Statistical Physics of Computation - Exercises

Emanuele Troiani, Vittorio Erba, Yizhou Xu

September 2024

Week 14

Important: The aim of the last exercise session is to give you an idea of what kind of questions we may ask at the exam, and what topics to focus on in your studying. Please note that:

- All the material of the lectures and exercise session is subject to examination, both concerning techniques and the phenomenology of the example problems we studied in class.
- The questions presented in this document cover just a selection of the topics you should know and of the kinds of questions you may be asked at the exam.
- The exam will include also step-by-step exercises of the same style as the various exercise sessions, even though they are not included in this document.

---

1. Consider a Bayesian inference problem with given prior, output channel, and posterior. State and prove the most general version of Nishimori's condition.
2. Consider a thermodynamic system with partition function
$$Z(\beta) = \sum_{s_1,\dots,s_N} e^{\beta H(s)} \tag{1}$$
where each variable s_i can assume values in a discrete set. Define the free entropy, and explain how to compute the entropy density as a function of the energy density $s(e)$ in the limit $N \to \infty$. Feel free to assume that all thermodynamic potentials/energy/entropy are invertible/well behaved if you need to assume that.
3. Consider a constraint satisfaction problem with N variables and P constraints, and assume that the correct thermodynamic scaling for the problem is $P = \alpha N$. Define the SAT/UNSAT transition point.
4. You are given a generic computational problem that is amenable to replica analysis. Describe (and motivate were appropriate) the main steps of the associated with the replica computation.
5. Describe the link between the overlap order parameter and the averaged overlap distribution of a given disordered system.
6. Consider a regression problem in which you generated data/label pairs $(x, y) \sim P_{\text{data}}$ with $x \in \mathbb{R}^d$ and $y \in \mathbb{R}$. A predictor/student function $f_w(x)$ parametrized by some weights w has been trained on a dataset. Write down the definition of the generalization error of the student. Use the mean-square-error in label space to assess the performance of the student.

7. Consider a teacher-student ERM problem amenable to replica computations in the RS ansatz. What are the fundamental order parameters describing the physics of the problem?

8. Give the definition of computational hard phase in a Bayes optimal inference problem. You can suppose that the phase diagram of the problem is given as a function of some signal-to-noise ratio (SNR) parameter, and you can assume that AMP is the best efficient algorithm for the problem at hand.

9. Is it possible for a Bayes optimal inference problem to always be in an impossible/hard phase? If so, provide an example.

10. Consider the factorized probability distribution

$$p(x, y, z, w) = f_1(x, y) f_2(z, w) \,. \tag{2}$$

    Is the corresponding factor graph connected? Justify your answer.

11. Under what condition on the factor graph is BP exact? Under what condition on the factor graph we can conjecture that BP will be exact in the thermodynamic limit?

12. Suppose that a given disordered system is replica symmetric, and has replica symmetric overlap $q_{\rm RS}$. Compute the averaged overlap distribution

$$\mathbb{E}_J \mathrm{Prob}(q(\sigma, \tau) = q) \tag{3}$$

    where σ and τ are independent draws from the Gibbs distribution associated to the system.

13. What is the assumption that we implicitly use whenever we average the free entropy over the disorder, and then claim that the averaged free entropy describes also the properties of specific disorder realizations?