Personalized Health and Africa

Neil D. Lawrence

University of Sheffield

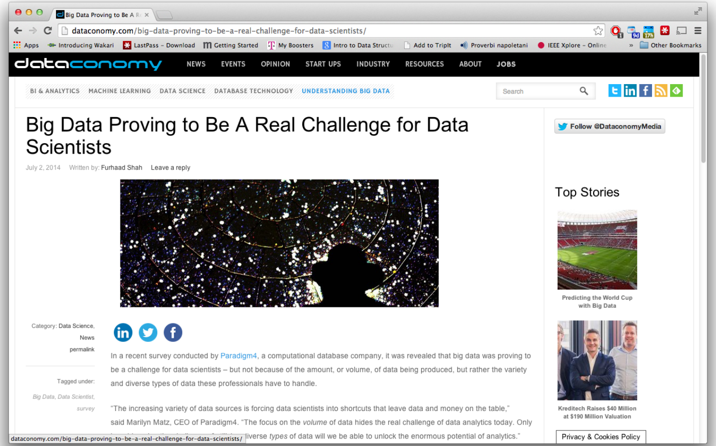
18th June 2015

Outline

Diversity of Data

Massively Missing Data

Not the Scale it's the Diversity



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,

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

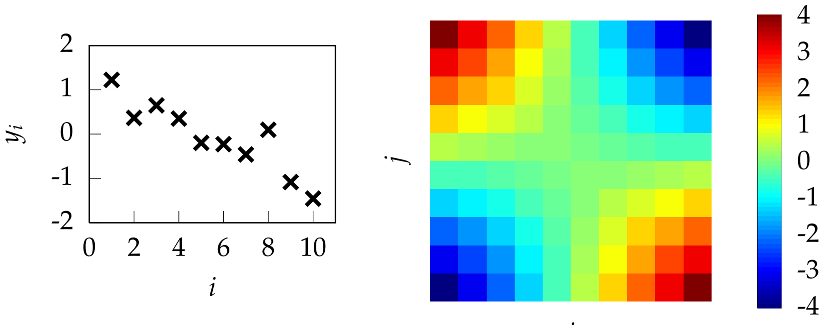
- Where $y_3, \ldots, y_p$ contains:
  1. all tests not applied to this patient
  2. all tests not yet invented!!

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

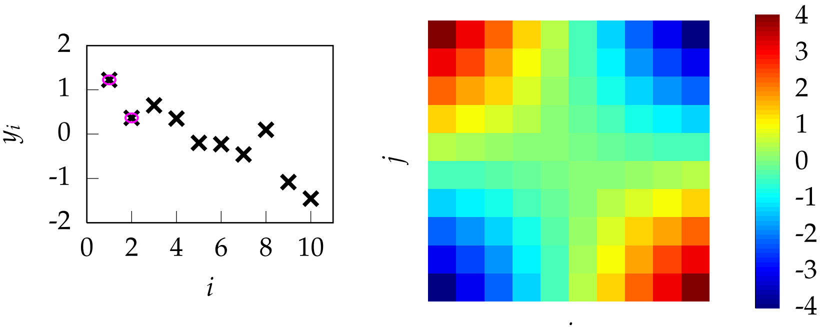


(a) A 10 dimensional sample

(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property

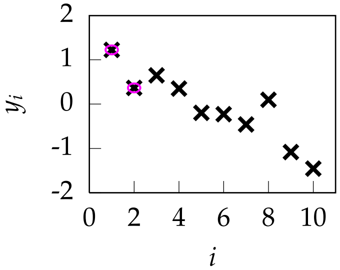


(a) A 10 dimensional sample

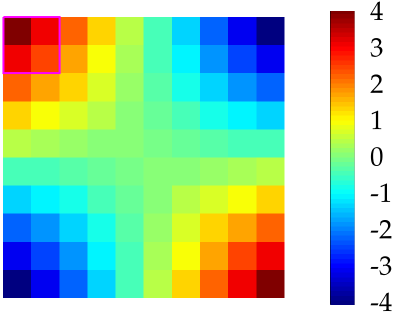
(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property



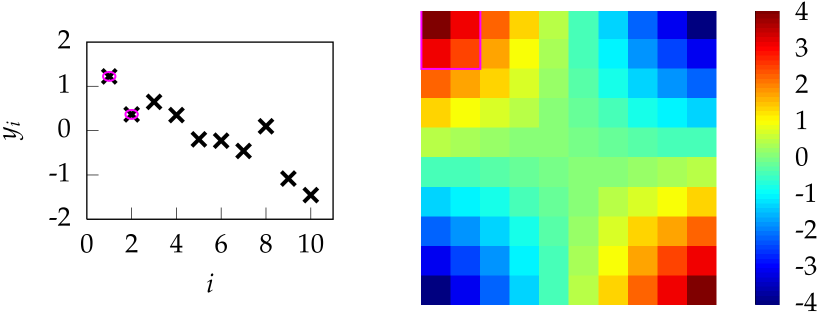
(a) A 10 dimensional sample



(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property

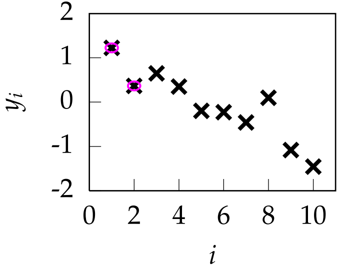


(a) A 10 dimensional sample

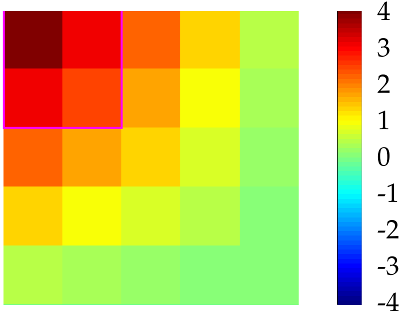
(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property



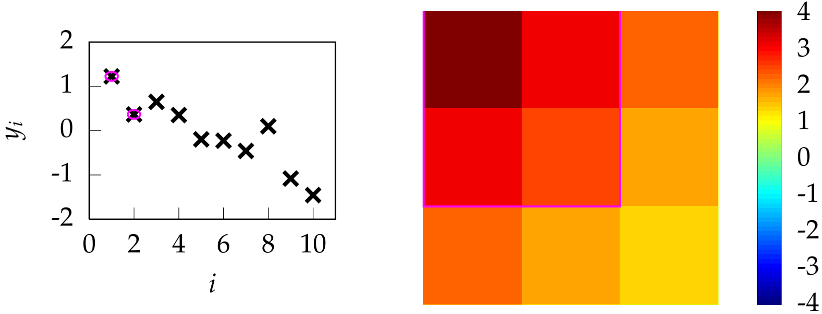
(a) A 10 dimensional sample



(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property

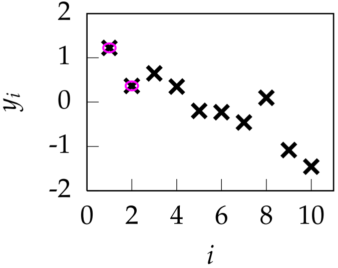


(a) A 10 dimensional sample

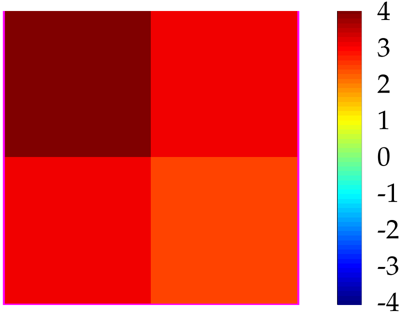
(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property



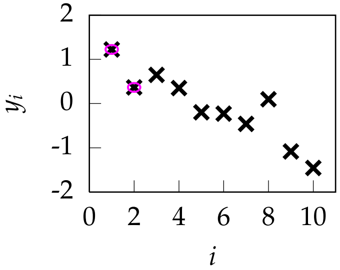
(a) A 10 dimensional sample



(b) colormap showing covariance between dimensions.

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

Gaussian Marginalization Property



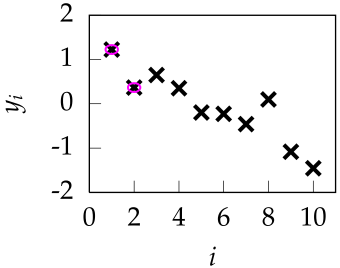
(a) A 10 dimensional sample

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

(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.

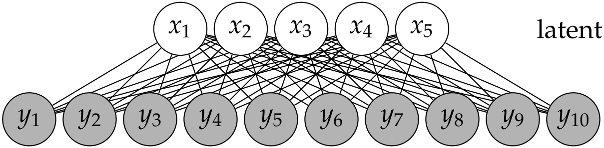
Figure : A sample from a 10 dimensional correlated Gaussian distribution.

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.

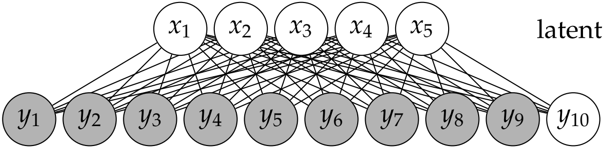
Work with Raquel Urtasun (Lawrence and Urtasun, 2009) and ongoing work with Max Zwießele and Nicoló Fusi.

Marginalization in Bipartite Undirected Graph



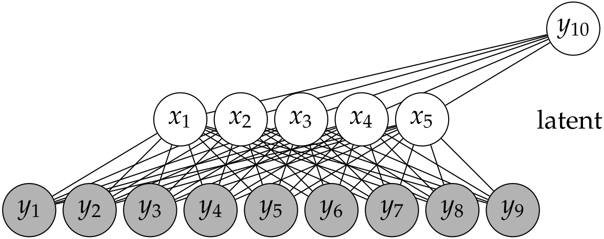
latent variables

Marginalization in Bipartite Undirected Graph



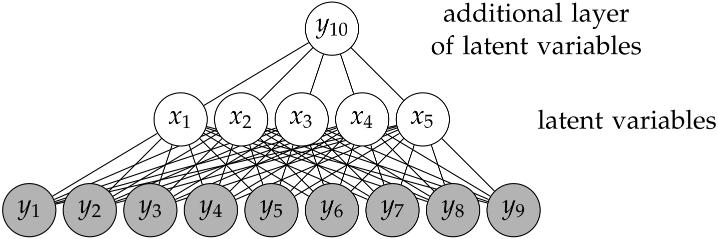
latent variables

Marginalization in Bipartite Undirected Graph

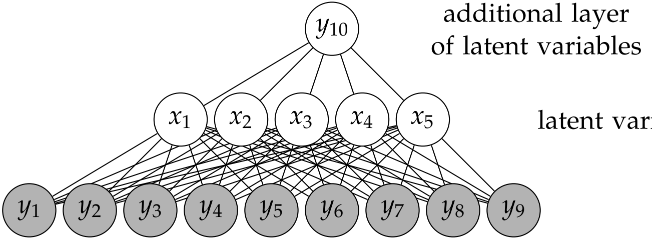


latent variables

Marginalization in Bipartite Undirected Graph



Marginalization in Bipartite Undirected Graph



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

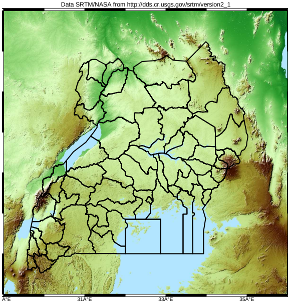
Methods that Interrelate Covariates

- 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!

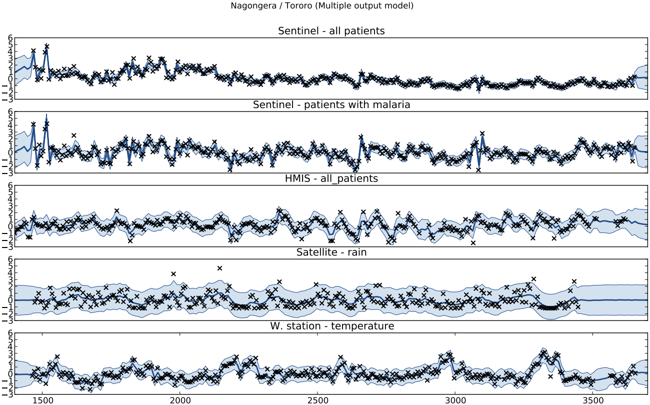
Example: Prediction of Malaria Incidence in Uganda

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

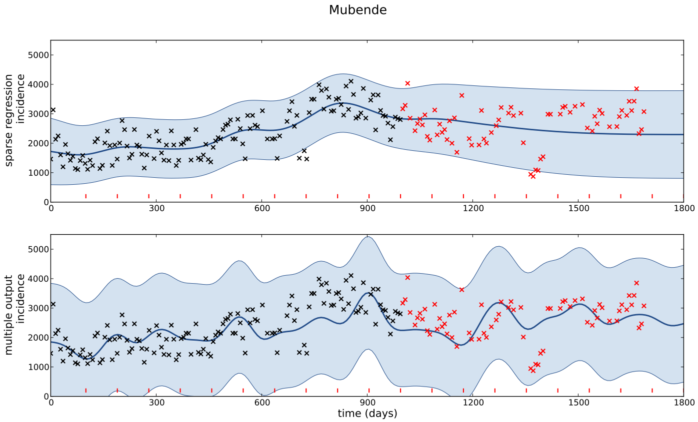
Malaria Prediction in Uganda



Malaria Prediction in Uganda

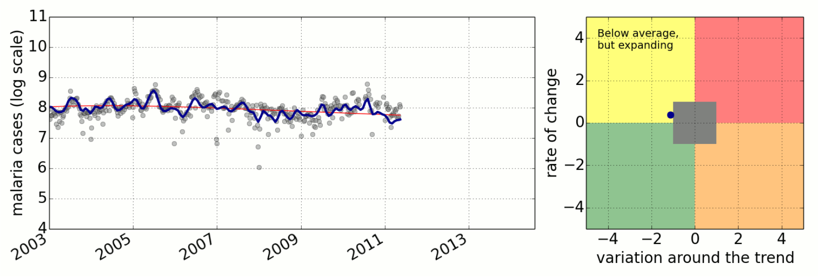


Malaria Prediction in Uganda

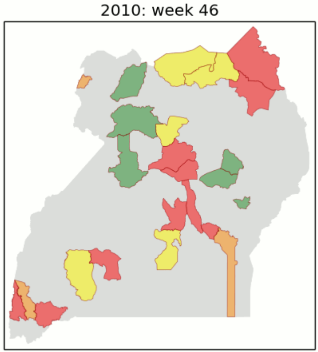


GP School at Makerere

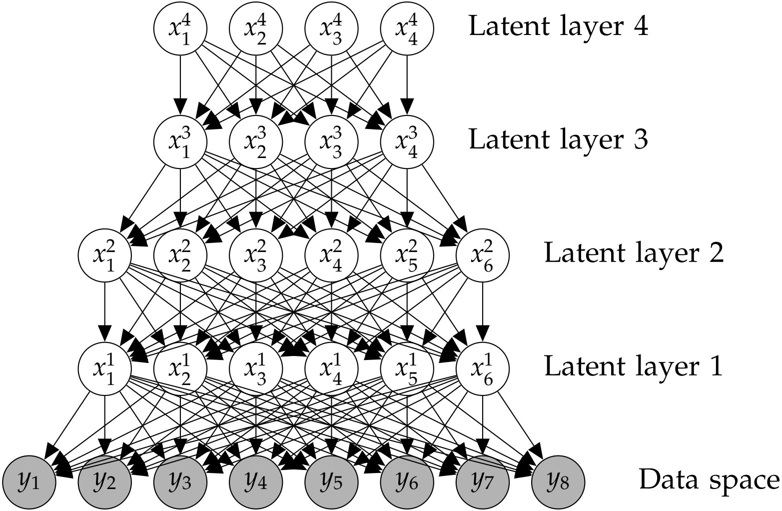
Early Warning Systems



Early Warning Systems

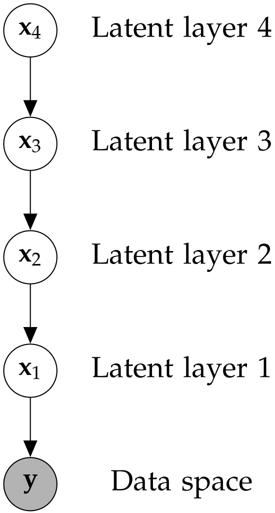


Deep Models



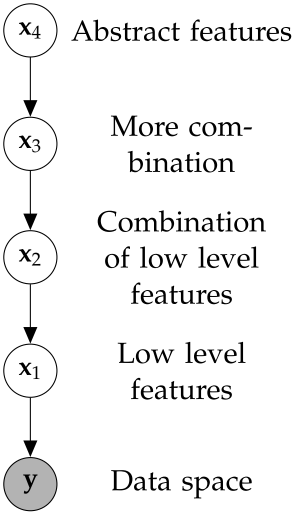
x_1^4
x_2^4
x_3^4
x_4^4
Latent layer 4
x_1^3
x_2^3
x_3^3
x_4^3
Latent layer 3
x_1^2
x_2^2
x_3^2
x_4^2
x_5^2
x_6^2
Latent layer 2
x_1^1
x_2^1
x_3^1
x_4^1
x_5^1
x_6^1
Latent layer 1
y_1
y_2
y_3
y_4
y_5
y_6
y_7
y_8
Data space

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



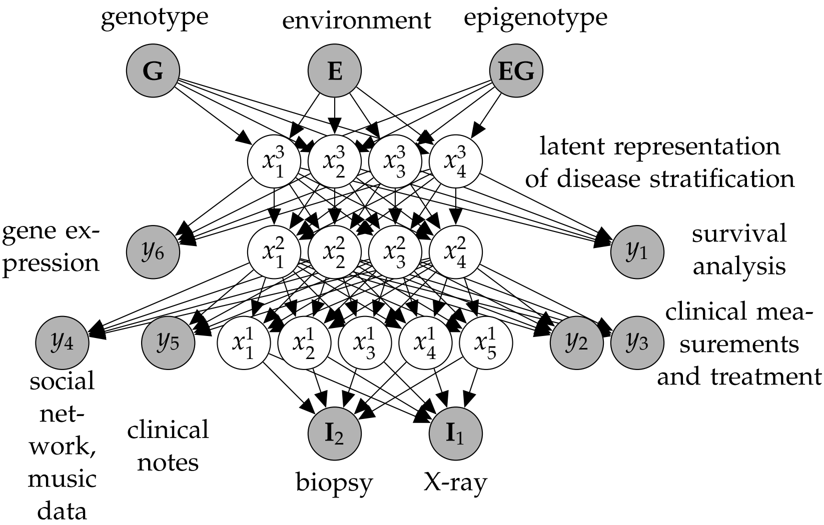
$\mathbf{x}_4$
Abstract features
$\mathbf{x}_3$
More com-
bination
$\mathbf{x}_2$
Combination
of low level
features
$\mathbf{x}_1$
Low level
features
$\mathbf{y}$
Data space

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.

Deep Health



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

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

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].

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.