Exercise Sheet #1

Due: 23.10.2023 11h00

Jan Pawlowski, Tilman Plehn
Jonas Spinner

General Regulations.

- Please hand in your solutions in groups of three people. A mix of attendees from different tutorials is fine.
- Your solutions to theoretical exercises can be either handwritten notes (scanned), typeset using LaTeX, or directly in the jupyter notebook using Markdown.
- For the practical exercises, the data and a skeleton for your jupyter notebook are available at https://github.com/heidelberg-hepml/mlph2023-Exercises. Always provide the (commented) python code as well as the output, and don't forget to explain/interpret the latter, we do not give points for code that does not run. Please hand in both the notebook (`.ipynb`) and an exported pdf.
- Submit all your files in the Übungsgruppenverwaltung, only once for your group of three.

1 KL-divergence

Compute the KL-divergence of two Gaussians

$$D_{\mathrm{KL}}\Big[\mathcal{N}_{\mu_1,\sigma_1}(x), \mathcal{N}_{\mu_2,\sigma_2}(x)\Big], \qquad \mathcal{N}_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big). \tag{1}$$

What happens in the case $\mu_1 = \mu_2, \sigma_1 = \sigma_2$? (2 pts)

2 Random numbers

(a) Given two probability density functions f and g and a sample $r \sim f$, one can generate another sample $x \sim g$ using the relation

$$x = G^{-1}(F(r)), \tag{2}$$

where F, G are the cumulative density functions of f, g. Derive this relation. (1 pts)

(b) Use Eq. 2 to generate random numbers from the distributions g_i using $r \sim \mathrm{Uniform}[0,1]$. If it is not possible, explain why.

- $g_1(x) = 3x^2, x \in [0,1]$
- $g_2(x) = e^{-x}, x > 0$
- $g_3(x) = \mathcal{N}_{0,1}(x)$
- $g_4(x) = \frac{2}{3\zeta_3}\frac{x^2}{e^x+1}, x > 0$

Hint: `scipy.special.erfinv`, `scipy.special.zeta` (2 pts)

(c) Repeat exercise (b) using rejection sampling. Restrict the region to $x < |5|$ whenever it extends to infinity. Compare the speed to your implementation from (b). Discuss pros and cons of this method, compared to the method from (a) and (b). (2 pts)

(d) These methods for generating gaussian random numbers are too slow. Can you do better? Hint: Box-Muller

(1 pts)

3 Neyman-Pearson Lemma

The Neyman-Pearson Lemma is a result from hypothesis tests that will show up a lot when constructing losses for neural networks. The goal of this exercise is to prove the Neyman-Pearson Lemma.

We compare two hypotheses, the null hypothesis H_0 and an alternative hypothesis H_1. They are quantified through statistical models $p(x|H_0), p(x|H_1)$. We use the test statistic $t(x)$ to accept/reject points if $t(x) < t_0 / t(x) > t_0$ with some cutoff t_0. The rejection region ω is the set of all rejected points $\omega = \{x|t(x) > t_0\}$. This allows us to define two basic properties of a hypothesis test, the size $p(\omega|H_0)$ and the power $p(\omega|H_1)$.

The Neyman-Pearson Lemma states that the likelihood ratio $t_{\mathrm{LR}}(x)$ is the optimal test statistic. This means that $t(x)$ has maximal power $p(\omega|H_1)$ for a test with given size $p(\omega|H_0)$. The likelihood ratio is defined as

$$t_{\mathrm{LR}}(x) = \frac{p(x|H_1)}{p(x|H_0)}. \tag{3}$$

Prove the Neyman-Pearson Lemma by showing that any modification of the rejection region does not increase the power of the test.
Hint: Modify the rejection region by adding A while removing B, where the regions A and B have to be chosen such that the size of the test remains constant. Then look at the change in power $p(A|H_1) - p(B|H_1)$ that this modification gives. (2 pts)