Visualizing the clustering transition using unsupervised learning

Alexandre G.R. Day[1,*] and Pankaj Mehta[1]

[1] Department of Physics, Boston University, 590 Commonwealth Ave., Boston, MA 02215, USA

(Dated: April 15, 2018)

We revisit the two random boolean satisfiability problems 3-SAT and XORSAT, with the latter equivalent to the p-spin model, using recently developed machine learning techniques. We use the non-linear embedding technique known as t-SNE to learn the local manifolds in which the solutions organize. In particular we provide a visualization of the so-called clustering transitions that are known to occur in those problems. Finally, using unsupervised clustering methods we are able to automatically extract informative quantities such as the distribution of entropy of clusters. This work highlights the potential use of common machine learning tools to study statistical physics problems.

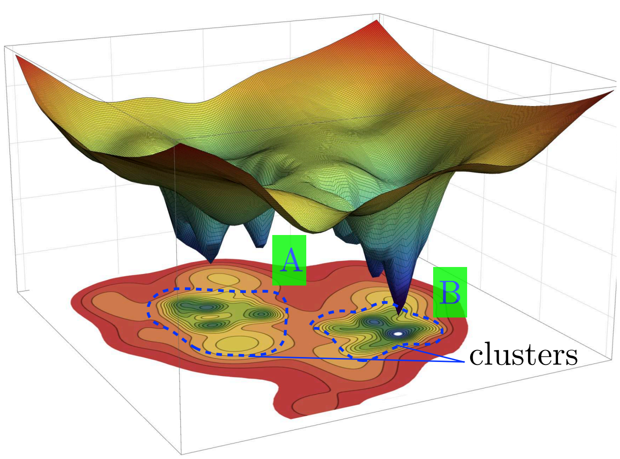

FIG. 1: Problem setting. (a) Rugged landscape cartoon picture. Optimizing rugged landscapes is hard ! Obviously.

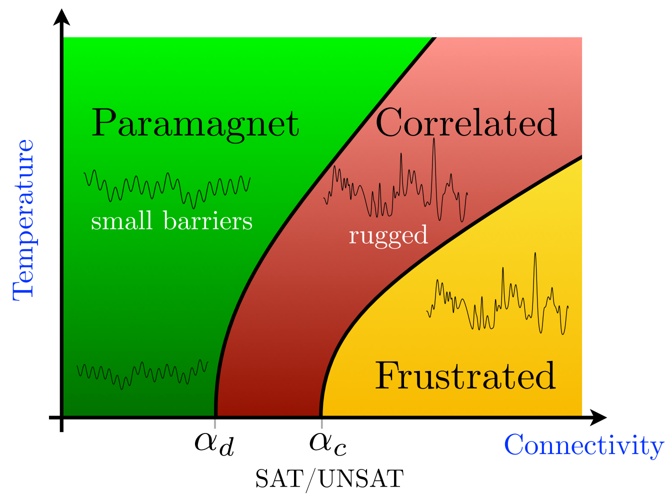

FIG. 2: (a) In 3 dimensions we project attractors to a 2D space

Landscape induces complexity, but not always : [1]

I. INTRODUCTION

Machine learning is finding more and more applications in the context of physics [2]. Chris Moore discussion, etc. cite Mzard, stat. phys. and glassy phases have been discussed in the context of quantum optimal control [3].

II. MODEL STUDIED

p-spin model—. We consider the diluted p-spin model with $p = 3$, defined as:

$$H = -\sum_{\langle ijk \rangle}^{\alpha N} J_{ijk} s_i s_j s_k, \tag{1}$$

where J_{ijk} take value $+1$ or -1. An instance of the diluted p-spin model is obtained by choosing αN such couplings uniformly at random. The zero-temperature limit of the p-spin model is equivalent to the p-XORSAT problem from computer science. The p-XORSAT corresponds to a linear system of equations in base 2 of the form $A\vec{x} = \vec{y}$ obtained from the mapping $s_i = (-1)^{x_i}$, $J_{ijk} = (-1)^{y_{ijk}}$, where A is a sparse $N \times M$ matrix encoding the interactions with the M constraints $\vec{y}$. As such, the problem of determining wether a XORSAT formula is satisfiable or equivalently if the p-spin model possesses a non-frustrated ground-state Eq.(1) can be solved efficiently using Gaussian elimination which scales like $\mathcal{O}(N^3)$. The statistical properties of XORSAT have been studied using powerful methods from spin-glass physics [4]. We sampled the XORSAT solutions uniformly at random using Gaussian elimination. The results are presented in figure 3.

k-SAT—. The k-SAT problem blablabla

*Electronic address: agrday@bu.edu

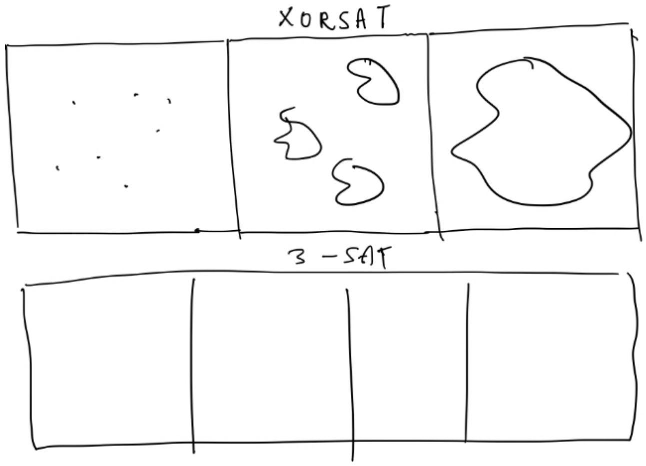

FIG. 3: t-SNE visualization of the solution space as a function of the constraint parameter α.

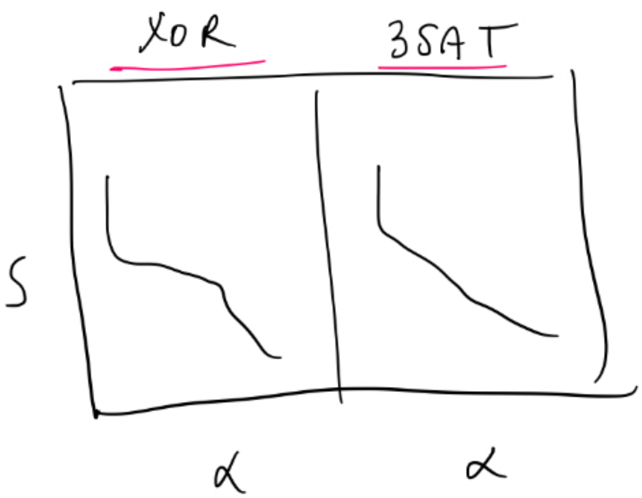

FIG. 4: Entropy of clusters vs. α.

A. Method used

a. Uniform XORSAT sampler

b. 3-SAT sampler Here we rely on prior work by Zecchina et al. and sophisticated implementation of decimation based survey propagation algorithm. In order to decorrelate samples as much as possible we perform trivial isomorphism of the CNF by a permutation of the litteral's labels. However we have no theoretical guarantees that the samples obtained are uniformly distributed in the space of 3-SAT solutions. We hope that we at least capture the clusters with the largest entropy.

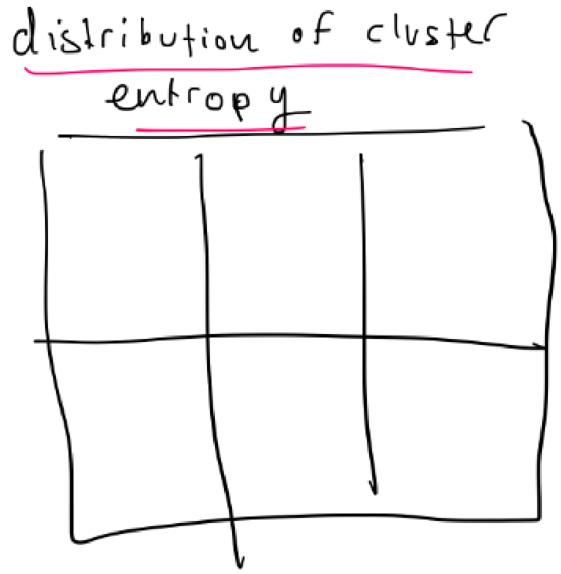

FIG. 5: Distribution of cluster entropies at different α.

III. RESULTS

A. Density clustering

Density clustering makes the intuitive assumption that clusters are high-density regions circumscribed by low-density regions in the configuration space. Density estimation is however notoriously hard in high-dimensional space due to large sampling noise. Here we use density clustering on the t-SNE map. This is usually sufficient to capture the main clusters in the data. Here we used a multi-scale variation of [5], for density clustering two-dimensional data. We also made a user-friendly package for this purpose. Density estimation is done efficiently and accurately using bootstrapped kernel density estimates see Appendix **??**.

B. Results and Discussion

Spin glass models such as the Sherrington-Kirkpatrick model or the Anderson model are known to be NP-complete [6]. Consequently it is believed that no algorithm can be devised in order to compute the ground-state in a time scaling polynomially with the system's size. Much work has been made in trying understanding what makes a problem hard from a statistical mechanics point of view. In particular, in random satisfiability problems such k-SAT and XORSAT the onset of a glass transition has been associated with the appearance of frozen variables and clustering in the solution space, which has been conjectured to induce failure of local search algorithms [7]. Note that glassiness does not necessarily imply hardness to solve. Phenomenologically, it appears that for glassy problems, any local search/stochastic method will fail, i.e. finding the ground-state is exponential in the system size. However, if one is able to devise some global method based on non-local updates, then some glassy problems are known to be in P. This is the case for instance of XORSAT, which is equivalent to solving a linear system mod 2, and thus it can be solved efficiently using gaussian elimination [1].

a. Clustering XORSAT, 3-SAT

b. Entropy via density clustering Relation to complexity and broader applications in stat. mechanics.

IV. CONCLUSION

Along with this work we provide simple python packages for the 3-SAT sampler (wrapper) and the XORSAT sampler along with a density clustering code, all of which are available at the author's GitHub.

V. ACKNOWLEDGEMENTS

Acknowlegements.— We thank C. Lauman and D. Sels. for insightful discussions. AD was support by an NSERC PGS-D. AD and PM acknowledge support from Simon's Foundation through the MMLS Fellow program. The authors are pleased to acknowledge that the computational work reported on in this paper was performed on the Shared Computing Cluster which is administered by Boston University's Research Computing Services. The authors also acknowledge the Research Computing Services group for providing consulting support which has contributed to the results reported within this paper.

[1] F. Ricci-Tersenghi, Science **330**, 1639 (2010).
[2] P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, ArXiv e-prints (2018), arXiv:1803.08823 [physics.comp-ph] .
[3] A. G. R. Day, M. Bukov, P. Weinberg, P. Mehta, and D. Sels, ArXiv e-prints (2018), arXiv:1803.10856 [quant-ph] .
[4] M. Mézard, F. Ricci-Tersenghi, and R. Zecchina, Journal of Statistical Physics **111**, 505 (2003).
[5] A. Rodriguez and A. Laio, Science **344**, 1492 (2014).
[6] D. Venturelli, S. Mandra, S. Knysh, B. O'Gorman, R. Biswas, and V. Smelyanskiy, Physical Review X **5**, 031040 (2015).
[7] C. Moore and S. Mertens, The nature of computation (OUP Oxford, 2011).

Supplemental Material

VI. DENSITY CLUSTERING VIA KERNEL DENSITY ESTIMATES