Institut for Theoretical Physics
University Leipzig

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

Statistical Mechanics of Deep Learning - Problem set 4

Winter Term 2023/24

Hand in: Friday, 10.11 at 09:00 am, you can upload your solutions to the course webpage on Moodle platform.

8. Volume of version space with spherical constraints *4+4 Points*

(a) Show that the surface of the unit sphere, occupied by vectors $\boldsymbol{J}$ with the normalization constraint $\boldsymbol{J}^2 = N$, is for large N to leading order given by

$$\Omega = \int d\boldsymbol{J}\ \delta(\boldsymbol{J}^2 - N) \sim \exp\left(\frac{N}{2}[1 + \ln 2\pi]\right).$$

Hint : start by introducing the integral representation of the delta function

$$\delta(x - a) = \int \frac{\hat{x}}{2\pi}\ e^{i\hat{x}(x-a)}$$

then compute the gaussian integral over $\boldsymbol{J}$ using

$$\int dx e^{-ax^2} = \sqrt{\frac{\pi}{a}}.$$

To obtain the final expression, use the saddle point method, see appendix A1.4 in A. Engel book "Statistical Mechanics of Learning".

(b) Evaluate the volume of version space defined by all vectors $\boldsymbol{J}$ that have an angle θ with a given direction defined by the vector $\boldsymbol{T}$, i.e. show that in the limit $N \to \infty$

$$\Omega_0(\varepsilon) = \int d\boldsymbol{J}\ \delta(\boldsymbol{J}^2 - N)\ \delta\left(\frac{\boldsymbol{J}\boldsymbol{T}}{N} - \cos(\pi\varepsilon)\right) \sim \exp\left(\frac{N}{2}\left[1 + \ln 2\pi + \ln \sin^2(\pi\varepsilon)\right]\right).$$

9. Gaussian joint probability density function *4+4 Points*

Consider the auxiliary variables

$$\lambda_\mu = \frac{1}{\sqrt{N}}\boldsymbol{J}\boldsymbol{\xi}^\mu\ ,\ u_\mu = \frac{1}{\sqrt{N}}\boldsymbol{T}\boldsymbol{\xi}^\mu,$$

where $\boldsymbol{J}$ and $\boldsymbol{T}$ are the student and teacher vectors as introduced in the lectures, while, $\boldsymbol{\xi}^\mu \in \mathcal{R}^N$ are the example vectors with components ξ_i^μ drawn from the distribution

$$P(\boldsymbol{\xi}) = \prod_j^N \left[\frac{1}{2}\delta(\xi_j + 1) + \frac{1}{2}\delta(\xi_j - 1)\right]$$

.

(a) Show that the joint probability density $P(\lambda, u)$ is indeed a Gaussian probability density . Start from

$$P(\lambda, u) = \left\langle \left\langle \delta\left(\lambda - \frac{1}{\sqrt{N}}\boldsymbol{J\xi}\right)\delta\left(u - \frac{1}{\sqrt{N}}\boldsymbol{T\xi}\right)\right\rangle\right\rangle_{\boldsymbol{\xi}}$$

The average in $P(\lambda, u)$ is with respect to a randomly chosen example $\boldsymbol{\xi}$.
Hint : you may use the Hubbard-Stratonovich transformation

$$\int Dt\; e^{bt} = e^{b^2/2}$$

$$\text{where } Dt := \frac{dt}{\sqrt{2\pi}}\exp(-t^2/2)$$

(b) Show that the distribution $P(\lambda, u)$ has the moments

$$\langle\langle\lambda\rangle\rangle = \langle\langle u\rangle\rangle = 0$$

$$\langle\langle\lambda^2\rangle\rangle = \langle\langle u^2\rangle\rangle = 1$$

$$\langle\langle\lambda u\rangle\rangle = \frac{\boldsymbol{JT}}{N} = R.$$