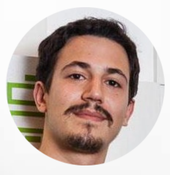
Thermodynamics properties of Restricted Boltzmann Machines

Aurélien Decelle – Giancarlo Fissore – Cyril Furtlehner

Tau team

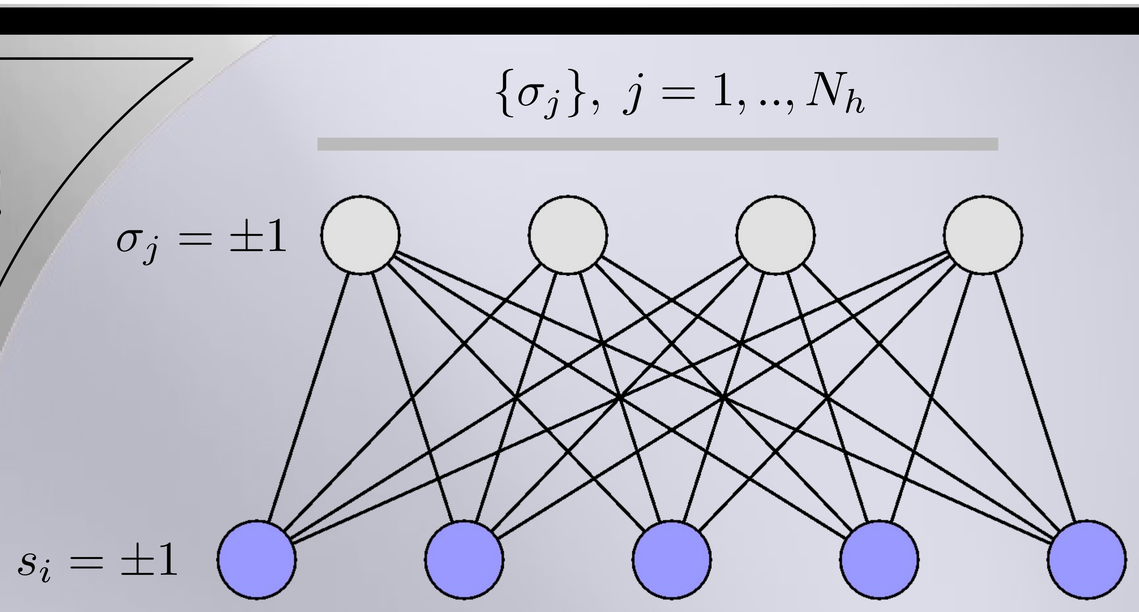




Generative models are gaining in popularity in Machine Learning, by their ability to sample images ! Yet they are still hard to train and understand. We propose an approach based on statistical physics to understand a simpler model but which is still mysterious :

the **Restricted Boltzmann Machine**

Many questions :
- *the landscape of the learned distribution*
- ***dynamics of the learning process***
- ***how the learned distribution is shaped by the data***
- ...



$$p(\underline{s}, \underline{\sigma}) = \frac{\exp\left(\sum_{ij} s_i w_{ij} \sigma_j + \sum_i a_i s_i + \sum_j b_j \sigma_j\right)}{Z}$$

Max Likelihood: $\dfrac{\partial \mathcal{L}}{\partial w_{ij}} = \langle s_i h_j \rangle_{\text{data}} - \langle s_i h_j \rangle_{\text{model}}$

Observations:

Mean-field equations

$$m_i^{(v)} = \tanh\left(\sum_j w_{ij} m_h^{(h)}\right)$$

$$m_j^{(h)} = \tanh\left(\sum_i w_{ij} m_h^{(v)}\right)$$

$\Longrightarrow$

SVD eqs:

$$\mathbf{m}^{(v)} = W \mathbf{m}^h$$

$$\mathbf{m}^{(h)} = W^T \mathbf{m}^v$$

Projection on the modes of the SVD

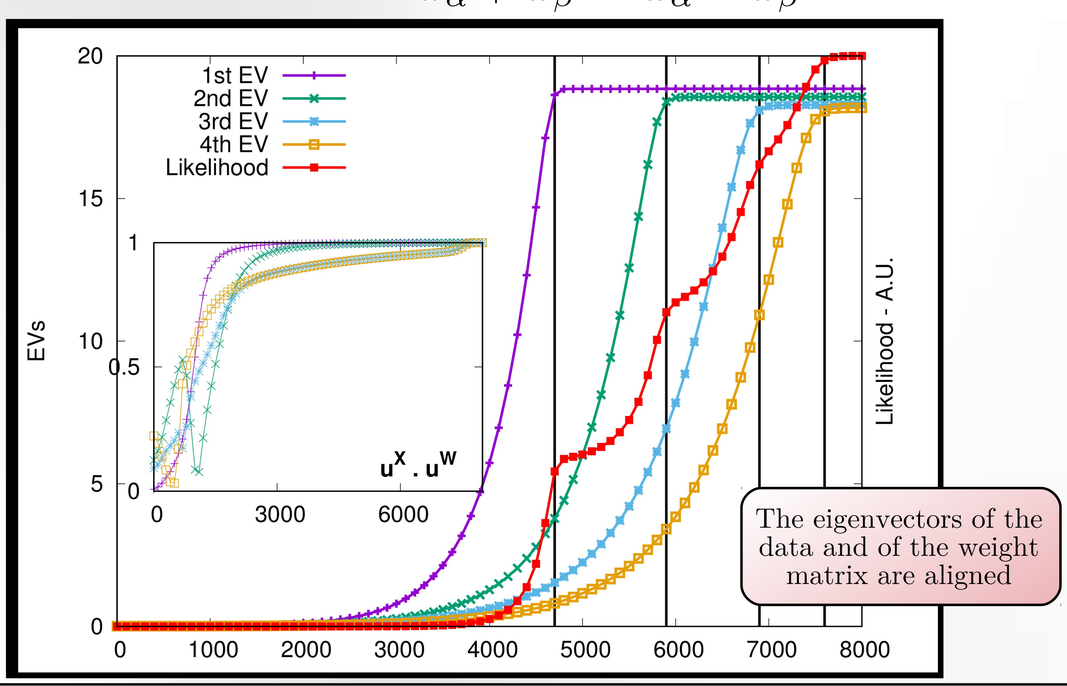
$$w_{ij} = \sum_\alpha u_i^\alpha w_\alpha v_j^\alpha$$

$$s_\alpha = \sum_i u_i^\alpha s_i \quad \text{and} \quad \sigma_\alpha = \sum_j v_j^\alpha \sigma_j$$

We can solve the dynamical equations of the gradient for the Gauss-Gauss RBM

$$\frac{dw_\alpha}{dt} = w_\alpha \sigma_h^2 \left(\langle s_\alpha^2 \rangle_{\text{Data}} - \frac{\sigma_v^2}{1 - \sigma_v^2 \sigma_h^2 w_\alpha^2}\right)$$

$$\Omega_{\alpha\beta}^{v,h} = (1 - \delta_{\alpha\beta}) \sigma_h^2 \left(\frac{w_\beta - w_\alpha}{w_\alpha + w_\beta} \mp \frac{w_\beta + w_\alpha}{w_\alpha - w_\beta}\right) \langle s_\alpha s_\beta \rangle_{\text{Data}}$$

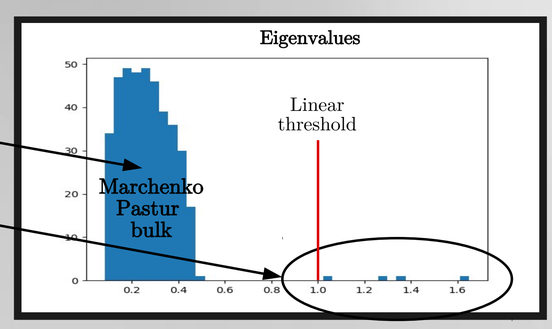


The eigenvectors of the data and of the weight matrix are aligned

Thermodynamics of the direct model:
we **assume K eigenmodes**

$$w_{ij} = \sum_{\alpha=1}^{K} u_i^\alpha w_\alpha v_j^\alpha + r_{ij}$$

u_i^α, v_j^α Quenched disorder

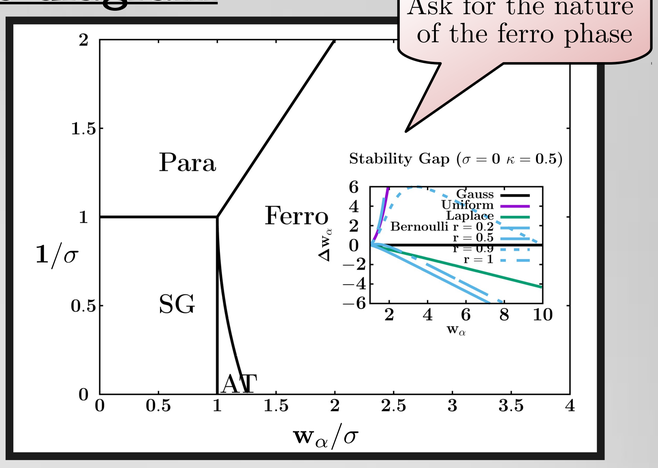


Quenched mean-field equations
&
Phase diagram

$$m_\alpha = (w_\alpha \bar{m}_\alpha - \theta_\alpha)(1 - q_\alpha)$$

$$\bar{m}_\alpha = (w_\alpha m_\alpha - \eta_\alpha)(1 - \bar{q}_\alpha)$$

$q_\alpha, \bar{q}_\alpha$ spin-glass order parameters



Numerical experiments:

Integration of the MF eqs on synthetic data

11 clusters in d=5 embedded in d=100

MNIST and SGD