Institut for Theoretical Physics
University Leipzig

Prof. Dr. B. Rosenow
A. Afanah , M. Staats

Statistical Mechanics of Deep Learning - Problem set 6

Winter Term 2023/24

Hand in: Friday, 24.11 at 09:00 am, you can upload your solutions to the course webpage on Moodle platform.

12. Information Gain

3+2 Points

Consider the expression for the so-called the average information gain " ΔI

$$\Delta I = S_{p-1} - S_p = -\frac{1}{N}\frac{dS(\alpha)}{d\alpha}$$
$$= -2\int Dt\ H\left(-\sqrt{\frac{R}{1-R}}\ t\right)\ \ln H\left(-\sqrt{\frac{R}{1-R}}\ t\right)$$

which describes the decrease of the version space volume induced by the $p-$th example in the training set. The relative reduction of the version space corresponds to the additional information gained about the teacher couplings. Hence, ΔI decreases as α increases, See Fig.(2.3) in Engel A. book " Statistical Mechanics of Learning".

(a) Show that the asymptotic behavior of ΔI in the large α limit is $\Delta I \sim \frac{const}{\alpha}$.
Hint : You may use the results we obtained from problem (10), i.e. $1 - R \sim \frac{1.926}{\alpha^2}$

(b) Why does ΔI decrease with increasing α? Could you think of a learning scenario that ensures constant information gain for all examples? What kind of scaling of ε with α would you expect for such a case?

13. Version Space

6 Points

The space of all student coupling vectors $\boldsymbol{J}$ which score on the examples exactly like the teacher is called the version space. The precise behavior on how the version space shrinks as we increase the number of examples can be calculated, but the task of this exercise is to simulate the behavior. We consider a simple teacher perceptron T which lies on the N-dimension sphere, i.e. is a normalized N-dimensional vector. The examples ξ^μ are also normalized vectors from the N-dimensional sphere, such that the labels σ^μ of the examples can be computed as

$$\sigma^\mu = \text{sign}(\xi^\mu \cdot T) \tag{1}$$

(a) Draw a random vector T in $N = 8$ dimensions that defines your teacher for this exercise. Now draw $p = 10$ vectors uniformly from the N-dimensional unit sphere that define your example set. Calculate the labels using Eq.1. Now estimate percentage that the version space takes from the whole sphere using Monte-Carlo sampling with 10^6 points. Monte-Carlo sampling works by drawing vectors repeatedly, if the drawn vector agrees with the teacher we increase the version space vectors count, otherwise we increase the count of the rejected vectors. Using these two values, give an estimate for the percentage of the version space vectors.

(b) Revisit your code and try to optimize it using matrix multiplication. Compute for $N = 8$ and $p = [4, 8, 16, 32]$ the percentage of the version space vectors and plot the result as a function of p. If your code is optimized, 10^7 Monte-Carlo samples for each set should run in less than a minute.