Saarland University
Faculty of Mathematics and Computer Science
Department of Mathematics
Prof. Dr. Roland Speicher
Dr. Johannes Hoffmann

Exercises for the lecture
High Dimensional Analysis: Random Matrices and Machine Learning
Summer term 2023
Sheet 6
Hand-in: Friday, 14.07.2023, 22:00 Uhr via CMS

Exercise 1 (5 + 5 points)**.**

(a) Let t be Poisson-distributed with rate $\lambda > 0$, i.e. t is a discrete random variable supported on $\mathbb{N}_0$ with distribution

$$\mathrm{P}(t = k) = \frac{\lambda^k \exp(-\lambda)}{k!}.$$

Compute the cumulants of t using their definition as coefficients in the logarithm of the characteristic function.

(b) Let t be χ^2-distributed with $k \in \mathbb{N}$ degrees of freedom, i.e. $t = \sum_{j=1}^{k} x_j^2$, where the $x_j \sim N(0, 1)$ are independent. Compute the cumulants of t using Theorem 7.13.

Exercise 2 (10 points)**.** Let $\{\alpha_n\}_{n\in\mathbb{N}}$ and $\{\kappa_n\}_{n\in\mathbb{N}}$ be two sequences that satisfy the relation

$$\alpha_n = \sum_{\pi \in \mathcal{P}(n)} k_\pi,$$

where $\kappa_\pi = \kappa_1^{r_1} \cdot \ldots \cdot \kappa_n^{r_n}$ and r_j is the number of blocks of π of size j. We want to show that, as formal power series,

$$\log\left(1 + \sum_{n=1}^{\infty} \alpha_n \frac{z^n}{n!}\right) = \sum_{n=1}^{\infty} \kappa_n \frac{z^n}{n!}. \tag{1}$$

(a) Show that by differentiating both sides of (1) it suffices to prove

$$\sum_{n=0}^{\infty} \alpha_{n+1} \frac{z^n}{n!} = \left(1 + \sum_{n=1}^{\infty} \alpha_n \frac{z^n}{n!}\right) \sum_{n=0}^{\infty} \kappa_{n+1} \frac{z^n}{n!}. \tag{2}$$

(b) By grouping the terms in $\sum_{\pi\in\mathcal{P}(n)} k_\pi$ according to the size of the block containing 1, show that

$$\alpha_n = \sum_{\pi \in \mathcal{P}(n)} k_\pi = \sum_{m=0}^{n-1} \binom{n-1}{m} \kappa_{m+1} \alpha_{n-m-1}.$$

(c) Use the result of (b) to prove (2).

please turn over

Exercise 3 (5 + 5 + 5 + 5 points). We consider, for $p = 1$, our 1 hidden layer neural network of width m,

$$f_m(x) = \frac{1}{\sqrt{m}} a^T \sigma(bx + c),$$

where a, b and c are independent standard Gaussian random vectors in $\mathbb{R}^m$. (Note that we include here also a bias c in the argument of σ). We want to use this to learn the function $g : \mathbb{R} \to \mathbb{R}$ given by

$$g(x) = \sqrt{|x|} + \sin(10x),$$

restricted to the interval $[-1, 1]$.

Choose randomly 15 data points x_i, drawn from the uniform distribution on the interval $[-1, 1]$, and let $y_i := g(x_i)$. From this data we try to recover g: Use gradient descent to train the parameters $\{a, b\}$ (we don't train the bias c, but keep this fixed) with respect to the loss function

$$\mathcal{L}(a, b) = \frac{1}{2} \sum_{i=1}^{15} \big(y_i - f_m(x_i)\big)^2,$$

for varying widths m. It is actually advisable to use *stochastic gradient descent*; that is, in each step one uses only the gradient of $\big(y_i - f_m(x_i)\big)^2$, with respect to a and to b, for a randomly chosen i. Train until the loss function is less than 0.01 (in the case $m > 15$) or until it does not decrease any more (in the case $m \leq 15$). Plot then the trained function $f_m(x)$ against the target function $g(x)$ for 2000 points x sampled evenly from the interval $[-1, 1]$, for $m \in \{1, 2, 5, 10, 15, 30, 100, 500\}$. Show also the 15 data points $(x_i, g(x_i))$ in this plot. As learning rate you might choose any $\eta \in (0.001, 0.01)$.

(a) Do this for $\sigma(x) = \sin(8x)$.

(b) Do this for $\sigma = \text{ReLU}$.

(c) Check in those cases also what happens if you switch off the bias (i.e., put $c = 0$).

(d) Explain why it is a bad idea to switch off the bias in the case of $\sigma(x) = \sin(8x)$. Explain why it is an even worse idea to do this in the case of $\sigma = \text{ReLU}$.