Mean-Field Framework for Unsupervised Learning with Boltzmann Machines

LRI – November 22nd, 2017

Marylou Gabrié

Éric Tramel, Andre Manoel, Francesco Caltagirone

& Florent Krzakala

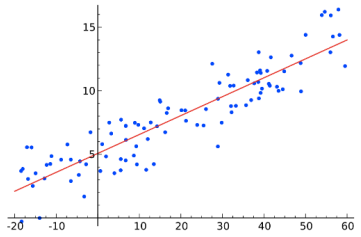
Machine Learning, Unsupervised Learning, Density Estimation

Machine learning: design algorithm able to learn from training data

Supervised learning:

learning a mapping between input and output (label)

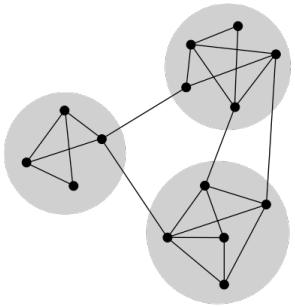
e.g. classification, regression

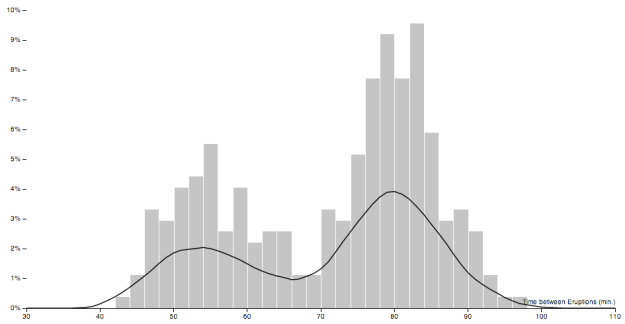


Unsupervised learning:

learning without labels, explaining key features of the data

e.g. clustering/community detection, dimensionality reduction, **generative models/density estimation**



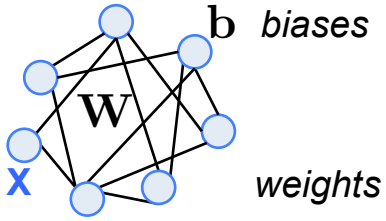


Boltzmann Machines
fundamental ML model borrowing from Physics

Generative model with pairwise interactions: an Ising model

(T. Sejnowski & G. Hinton 1985)

Set of binary correlated random variables $\mathbf{x} \in \{0,1\}^d$



Boltzmann measure $p_{\mathbf{W},\mathbf{b}}(\mathbf{x}) = \dfrac{e^{-E_{\mathbf{W},\mathbf{b}}(\mathbf{x})}}{Z_{\mathbf{W},\mathbf{b}}}$

$\theta = \{\mathbf{W}, \mathbf{b}\}$

Ising energy

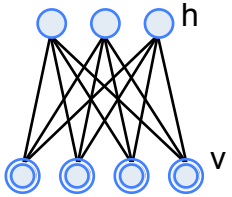
$$E_{\mathbf{W},\mathbf{b}}(\mathbf{x}) = -\sum_{i,j} W_{ij} x_i x_j - \sum_i b_i x_i$$

$$E_{\mathbf{W},\mathbf{b}}(\mathbf{x}) = -\mathbf{x}^T \mathbf{W} \mathbf{x} - \mathbf{b}^T \mathbf{x}$$

Boltzmann Machines
fundamental ML model borrowing from Physics

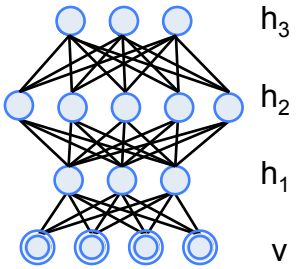
Introducing latent variables for representational power

Restricted Boltzmann Machine (RBM)
(bipartite)



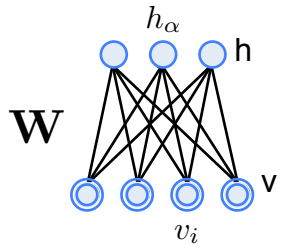
$$p_\theta(\mathbf{v}) = \sum_{\{\mathbf{h}\}} \frac{e^{-E_\theta(\mathbf{v},\mathbf{h})}}{Z_\theta}$$

Representation learning from hidden state with Deep BM



(R. Salakhutdinov & G. Hinton 2009)

RBM learning requires many approximations



Maximum likelihood learning has no closed form solution

$$\max_{\theta} \prod_{\mathbf{v}^* \in \text{training set}} p_\theta(\mathbf{v}^*)$$

$$\mathbf{v}^* \in \textbf{training set} \quad \ell(\theta|\mathbf{v}^*) = \ln p_\theta(\mathbf{v}^*) = \underbrace{\ln \left(\sum_{\{\mathbf{h}\}} e^{-E_\theta(\mathbf{v}^*, \mathbf{h})} \right)}_{\text{data dependent easy}} \underbrace{- \ln Z_\theta}_{\text{log partition hard}}$$

Iterative gradient ascent training algorithm $W_{i\alpha} \leftarrow W_{i\alpha} + \Delta W_{i\alpha}$

With approximated gradients

$$\Delta W_{i\alpha} \propto \frac{\partial \ell(\theta|\mathbf{v}^*)}{\partial W_{i\alpha}} = \langle v_i^* h_\alpha \rangle_{\mathbf{v}^*} - \langle v_i h_\alpha \rangle$$

easy hard

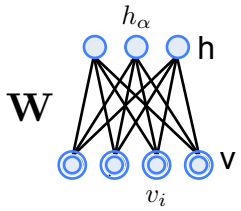
$p_\theta(\mathbf{h}|\mathbf{v}^*)$ $p_\theta(\mathbf{v}, \mathbf{h})$

$\langle \ldots \rangle \equiv \mathbb{E}[\ldots]$

Monte Carlo heuristics for RBMs

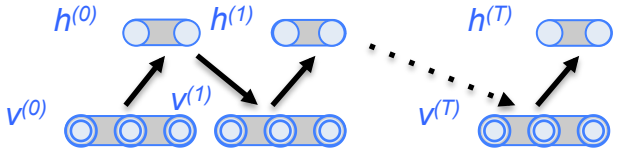
Markov Chains

Conditional independence



$$\begin{cases} h_\alpha^{(t)} & \sim p(h_\alpha|\mathbf{v}^{(t)}, \mathbf{h}_{\setminus\alpha}^{(t-1)}) = p(h_\alpha|\mathbf{v}^{(t)}) \\ v_i^{(t+1)} & \sim p(v_i|\mathbf{h}^{(t)}) \end{cases}$$

Efficient Gibbs sampling by blocks



Contrastive-divergence Monte Carlo with non-thermalized samples

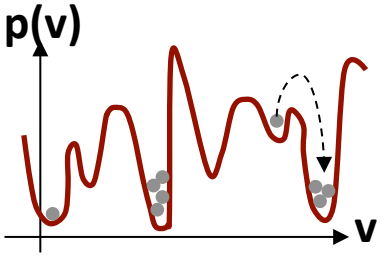
$$\langle v_i h_\alpha \rangle \simeq \frac{1}{K} \sum_{k=1}^{K} v_k^{(T)} h_k^{(T)}$$

with T = O(1) !

(G. Hinton 2002)

Successful training but …

Gibbs sampling unlikely to be accurate for
complicated real valued landscapes
(more sophisticated schemes:
parallel tempering, annealed importance sampling ..)



$\langle \ldots \rangle \equiv \mathbb{E}[\ldots]$

A mean-field framework instead ?

Previous works

On RBMs and DBMs

- Tieleman et al. (2009): Naive (or Weiss, or fully factorized) mean-field unsatisfactory
- Salakhutdinov & Hinton (2009): Using naive mean-field for training would be wrong

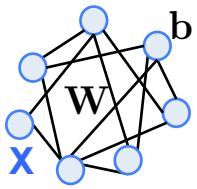
On fully visible Ising

- Kappen (99): Successful learning with linear response corrections on fully visible BM
- Cocco & Monasson (2012): Cluster expansion including higher orders for inverse Ising problem
- Ricci-Tersenghi (2012): Comparison of the Bethe approximation and other mean-field methods for inverse Ising problem

Can we use extended mean field methods

to efficiently train latent variables BMs ?

Thouless-Anderson-Palmer free energy



Intractable distribution and log-partition function = free energy

$$\mathbf{s} \in \{0,1\}^d \qquad E(\mathbf{s}) = -\mathbf{s}^T\mathbf{W}\mathbf{s} - \mathbf{b}^T\mathbf{s} \qquad p_\theta(\mathbf{s}) = \frac{1}{Z_\theta} e^{\mathbf{s}^T\mathbf{W}\mathbf{s}+\mathbf{b}^T\mathbf{s}} \qquad \ln Z_\theta = -F$$

Plefka (82), Georges-Yedidia (99) derivation

Legendre transform $-\beta F(\mathbf{q}) = \ln \sum_{\{\mathbf{s}\}} e^{-\beta(E(\mathbf{s})+\mathbf{q}^T\mathbf{s})} \Rightarrow \begin{cases} -\beta G(\mathbf{m}) = -\beta \max_{\mathbf{q}}[F(\mathbf{q}) + \mathbf{q}^T\mathbf{m}] \\ \langle \mathbf{s} \rangle = \mathbf{m} \end{cases}$ Mean-field parameters = magnetizations

Large temperature expansion up to second order $G(\mathbf{m}) \approx G_0 + \beta \left.\frac{\partial G}{\partial \beta}\right|_{\beta=0} + \frac{\beta^2}{2} \left.\frac{\partial^2 G}{\partial \beta^2}\right|_{\beta=0}$

Self consistency $\left.\frac{dG}{dm_i}\right|_{m_i} = 0 \Rightarrow m_i = \sigma\left[\beta b_i + \sum_j \beta W_{ij} m_j - \beta^2 W_{ij}^2 \left(m_i - \frac{1}{2}\right)\left(m_j - {m_j}^2\right)\right]$ NMF TAP

$\sigma(x) = \frac{1}{1+e^{-x}}$

TAP free energy as a function of magnetizations

$$F_{\mathrm{TAP}} = S_{\mathrm{MF}} - \sum_i \beta b_i m_i - \sum_{i,j} \beta W_{ij} m_i m_j - \sum_{i,j} \beta^2 \frac{W_{ij}^2}{2} (m_i - {m_i}^2)(m_j - {m_j}^2)$$

NMF TAP

$\langle \ldots \rangle \equiv \mathbb{E}[\ldots]$

TAP free energy on the restricted graph

Intractable joint distribution

$$\mathbf{v} \in \{0,1\}^{d^{\mathbf{v}}} \quad \mathbf{h} \in \{0,1\}^{d^{\mathbf{h}}} \quad p_\theta(\mathbf{v},\mathbf{h}) = \frac{1}{Z_\theta} e^{\mathbf{v}^T \mathbf{W}\mathbf{h} + \mathbf{b_v}^T \mathbf{v} + \mathbf{b_h}^T \mathbf{h}} \qquad \ln Z_\theta = -F$$

TAP equations

- magnetizations parameters $\mathbf{m}^{\mathbf{v}} = \langle \mathbf{v} \rangle,\ \mathbf{m}^{\mathbf{h}} = \langle \mathbf{h} \rangle \in [0,1]$
- small weights expansion of free energy F $(W \leftarrow \beta W)$

$$F_{\mathrm{TAP}} = S_{\mathrm{MF}}(\theta) - \mathbf{b_v}^T \mathbf{m}^{\mathbf{v}} - \mathbf{b_h}^T \mathbf{m}^{\mathbf{h}} - \mathbf{m}^{\mathbf{v}T} \mathbf{W} \mathbf{m}^{\mathbf{h}} - \sum_{i,\alpha} \frac{W_{i\alpha}^2}{2} (m_i^{\mathbf{v}} - m_i^{\mathbf{v}\,2})(m_\alpha^{\mathbf{h}} - m_\alpha^{\mathbf{h}\,2})$$

NMF — TAP

- self consistency relations for TAP free energy modes

$$m_i^{\mathbf{v}} = \sigma\left[b_{\mathbf{v}i} + \sum_\alpha W_{i\alpha} m_\alpha^{\mathbf{h}} - W_{i\alpha}^2 \left(m_i^{\mathbf{v}} - \frac{1}{2} \right) \left(m_\alpha^{\mathbf{h}} - m_\alpha^{\mathbf{h}\,2} \right) \right]$$

NMF — TAP

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Unsupervised learning with a TAP machine

(M.G, E.Tramel, F. Krzakala NIPS 2015)

We propose to define a TAP machine as:

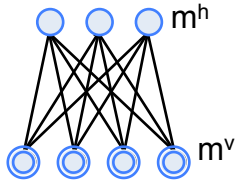
A bipartite graphical model parametrized by $\theta = \{\mathbf{W}, \mathbf{b_v}, \mathbf{b_h}\}$

Implicitely defining a set of magnetizations $\{(\mathbf{m_1^v}, \mathbf{m_1^h}), (\mathbf{m_2^v}, \mathbf{m_2^h}), ...\}$

Computed as the fixed point of TAP equations

$$m_i^{\mathbf{v}} \leftarrow \sigma\left[\mathbf{b_v}_i + \sum_\alpha W_{i\alpha} m_\alpha^{\mathbf{h}} - W_{i\alpha}^2 \left(m_i^{\mathbf{v}} - \frac{1}{2}\right)\left(m_\alpha^{\mathbf{h}} - m_\alpha^{\mathbf{h}\,2}\right)\right]$$

$$m_\alpha^{\mathbf{h}} \leftarrow \sigma\left[\mathbf{b_h}_\alpha + \sum_i W_{i\alpha} m_i^{\mathbf{v}} - W_{i\alpha}^2 \left(m_\alpha^{\mathbf{h}} - \frac{1}{2}\right)\left(m_i^{\mathbf{v}} - m_i^{\mathbf{v}\,2}\right)\right]$$



We define a natural and tractable objective for unsupervised learning:

The TAP log-likelihood

TAP log-likelihood

$$\ln Z_\theta \simeq -F_{\mathrm{TAP}} \quad \Rightarrow \quad \ell_{\mathrm{TAP}}(\theta|\mathbf{v}^*) = \ln\left(\sum_{\{\mathbf{h}\}} e^{-E_\theta(\mathbf{v}^*,\mathbf{h})}\right) + F_{\mathrm{TAP}}$$

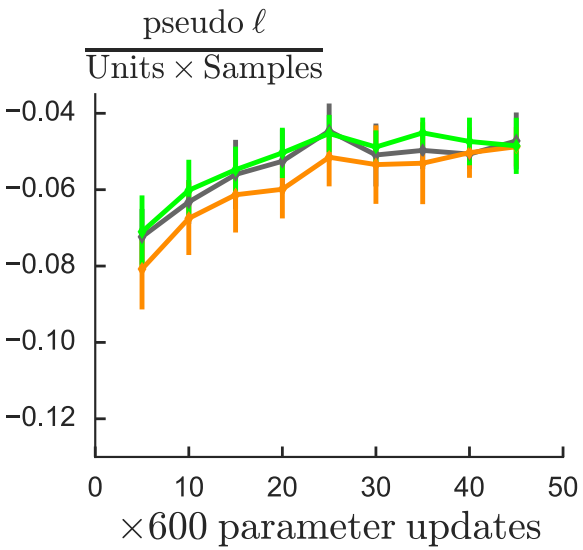
Gradient ascent of the TAP likelihood: a tractable training scheme for TAP machines

TAP log-likelihood $\ell_{\mathrm{TAP}}(\theta|\mathbf{v}^*) = \ln \sum_{\{\mathbf{h}\}} e^{-E_\theta(\mathbf{v}^*,\mathbf{h})} + F_{\mathrm{TAP}}$

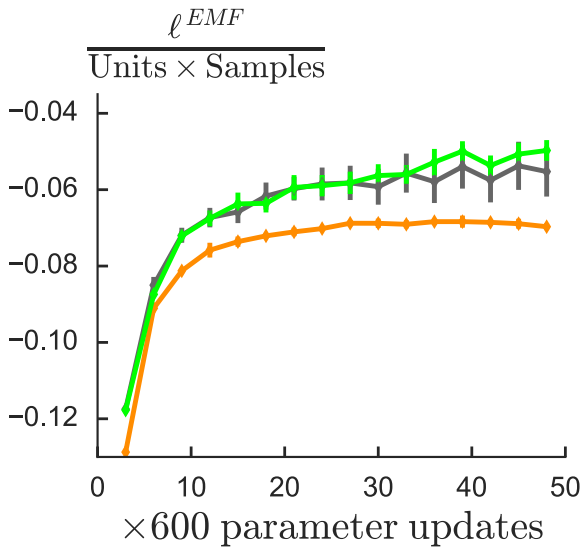
TAP gradients $\Delta W_{i\alpha} \propto \frac{\partial \ell_{\mathrm{TAP}}(\theta|\mathbf{v}^*)}{\partial W_{i\alpha}} = \underbrace{\langle v_i^* h_\alpha \rangle_{p(h|v^*)}}_{\text{easy}} + \underbrace{\frac{\partial F_{TAP}}{\partial W_{i\alpha}}}_{\text{explicit expression in terms of magnetizations}}$

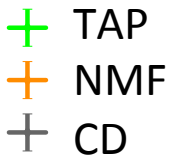
Binarized MNIST

pseudo log-likelihood



TAP log-likelihood





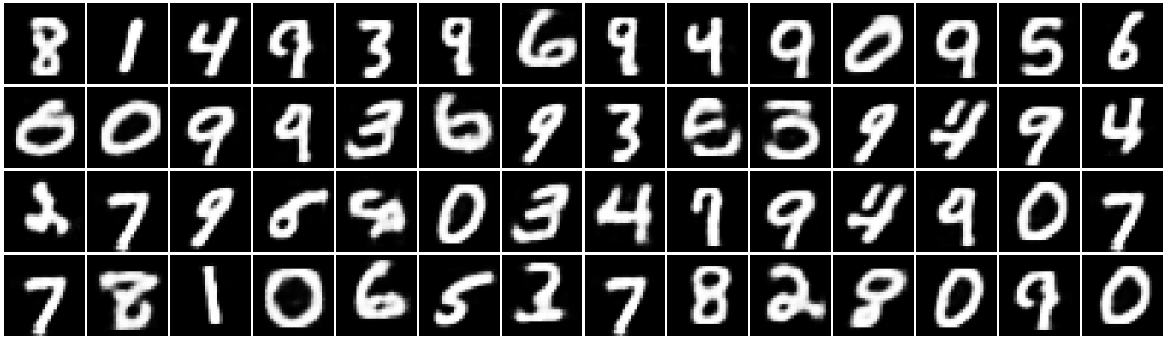
Magnetizations of TAP machines after learning

$$m_i^{\mathbf{v}} \leftarrow \sigma\left[b_{\mathbf{v}i} + \sum_{\alpha} W_{i\alpha} m_{\alpha}^{\mathbf{h}} - {W_{i\alpha}}^2 \left(m_i^{\mathbf{v}} - \frac{1}{2}\right)\left(m_{\alpha}^{\mathbf{h}} - {m_{\alpha}^{\mathbf{h}}}^2\right)\right]$$

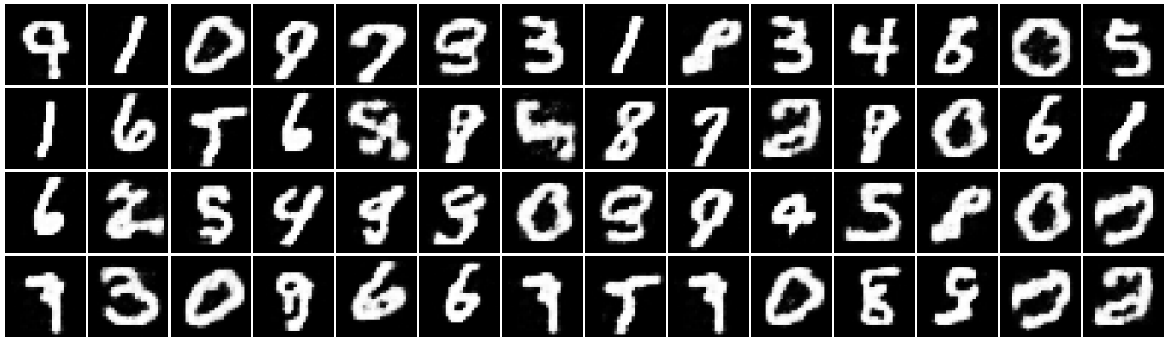
NMF

TAP

TAP fixed points with trained weights and biases for different initializations

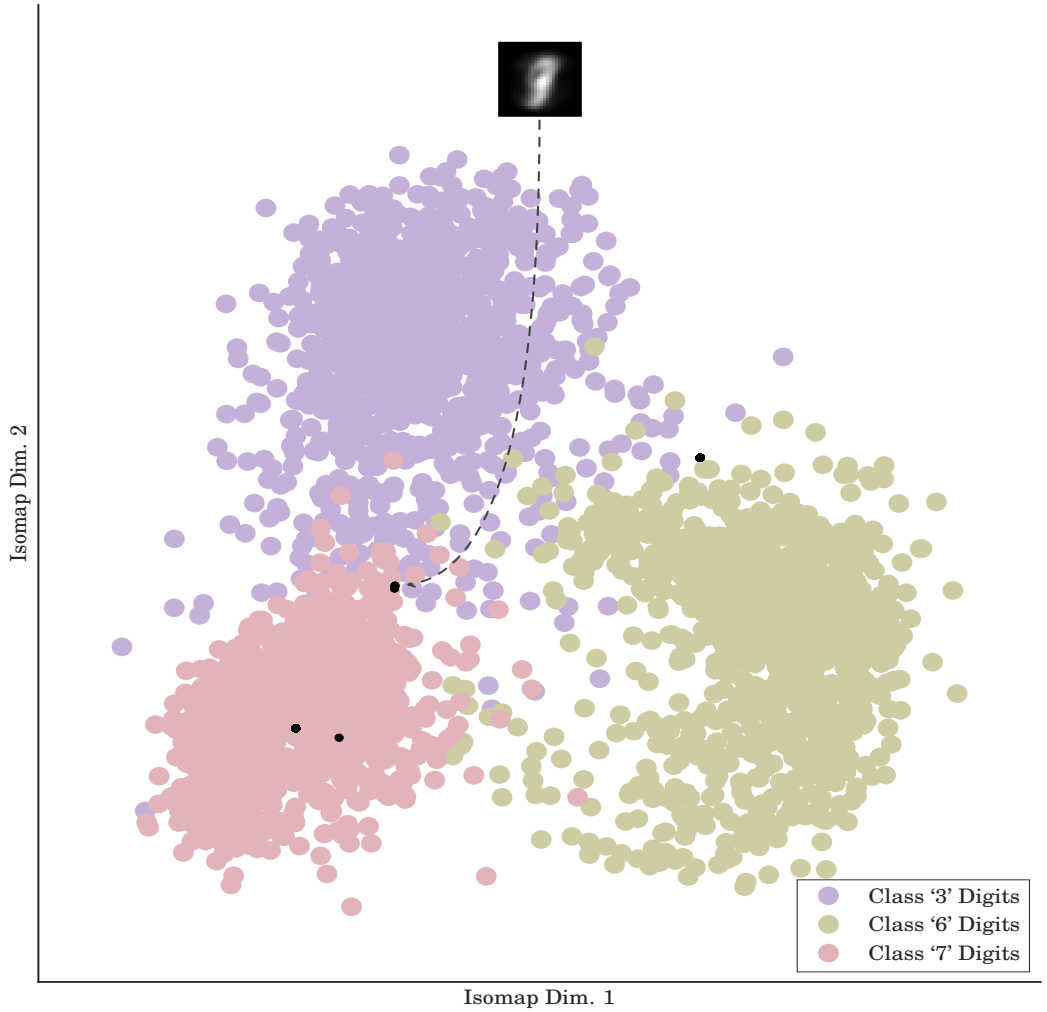


Now for a naive-MF machine



Magnetizations of a TAP machine during learning

Evolution of magnetizations in « phase space »

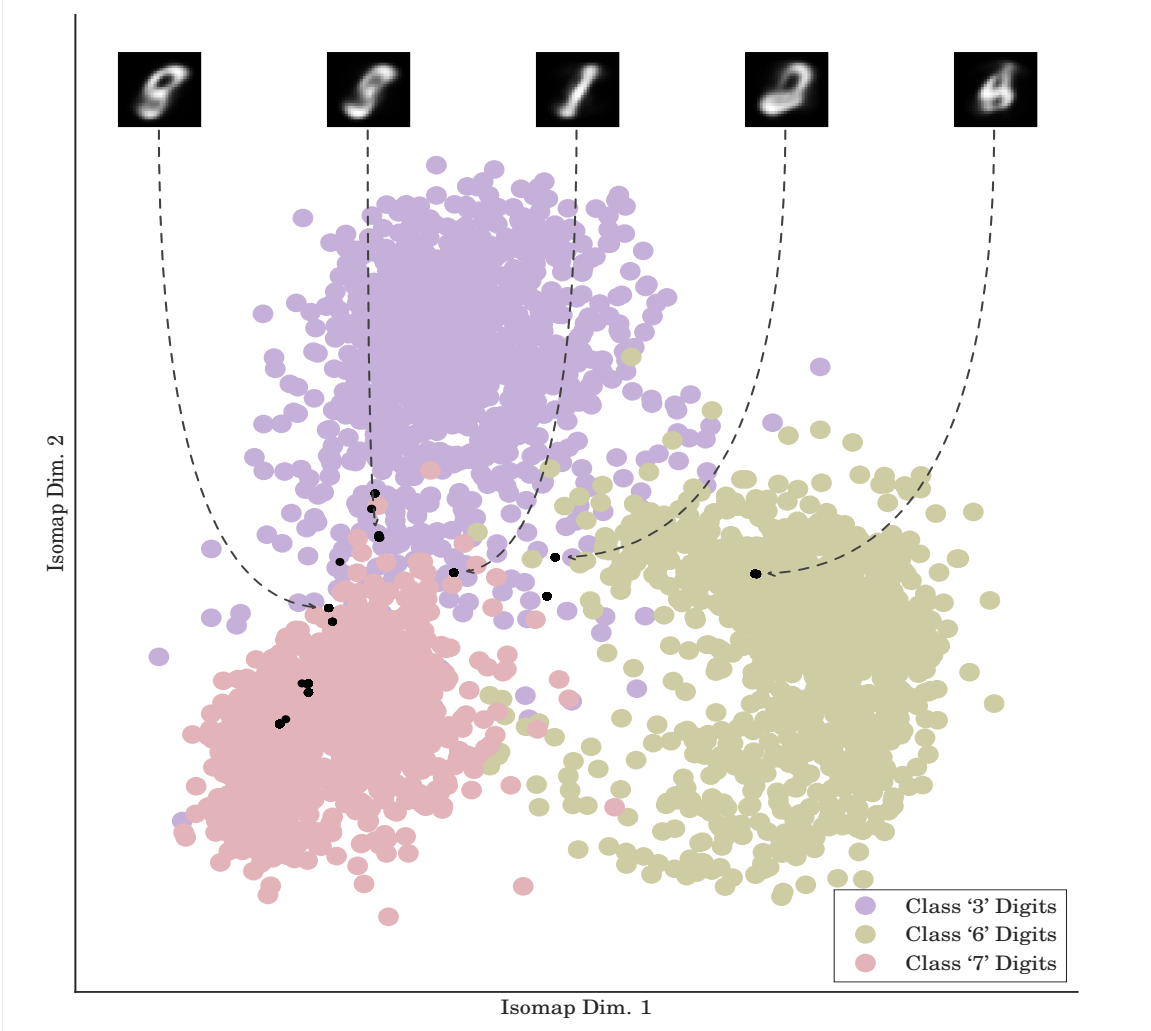


Epoch 1

(E.Tramel, M.G., A. Manoel, F. Caltagirone, F. Krzakala 2017)

Magnetizations of a TAP machine during learning

Evolution of magnetizations in « phase space »

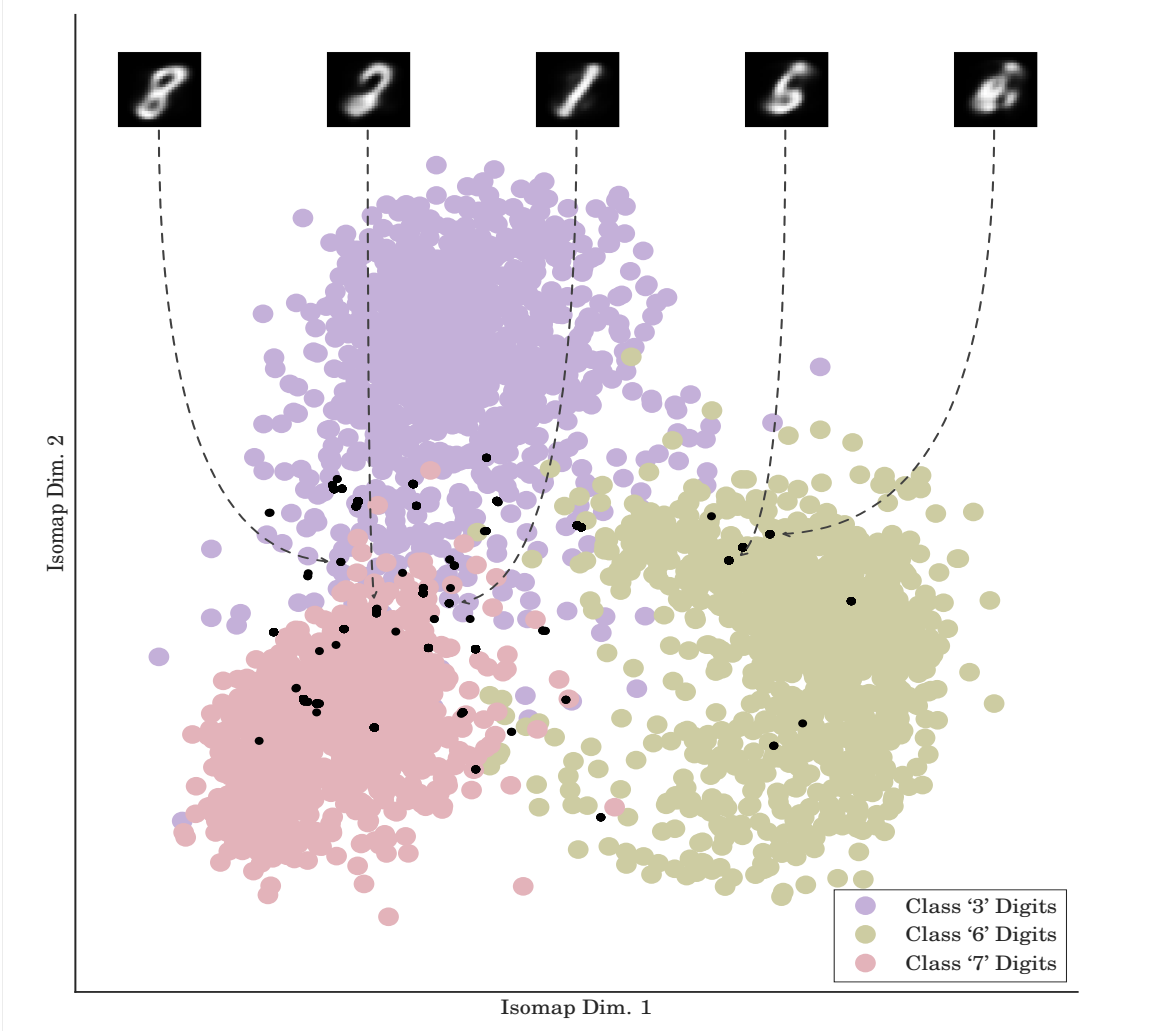


Epoch 5

(E.Tramel, M.G., A. Manoel, F. Caltagirone, F. Krzakala 2017)

Magnetizations of a TAP machine during learning

Evolution of magnetizations in « phase space »

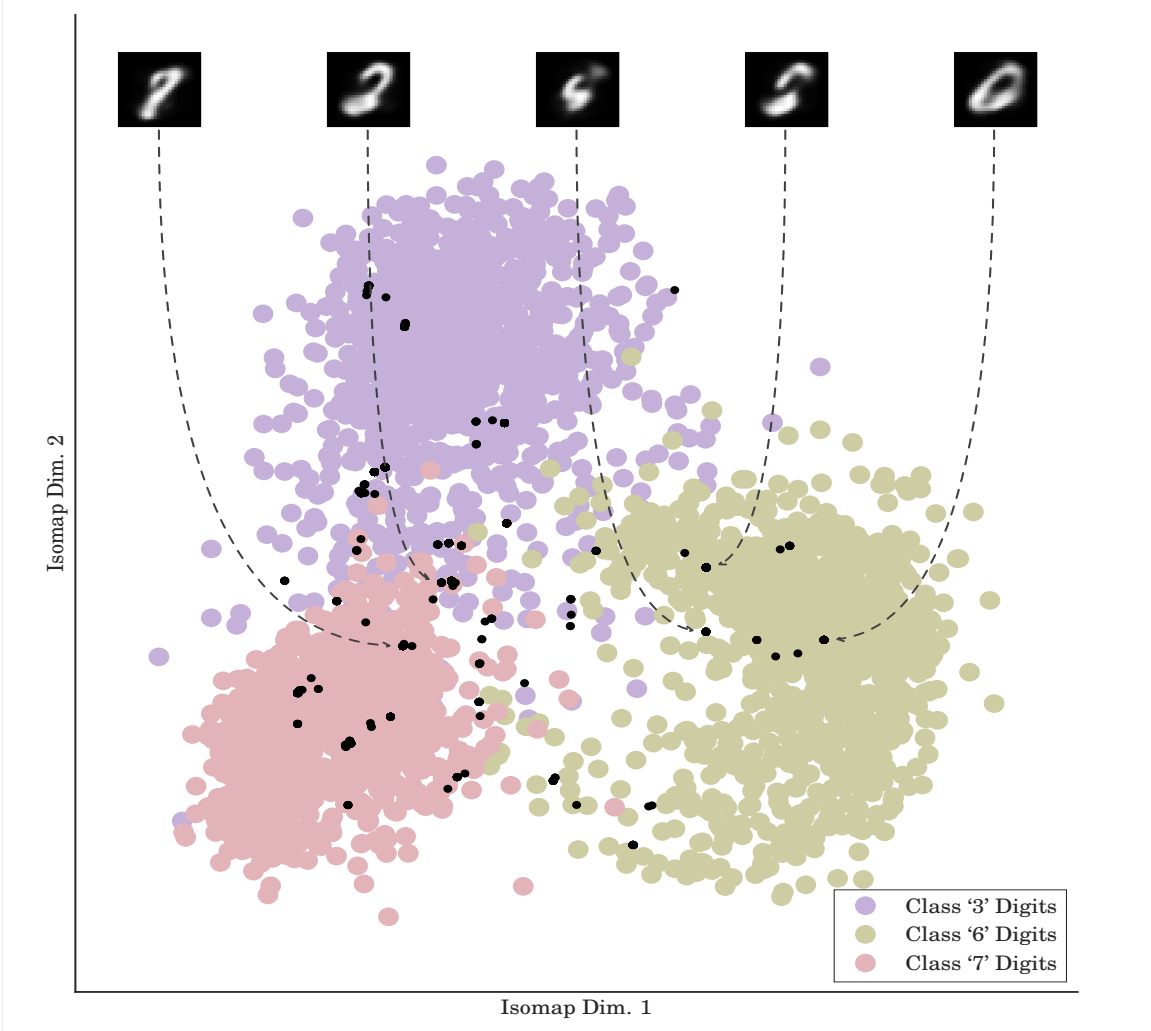


Epoch 25

(E.Tramel, M.G., A. Manoel, F. Caltagirone, F. Krzakala 2017)

Magnetizations of a TAP machine during learning

Evolution of magnetizations in « phase space »

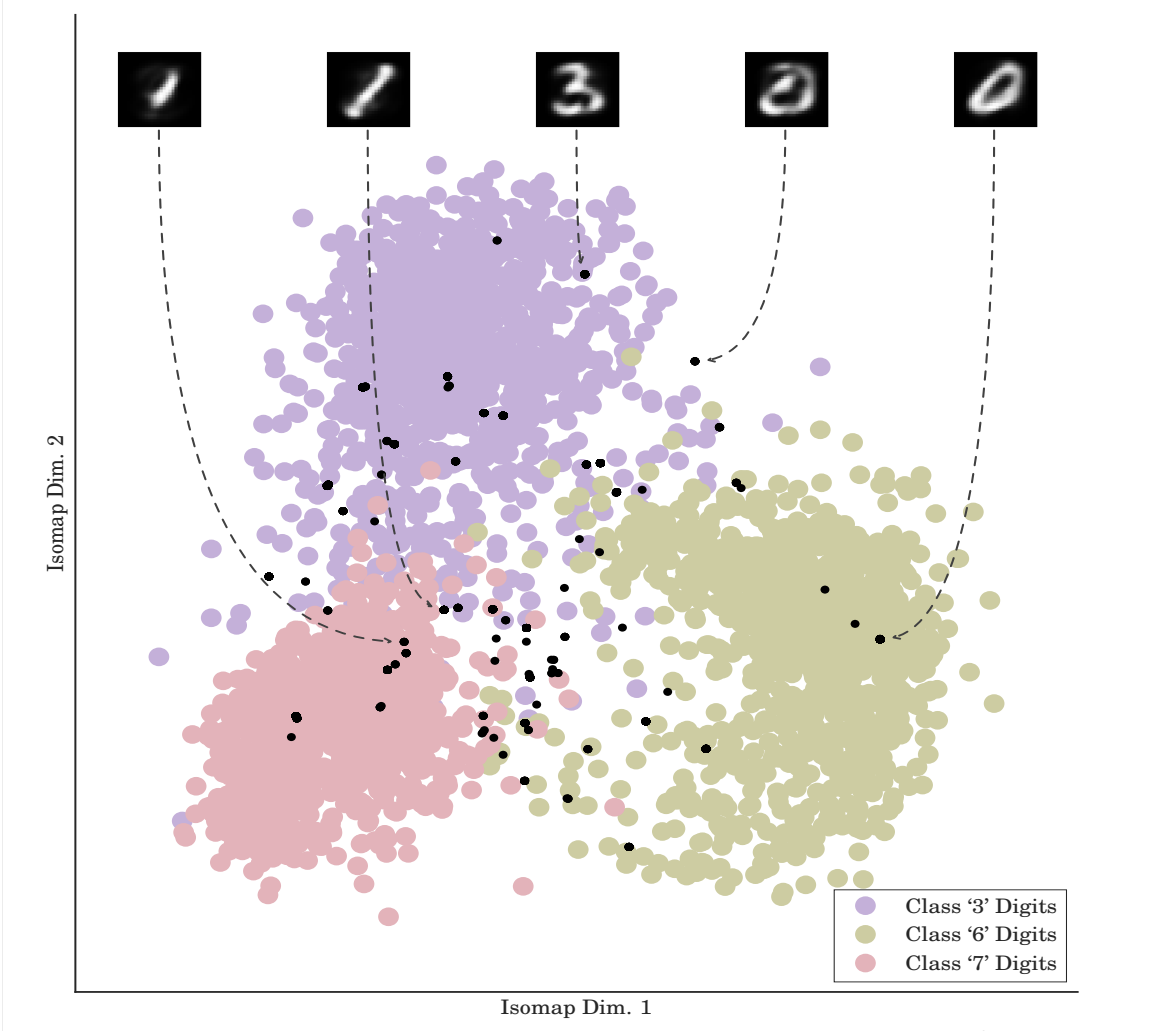


Epoch 50

(E.Tramel, M.G., A. Manoel, F. Caltagirone, F. Krzakala 2017)

Magnetizations of a TAP machine during learning

Evolution of magnetizations in « phase space »



Epoch 100

(E.Tramel, M.G., A. Manoel, F. Caltagirone, F. Krzakala 2017)

Mean-Field Framework for Unsupervised Learning

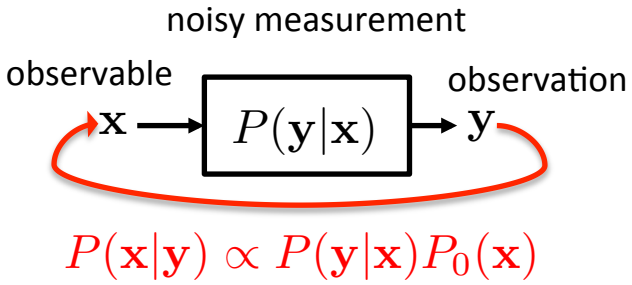
1 – TAP formalism to define TAP machines

- Deterministic model inspired by Boltzmann Machines
- Endowed with a tractable training algorithm
- Acheiving similar quality of training than CD
- Providing new ways of monitoring learning (TAP likelihood, magnetizations)

2 - An application of TAP machines to Bayesian inference

Learning complex structured priors for compressed sensing (CS)

Bayesian inference problem



Machine learning construction of priors for Bayesian inference

- Learn an RBM from data samples to estimate underlying P_0
- But for an RBM $P_0(x)$ is intractable
- Resort instead to a MF latent variable model

Prior learning for compressed sensing

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{w}$$

$\mathbf{F} \in \mathbb{R}^{M \times N}$, $\mathbf{y} \in \mathbb{R}^M$, $\mathbf{x} \in \mathbb{R}^N$

$\mathbf{x}$: sparse

$$M << N$$

$$w_i \sim \mathcal{N}(0, \Delta)$$

Learn to locate the non-zero entries of x with a TAP machine for the reconstruction ?

Learning the support of MNIST digits for CS

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{w}$$

(S. Rangan et al. 2011)

(E.Tramel, A. Drémeau, F. Krzakala J. Stat. Mech 2016)

Binarized MNIST

Priors tested

(a) Gauss-Bernouli i.i.d.: $P_0(x_i) = (1-\rho)\delta(x_i) + \rho\mathcal{N}(x_i;\mu,\sigma^2)$

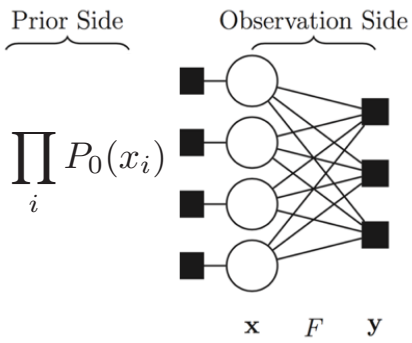
(b) Gauss-Bernoulli non-i.i.d.: $P_0(x_i) = (1-\rho_i)\delta(x_i) + \rho_i\mathcal{N}(x_i;\mu,\sigma^2)$

(c) NMF latent graph. model: $m_i^{\mathbf{v}} = \sigma\left[\theta_{\mathbf{v}i} + \sum_\alpha W_{i\alpha} m_\alpha^{\mathbf{h}}\right]$

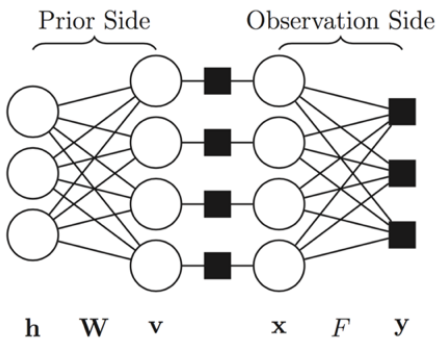
(d) TAP latent graph. Model: $m_i^{\mathbf{v}} = \sigma\left[\theta_{\mathbf{v}i} + \sum_\alpha W_{i\alpha} m_\alpha^{\mathbf{h}} - W_{i\alpha}^2\left(m_i^{\mathbf{v}} - \frac{1}{2}\right)\left(m_\alpha^{\mathbf{h}} - m_\alpha^{\mathbf{h}^2}\right)\right]$

Factor graph representation of posterior $P(\mathbf{x}|\mathbf{y}) \propto P(\mathbf{y}|\mathbf{x})P_0(\mathbf{x})$

Fully factorized priors



MF latent variable priors



An efficient way to incorporate ML to CS ?

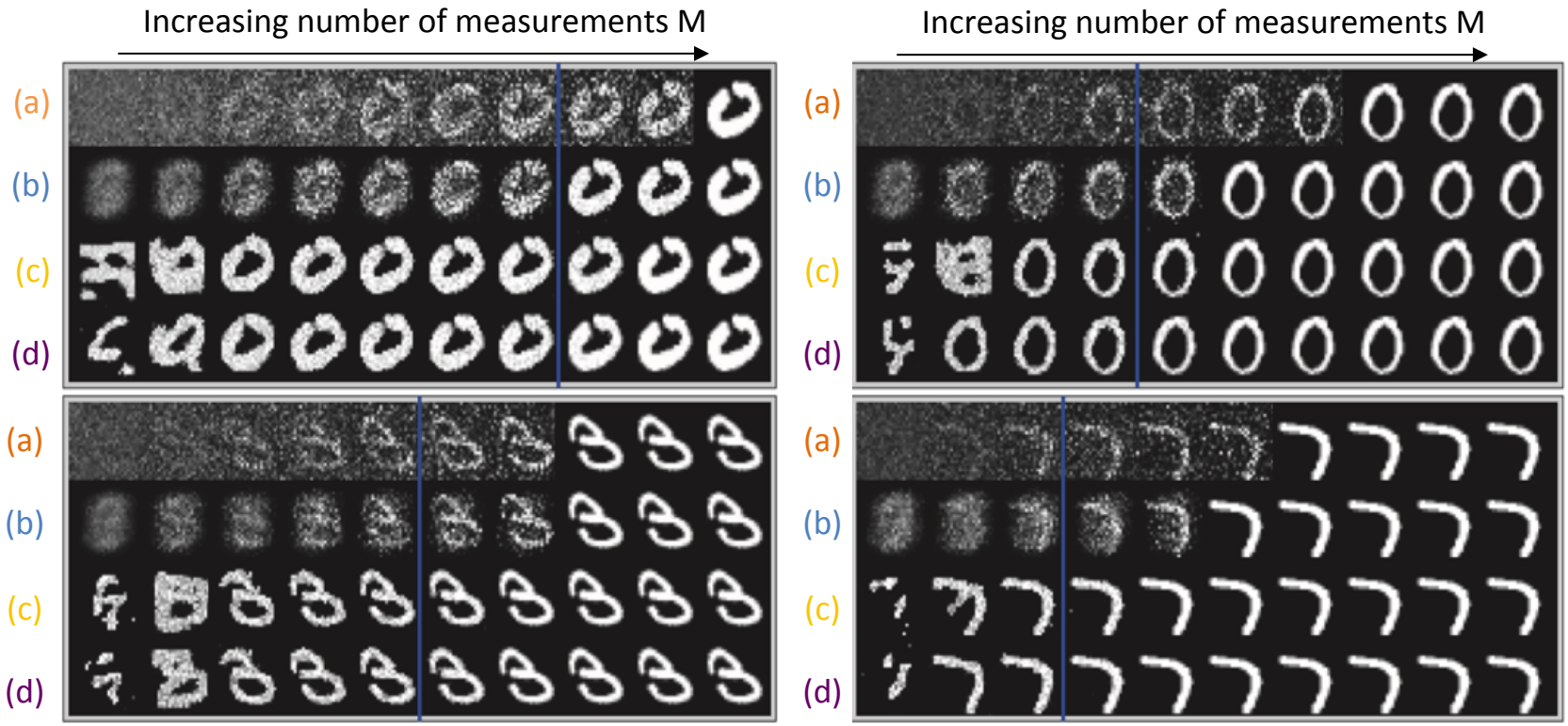
Priors tested

(a) Gauss-Bernouli i.i.d.

(b) Gauss-Bernoulli non-i.i.d.

(c) NMF inference on graph. model

(d) TAP inference on graph. model



Mean-Field Framework for Unsupervised Learning

1 - TAP formalism to define TAP machines

2 - An application of TAP machines to Bayesian inference

- Structured prior can be incorporated thanks to MF/TAP inference on MF/TAP machines.

- Reconstruction is possible from a lot fewer measurements.

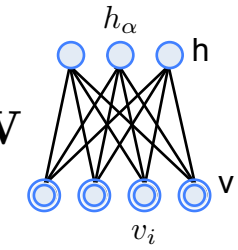
3 – Generalization to arbitrary variables and deep networks

Generalized Restricted Boltzmann Machines

(E.Tramel, M.G., A. Manoel, F. Caltagirone, F. Krzakala arxiv.org/abs/1702.032602017)

Arbitrary continuous priors for real valued units

$$\begin{matrix}\mathbf{v} \in \mathbb{R}^{d^{\mathbf{v}}} \\ \mathbf{h} \in \mathbb{R}^{d^{\mathbf{h}}}\end{matrix} \qquad p_\theta(\mathbf{v},\mathbf{h}) = \frac{1}{Z_\theta} e^{\mathbf{v}^T \mathbf{W} \mathbf{h}} \prod_i P_{\theta_i^{\mathbf{v}}}(v_i) \prod_\alpha P_{\theta_\alpha^{\mathbf{h}}}(h_\alpha)$$



Log – likelihood

$$\ell(\theta|\mathbf{v}^*) = \ln p_\theta(\mathbf{v}^*) = \sum_\alpha \ln \int \mathrm{d}h_\alpha P_{\theta_\alpha^{\mathbf{h}}}(h_\alpha) e^{\sum_i W_{i\alpha} v_i^* h_\alpha} + \sum_i \ln P_{\theta_i^{\mathbf{v}}}(v_i^*) - \ln Z_\theta$$

TAP approximation using « mean » and « variance » variational parameters

This again recovers Approximate Message Passing algorithm

$$\begin{matrix}\mathbf{m}^{\mathbf{v}} = \langle \mathbf{v} \rangle, \ \mathbf{c}^{\mathbf{v}} = \langle \mathbf{v}^2 \rangle - \langle \mathbf{v} \rangle^2 \\ \mathbf{m}^{\mathbf{h}} = \langle \mathbf{h} \rangle, \ \mathbf{c}^{\mathbf{h}} = \langle \mathbf{h}^2 \rangle - \langle \mathbf{h} \rangle^2\end{matrix} \quad \Longrightarrow \quad F_{\mathrm{TAP}}(\mathbf{m}^{\mathbf{v}}, \mathbf{c}^{\mathbf{v}}, \mathbf{m}^{\mathbf{h}}, \mathbf{c}^{\mathbf{h}})$$

Training continuous TAP machines

CBCL Faces

Distributions

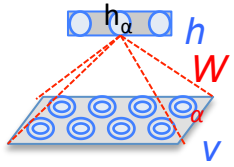
Binary hidden $$P_0(h_\alpha) = \frac{e^{\theta^{\mathbf{h}}_\alpha h_\alpha}}{1 + e^{\theta^{\mathbf{h}}_\alpha}}$$

Truncated Gaussian visibles $$P_0(v_i) = \mathcal{N}(v_i; \theta^{\mathbf{v}}_i, \sigma^2_i)\mathbb{1}_{0 \le v_i \le 1}$$

Log-likelihood

$$\ell_{\mathrm{TAP}}(\mathbf{W}, \theta^{\mathbf{h}}_\alpha, \theta^{\mathbf{v}}_i, \sigma^2_i | \mathbf{v})$$

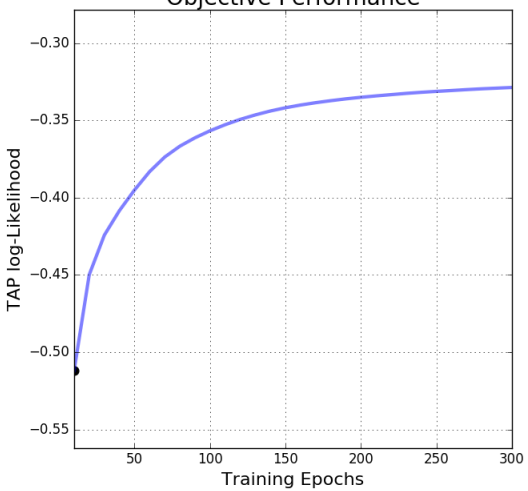
Learned weights



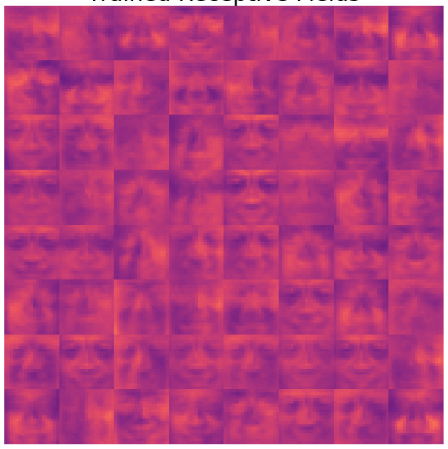
Marginal means

$$\mathbf{m}^{\mathbf{v}} = \langle \mathbf{v} \rangle$$

Objective Performance

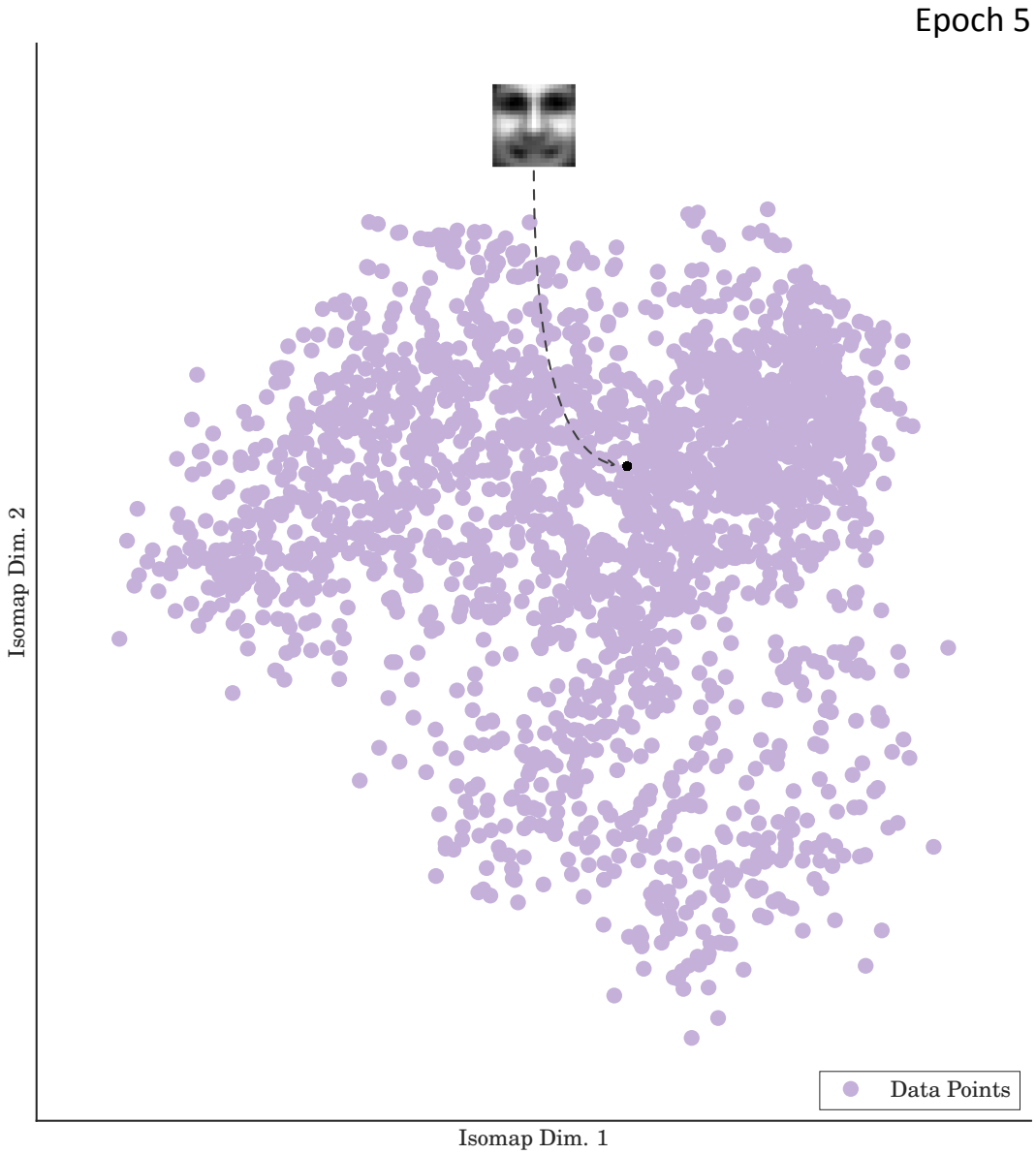


Trained Receptive Fields

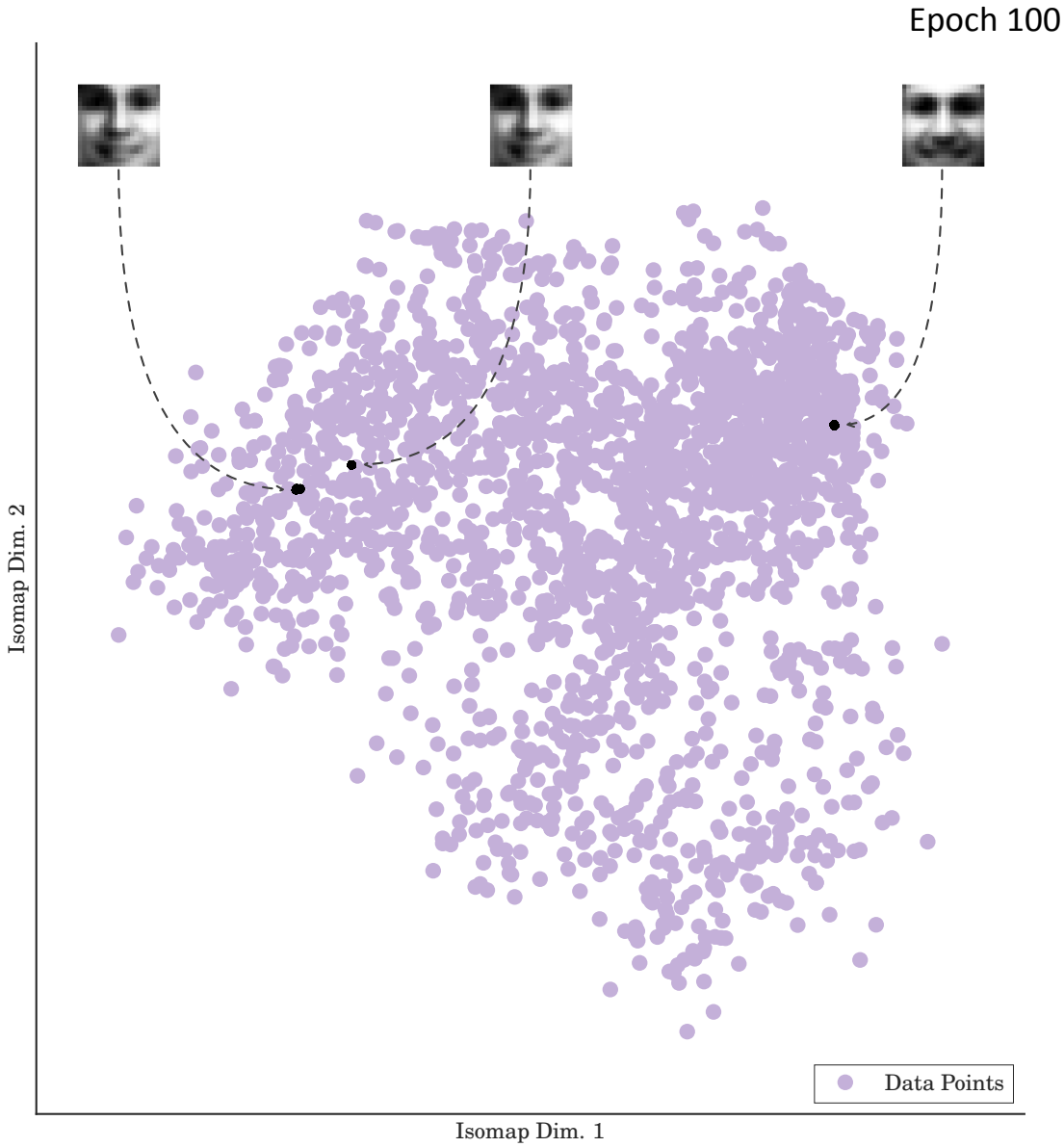


Equilibrium Mean Values

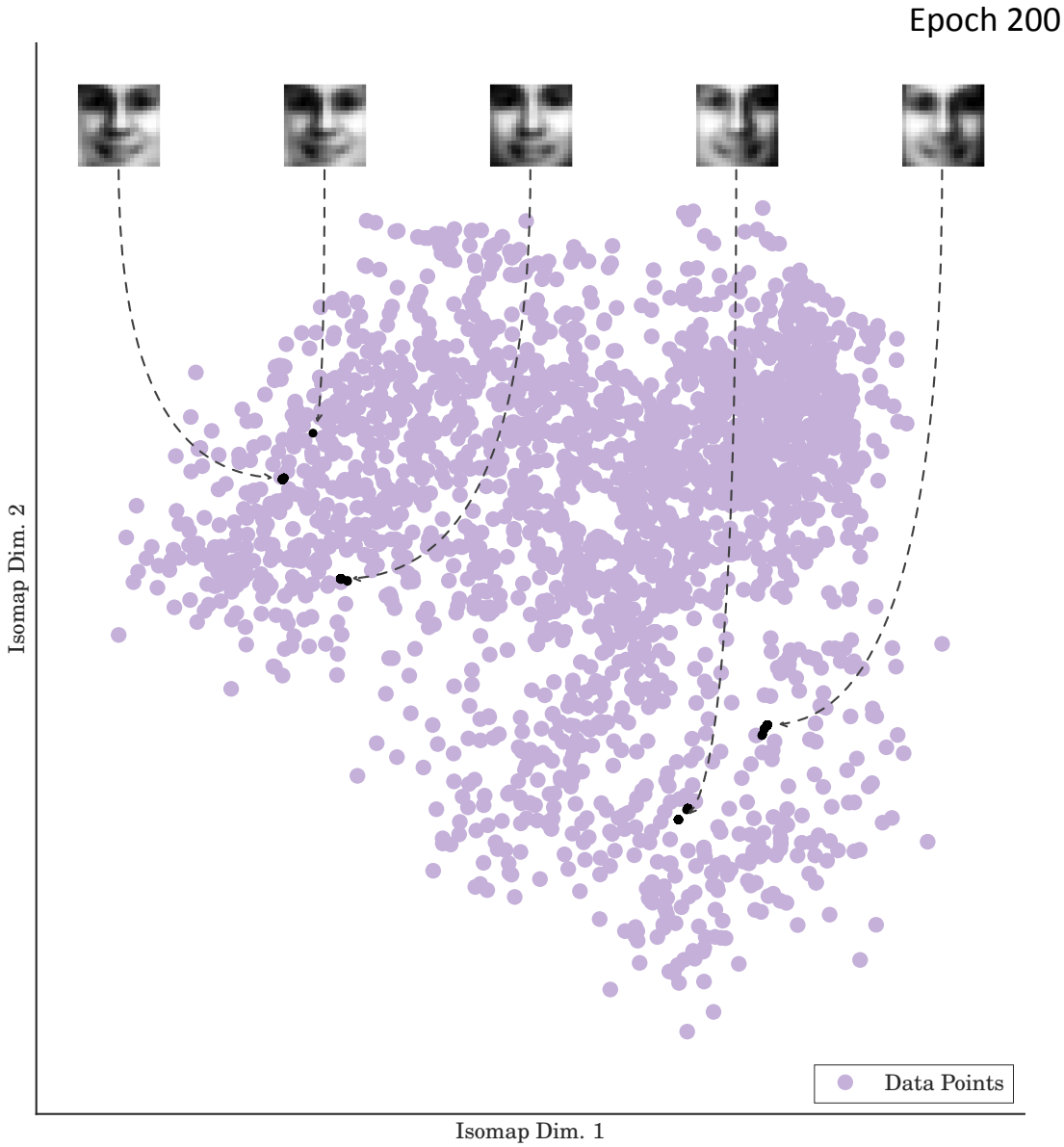
Evolution of TAP magnetizations in « phase space »



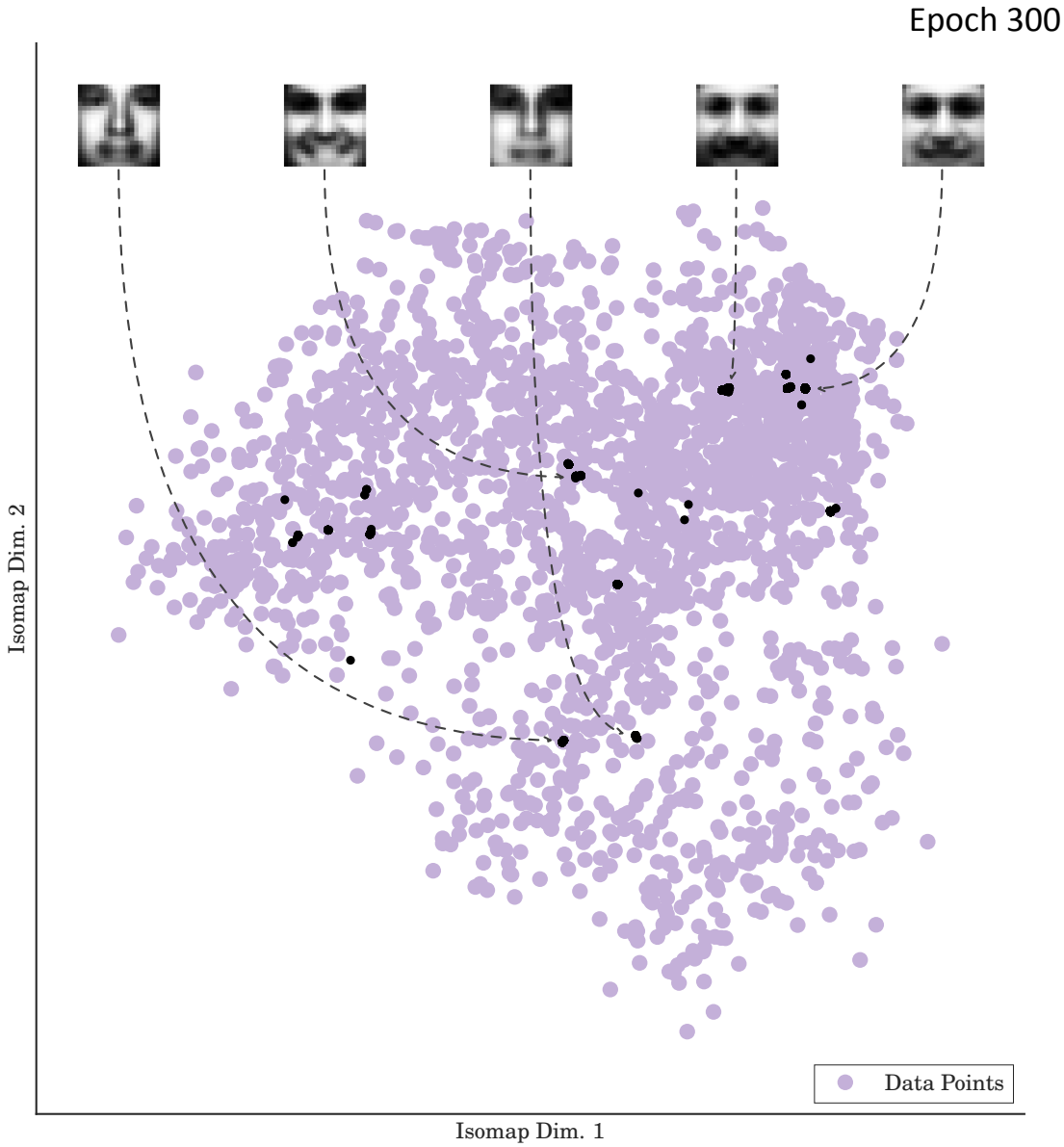
Evolution of TAP magnetizations in « phase space »



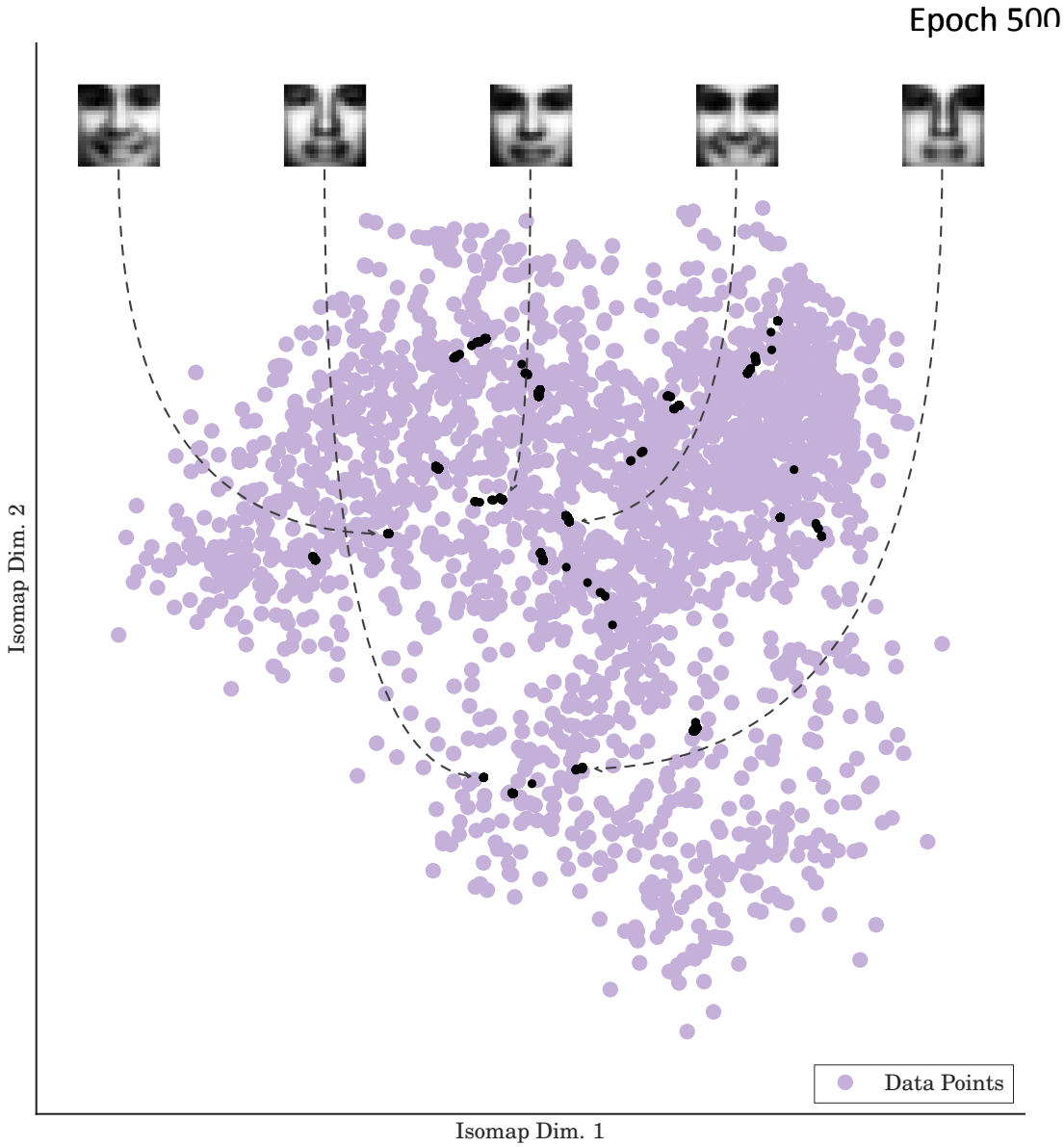
Evolution of TAP magnetizations in « phase space »



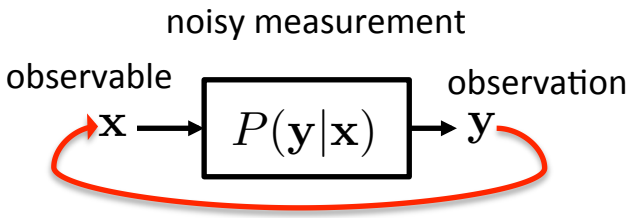
Evolution of TAP magnetizations in « phase space »



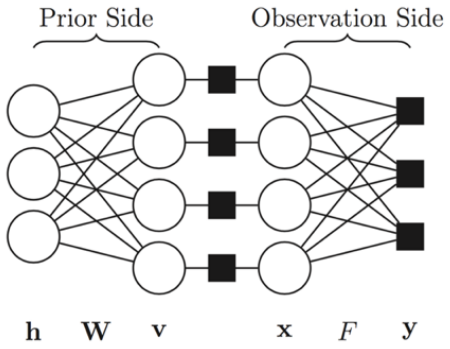
Evolution of TAP magnetizations in « phase space »



Prior learning and inference with real values



$$P(\mathbf{x}|\mathbf{y}) \propto P(\mathbf{y}|\mathbf{x})P_0(\mathbf{x})$$



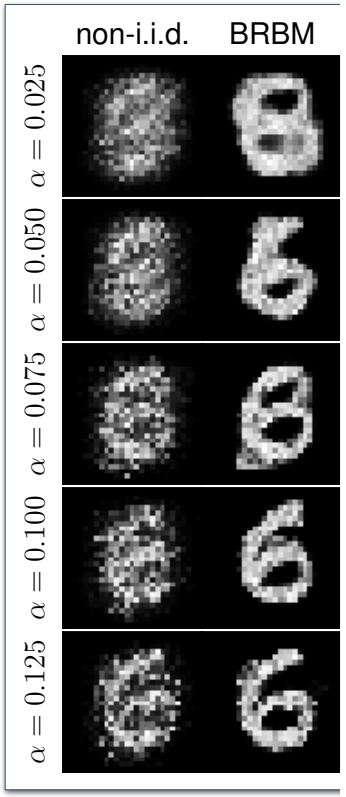
Binary model for the support only (BRBM)

$$\begin{array}{l}\mathbf{v} \in \{0,1\}^{d^{\mathbf{v}}} \\ \mathbf{h} \in \{0,1\}^{d^{\mathbf{h}}}\end{array} \quad p_\theta(\mathbf{v},\mathbf{h}) = \frac{1}{Z_\theta} e^{\mathbf{v}^T \mathbf{W} \mathbf{h}} \prod_i e^{\theta_i^{\mathbf{v}T} v_i} \prod_\alpha e^{\theta^{\mathbf{h}T} h_\alpha}$$

Generalized model on real valued digits (GRBM)

$$\begin{array}{l}\mathbf{h} \in \{0,1\}^{d^{\mathbf{h}}} \\ \mathbf{v} \in \mathbb{R}^{d^{\mathbf{v}}}\end{array} \quad p_\theta(\mathbf{v},\mathbf{h}) = \frac{1}{Z_\theta} e^{\mathbf{v}^T \mathbf{W} \mathbf{h}} \prod_i P_{\theta_i^{\mathbf{v}}}(v_i) \prod_\alpha e^{\theta_\alpha^{\mathbf{h}T} h_\alpha}$$

$$P_{\theta_i^{\mathbf{v}}}(v_i) = (1-\rho_i)\delta(v_i) + \rho_i \mathcal{N}(v_i; \mu_i, \sigma_i^2)$$

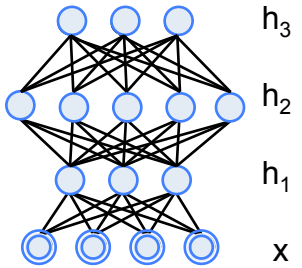


(E.Tramel, A. Manoel, F. Caltagirone, M.G., F. Krzakala ITW 2016)

Deep Boltzmann machine learning

Data dependent terms becomes intractable because of direct interactions between hidden units

- log-likelihood



log partition hard data dependent easy

$$\ell(\theta|\mathbf{x}) = \ln P(\mathbf{x};\theta) = -\ln \mathcal{Z}[\theta] + \sum_i \ln P_i(x_i;\theta_i)$$

$$+ \ln \int \left(\prod_l \prod_j \mathrm{d}h_j^{(l)} \right) e^{\sum_{i,j} x_i W_{ij} h_j^{(1)} + \sum_l \sum_{i,j} h_i^{(l)} W_{ij}^{(l)} h_j^{(l+1)}} \prod_l \prod_j P_j^{\mathrm{h}(l)}(h_j^{(l)};\theta_j{}^{(l)})$$

data dependent hard !

- gradients w.r.t. trained parameters

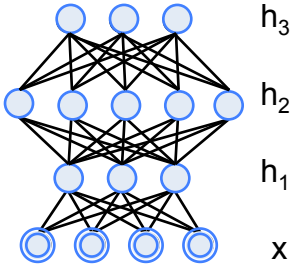
$$\Delta W^{(l)}{}_{i\alpha} \propto \frac{\partial \ell(\theta|\mathbf{x}^*)}{\partial W_{i\alpha}} = \langle h_i^{(l-1)} h_\alpha^{(l)} \rangle_{\mathbf{x}^*} - \langle h_i^{(l-1)} h_\alpha^{(l)} \rangle$$

hard hard

$p_\theta(\mathbf{h}|\mathbf{v}^*)$ $p_\theta(\mathbf{v},\mathbf{h})$

Deep Boltzmann machine learning

$$\Delta W^{(l)}{}_{i\alpha} \propto \frac{\partial \ell(\theta|\mathbf{x}^*)}{\partial W_{i\alpha}} = \langle h_i^{(l-1)} h_\alpha^{(l)} \rangle_{\mathbf{x}^*} - \langle h_i^{(l-1)} h_\alpha^{(l)} \rangle$$

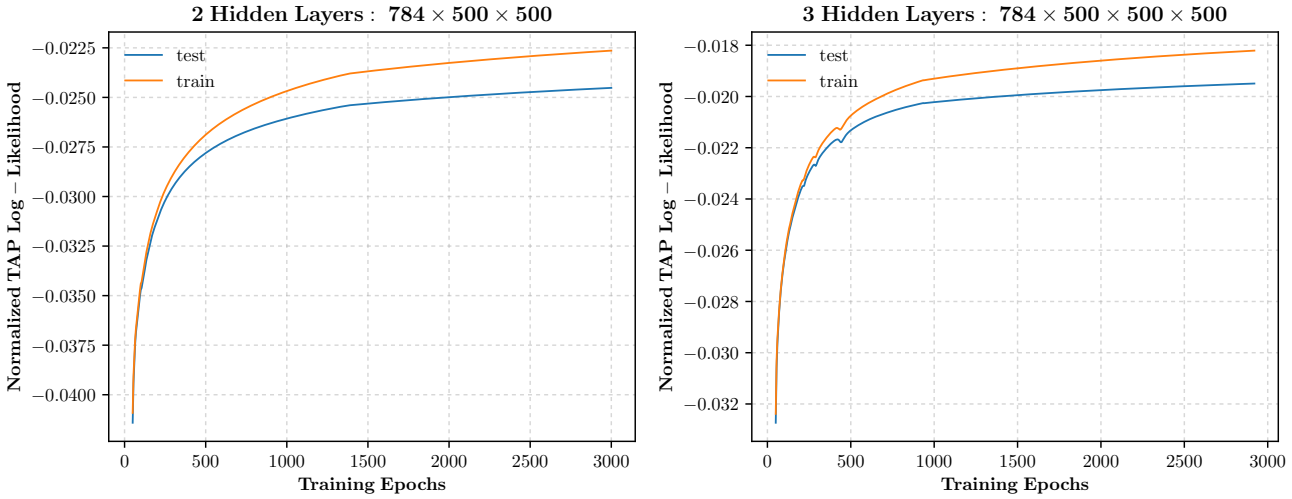


Usual training algorithm:

- NMF approximation of data dependent terms
- MCMC approximation of second term

The mean field framework suggests instead the consistent scheme

- TAP approximation of data dependent term
- TAP approximation of second term



Conclusion & Perspectives

- **TAP equations allow for new unsupervised learning with latent variables**
  - New way of monitoring an unsupervised training
  - New way of using ML for Bayesian inference
- **Mean field framework deals with arbitrary variables and architecture**

PERSPECTIVES

- Bayesian inference problems with very meaning full priors ?
- Further analytical analysis of the TAP machine ? Spectral analysis ?

THANK YOU !