The Amount of Data Needed to Train Dense Hopfield Networks

Robin Thériault [1], Daniele Tantari [2]

[1]Scuola Normale Superiore di Pisa

[2]University of Bologna

Background

Ordinary Hopfield networks

Ordinary Hopfield networks [1] retrieve examples σ^a of memories ξ^b by finding the statistical equilibrium of

$$H\left[\sigma^a|\xi\right] = -\frac{1}{N}\sum_{b=1}^{M}\left(\sum_{i=1}^{N}\xi_i^b\sigma_i^a\right)^2,$$

... or conversely store memories ξ^b of examples σ^a using

$$H\left[\xi^b|\sigma\right] = -\frac{1}{N}\sum_{a=1}^{M}\left(\sum_{i=1}^{N}\xi_i^b\sigma_i^a\right)^2.$$

Pros:

- Biologically plausible [1, 2].
- Simple to implement and well-studied.

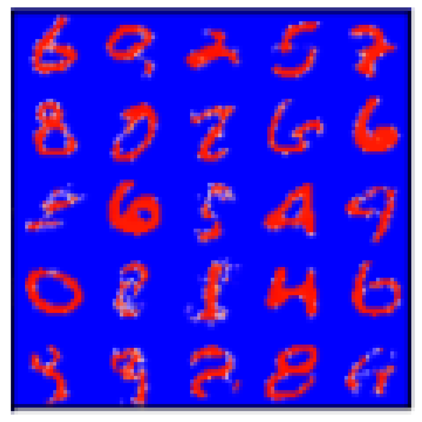
Cons:

- Correlated memories produce spurious examples [3].
- Retrieval fails with $M \gtrsim 0.14N$ memories [1, 3].

Dense Hopfield networks

Dense Hopfield networks [4] overcome these limitations by using

$$H\left[\sigma^a|\xi\right] = -\frac{1}{N^{p-1}}\sum_{b=1}^{M}\left(\sum_{i=1}^{N}\xi_i^b\sigma_i^a\right)^p.$$



Pros [4]:

- No spurious states when $p \gg 1$.
- Can be made into an explainable and robust classifier.
- Can store up to $O\left(\frac{N^{p-1}}{p!}\right)$ memories.

Cons:

- Biologically implausible.
- Not very well studied yet.

We want to know many examples are needed to learn a memory.

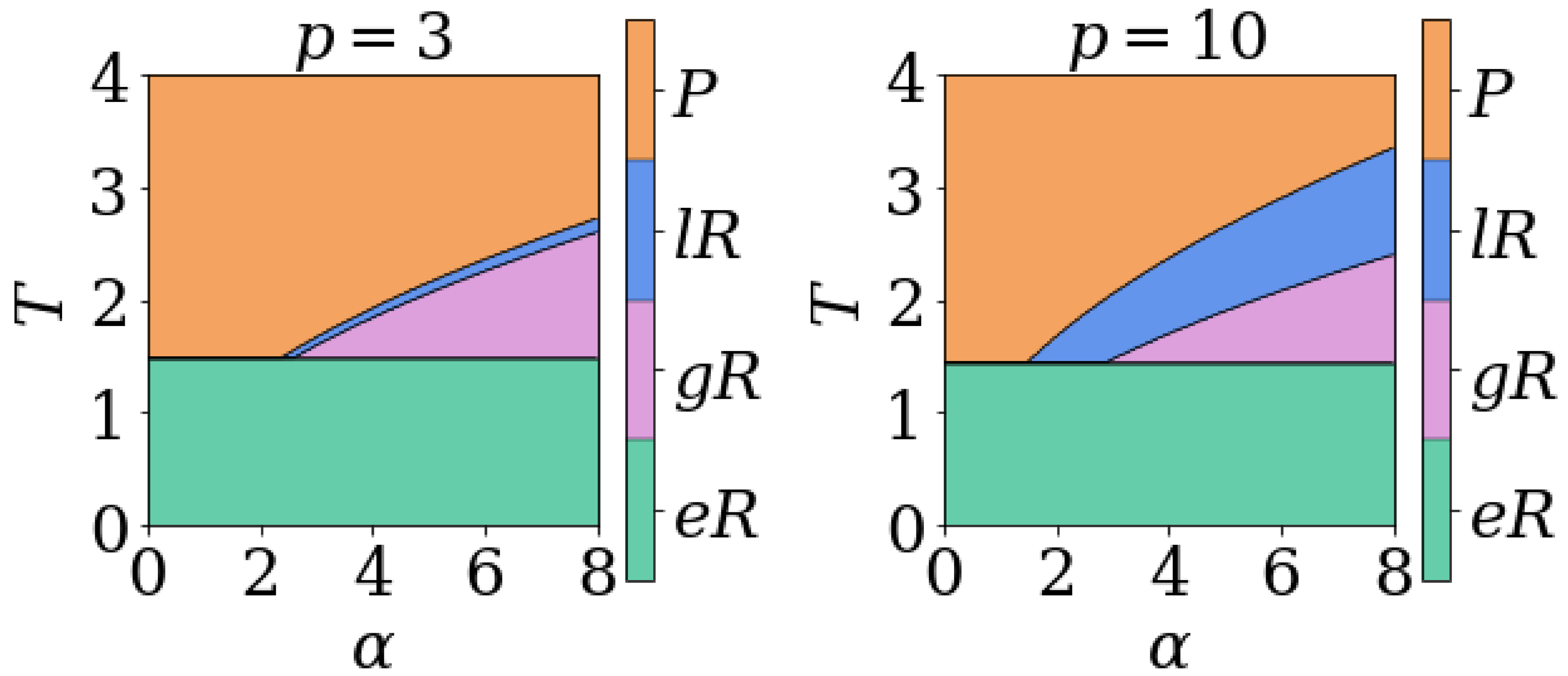
Teacher-student setting

We train a student model $H\left[\xi^b|\sigma\right]$ with M teacher examples $\sigma^a \sim H\left[\sigma^a|\xi^*\right]$. In other words, the student tries to infer the pattern ξ^* of the teacher using a large structured set of examples σ^a. In this setting, we use the overlaps $q^* = \frac{1}{N}\sum_i \xi_i^*\xi_i^b$ and $m = \frac{1}{N}\sum_i \xi_i^b\sigma_i^a$ to measure inference quality.

Results

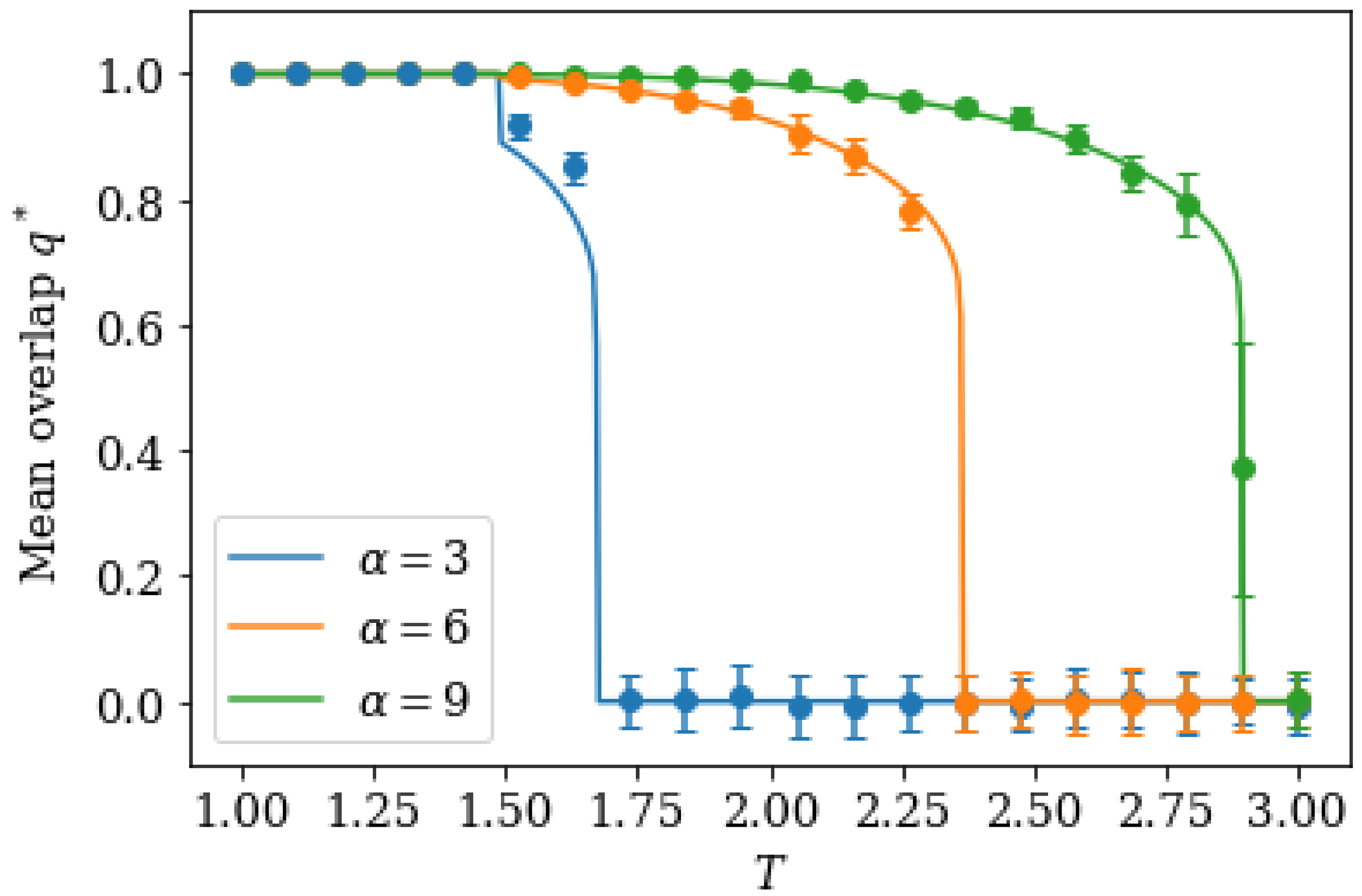
Given noise T and $\alpha = \frac{Mp!}{N^{p-1}}$, the replica method yields a phase diagram with four regimes:

- P: Paramagnetic phase with $q^* = m = 0$.
- lR: Local retrieval with $q^* > 0$ and $m = 0$.
- gR: Global retrieval phase with $q^* > 0$ and $m = 0$.
- eR: Example retrieval phase with $q^* > 0$ and $m > 0$.



The phase diagram passes three benchmarks:

- The entropy is always positive.
- The $p = 10$ gR boundary matches its analytical $p \to \infty$ counterpart.
- The $p = 3$ overlap landscape agrees with Monte-Carlo simulations.



Discussion

We need $M \sim O\left(\frac{N^{p-1}}{p!}\right)$ examples to reach the onset between eR and gR/lR. When $p \gg 1$, it is intractable to manage so many examples, so high overlap is only possible in the eR phase. In consequence, ξ^* cannot be retrieved when $p \gg 1$ and $T > T_{eR} \approx \frac{1}{\log 2}$. On the other hand, it is possible to reach gR when $p = 3$. However, it still requires significant computer time and resources with a Monte-Carlo simulation.

In the eR phase, the student memorizes examples that are strongly correlated with ξ^*. In the lR and gR phases, on the other hand, retrieval is done by learning subtle cues from weakly-correlated examples. These two types of examples are called prototypes and features, respectively. Suppose the initial value of ξ^b given to the student is a corrupted copy of ξ^*, then recovery is much faster with prototypes. Essentially, we are exchanging resistance to teacher noise against simplicity of the optimization landscape. This behavior is similar to the robustness-accuracy trade-off observed in machine learning, notably in classifiers built upon dense Hopfield networks [5].

Summary

Using the replica method, we compute the phase diagram of dense Hopfield networks in the teacher-student setting and find:

- Our benchmarks suggest that the phase diagram is exact.
- The student can retrieve ξ^* by memorizing prototypes or learning features.
- The feature regime is intractable when $p \gg 1$.
- There is a trade-off between the two regimes that shares some similarities with the robustness-accuracy trade-off of machine learning.

References

[1] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities." *Proceedings of the national academy of sciences*, vol. 79, no. 8, pp. 2554–2558, 1982.

[2] D. O. Hebb, *The organization of behavior: A neuropsychological theory*. Psychology press, 2005.

[3] D. J. Amit, H. Gutfreund, and H. Sompolinsky, "Storing infinite numbers of patterns in a spin-glass model of neural networks," *Physical Review Letters*, vol. 55, no. 14, p. 1530, 1985.

[4] D. Krotov and J. J. Hopfield, "Dense associative memory for pattern recognition," *Advances in neural information processing systems*, vol. 29, 2016.

[5] D. Krotov and J. Hopfield, "Dense associative memory is robust to adversarial inputs," *Neural computation*, vol. 30, no. 12, pp. 3151–3167, 2018.