

Statistical Physics of Computation - Exercises

Emanuele Troiani, Vittorio Erba, Yizhou Xu

September 2024

Week 7

7.1 Empirical risk minimisation for the binary perceptron

A simple algorithm to train a classifier that separates two sets of labelled points is to choose a student vector minimizing the risk (or energy, or cost) $E(J)$:

$$\hat{J} = \operatorname{argmin}_{\|J\|^2=N} E(J) = \operatorname{argmin}_{\|J\|^2=N} \sum_{\mu=1}^P v\left(\frac{\sigma^\mu}{\sqrt{N}} J^T \xi^\mu - \kappa\right). \quad (1)$$

The training dataset is composed of $P = \alpha N$ points $\{\xi^\mu\}_{\mu=1}^P$ of dimension N , each with a label $\sigma^\mu = \pm 1$ taken at random.

The random labels can be modified into teacher-student generated labels as we saw in the previous lectures, allowing us to discuss the generalization and estimation tasks. This modifies the replica computation making it a bit more cumbersome, adding more order parameters, so we don't do it here. Instead, we focus our attention on the Empirical Risk Minimisation task, i.e. the fact that our student will be trained by minimizing a cost function.

We will take $v(x) = 0$ for $x \geq 0$, and $v(x) > 0$ for $x < 0$ in all following points.

1. Intuitively, why is it a good idea to pick $v(x)$ to be convex and differentiable when on $x < 0$?
2. We know that in the high dimensional limit the space of solutions of this classification problem has a SAT/UNSAT transition at a critical value α_C . How many global minima of $E(J)$ do you expect to find in the SAT region? And how many in the UNSAT region?
3. What do you expect the ground state energy density to be in the SAT and UNSAT phases?

We will now study the empirical risk minimisation of the energy function $E(J)$ for a generic per-point cost $v(x) = 0$ for $x \geq 0$, and $v(x) > 0$ for $x < 0$.

4. Write the Gibbs distribution associated to the cost function $E(J)$, thought of as an energy function. Is the energy function properly scaled for large N ?
5. Write the averaged replicated partition function for the Gibbs measure of point 2. In which limit should we consider this quantity to access properties of the global minimum?

6. Argue without many computations that the averaged free energy is

$$f(\beta) = -\frac{1}{2\beta} \left[\log(1-q) + \frac{q}{1-q} \right] - \frac{\alpha}{\beta} \int dz \frac{e^{-\frac{z^2}{2q}}}{\sqrt{2\pi q}} \log \left[\int dr \frac{e^{-\frac{r^2}{2(1-q)}}}{\sqrt{2\pi(1-q)}} e^{-\beta v(r+z-\kappa)} \right] \quad (2)$$

to be optimized over q .

Recall that the free energy is defined as $f = -\phi/\beta$, where ϕ is the free entropy $\phi = N^{-1} \log Z$. The two quantities are interchangeable at any finite β , but in the limit $\beta \rightarrow \infty$ the free entropy $\phi = O(-\beta)$, so it is useful to divide by $-\beta$ to access a finite quantity. Also, recall that at the extremiser q^* the free entropy has the decomposition $\phi = s - \beta e$, where e and s are the average values of the energy density and of the entropy density, hence $f = e - s/\beta$. The last equation shows us that $\lim_{\beta \rightarrow \infty} f(\beta) = e_{\text{ground state}}$, which in our problem is the training error, i.e. the value of the cost function $E(J)/N$ at its global minimum.

In the following, assume that $v(x)$ is convex for $x < 0$.

7. We already know that there is going to be a critical value of α_C such that the energy is zero for $\alpha < \alpha_C$. Argue that $q \rightarrow 1$ if we take the limit $\beta \rightarrow \infty$ in the UNSAT region.

8. Show that

$$\lim_{\beta \rightarrow \infty} \frac{-1}{\beta} \frac{e^{-\frac{z^2}{2q}}}{\sqrt{2\pi q}} \log \left[\int dr \frac{e^{-\frac{r^2}{2(1-q)}}}{\sqrt{2\pi(1-q)}} e^{-\beta v(r+z-\kappa)} \right] = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} \min_r \left[\frac{r^2}{2x} + v(r+z-\kappa) \right]. \quad (3)$$

Use the ansatz $q = 1 - \chi/\beta + \dots$ with $\chi > 0$.

Thus, we found that the free energy simplifies quite drastically in the large β limit, as one for the Gaussian integrals can be substituted by a minimisation problem, usually easier to perform analytically or numerically. This is a recurrent point: when considering empirical risk minimisation problems, the $\beta \rightarrow \infty$ limit will lead to some simplifications.

We now focus on the case $v(x) = 0$ for $x > 0$ and $v(x) = x^2/2$ for $x < 0$. This per-point cost function is differentiable, and convex for $x < 0$, so it is a good candidate of an actual cost function we could use in practice to algorithmically solve the problem.

9. Solve the minimisation problem over r in the case $v(x) = 0$ for $x > 0$ and $v(x) = x^2/2$ for $x < 0$. You should obtain

$$\min_r \left[\frac{r^2}{2\chi} + v(r+z-\kappa) \right] = \theta(\kappa-z) \frac{(z-\kappa)^2}{2(1+x)}. \quad (4)$$

10. Show that the leading order for large β of the energy density in the UNSAT phase is given by

$$e(\chi) = -\frac{1}{2\chi} + \frac{\alpha}{2\alpha_c(1+\chi)} \quad (5)$$

where α_c is the SAT/UNSAT threshold we computed in the previous lectures, i.e.

$$\frac{1}{\alpha_c} = \int_{-\infty}^{\kappa} dz \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} (z-\kappa)^2. \quad (6)$$

11. Find the state equation for χ in the UNSAT phase, and use it to find the energy as a function of α both in the SAT and in the UNSAT phase

$$e(\alpha) = \frac{1}{2} \left(\sqrt{\frac{\alpha}{\alpha_c}} - 1 \right)^2 \theta(\alpha - \alpha_c). \quad (7)$$

Thus, we showed that

- the large β limit leads in general to simplifications in the computations.
- the critical α_c delimiting the SAT and UNSAT phases arises naturally in the empirical risk minimisation computation as the point at which the overlap converges to 1 in the large β limit.

Analogous considerations would arise in the teacher-student version of the classification problem, with the added difficulty of having to deal with at least 3 order parameters.