Institut for Theoretical Physics
University Leipzig

Prof. Dr. B. Rosenow
M. Kühn

Statistical Mechanics of Deep Learning - Problem set 7

Winter Term 2024/25

Hand in Python code: Before **Monday 02.12.2024, 9:15**, only submit the Python code you have written. Share a Google Colab Notebook with your code and send the link via email to itpleipzig@gmail.com.

14. The Gardner Analysis

3+3+2 Points

(a) Determine the asymptotic behaviour of the quantities describing Gibbs learning for large training set size α. To this end, start from

$$\frac{R}{\sqrt{1-R}} = \frac{\alpha}{\pi} \int \boldsymbol{D}t \, \frac{\exp(-Rt^2/2)}{H(-\sqrt{R}t)}$$

and show that

$$1 - R \sim \left[\frac{\alpha}{\pi} \int \boldsymbol{D}t \, \frac{\exp(-t^2/2)}{H(-t)} \right]^{-2} \sim \frac{1.926}{\alpha^2}$$

Hint : You may use numerical methods to compute the final result.

(b) Use the result you obtain from part (a) to show that

$$\varepsilon \sim \frac{\sqrt{2}}{\int \boldsymbol{D}t \, [\exp(-t^2/2)]/H(t)} \frac{1}{\alpha} \sim \frac{0.625}{\alpha},$$

as follows from

$$\varepsilon = \frac{1}{\pi} \arccos(R)$$

(c) Show also that the quenched entropy $S \sim \frac{1}{2}\ln(1-R)$ for large α is

$$S \sim -\ln(\alpha).$$

15. 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 $\boldsymbol{\xi}^\mu$ are also normalized vectors from the N-dimensional sphere, such that the labels σ^μ of the examples can be computed as

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

(a) Draw a random vector $\boldsymbol{T}$ 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 the 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.