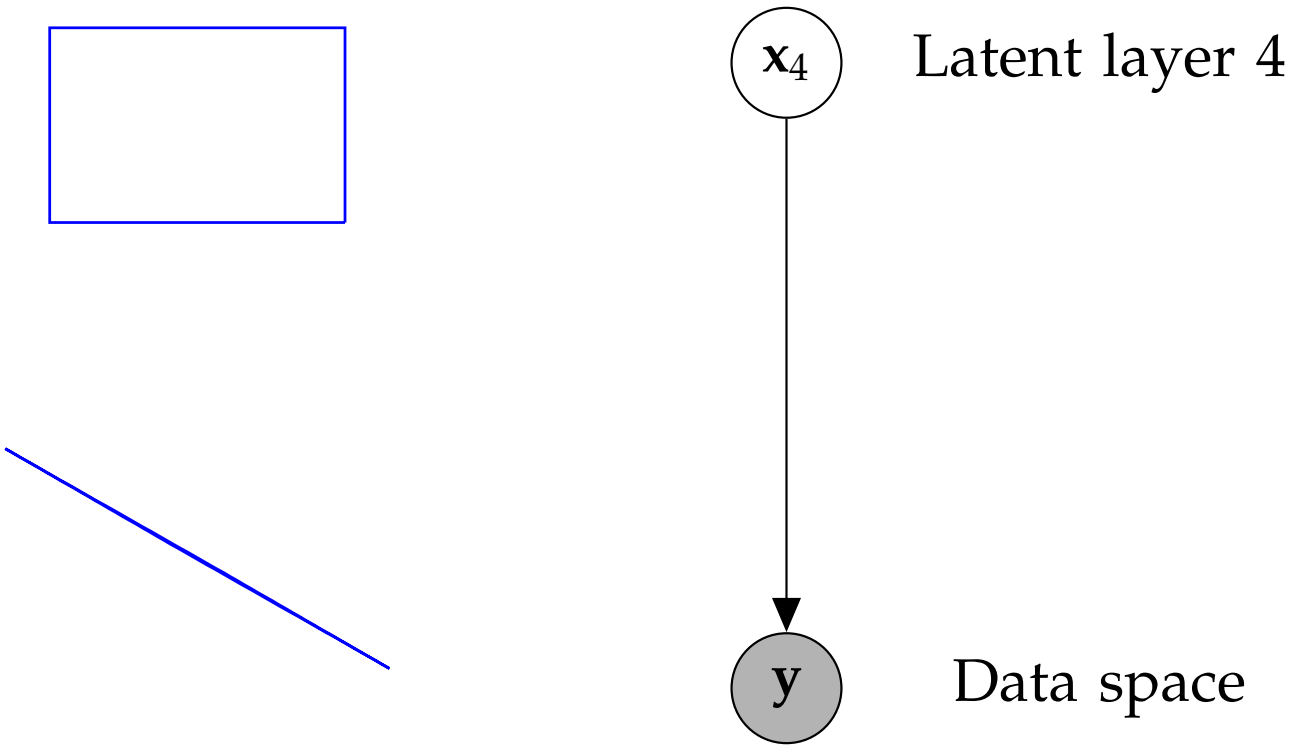
- Deep architectures allow abstraction of features (Bengio, 2009; Hinton and Osindero, 2006; Salakhutdinov and Murray, 2008).
- We use variational approach to stack GP models.

Stacked PCA

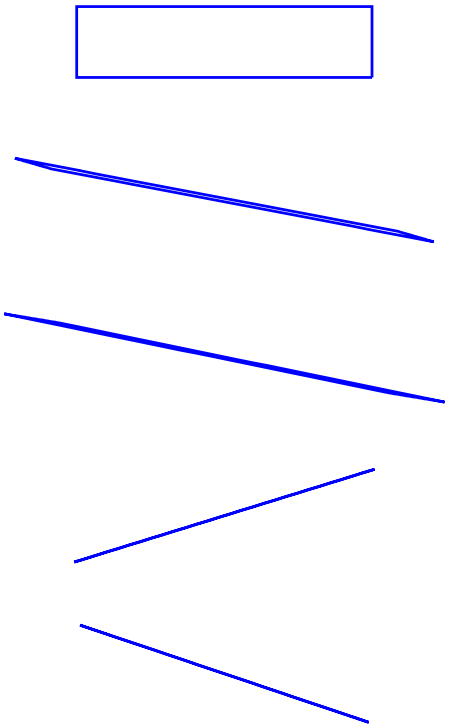


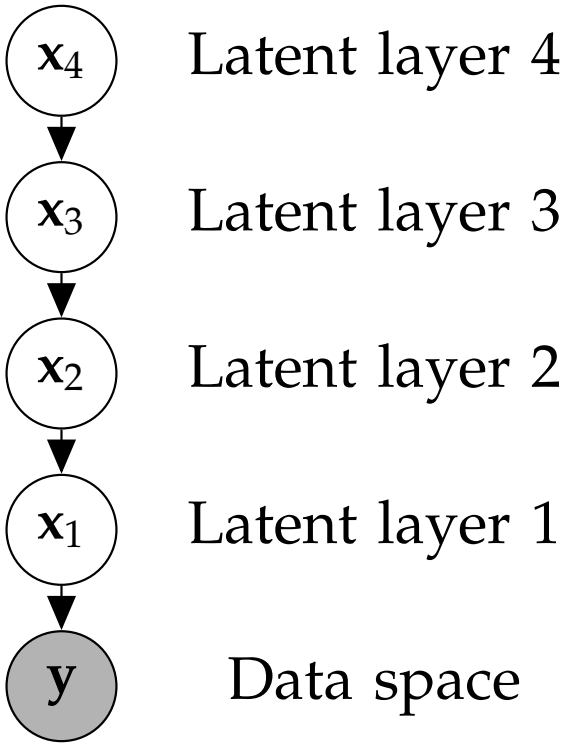

$\mathbf{x}_4$
Latent layer 4
$\mathbf{x}_3$
Latent layer 3
$\mathbf{x}_2$
Latent layer 2
$\mathbf{x}_1$
Latent layer 1
$\mathbf{y}$
Data space

Stacked PCA

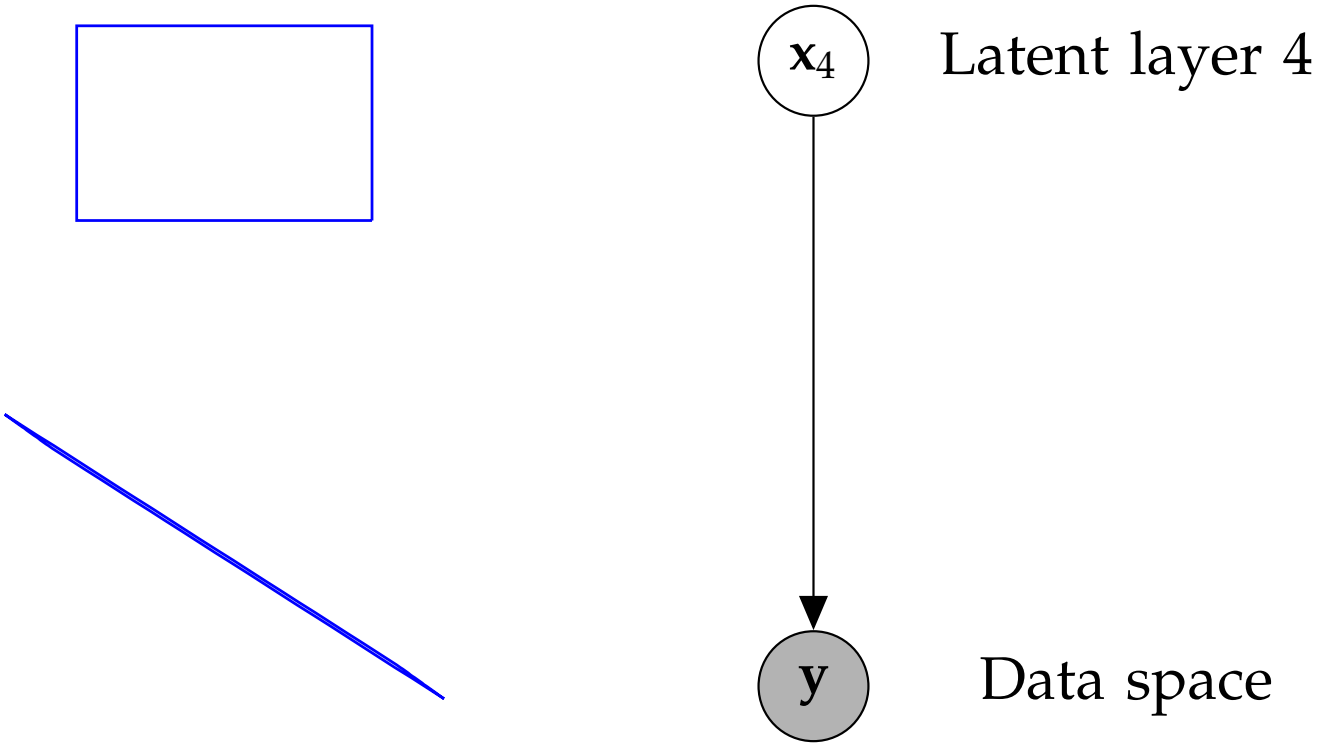


Stacked PCA

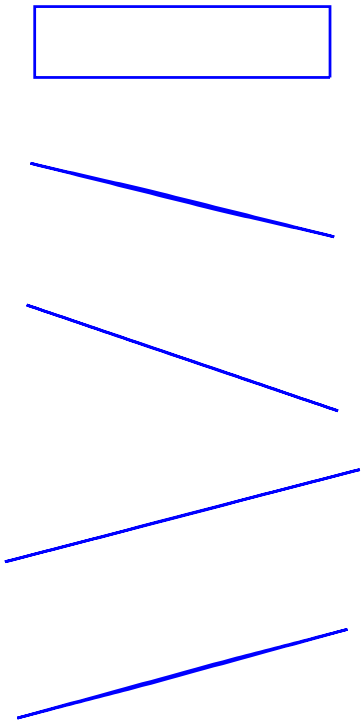


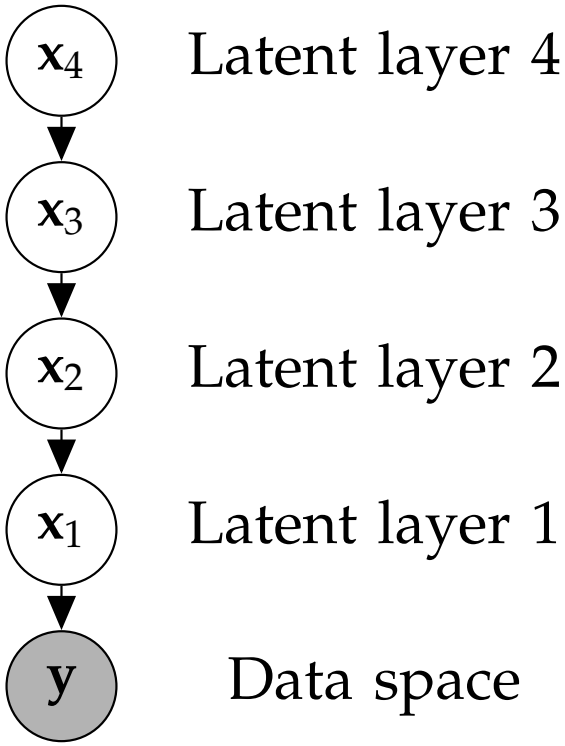

$\mathbf{x}_4$
Latent layer 4
$\mathbf{x}_3$
Latent layer 3
$\mathbf{x}_2$
Latent layer 2
$\mathbf{x}_1$
Latent layer 1
$\mathbf{y}$
Data space

Stacked PCA

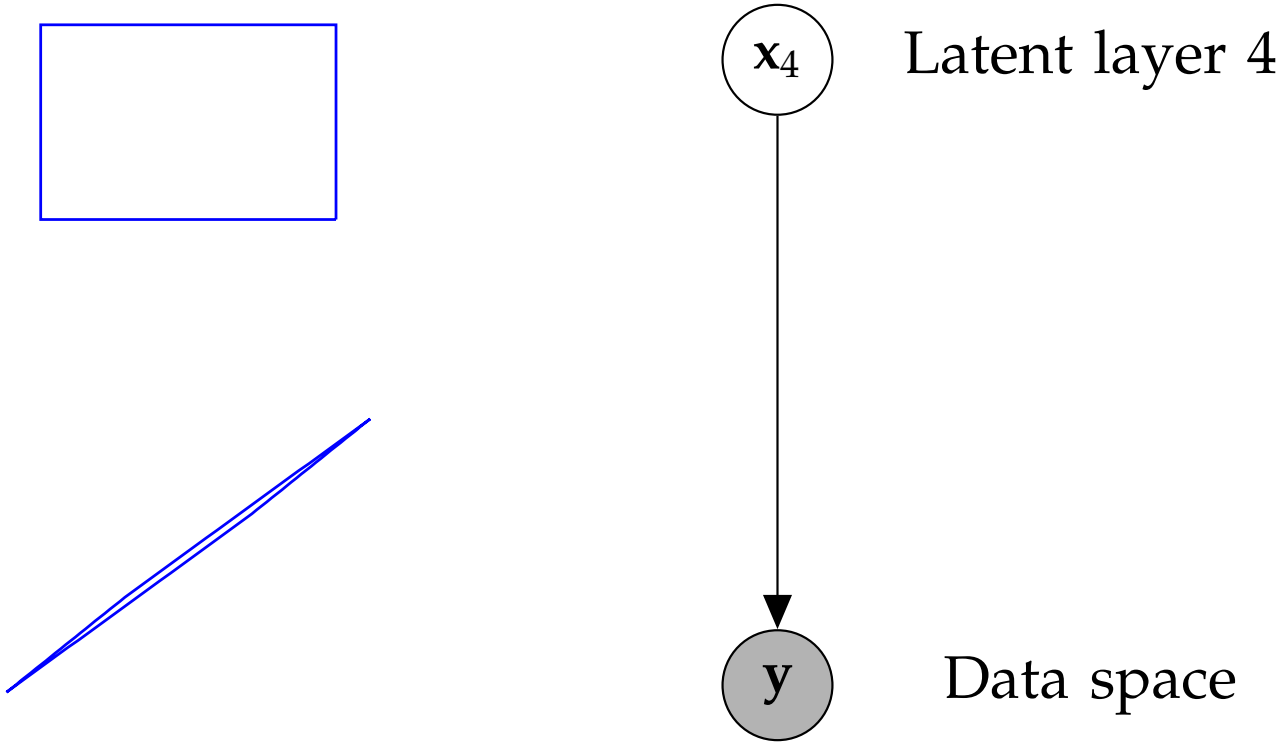


Stacked PCA

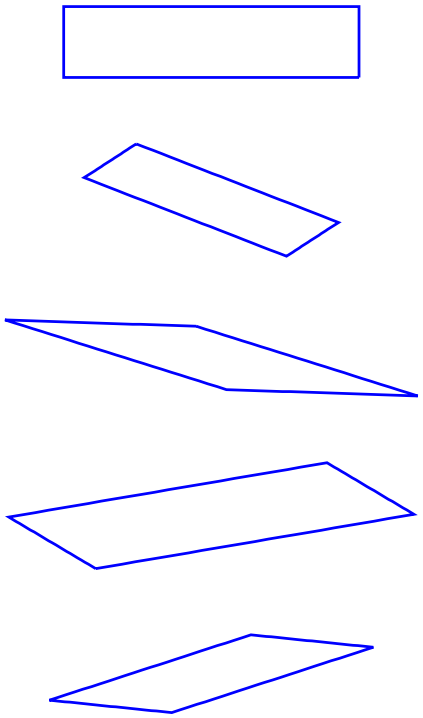


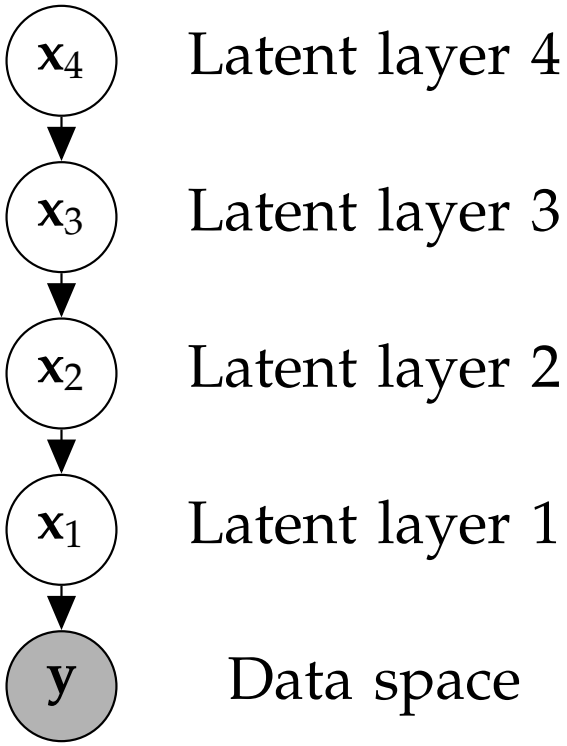

$\mathbf{x}_4$
Latent layer 4
$\mathbf{x}_3$
Latent layer 3
$\mathbf{x}_2$
Latent layer 2
$\mathbf{x}_1$
Latent layer 1
$\mathbf{y}$
Data space

Stacked PCA

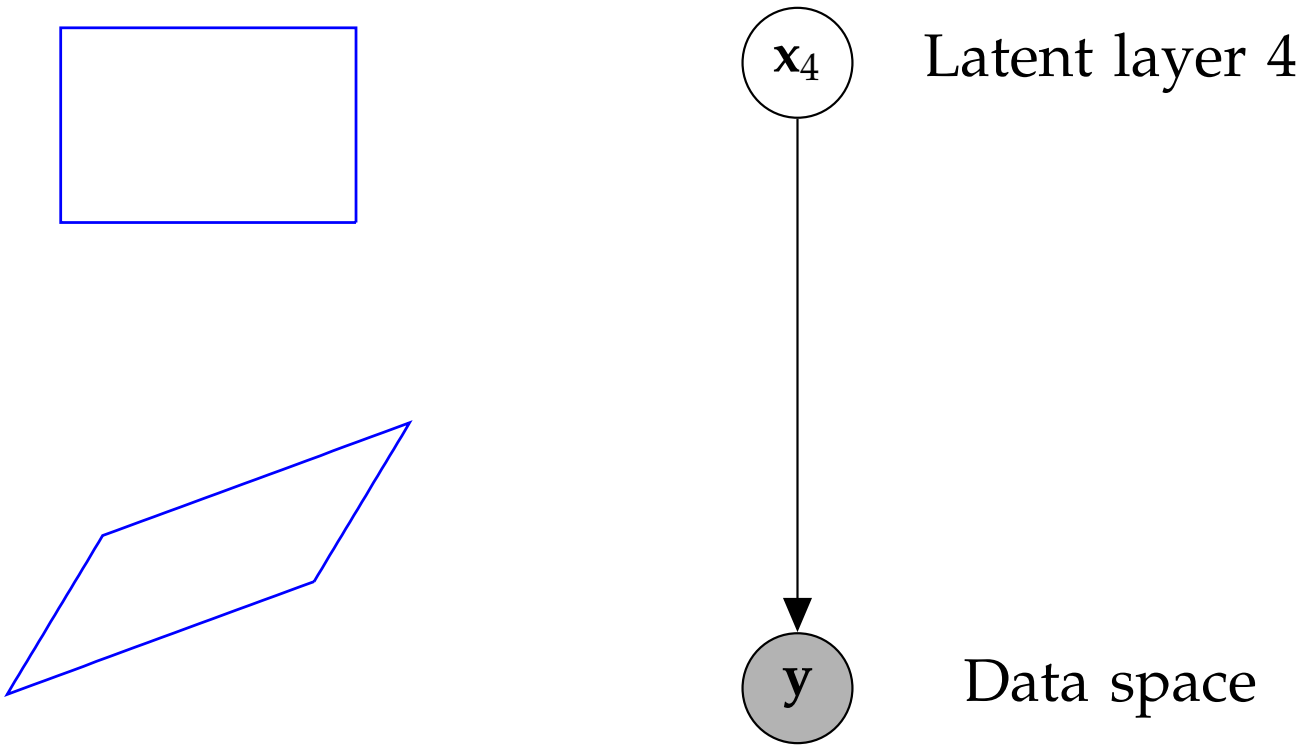


Stacked PCA

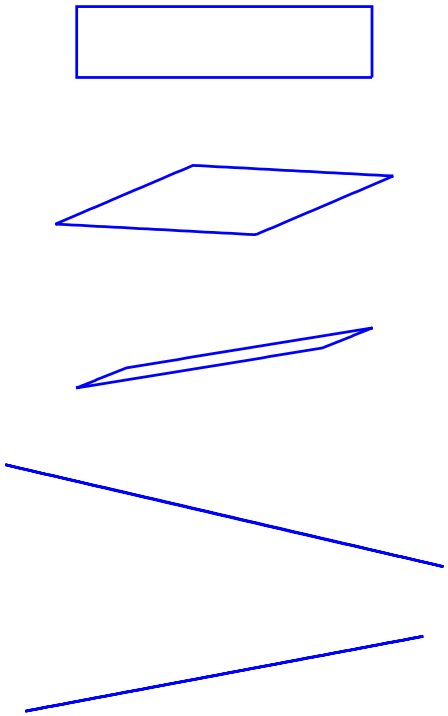


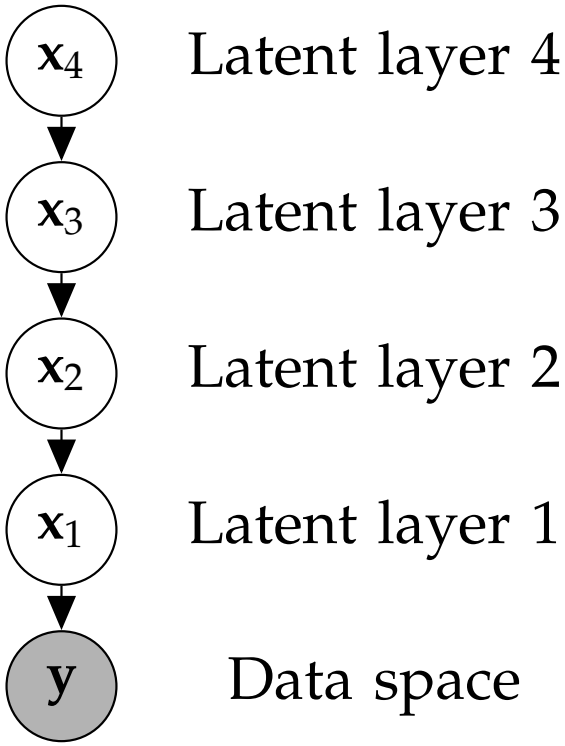

$\mathbf{x}_4$
Latent layer 4
$\mathbf{x}_3$
Latent layer 3
$\mathbf{x}_2$
Latent layer 2
$\mathbf{x}_1$
Latent layer 1
$\mathbf{y}$
Data space

Stacked PCA

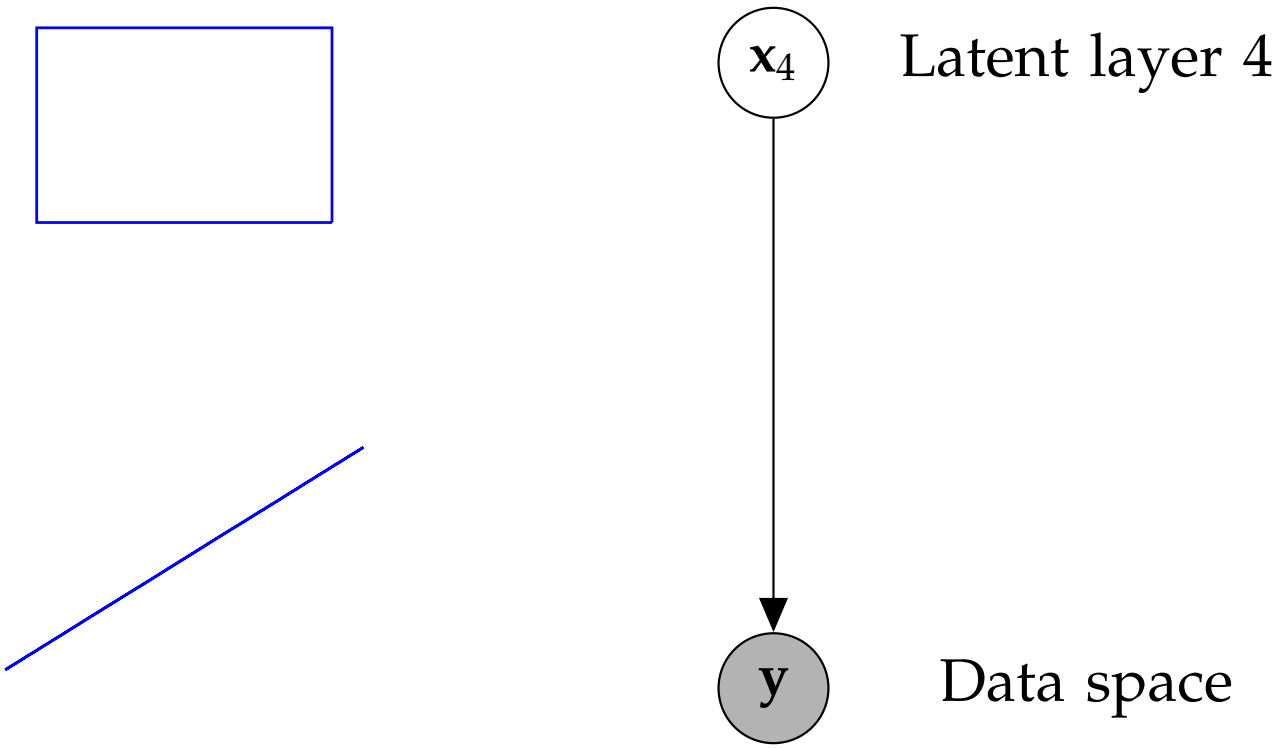


Stacked PCA

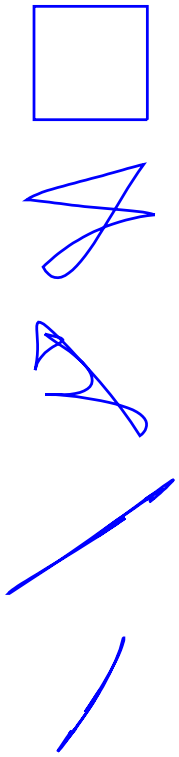


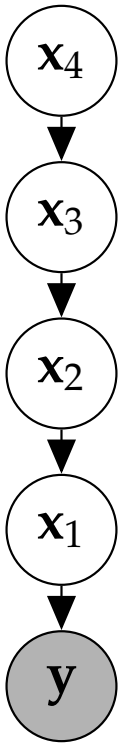

$\mathbf{x}_4$
Latent layer 4
$\mathbf{x}_3$
Latent layer 3
$\mathbf{x}_2$
Latent layer 2
$\mathbf{x}_1$
Latent layer 1
$\mathbf{y}$
Data space

Stacked PCA



Stacked GPs





Latent layer 4

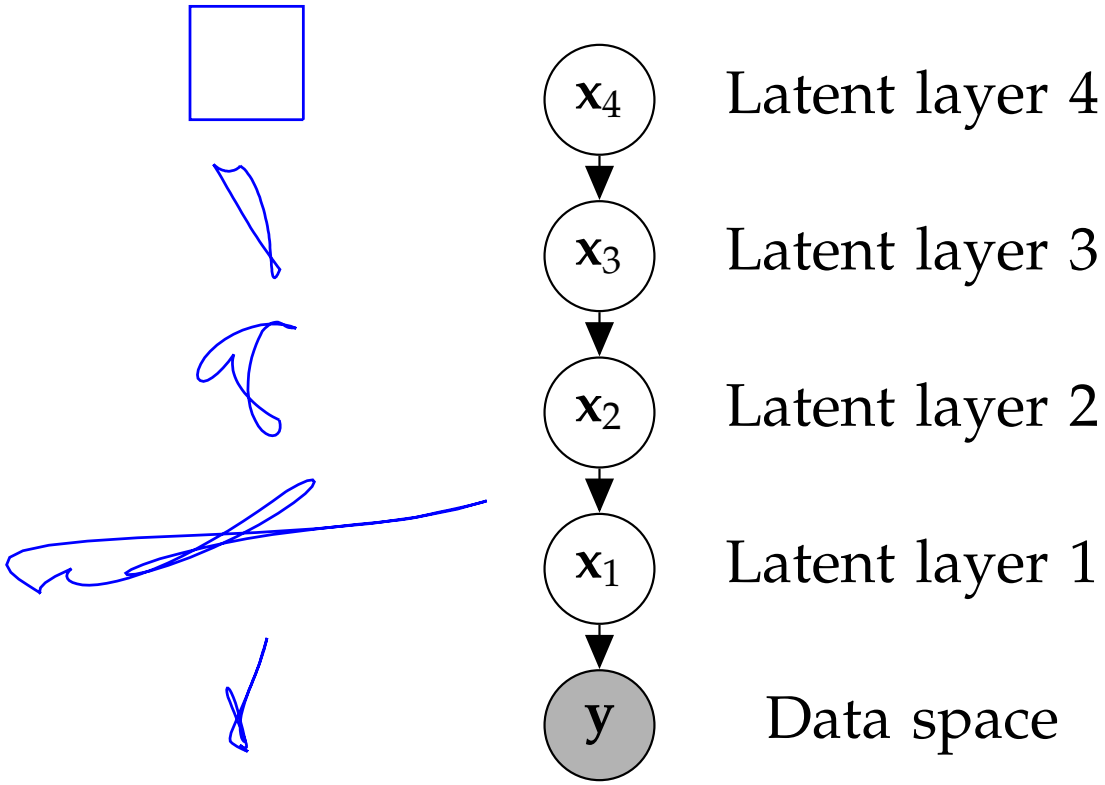
Latent layer 3

Latent layer 2

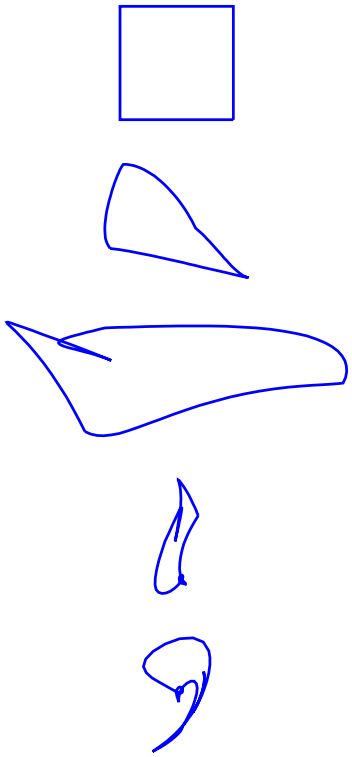
Latent layer 1

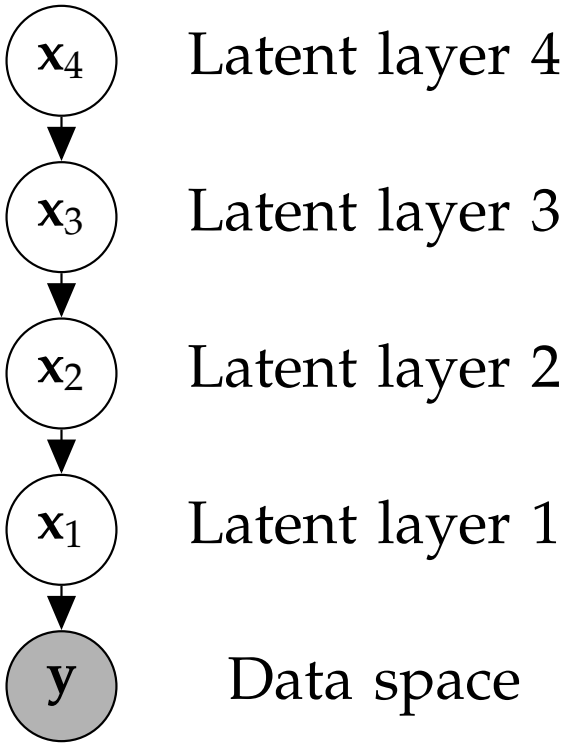
Data space

Stacked GPs

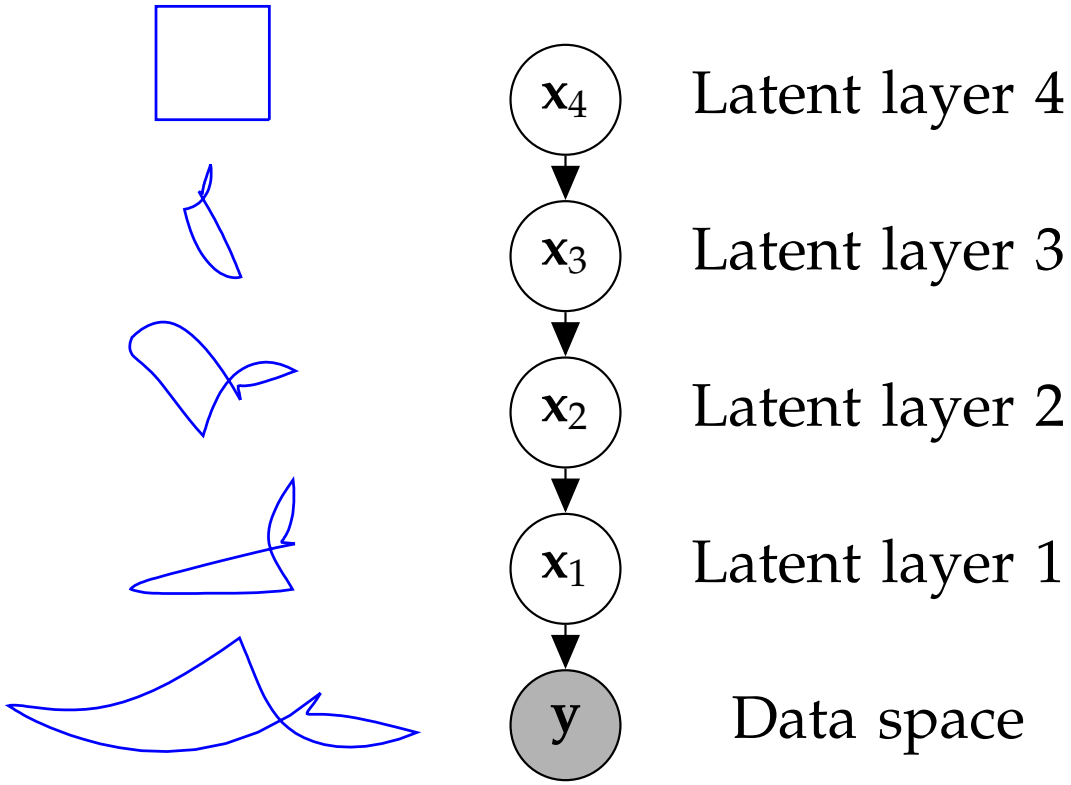

$\mathbf{x}_4$
Latent layer 4
$\mathbf{x}_3$
Latent layer 3
$\mathbf{x}_2$
Latent layer 2
$\mathbf{x}_1$
Latent layer 1
$\mathbf{y}$
Data space

Stacked GPs

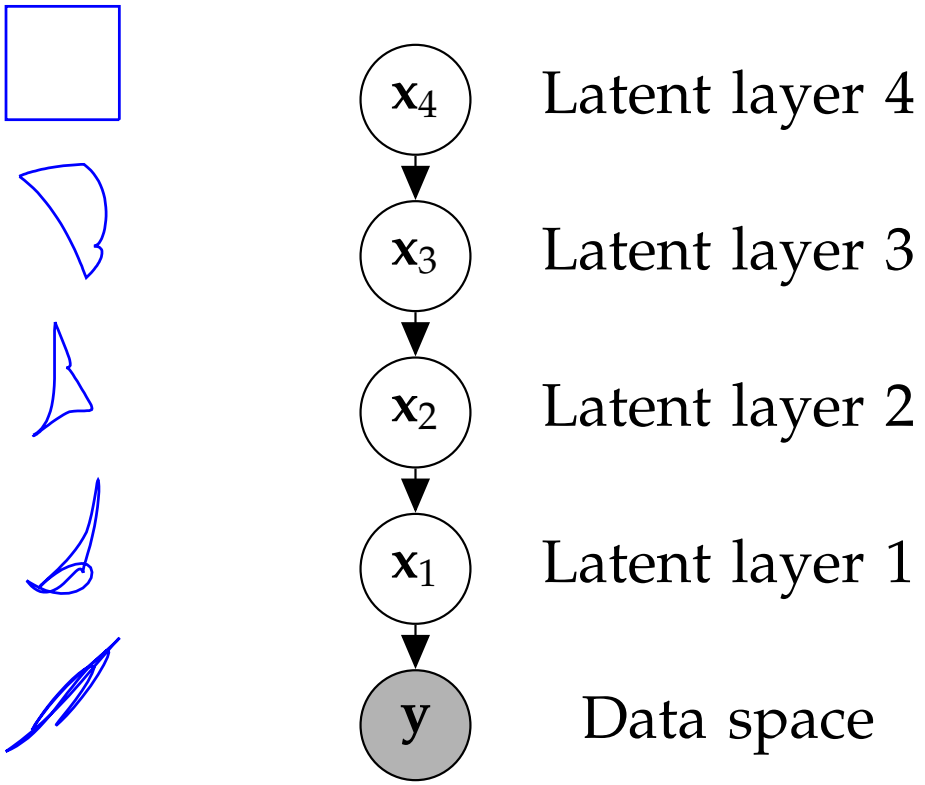



$\mathbf{x}_4$
Latent layer 4
$\mathbf{x}_3$
Latent layer 3
$\mathbf{x}_2$
Latent layer 2
$\mathbf{x}_1$
Latent layer 1
$\mathbf{y}$
Data space

Stacked GPs


$\mathbf{x}_4$
Latent layer 4
$\mathbf{x}_3$
Latent layer 3
$\mathbf{x}_2$
Latent layer 2
$\mathbf{x}_1$
Latent layer 1
$\mathbf{y}$
Data space

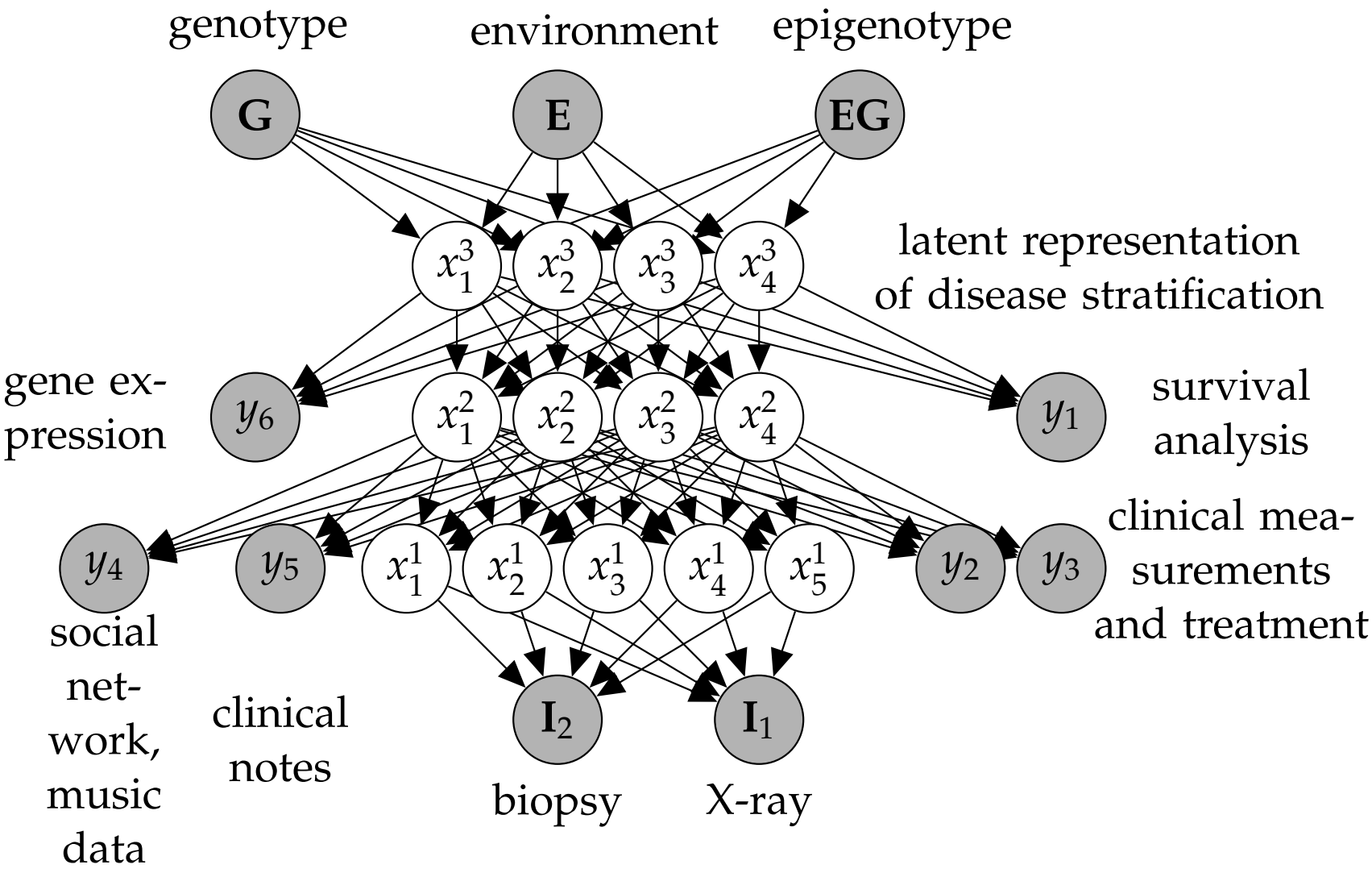
Stacked GPs


$\mathbf{x}_4$
Latent layer 4
$\mathbf{x}_3$
Latent layer 3
$\mathbf{x}_2$
Latent layer 2
$\mathbf{x}_1$
Latent layer 1
$\mathbf{y}$
Data space

What Can We 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


genotype
environment
epigenotype
G
E
EG
x_1^3 x_2^3 x_3^3 x_4^3
latent representation of disease stratification
gene expression
y_6
x_1^2 x_2^2 x_3^2 x_4^2
y_1
survival analysis
y_4
y_5
x_1^1 x_2^1 x_3^1 x_4^1 x_5^1
y_2
y_3
clinical measurements and treatment
social network, music data
clinical notes
$\mathbf{I}_2$
$\mathbf{I}_1$
biopsy
X-ray

Deep Health: Power Ranger Model of Research

Thanks to Alan Saul for creating the image.

GPy: A Gaussian Process Framework in Python

- BSD Licensed software base.
- Wide availability of libraries, "modern" scripting language.
- Allows us to set projects to undergraduates in Comp Sci that use GPs.
- Available through GitHub `https://github.com/SheffieldML/GPy`
- Reproducible Research with IPython Notebook.

Probable Features for Next Release

- Non-Gaussian likelihoods.
- Multivariate outputs.
- Dimensionality reduction.
- Approximations for large data sets.
- Probabilistic-style programming (specify the model, not the algorithm).
- A range of covariance functions, ability to implement your own symbolically.
- Data missing at random.

Planned Features for Future Releases

- Deep models.
- Massive data: millions of points.
- Data missing not at random.
- Deep models for massive data.

GPSS: Gaussian Process Summer School

- http://ml.dcs.shef.ac.uk/gpss/
- Next one is in Sheffield in September 2014.
- Ran a 'Gaussian Process Roadshow' at Pereira last week.
- Tentative plans for a Roadshow in Kenya in 2015.

Summary

- Gaussian models good for missing data.
- Disparate data types handled with EP and Laplace.
- Deep models allow complex abstract representation of data sets at higher levels.
- Current limitation is on data set size.
- Addressing this through work by James Hensman on Stochastic Variational Inference for GPs (recent UAI paper).
- Intention is to deploy these models for assimilating a wide range of data types in personalized health (text, survival times, images, genotype, phenotype).
- Requires population scale models with millions of features.

References I

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].

J. Hensman, N. Fusi, and N. D. Lawrence. Gaussian processes for big data. In A. Nicholson and P. Smyth, editors, *Uncertainty in Artificial Intelligence*, volume 29. AUAI Press, 2013. [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].

D. J. C. MacKay. Introduction to Gaussian Processes. In C. M. Bishop, editor, *Neural Networks and Machine Learning*, volume 168 of *Series F: Computer and Systems Sciences*, pages 133–166. Springer-Verlag, Berlin, 1998.

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.

M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, B*, 6 (3):611–622, 1999. [PDF]. [DOI].

A. G. Wilson, D. A. Knowles, and Z. Ghahramani. Gaussian process regression networks. In J. Langford and J. Pineau, editors, *Proceedings of the International Conference in Machine Learning*, volume 29, San Francisco, CA, 2012. Morgan Kauffman.