Centro AlmaAI - Hard Sciences: Kick-off Workshop

Bologna, April 28, 2021

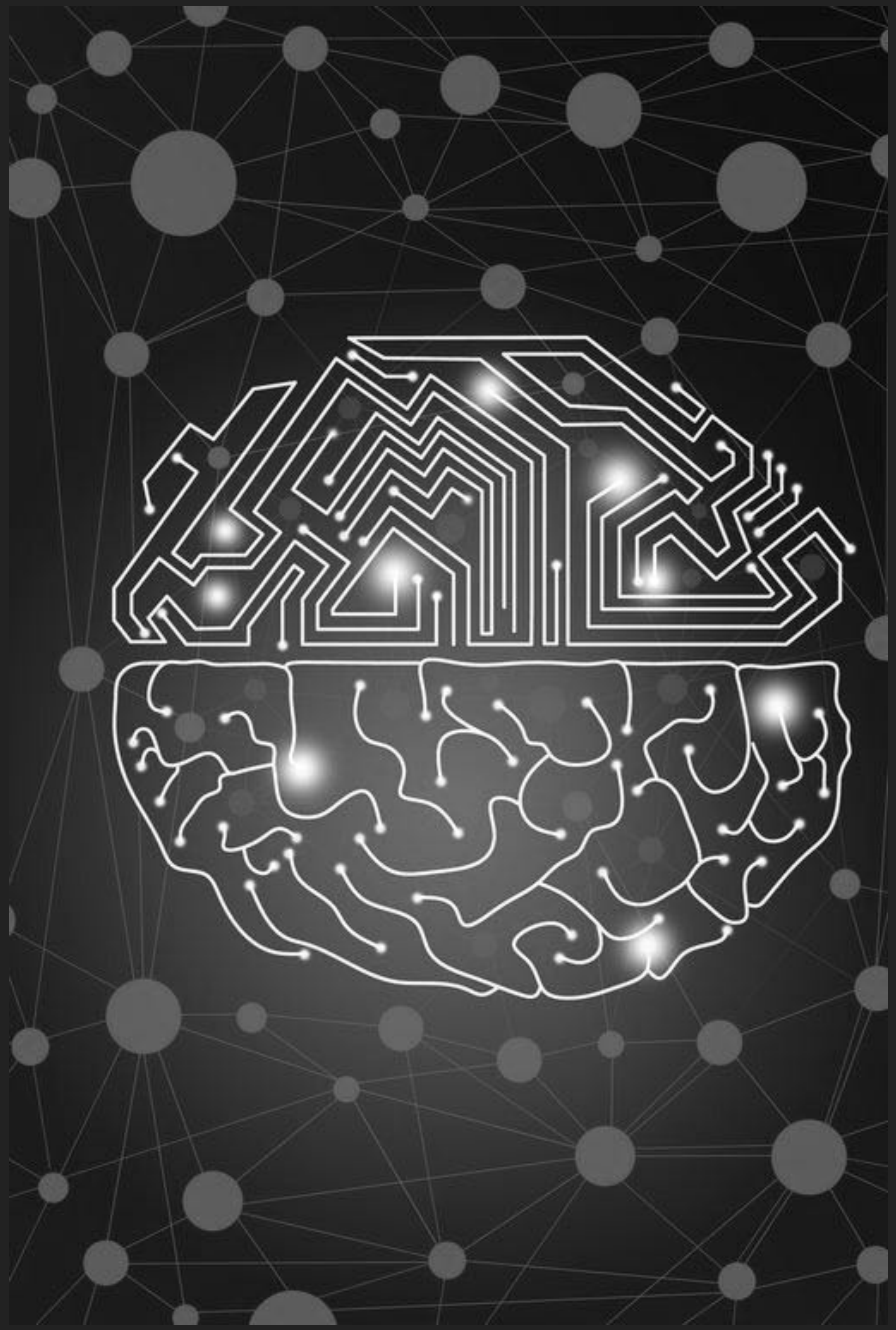
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

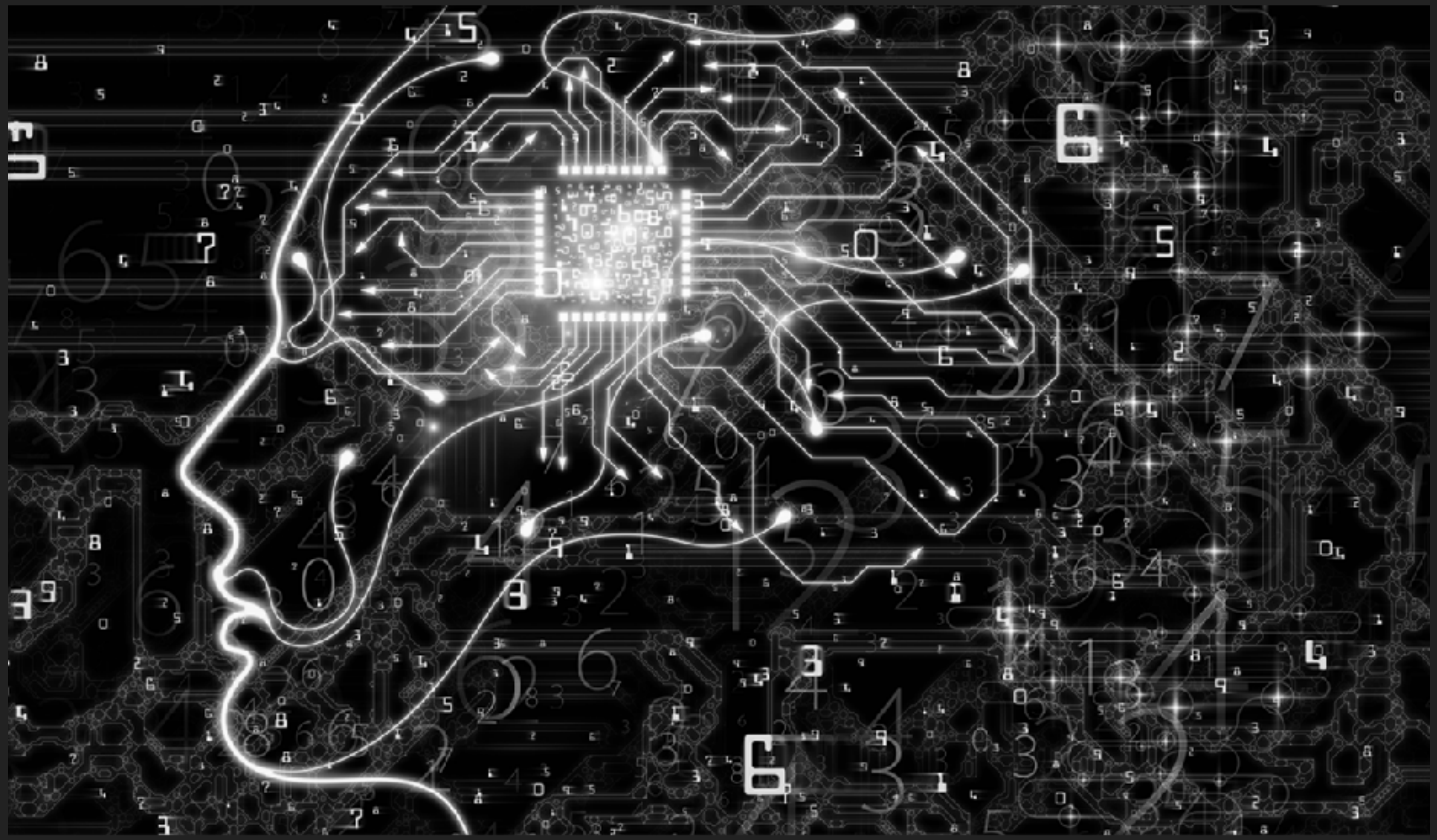
DANIELE TANTARI

STATISTICAL MECHANICS FOR MACHINE LEARNING: RECENT ADVANCES

OUTLINE OF THE TALK

- Introduction: Restricted Boltzmann Machines
- Statistical Mechanics and data representation;
- Statistical Mechanics of the Learning Process;
- Statistical Mechanics for new algorithms;

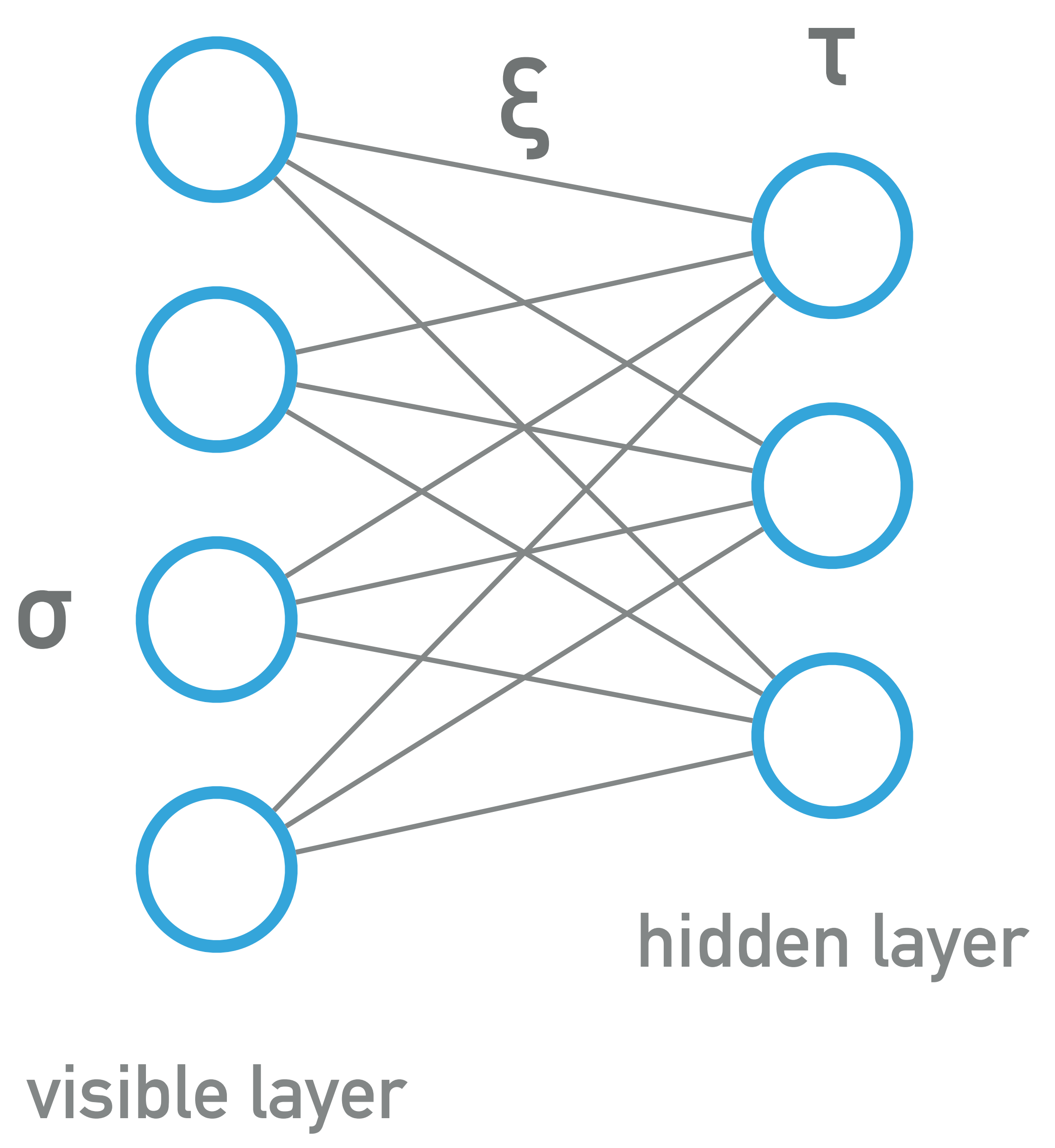




INTRODUCTION

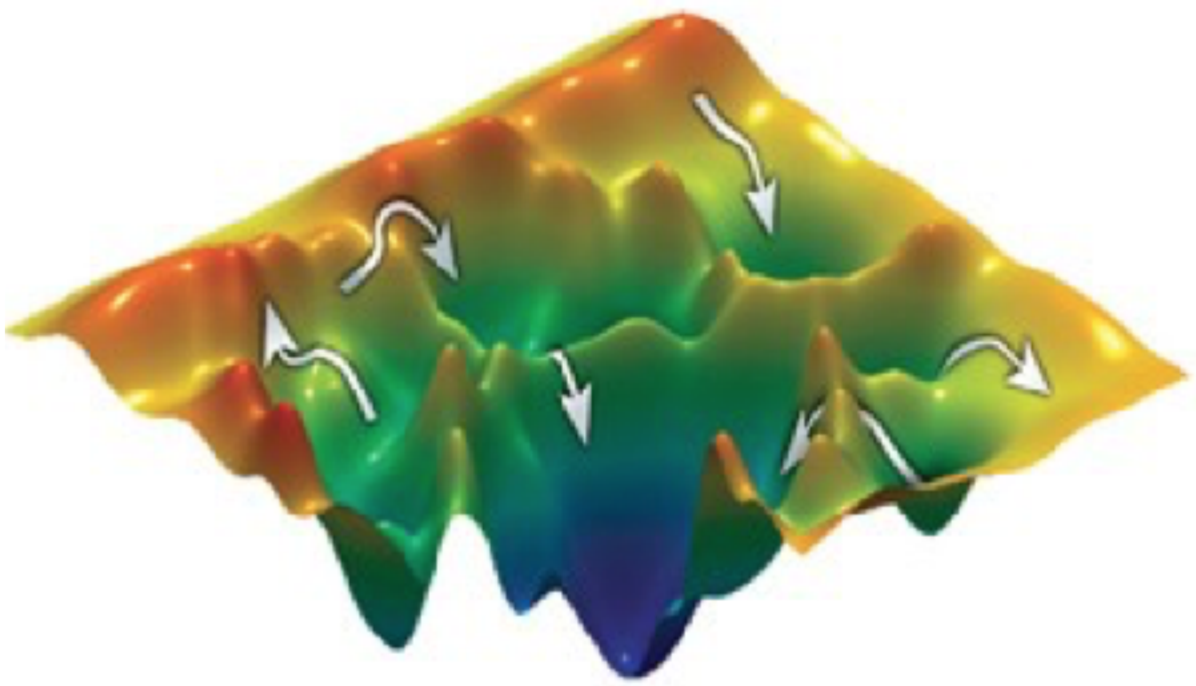
RESTRICTED BOLTZMANN MACHINES

RESTRICTED BOLTZMANN MACHINES



Training set (input)

$$\{\sigma_a\} \quad a = 1, \ldots, M$$



Probabilistic model

$$P_\xi(\sigma, \tau) = Z^{-1} P_\sigma(\sigma) P_\tau(\tau) e^{-E(\sigma,\tau;\xi)}$$

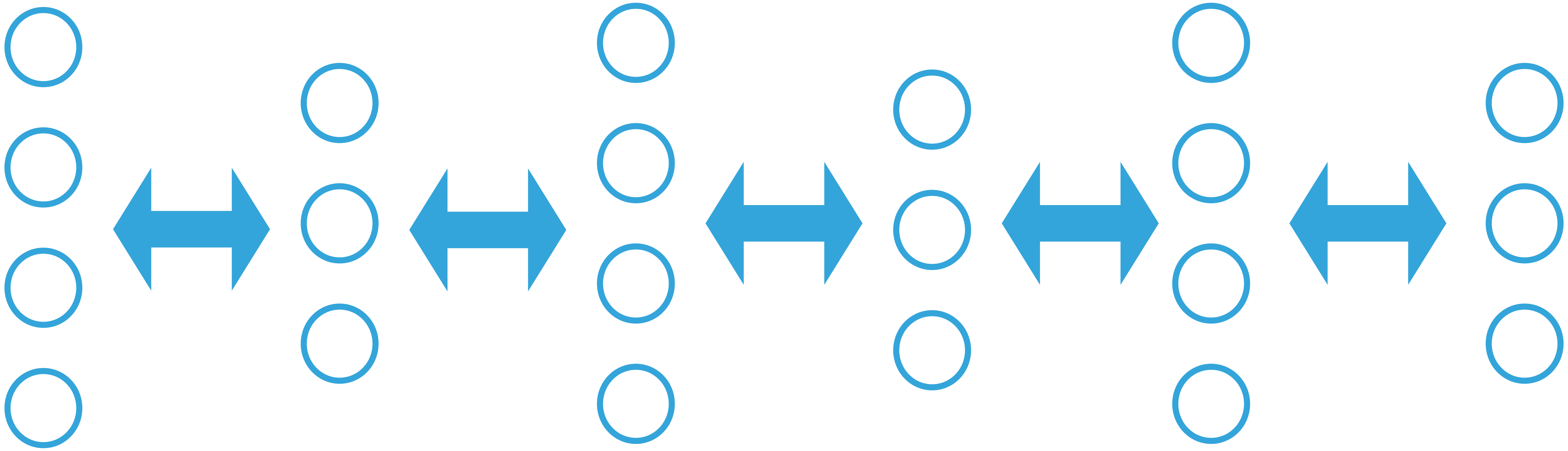
$$Z = \int d\sigma d\tau \, P_\sigma(\sigma) P_\tau(\tau) e^{-E(\sigma,\tau;\xi)}$$

$$\boxed{E(\sigma, \tau; \xi) = -\sum_{i,\mu} \xi_i^\mu \sigma_i \tau^\mu}$$

UNSUPERVISED LEARNING

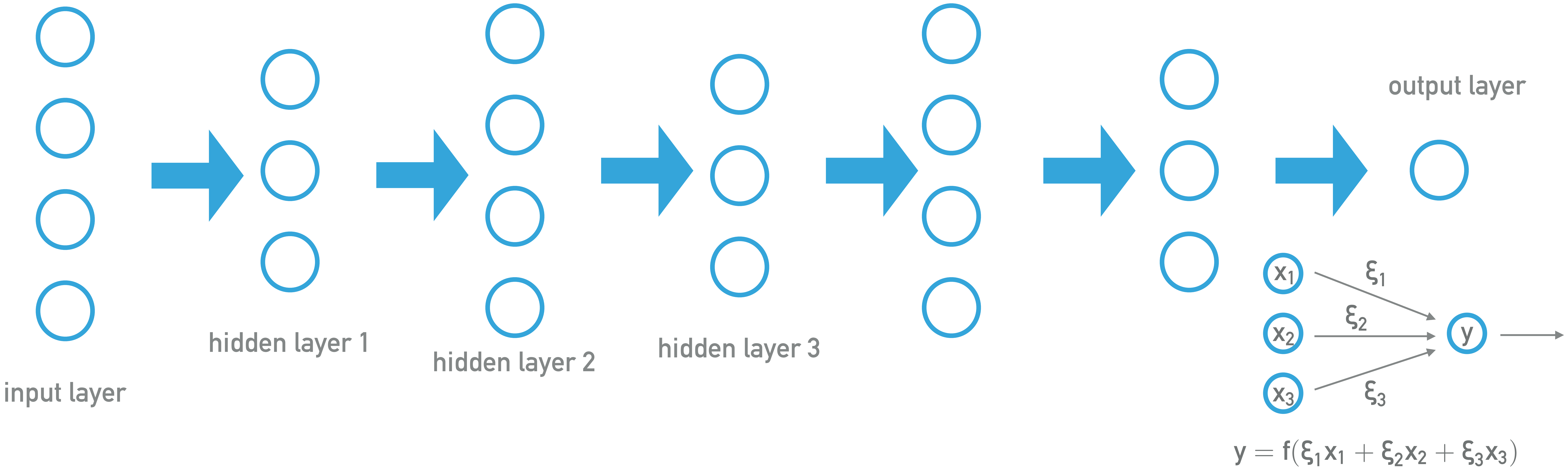
- Present a series of inputs
- Learn the ξ so that the inputs will be low energy configuration of the visible units

DEEP BOLTZMANN MACHINES



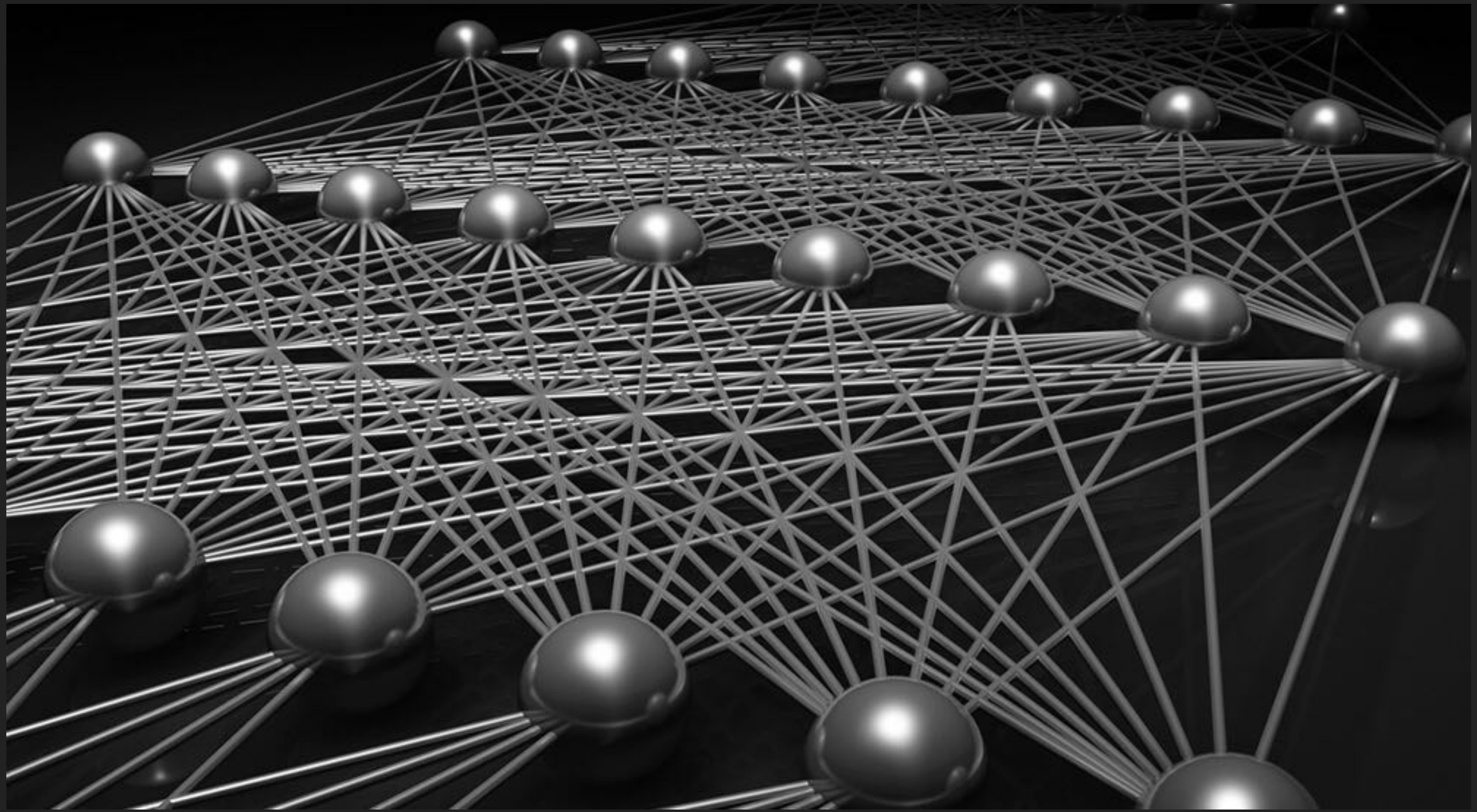
- Obtain an internal representation of data
- Storing patterns of information
- Disentangling and organizing different levels of correlations

FEED FORWARD NEURAL NETWORKS



...SUPERVISED LEARNING ...learn a mapping between input and output

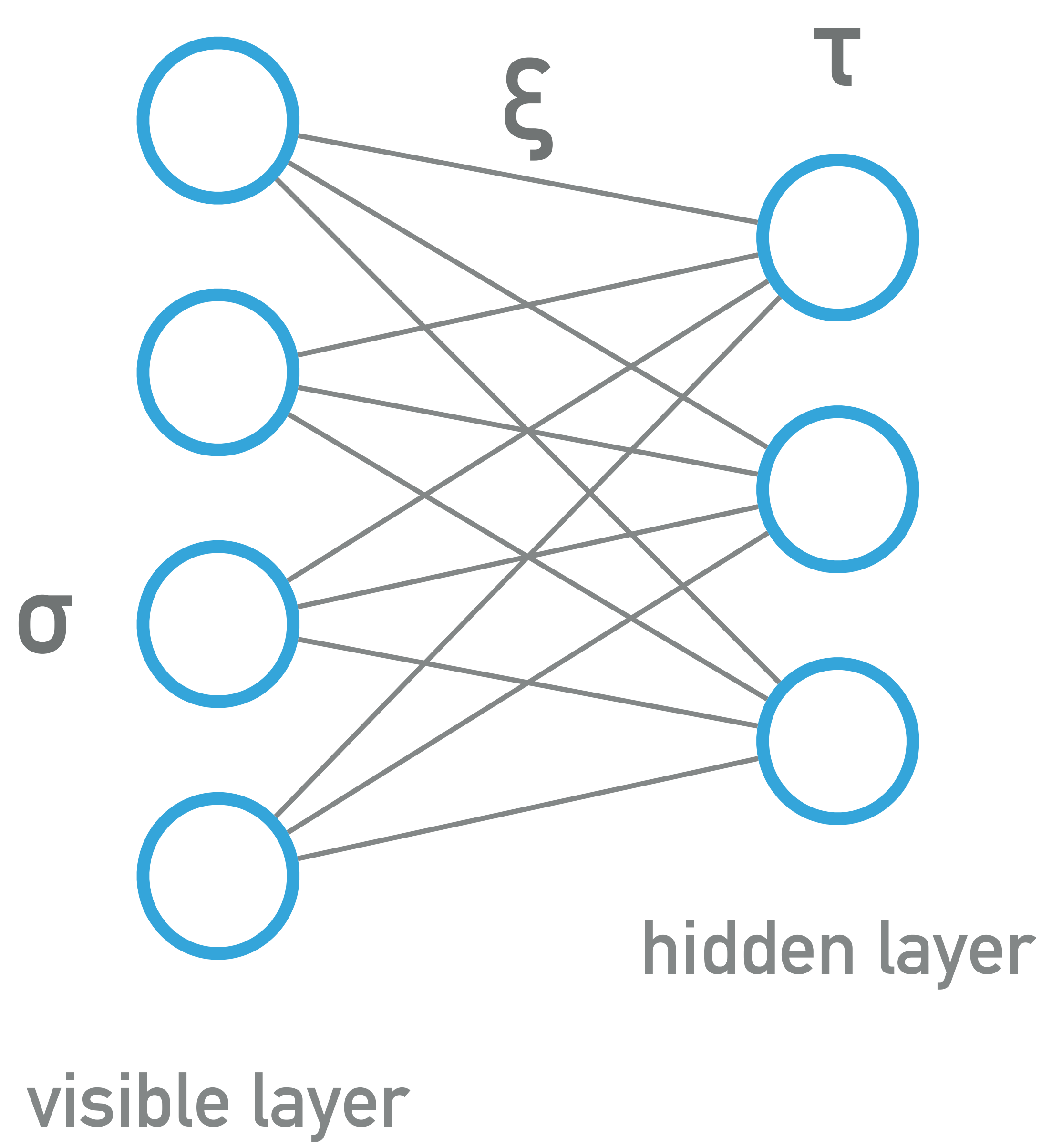
- Supervised fine tuning (Gradient Descent)



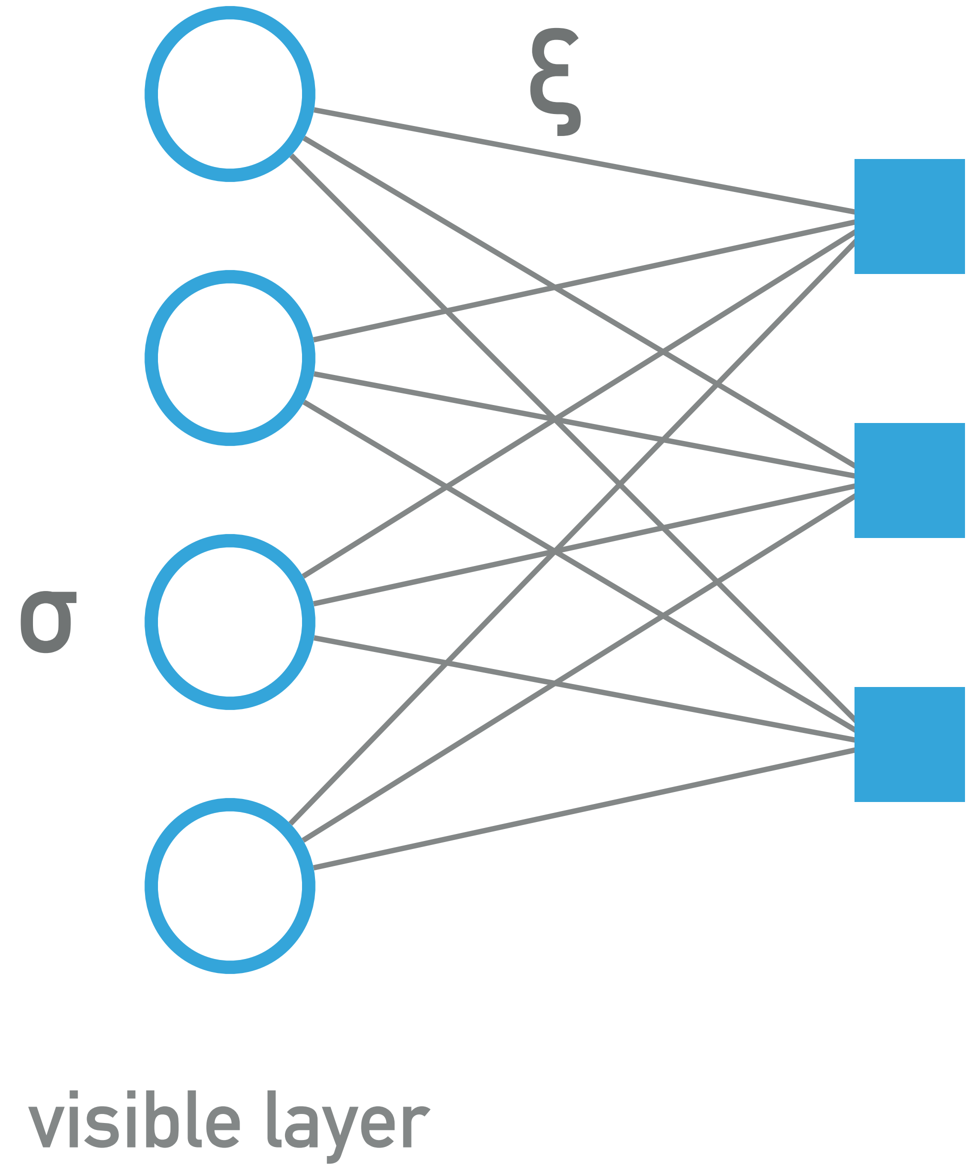
STATISTICAL MECHANICS

…AND DATA REPRESENTATION

RBM WITH RANDOM WEIGHTS



$$\xi_i^\mu \text{ i.i.d. } \in \{-\beta, \beta\}$$



$$\boxed{P(\sigma,\tau;\xi) = Z^{-1} P_\sigma(\sigma) P_\tau(\tau) e^{\sum_{i,\mu} \xi_i^\mu \sigma_i \tau^\mu}}$$

- Multi-species spin glass (SK) model

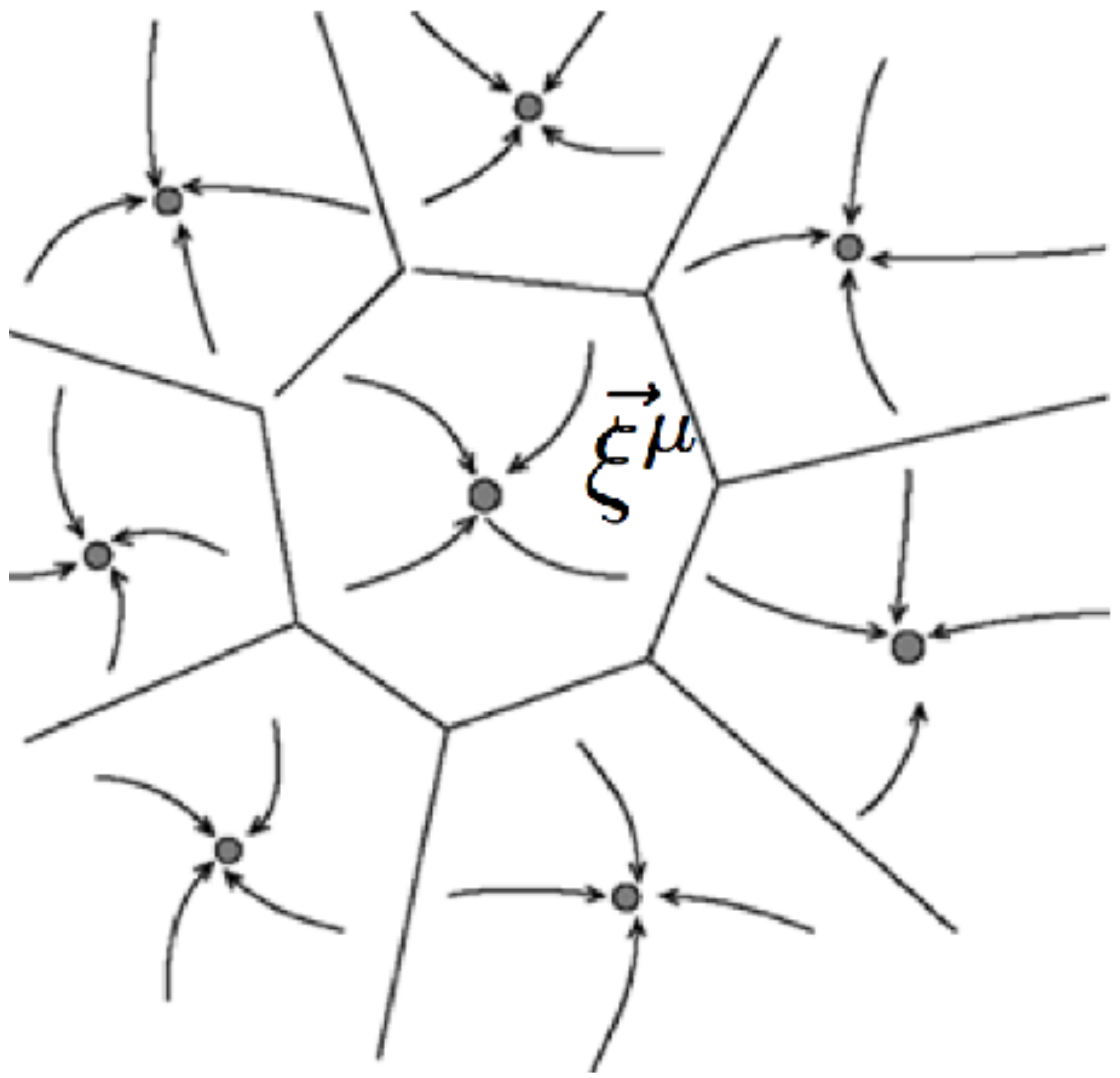
Barra et al (2015)

$$\boxed{P(\sigma;\xi) = Z^{-1} P_\sigma(\sigma) e^{\sum_\mu \psi(\xi^\mu \cdot \sigma)}}$$

- Generalized Hopfield Model

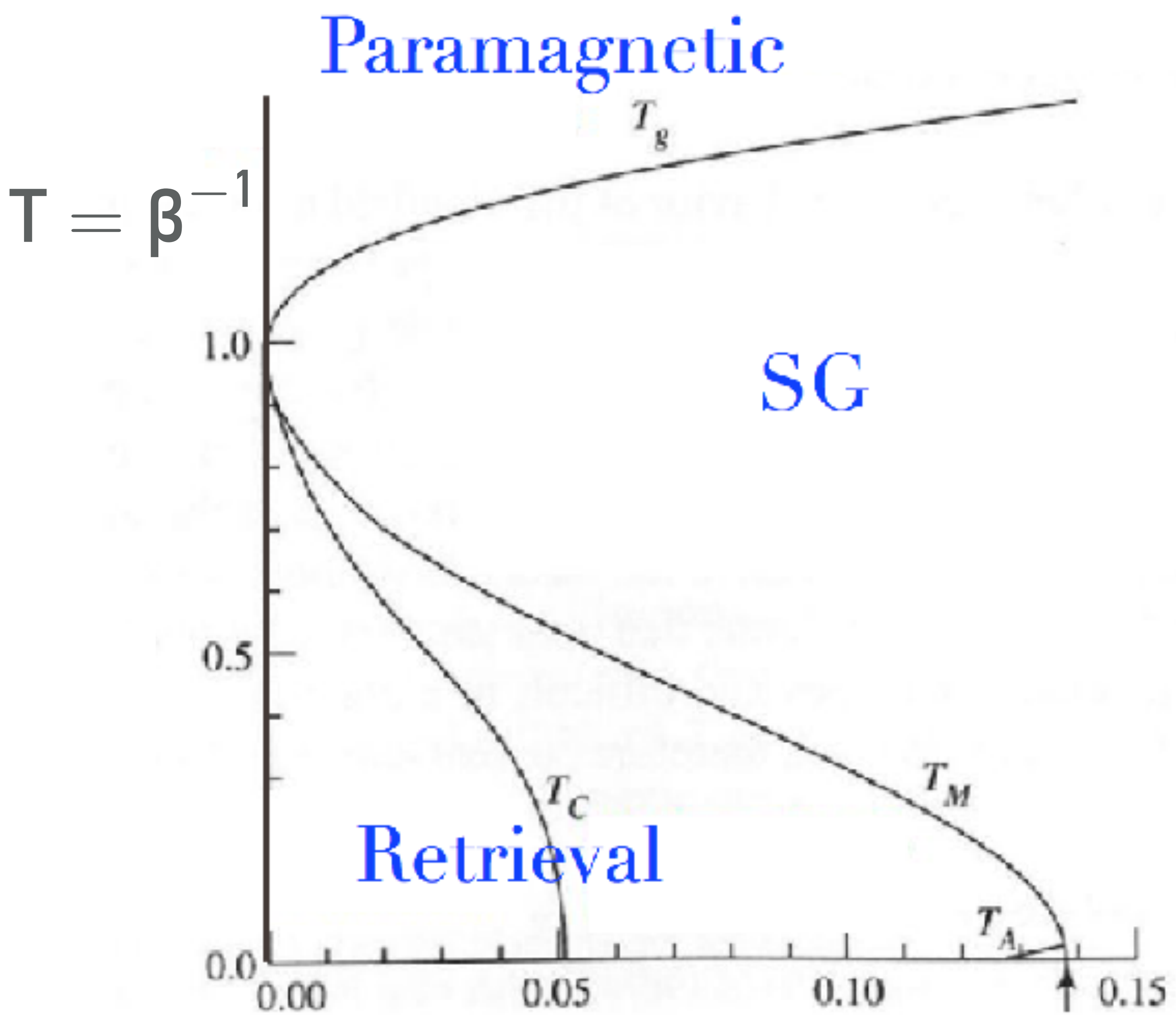
Amit et al (1985)

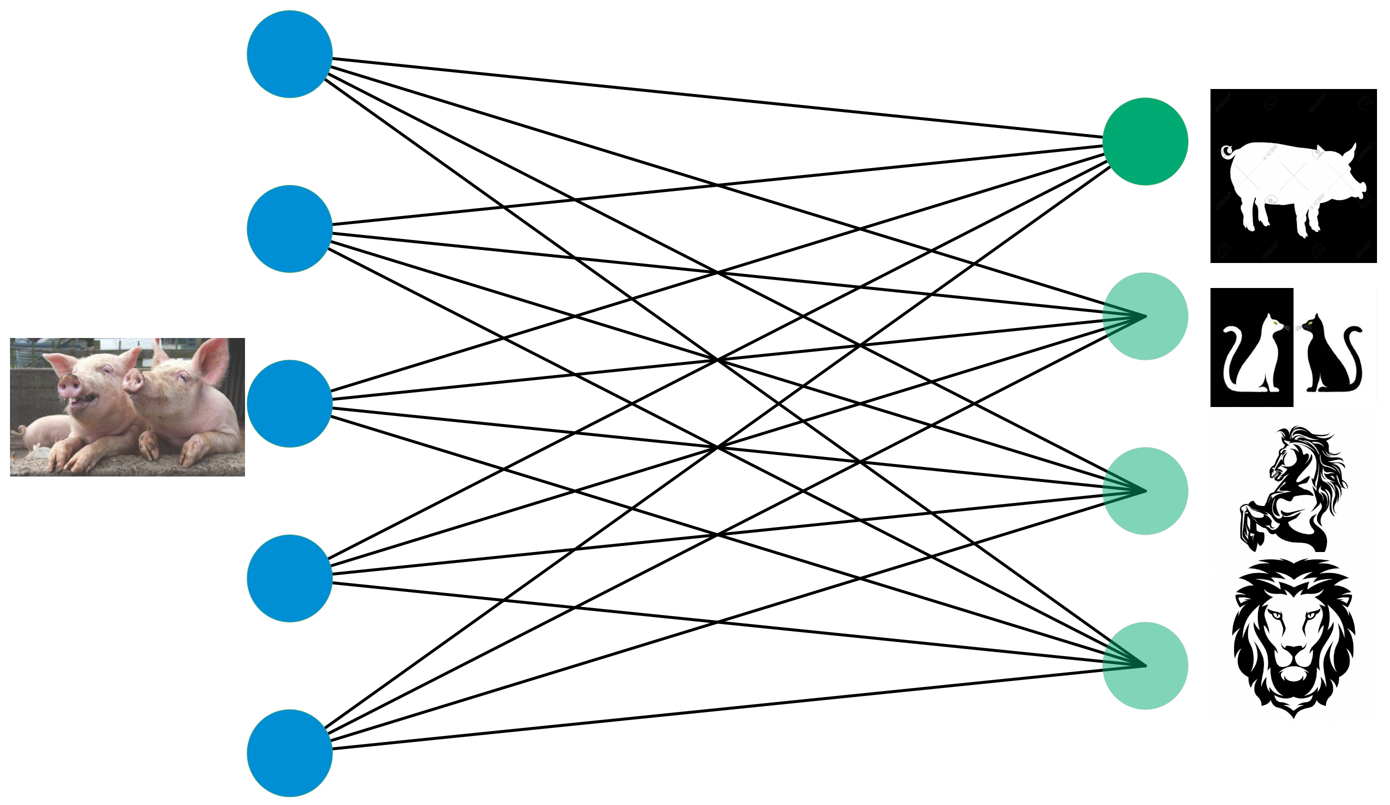
RBM WITH RANDOM WEIGHTS



- global feature (Prototype) $\vec{\xi}^{\mu}$

Phase diagram (Amit Gutfreund Sompolinsky 1985)





Order parameters

$$M^{\mu} = \frac{1}{N} \sum_i \xi_i^{\mu} \langle \sigma_i \rangle$$

$$q = \frac{1}{N} \sum_i \langle \sigma_i \rangle^2$$

Paramagnetic

$$q = M^{\mu} = 0$$

SG

$$M^{\mu} = 0 \qquad q \neq 0$$

Retrieval

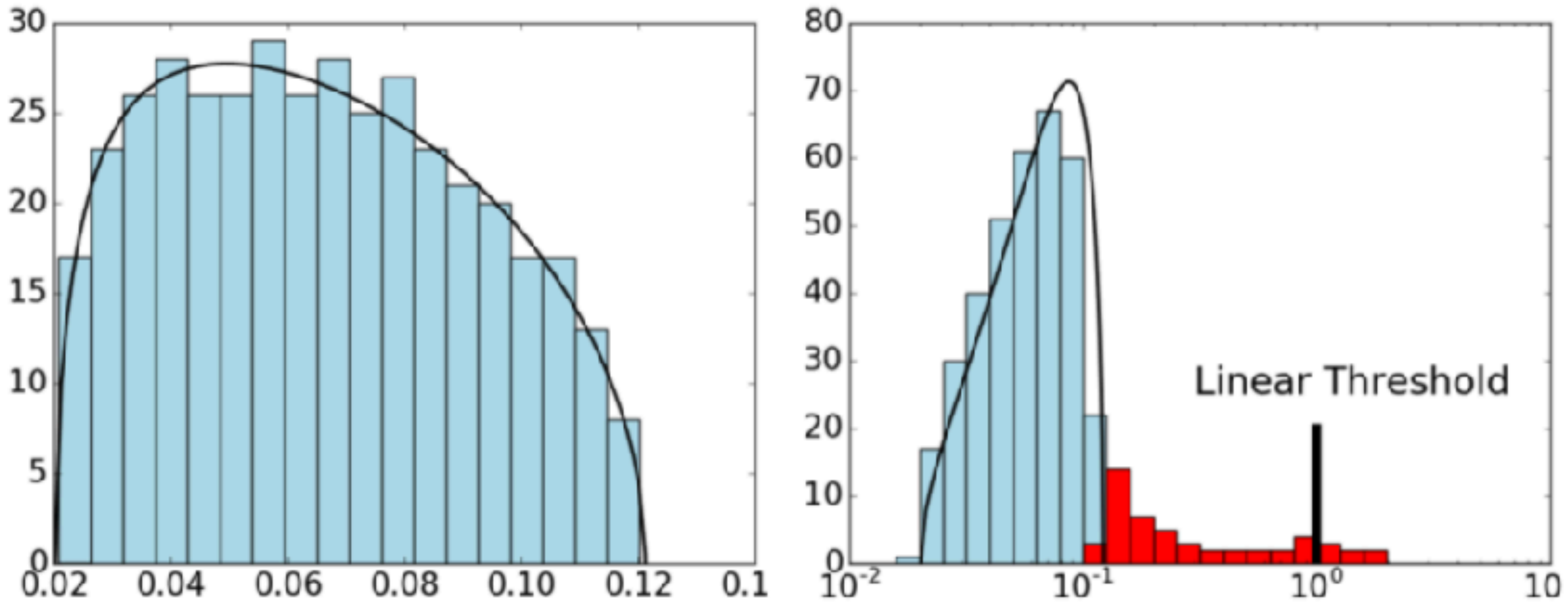
$$M^1 \neq 0 \qquad q \neq 0$$

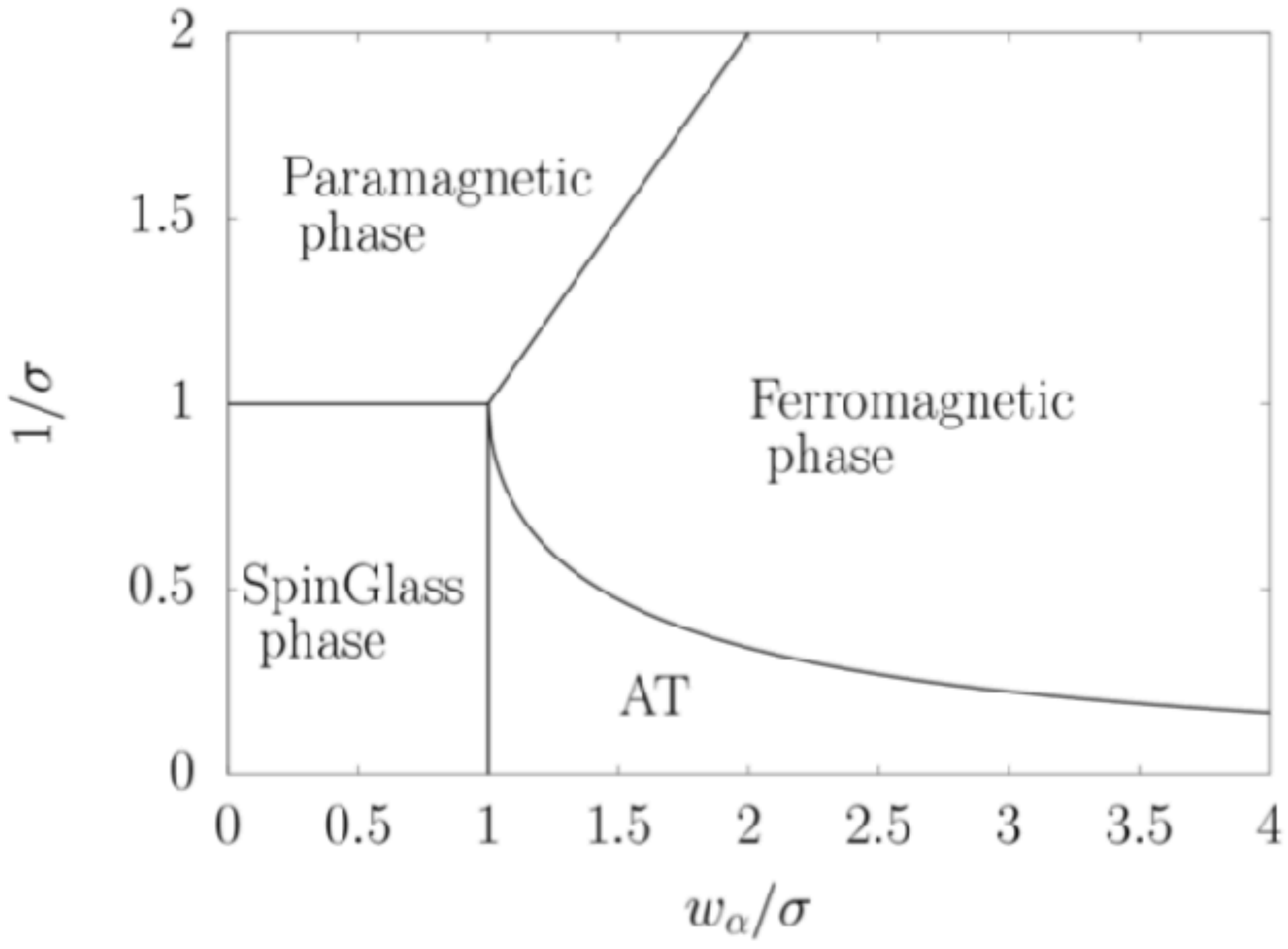
RBM WITH LOW RANK SIGNAL WEIGHTS

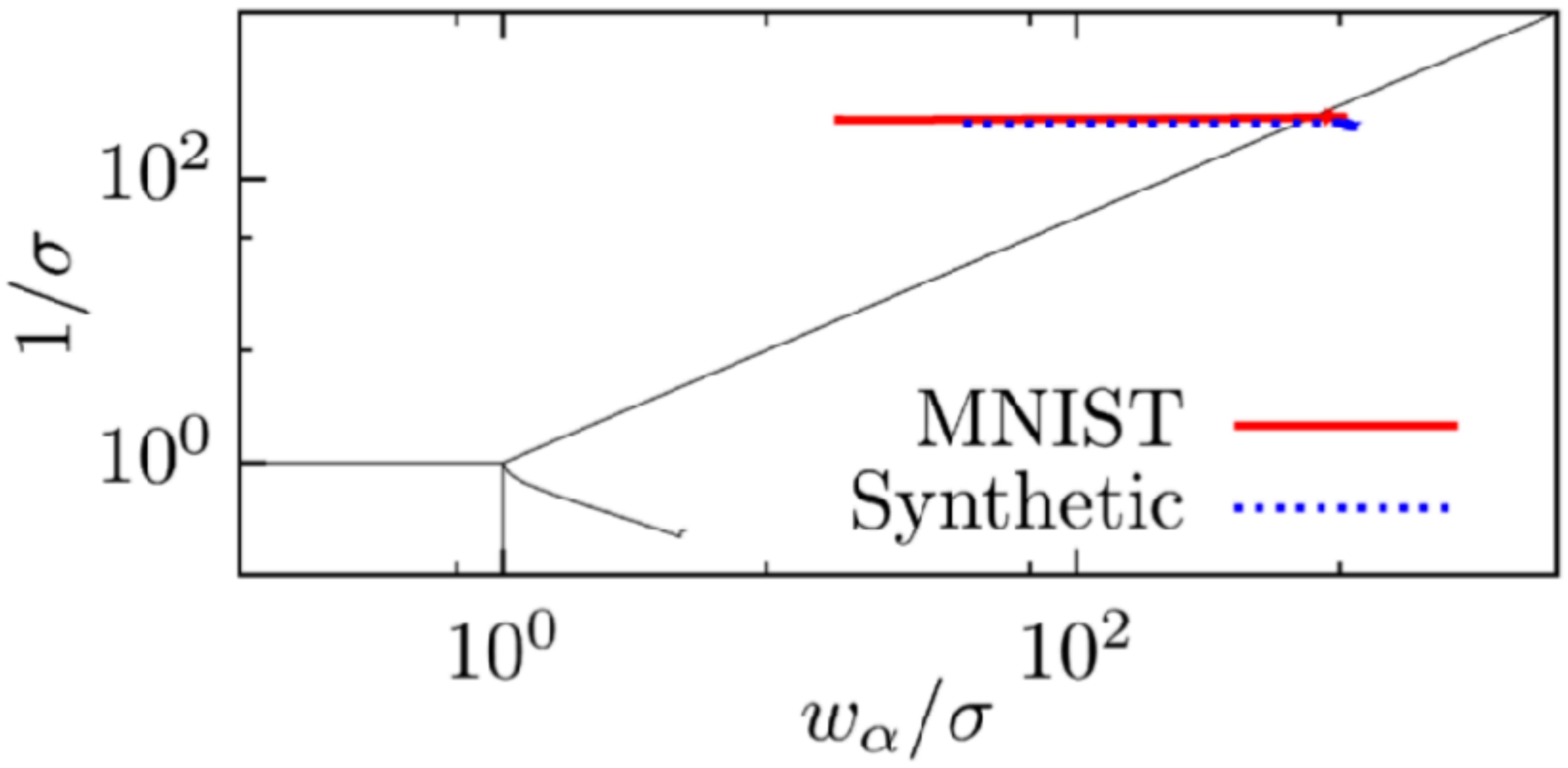
$$\vec{\xi^{\mu}} = \sum_{a=1}^{K} v_{\mu}^{a} w_a \vec{u^{\mu}} + \vec{r^{\mu}}$$

low rank signal

noise







RBM WITH DILUTED WEIGHTS

- localized feature $\vec{\xi^{\mu}}$ → ξ_i^{μ} i.i.d $\in \{-\beta, \beta, 0\}$ $P(\xi_i^{\mu} = 0) = d$

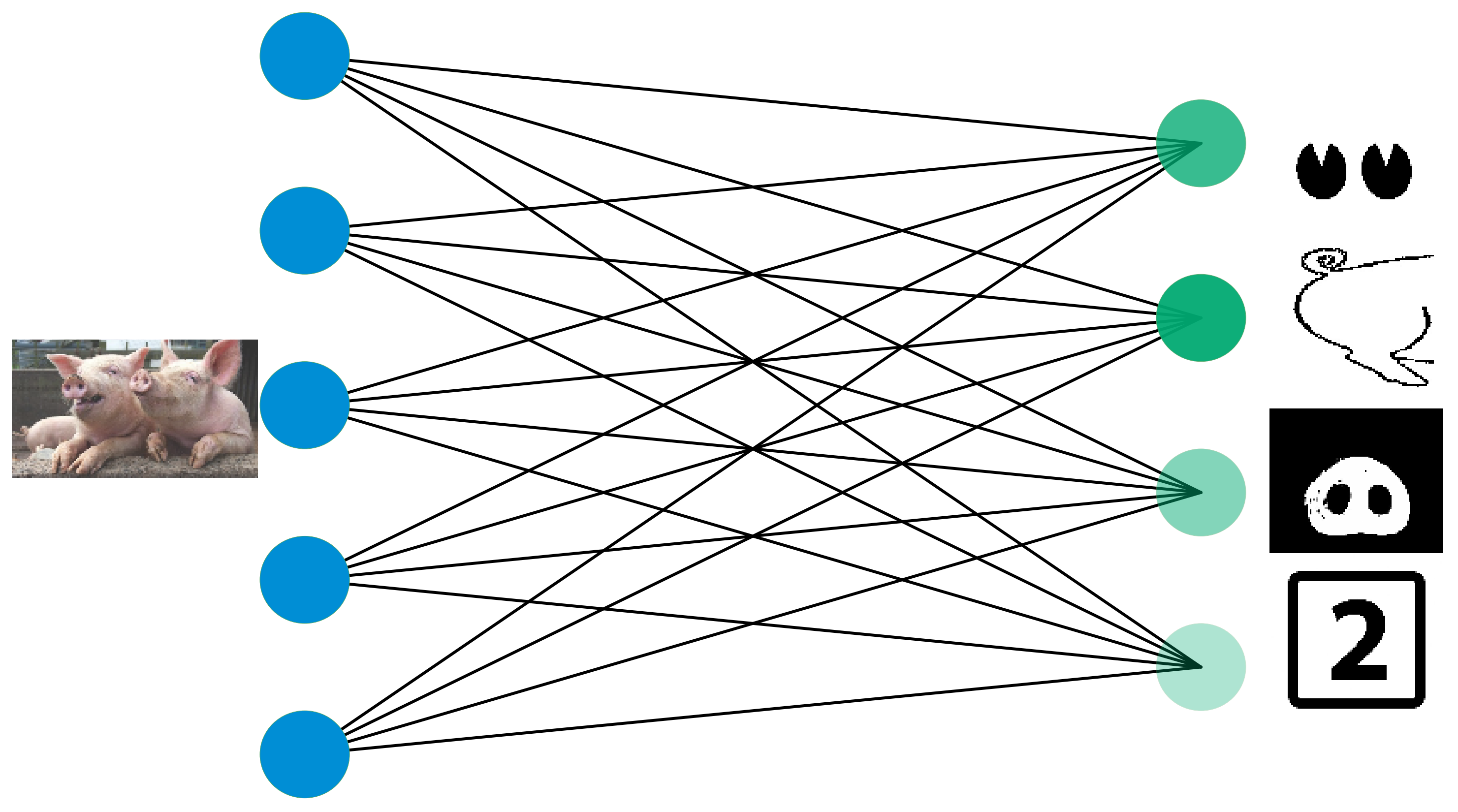
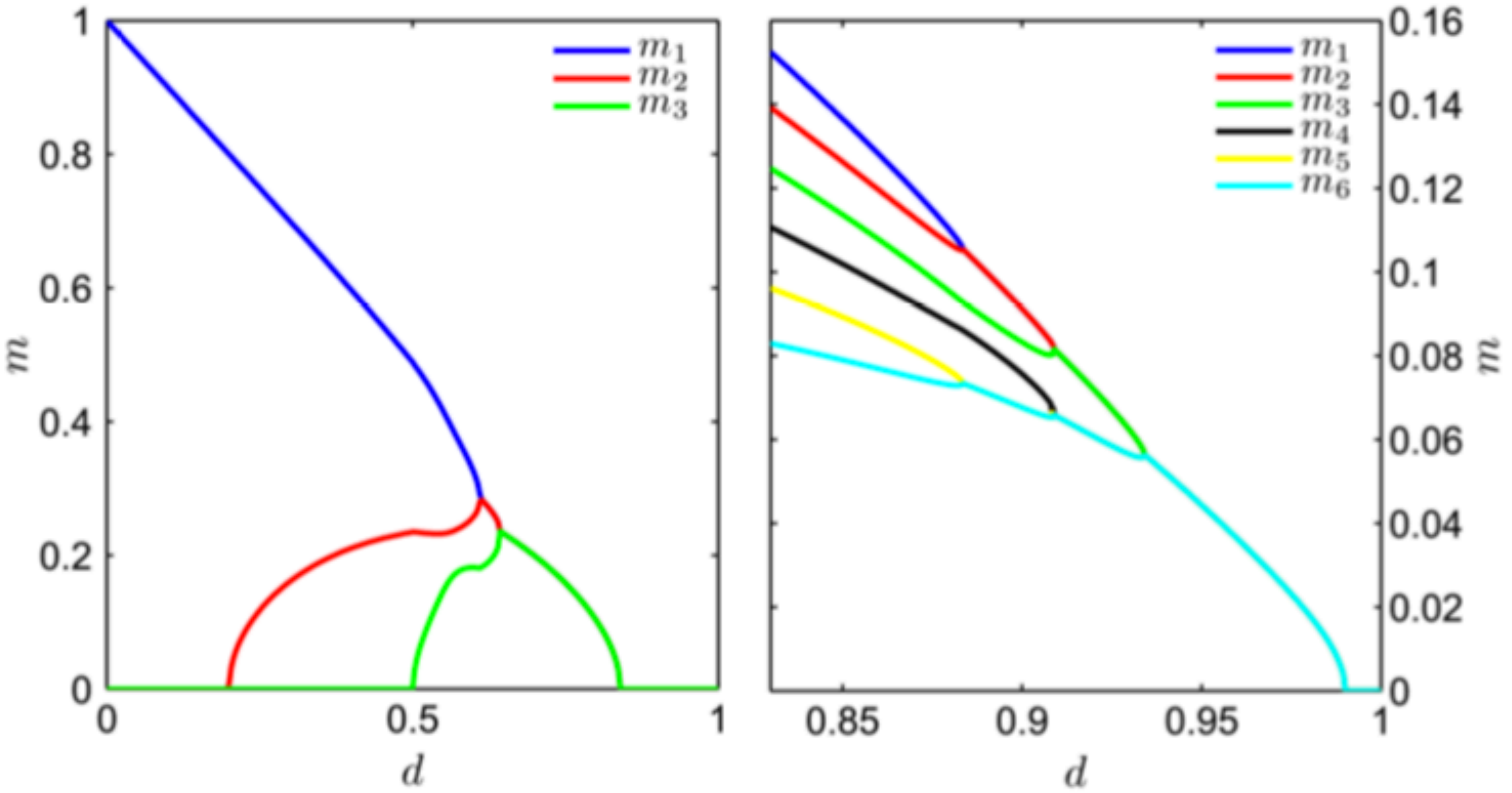


Figure from Agliari et al (2012)

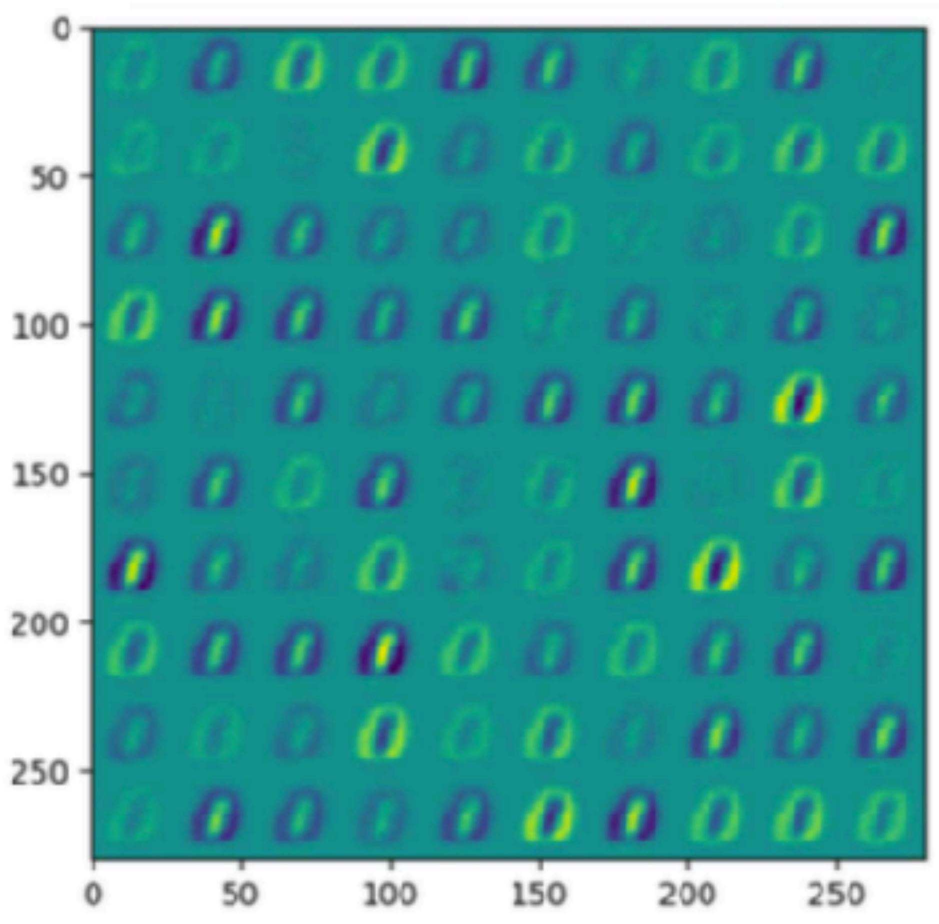


STATISTICAL MECHANICS

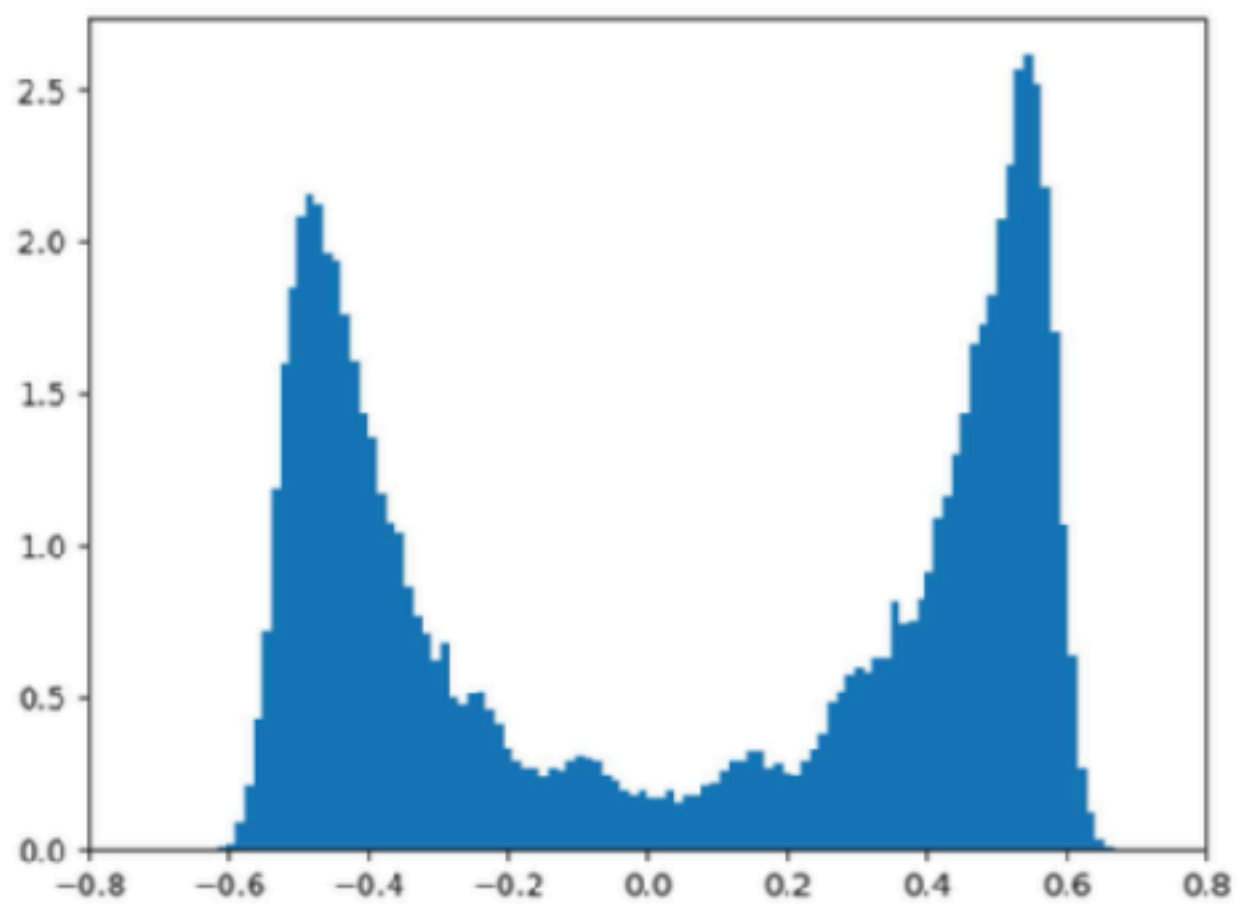
…AND THE LEARNING PROCESS

PHASES OF LEARNING (STAGE 1)

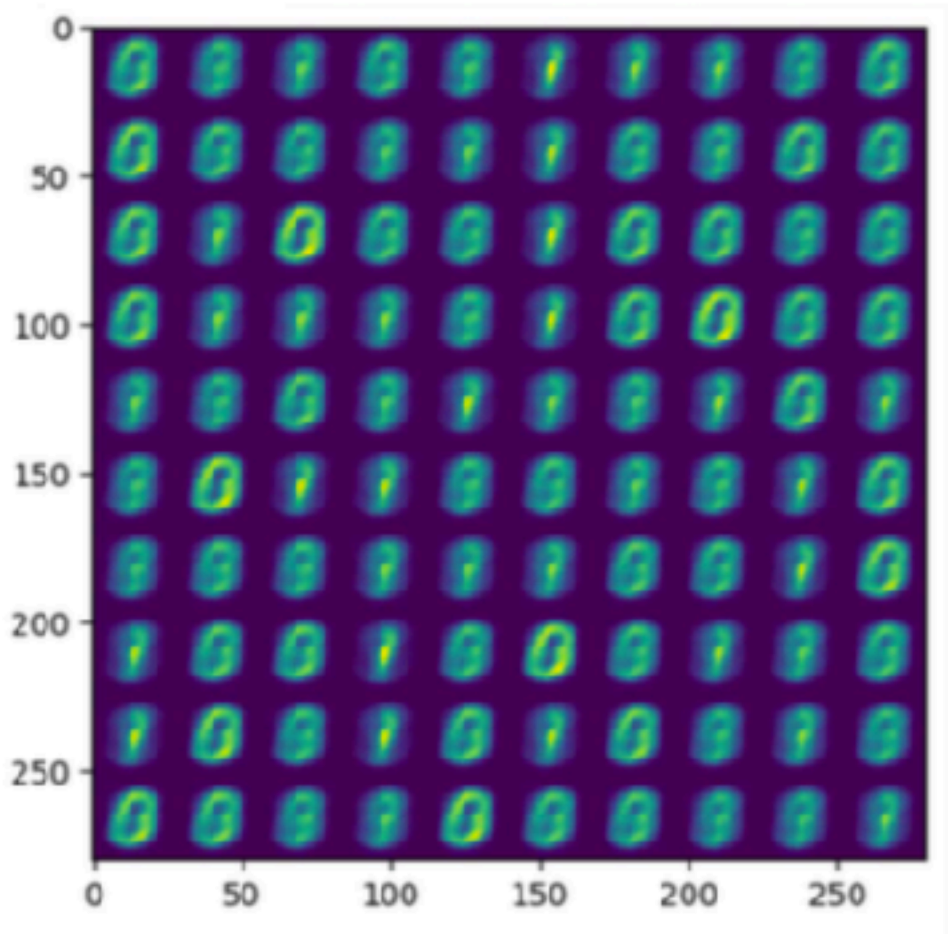
FEATURES



FEATURE SIMILARITY



SAMPLES



FERROMAGNETIC PHASE

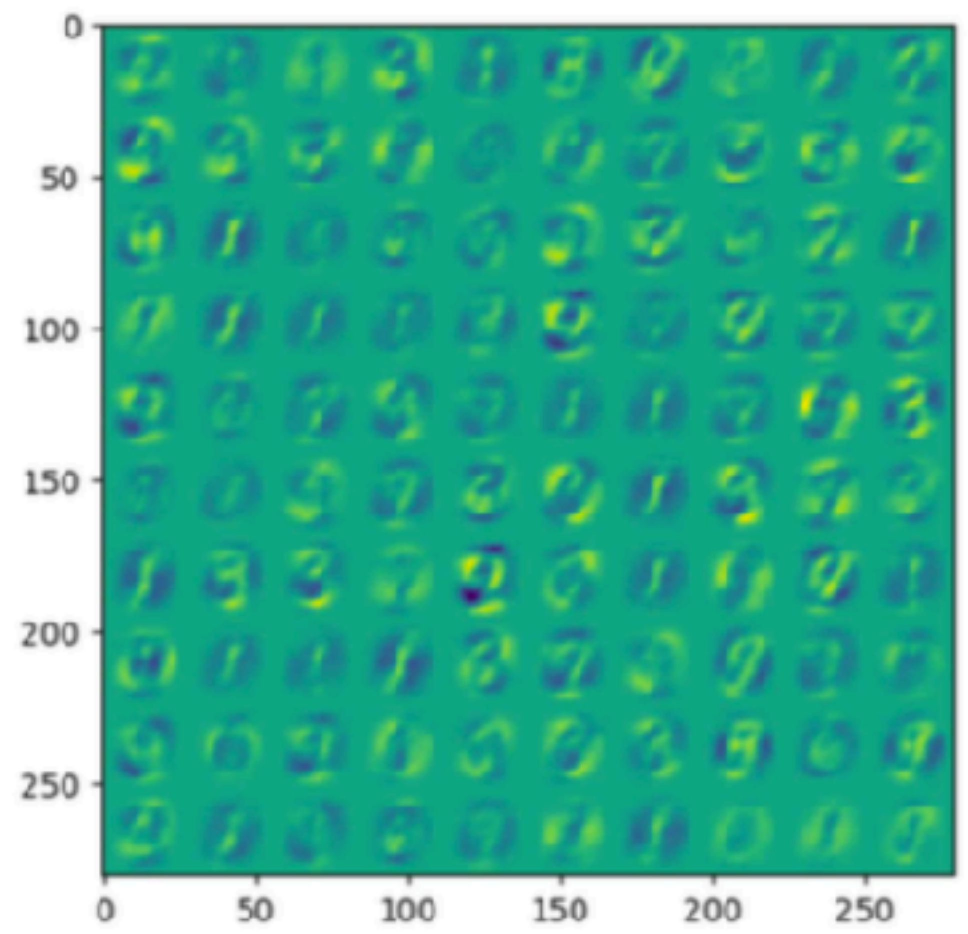
- the first strongest mode of the data is learned by all features;
- high positive (or negative) feature similarity;
- the generated samples have a high overlap with the learned features;

Figures from Decelle et al (2020)

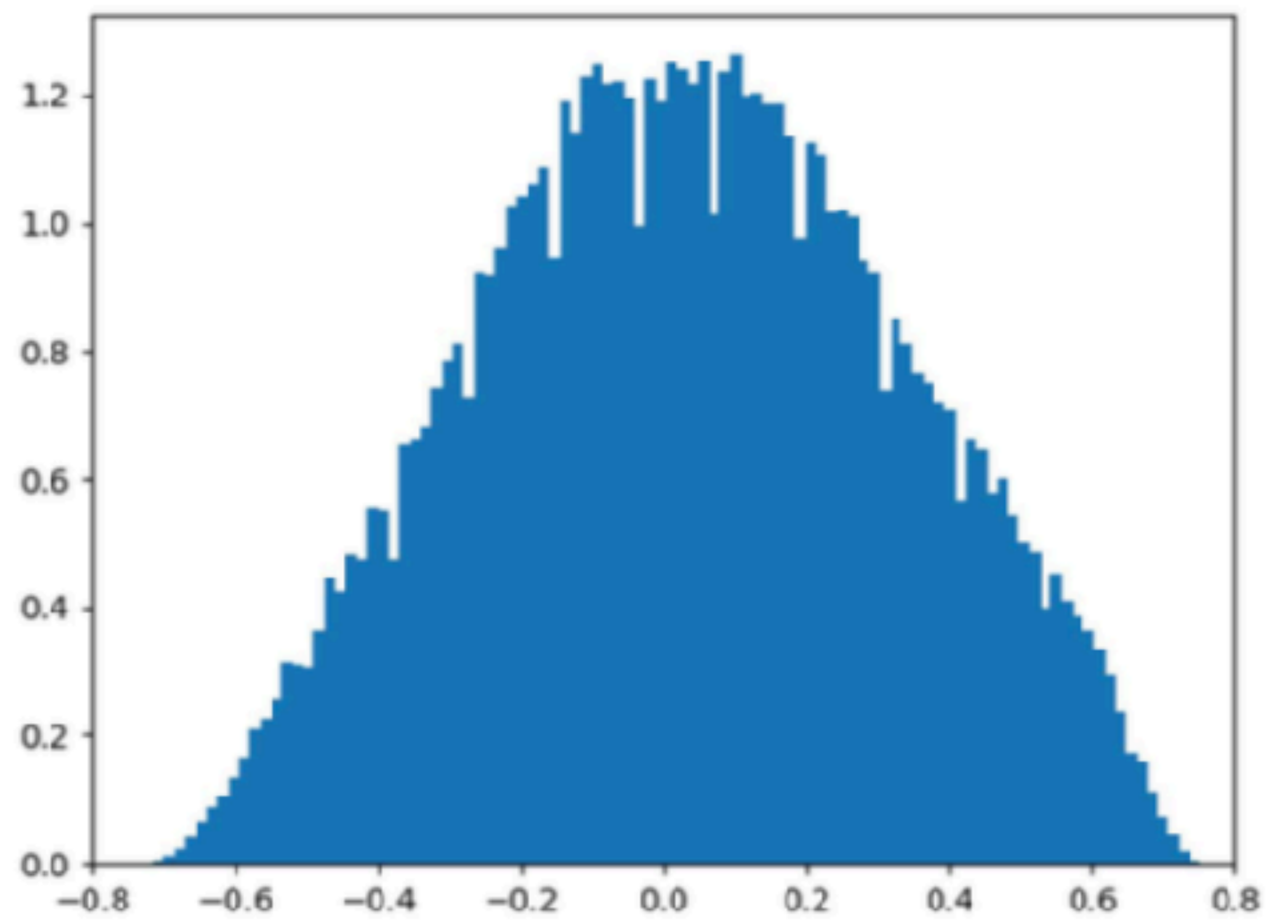
PHASES OF LEARNING (STAGE 2)

FERROMAGNETIC MATTIS PHASE

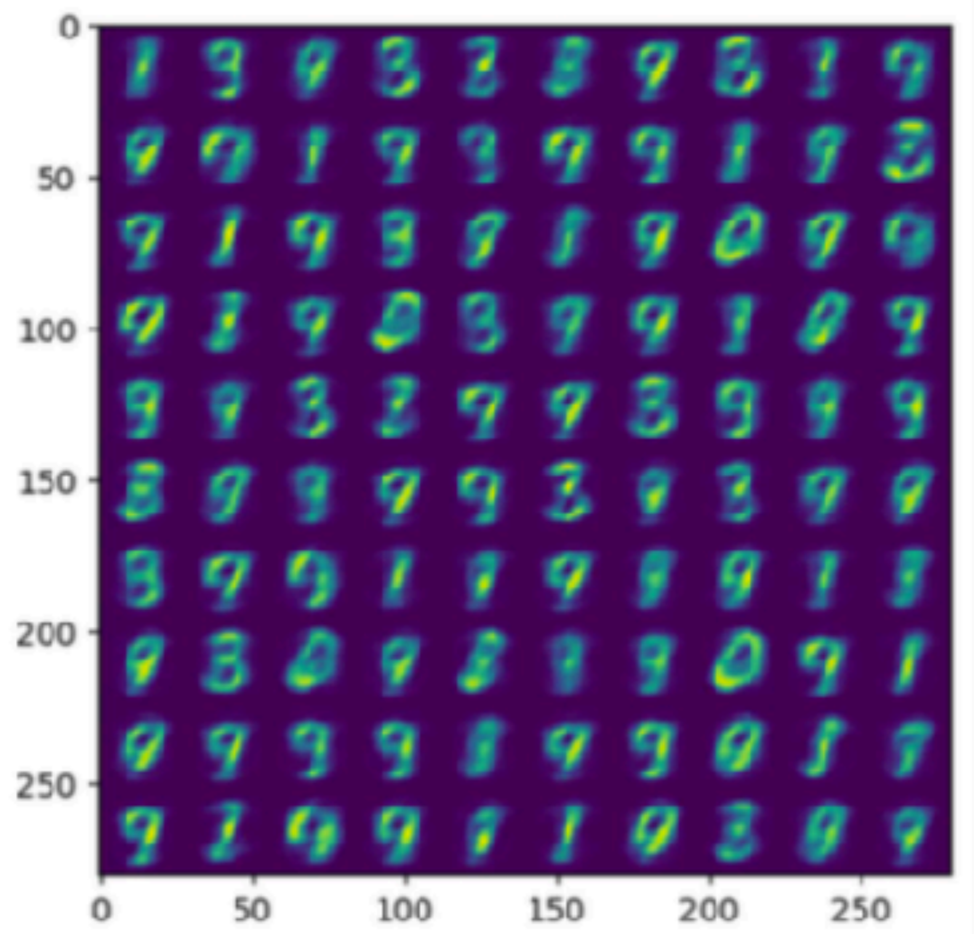
FEATURES

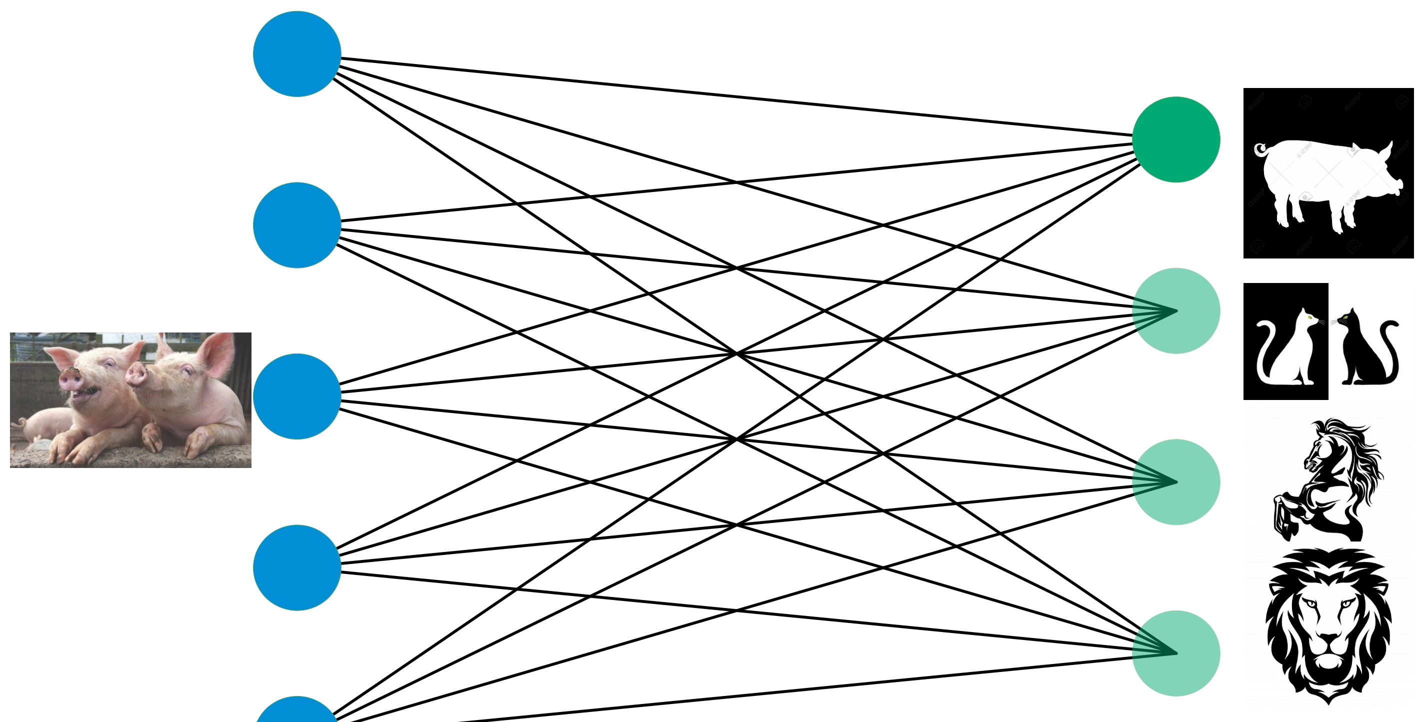


FEATURE SIMILARITY



SAMPLES



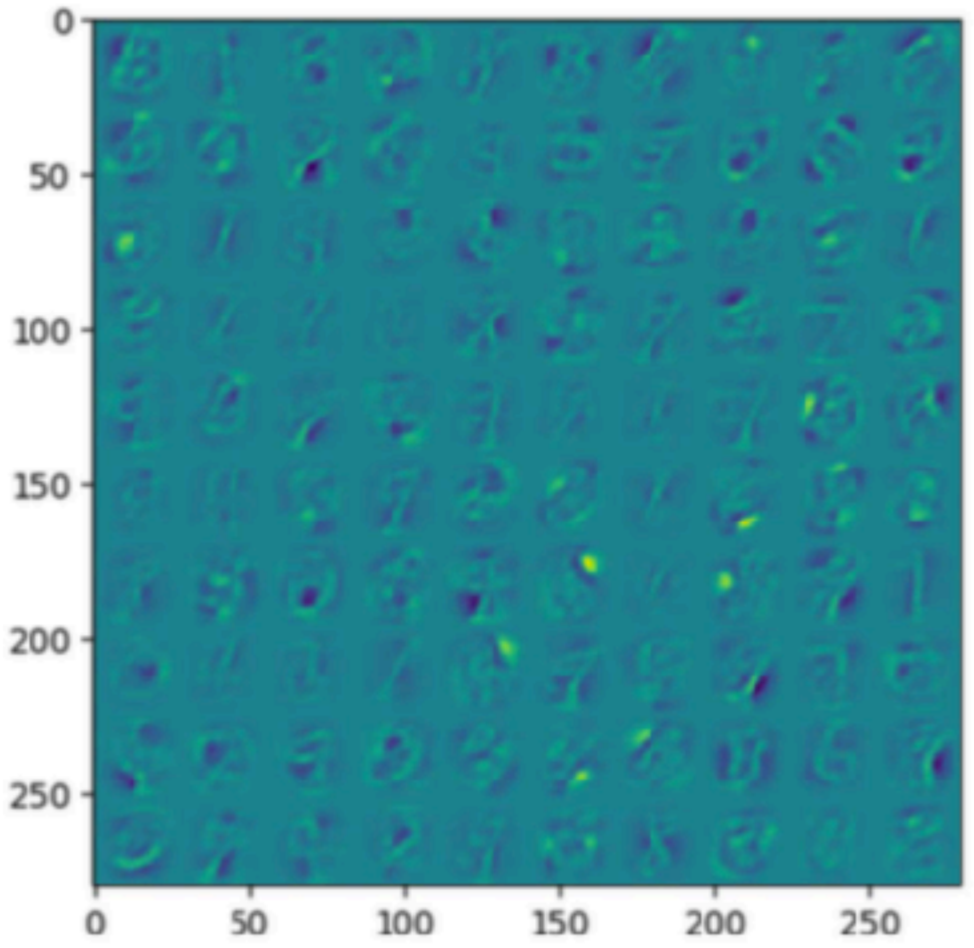


- many modes emerged: features are global and close to modes ;
- smaller similarity but broad similarity distribution;
- the generated samples correspond basically to the learned features with few variety.

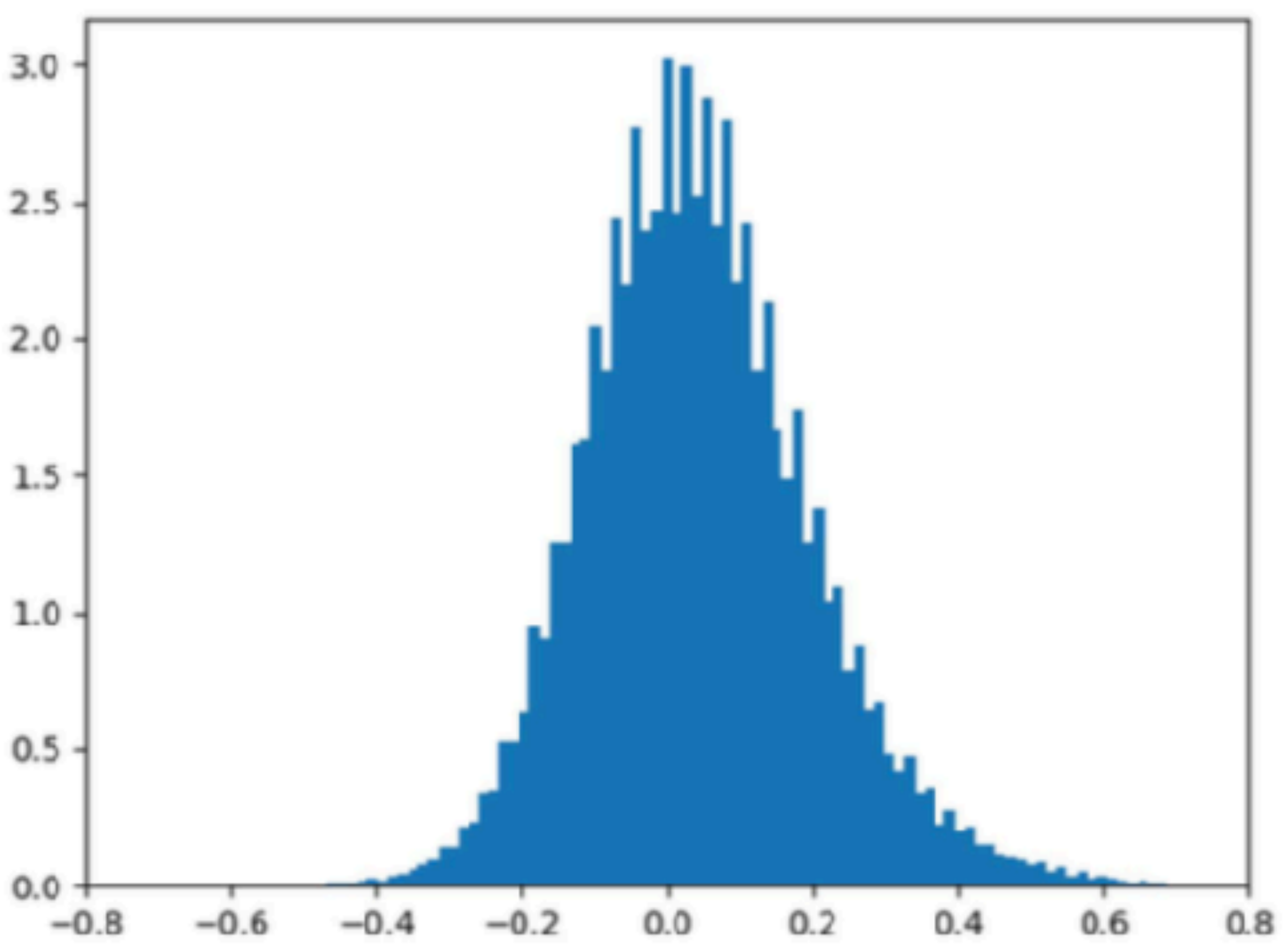
PHASES OF LEARNING (STAGE 3)

FERROMAGNETIC COMPOSITIONAL PHASE

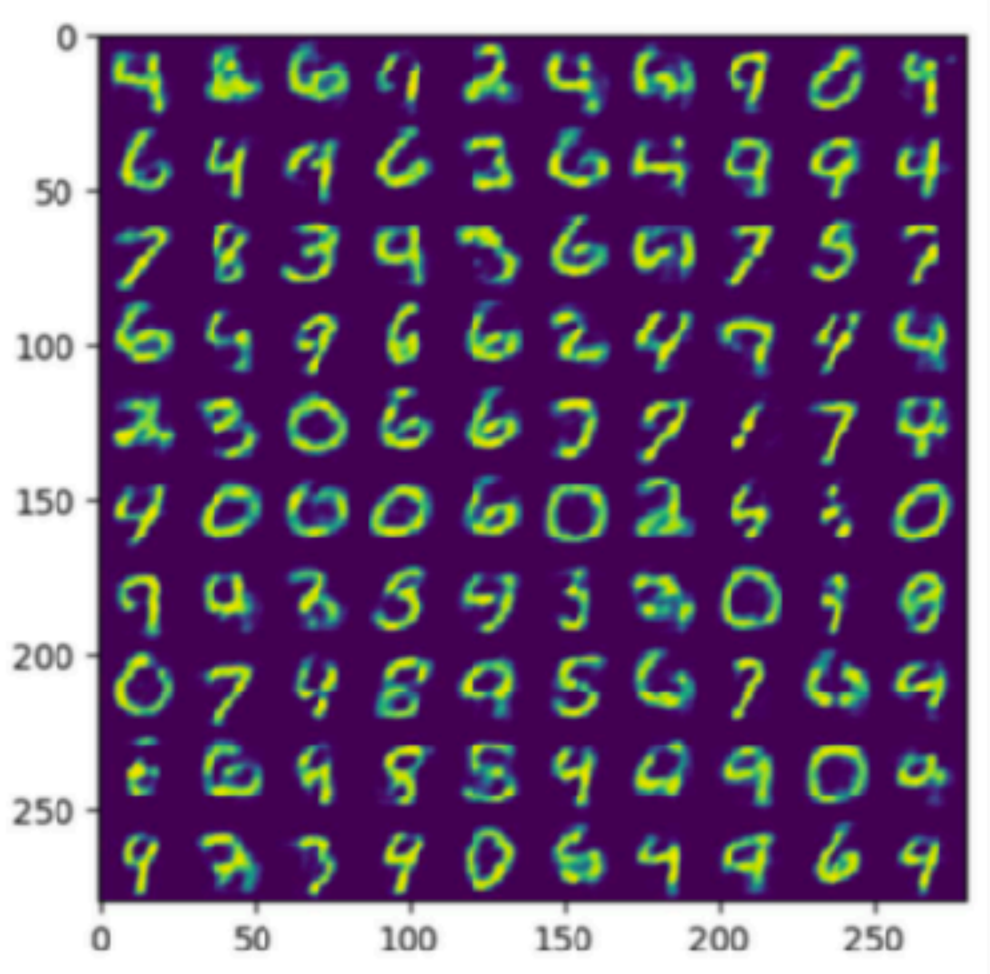
FEATURES

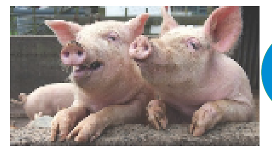


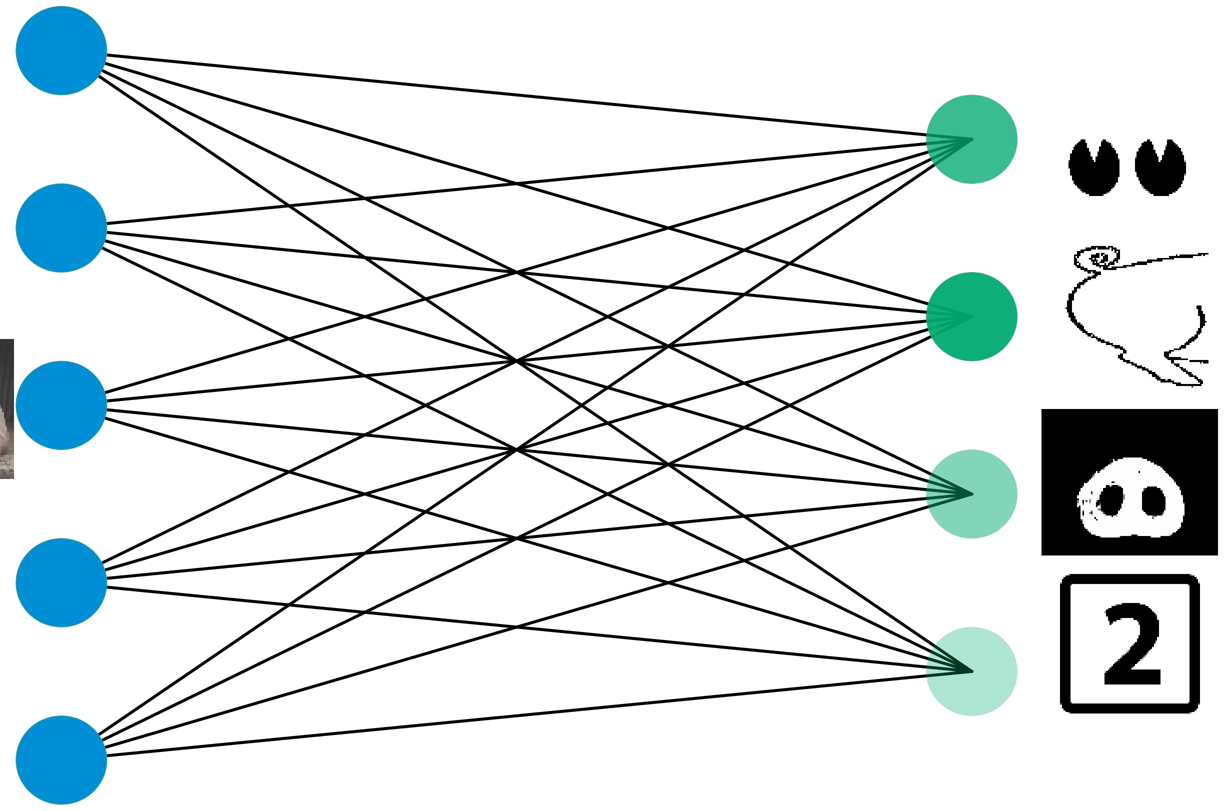
FEATURE SIMILARITY



SAMPLES







- features are much more localized (like the case study with diluted weights);
- feature similarity distribution around zero with smaller variance
- the generated samples look very similar to the provided dataset

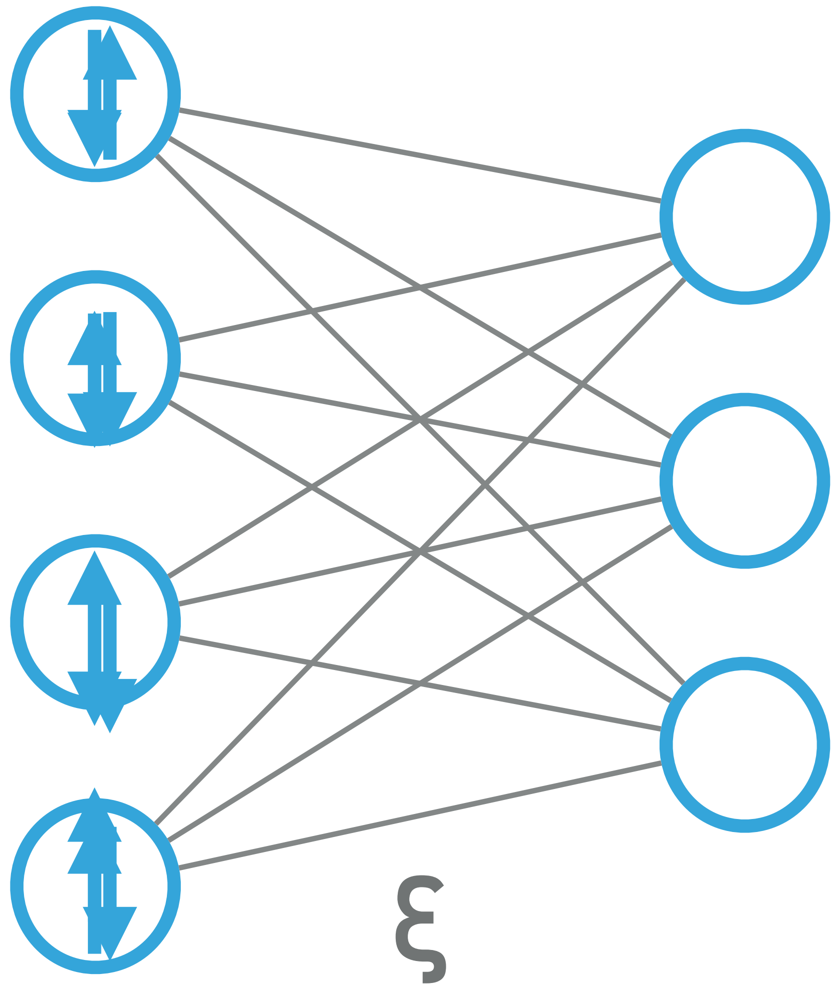
SIZE OF THE DATASET AND LEARNING FEASIBILITY

TEACHER RBM

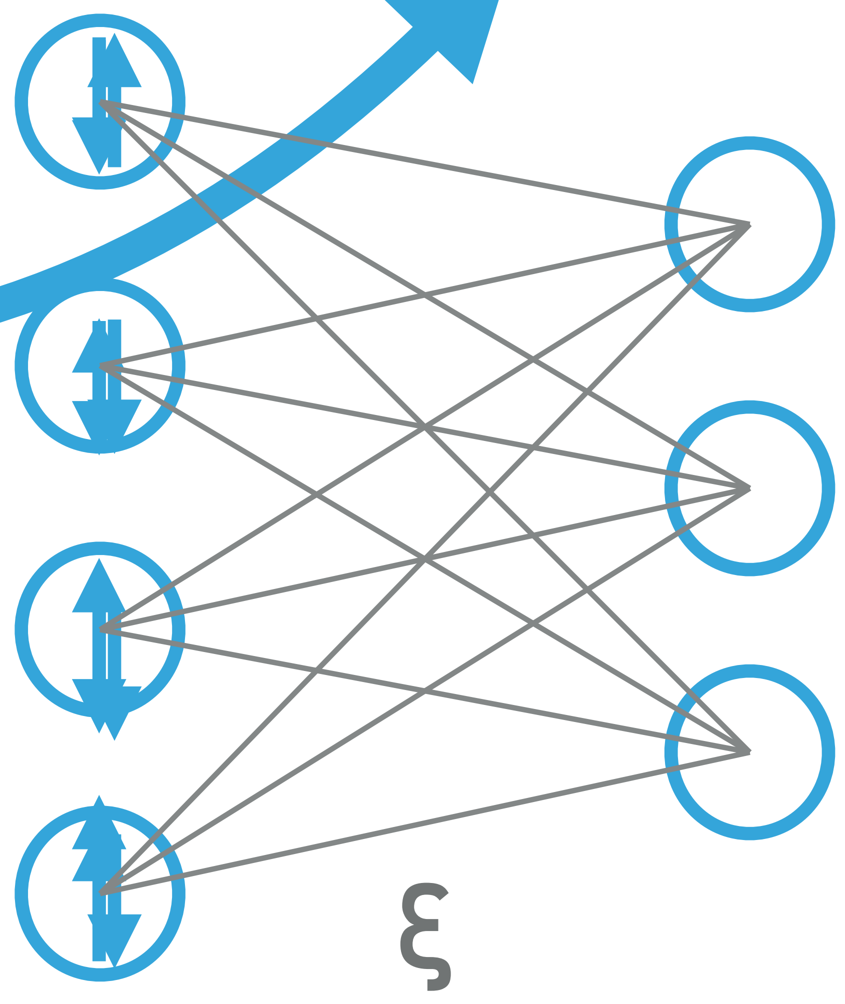
STUDENT RBM

$$\boxed{\sigma^a \sim P(\sigma|\xi) = Z^{-1}\mathbb{E}_{\tau} e^{\sum_{\mu,i} \xi_i^{\mu}\sigma_i\tau^{\mu}}}$$

- **Direct Problem**: given $\{\xi_i^{\mu}\}$ sample $\{\sigma^a\}_{a=1}^{M}$
- **Inverse Problem**: given $\{\sigma^a\}_{a=1}^{M}$ find $\{\xi_i^{\mu}\}$



RBM Unsupervised Learning
in a controlled setting

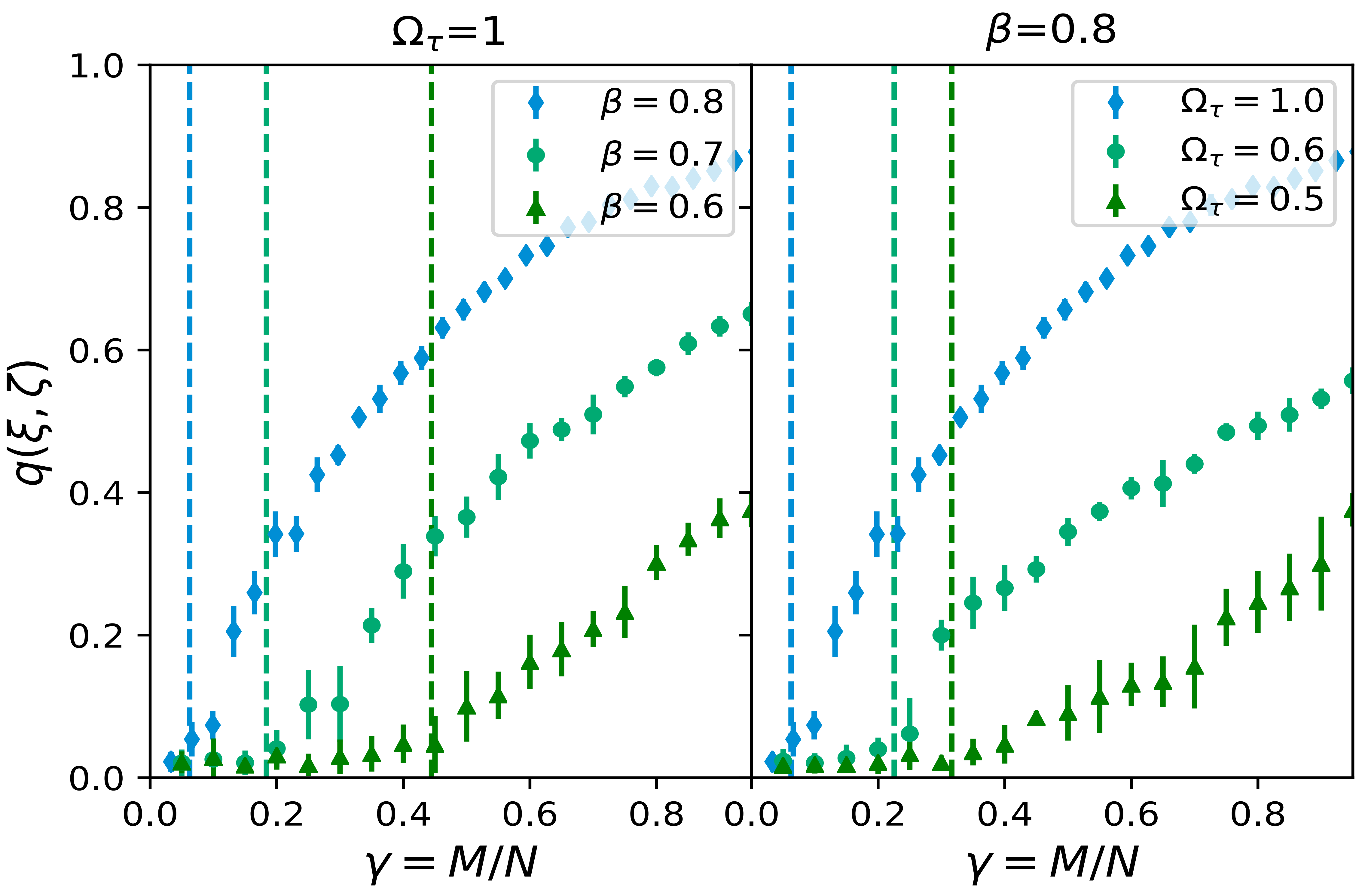


SIZE OF THE DATASET AND LEARNING FEASIBILITY

- How many samples M are necessary to reconstruct the teacher weights?

Posterior Prior Likelihood

INVERSE PROBLEM $$P(\xi|\{\sigma^a\}_{a=1}^M) \propto P(\xi) \prod_a P(\sigma^a|\xi)$$ **DIRECT MODEL**



$$q = \frac{1}{N}\sum_{i=1}^{N} \xi_i \zeta_i$$ Overlap between student and teacher weights

PHASE TRANSITION AND FEASIBILITY THRESHOLD

$$\beta_c^{-1} = \Omega_\tau + \frac{\sqrt{\gamma_c}}{2} + \frac{1}{2}\left[\gamma_c + 4\Omega_\tau(1-\Omega_\tau)\sqrt{\gamma_c}\right]^{1/2}$$

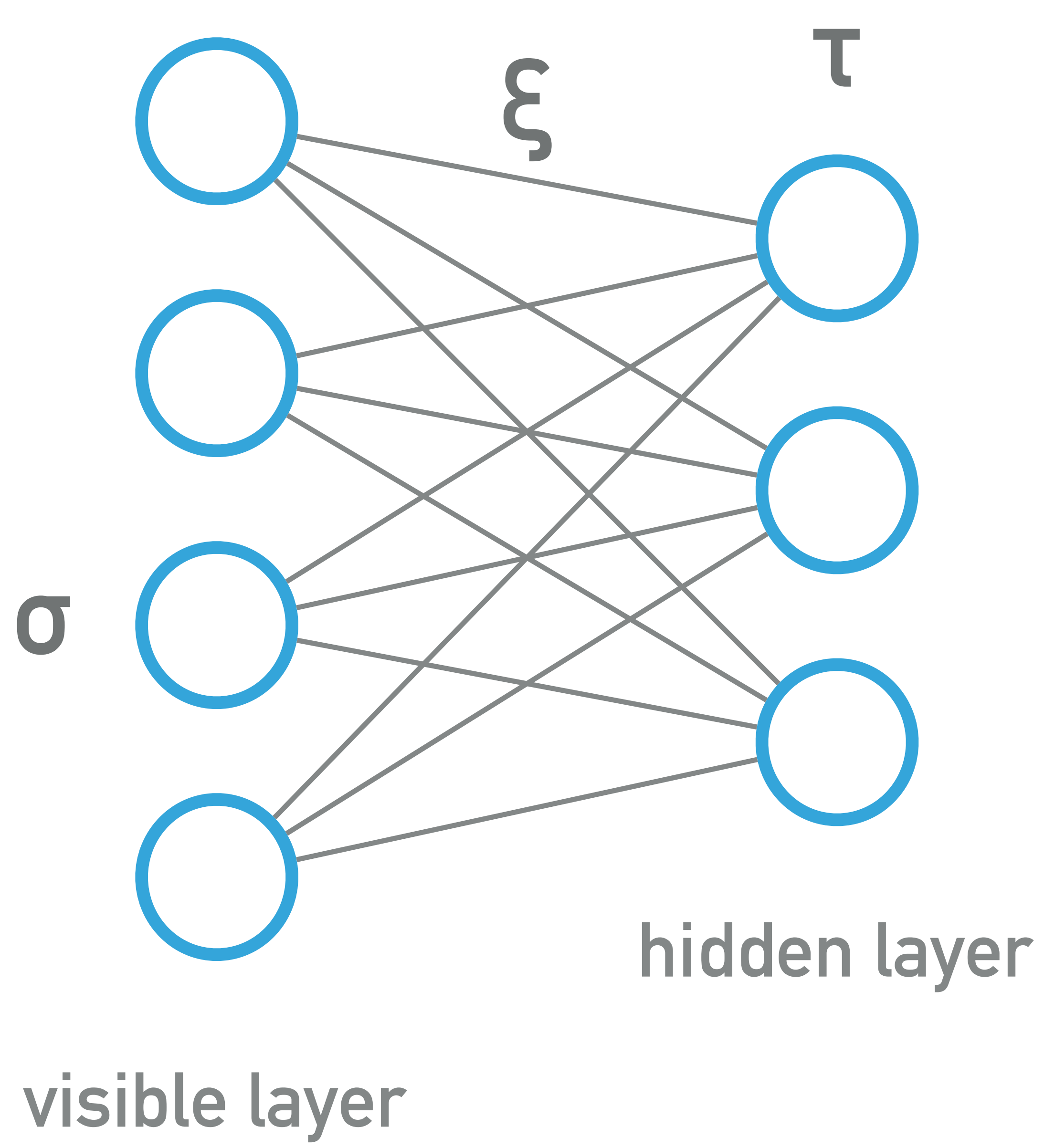
STATISTICAL MECHANICS

...FOR NEW ALGORITHMS

RBM LEARNING

- Weights can be found maximizing the log-likelihood of a training set of data

$$\mathcal{L} = \frac{1}{M}\sum_{a=1}^{M}\left(-\log Z(\xi) + \log \int dP(\tau) e^{-E_\xi(\sigma^a,\tau)}\right)$$



GRADIENT ASCENT

$$\boxed{\xi_i^\mu = \xi_i^\mu + \left(\langle \tau_\mu \sigma_i \rangle_{\text{sample}} - \langle \tau_\mu \sigma_i \rangle_{\text{RBM}}\right)}$$

$$\langle \tau_\mu \sigma_i \rangle_{\text{sample}} = M^{-1} \sum_{a=1}^{M} \int dP(\tau^\mu) P(\tau^\mu | \sigma^a; \xi) \tau_\mu \sigma_i^a$$

$$\langle \tau_\mu \sigma_i \rangle_{\text{RBM}} = \frac{\partial}{\partial \xi_i^\mu} \log Z(\xi)$$

Contrastive Divergence is used to evaluate the difficult term of the momentum-matching condition

DIFFERENT IDEAS FROM STATISTICAL MECHANICS

$\log Z(\xi)$

BOLTZMANN LEARNING

Free energy approximation: (Zero temperature optimization)

- Pseudo-Likelihood optimization

  Decelle et al (2014)

- O(1/N) expansion of the log-likelihood

  Cocco et al. (2011)

- Belief Propagation and Bethe free energy

  Tramel et al. (2018)

- High temperature expansion and TAP equations

  Gabriè et al (2015)

Finite temperature optimization

- Inverse problem and dual TAP equations

  Decelle et al (2019)

$P(\xi | \{\sigma^a\})$

POSTERIOR DISTRIBUTION

HIGH TEMPERATURE EXPANSION AND TAP EQUATIONS

$$\boxed{\log Z(\xi) = A(\beta = 1)}$$

Free energy $$A(\beta) = \log \sum_{\sigma,\tau} e^{-\beta E_\xi(\sigma,\tau)} = \inf_m \Lambda(m, \beta)$$

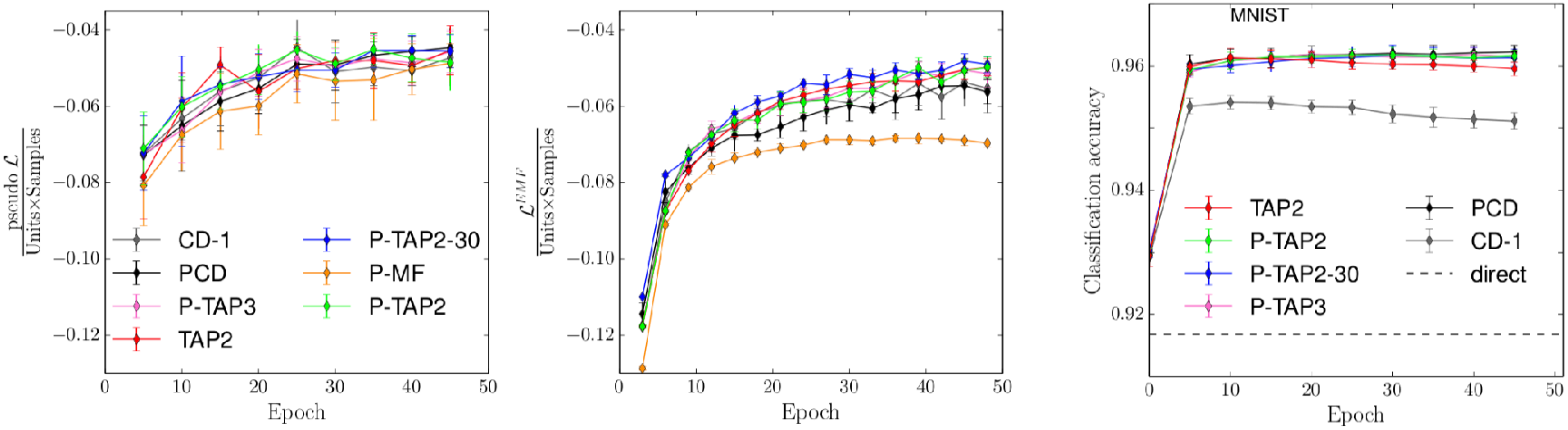
- Expansion around $\beta = 0$
- Find the minimum $m = \{m_i^\sigma, m_\mu^\tau\}$

Local magnetizations

$$\boxed{\begin{aligned} m_i^\sigma &= \psi\left(b_i + \sum_\mu \xi_i^\mu m_\mu^\tau + \cdots\right) \\ m_\mu^\tau &= \psi\left(c_i + \sum_i \xi_i^\mu m_i^\sigma + \cdots\right) \end{aligned}}$$

TAP EQUATIONS

HIGH TEMPERATURE EXPANSION AND TAP EQUATIONS



Figures from Gabriè et al (2015)

DUAL TAP EQUATIONS

- Assuming the data generated from un unknown RBM (teacher), we can consider the posterior distribution

GENERALIZED HOPFIELD MODEL

$$P(\xi | \{\sigma^a\}_{a=1}^M) = Z^{-1} e^{-\hat{E}_\sigma(\xi)}$$
$$= P_0(\xi) \cdot Z^{-1} e^{-\hat{E}_\sigma(\xi^1)} \, . \, Z^{-1} e^{-\hat{E}_\sigma(\xi^2)} \, . \, \ldots Z^{-1} e^{-\hat{E}_\sigma(\xi^P)}$$

- Mezard 2017

$$\boxed{m_i^{t+1} = \tanh\left(\beta \sum_{j=1}^N J_{ij} m_j^t - \frac{\alpha\beta}{1-\beta(1-q)} m_i^t\right)}$$

FINITE TEMPERATURE MMAP APPROACH

$$\xi_i^\mu = \mathrm{sign}(m_i^\mu)$$

DUAL TAP EQUATIONS

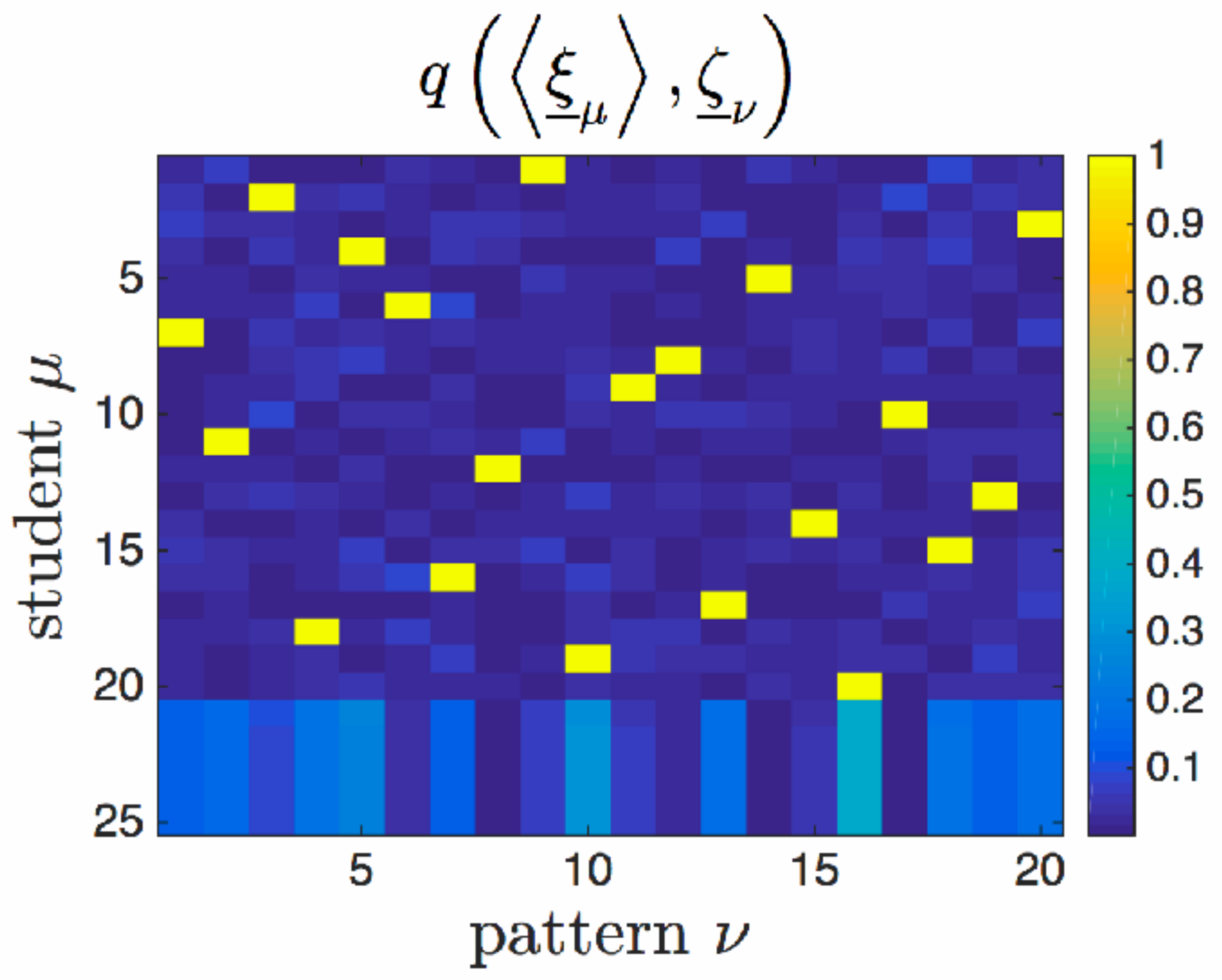
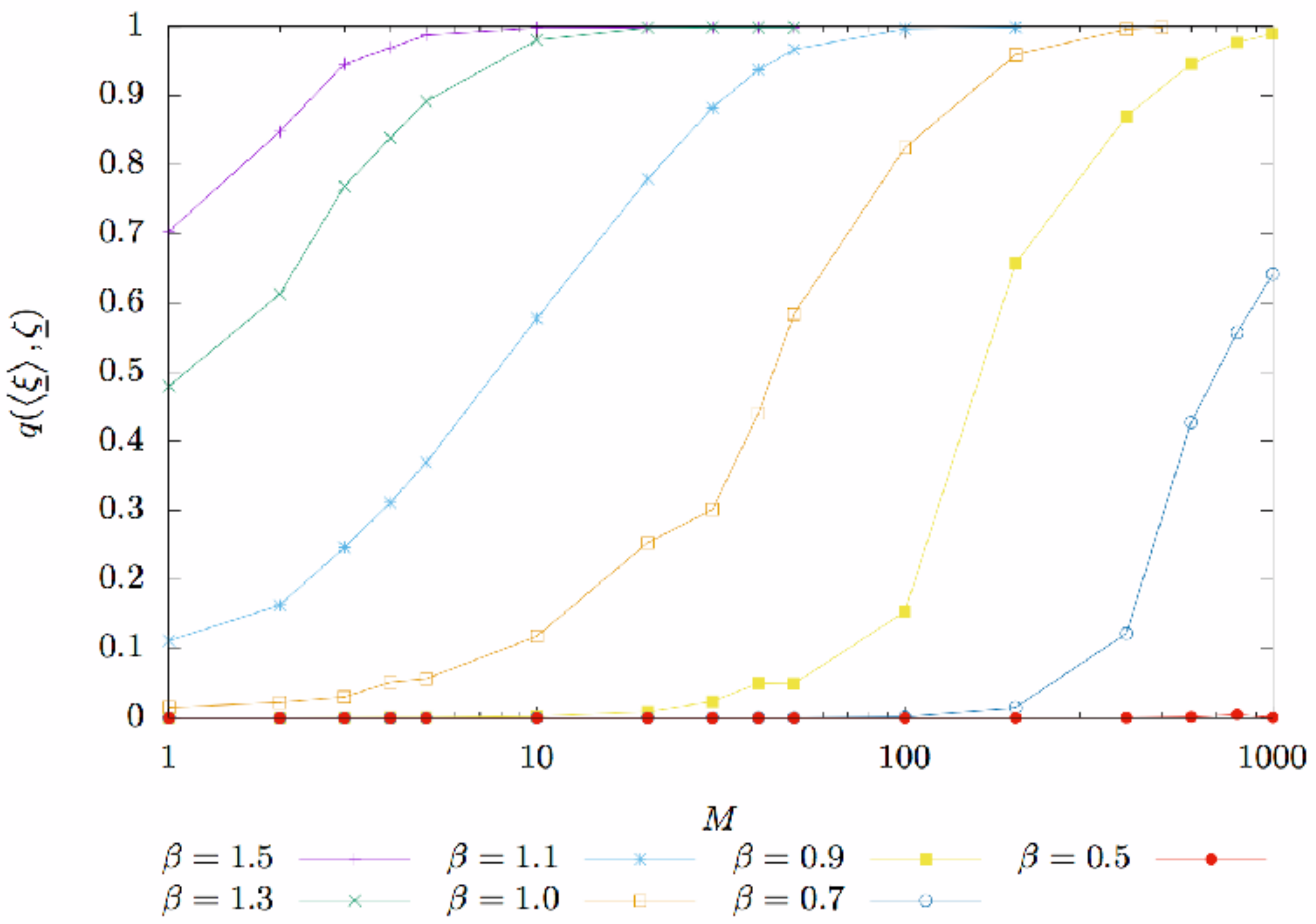


FIG. 2. Overlap between the TAP solutions and the teacher's patterns. The system size is $N = 1000$, $P = 20$, $\beta = 2$, i.e. data is generated in the retrieval phase. Inference is done with $P' = 25$ students observing $M = 200$ samples.

- A teacher-student experiment



P=1

REFERENCES

- Decelle, A., Hwang S., Rocchi J., Tantari D. (2019) Inverse problems for structured data sets using wisely TAP equations and RBM, preprint arXiv:1906.11988

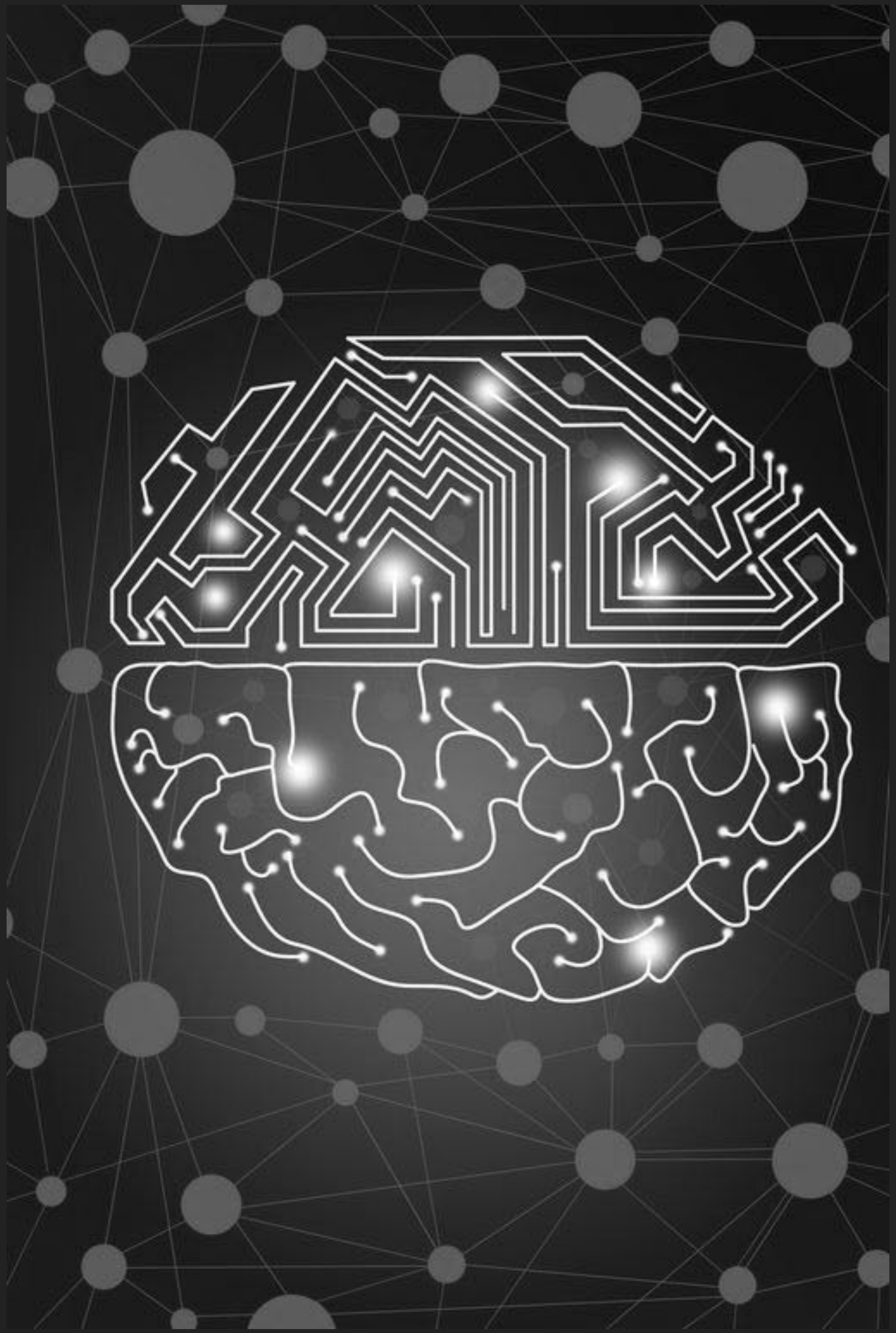
- Barra, A., Genovese, G., Sollich, P., Tantari, D. (2017) Phase transitions in Restricted Boltzmann Machines with generic priors. Physical Review E, 96(4), 042156.

- Barra, A., Genovese, G., Sollich, P., Tantari, D. (2018) Phase diagram of restricted Boltzmann machines and generalized Hopfield networks with arbitrary priors. Physical Review E, 97(2), 022310.

- Barra, A., Contucci, P., Mingione, E., & Tantari, D. (2015, March). Multi-species mean field spin glasses. Rigorous results. In Annales Henri Poincaré (Vol. 16, No. 3, pp. 691-708). Springer Basel.

- M Gabrié, EW Tramel, F Krzakala, Training Restricted Boltzmann Machine via the Thouless-Anderson-Palmer free energy. Advances in Neural Information Processing Systems, 640-648

- Decelle, A., & Furtlehner, C. (2020). Restricted Boltzmann Machine, recent advances and mean-field theory. Chinese Physics B.



THANK YOU FOR THE ATTENTION!