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 5
Hand-in: Friday, 23.06.2023, 22:00 Uhr via CMS

Recall that the ReLU function is defined as $\mathrm{ReLU}(t) = \max(0, t)$.

Exercise 1 (10 points). We now investigate a one-layer perceptron with random features and n parameters: given an input $x \in \mathbb{R}$, the neural network computes $y = w\sigma(ax + b)$, where

- $a \sim N(0, I_n)$ is the *weight* and $b \sim N(0, I_n)$ is the *bias*,
- $\sigma : \mathbb{R} \to \mathbb{R}$ is the (non-linear) activation, applied component-wise,
- $w \in \mathbb{R}^{1\times n}$ is the linear regression of the training data.

Consider the following eleven (training) data points:[1]

x_k	-5	-4	-3	-2	-1	0	1	2	3	4	5
y_k	-3	-3	-4	1	-0.2	0.1	2	1.8	1.9	-0.2	2

For each $n \in \{5, 10, 11, 30, 300, 1000\}$ and each $\sigma \in \{\mathrm{ReLU}, \sin\}$ do the following:

(a) For two d-dimensional standard Gaussian vectors $a, b \sim N(0, I_n)$, compute the feature matrix

$$F = \begin{pmatrix} f_1 & \dots & f_{11} \end{pmatrix} \in \mathbb{R}^{n\times 11}, \quad \text{where} \quad f_k = \sigma(a \cdot x_k + b).$$

(b) Perform linear regression on the so-obtained features in order to fit the data given above: $w = YF^T(FF^T)^+$, where $Y = \begin{pmatrix} y_1 & \dots & y_{11} \end{pmatrix} \in \mathbb{R}^{1\times 11}$ and A^+ is the pseudo-inverse of A.

(c) Plot the output of your neural network on the grid from -5 to 5 with step size 0.1. For comparison, also plot the original data points. It suffices to hand in the plots, no need to print out all the intermediate data.

Compare the plots and describe what you see. This is an instance of the so-called double-descent!

please turn over

[1]Copy-friendly version of the y_k: `[-3, -3, -4, 1, -0.2, 0.1, 2, 1.8, 1.9, -0.2, 2]`

Exercise 2 (7 points). Consider the entries x_{ij} of our matrix $X = (x_{ij}) \in \mathbb{R}^{p\times n}$ as formal variables. For fixed $z \in \mathbb{C}$, we put

$$R = R(z) = \left(\frac{1}{n}XX^T - zI_p\right)^{-1} \in \mathbb{R}^{p\times p}.$$

Show that we have

$$\left[\frac{\partial R}{\partial x_{ij}}\right]_{kl} = -\frac{1}{n}\Big(R_{ki}[X^T R]_{jl} + [RX]_{kj}R_{il}\Big).$$

Exercise 3 ((3 + 4) + (2 + 3) points). For a function $\sigma : \mathbb{R} \to \mathbb{R}$ we denote

$$\theta_1(\sigma) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \sigma(t)^2 \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t$$

and

$$\theta_2(\sigma) := \left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \sigma'(t) \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t\right)^2 = \left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t\sigma(t) \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t\right)^2.$$

(a) Let $\sigma : \mathbb{R} \to \mathbb{R}$ be such that $\theta_1(\sigma)$ and $\theta_2(\sigma)$ are finite.
   (i) Show that $\theta_2(\sigma) \le \theta_1(\sigma)$.
   (ii) Show that $\theta_2(\sigma) = \theta_1(\sigma)$ if and only if σ is a linear function, i.e., $\sigma(t) = \beta t$ for some $\beta \in \mathbb{R}$.

(b) Let $\alpha \in \mathbb{R}$ be a constant and consider the shifted ReLU function

$$\sigma(t) = \mathrm{ReLU}(t) - \alpha.$$

   (i) Determine α such that

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \sigma(t) \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t = 0.$$

   (ii) Determine for this σ the quantities $\theta_1(\sigma)$ and $\theta_2(\sigma)$.

please turn over

Exercise 4 ((4 + 4) + 3 points). Like in class, consider standard Gaussian random matrices $X \in \mathbb{R}^{p\times n}$ and $W \in \mathbb{R}^{p\times p}$ together with a non-linearity $\sigma : \mathbb{R} \to \mathbb{R}$. Let

$$F := \sigma\left(\frac{1}{\sqrt{p}}WX\right) \in \mathbb{R}^{p\times p} \quad \text{and} \quad M := \frac{1}{n}FF^T \in \mathbb{R}^{p\times p}.$$

(a) Consider $\sigma_1(t) = t^2 - 1$ and $\sigma_2(t) = t^3 - 3t$. For each $\sigma \in \{\sigma_1, \sigma_2\}$ do the following:

   (i) Compute $\theta_1(\sigma)$ and show that $\theta_2(\sigma) = 0$.

   (ii) For $p = 2000$ and each $\gamma \in \left\{1, \frac{1}{2}, \frac{1}{4}\right\}$, draw a diagram including a histogram of the eigenvalues of M and the corresponding Marchenko-Pastur distribution. Re-scale σ such that the distribution matches the histogram.

(b) From class we know that in general, F behaves like

$$\tilde{F} = \frac{\sqrt{\theta_2}}{\sqrt{p}}WX + \sqrt{\theta_1 - \theta_2}Z$$

for (independent) standard Gaussian matrices $W \in \mathbb{R}^{p\times p}$ and $X, Z \in \mathbb{R}^{p\times n}$. For $\sigma(t) = \mathrm{ReLU}(t) - \alpha$ from the previous exercise, compare a histogram of the eigenvalues of M with a histogram of the eigenvalues of $\tilde{M} := \frac{1}{n}\tilde{F}\tilde{F}^T$. Again, use $p = 2000$ and consider each $\gamma \in \left\{1, \frac{1}{2}, \frac{1}{4}\right\}$.