Math of ML : Exercises 7 *

November 3, 2025

Before attempting the exercises, read the following blog post by Francis Bach: https://francisbach.com/quest-for-adaptivity/.

In this exercise sheet, let $\tau \in \mathcal{P}(\mathcal{S})$ be the law of the first neural network layer initialization weights (w_j, b_j), where $\mathcal{S} \subseteq \mathbb{R}^d \times \mathbb{R}$ is the support of τ. Further, let σ be the activation function and let the input space be $\mathcal{X}_r = \{x \in \mathbb{R}^d : \|x\|_2 \leq r\}$ for some constant $r > 0$.

Let $\mathcal{M}(\mathcal{S})$ denote the set of signed measures on $\mathcal{S}$ equipped with Borel sigma algebra. The total variation norm of a measure $\mu \in \mathcal{M}(\mathcal{S})$ is defined by

$$\|\mu\|_{\mathrm{TV}} = \sup_{|f| \leq 1} \int_{\mathcal{S}} f(w,b) d\mu(w,b),$$

where the above supremum runs over all measurable functions f.

Following the 6-th week lecture notes, define the normed function spaces $(\mathcal{F}_1, \|\cdot\|_{\mathcal{F}_1})$ and $(\mathcal{F}_2, \|\cdot\|_{\mathcal{F}_2})$ by

$$\|f\|_{\mathcal{F}_1} = \inf_{\mu \in \mathcal{M}(\mathcal{S})} \left\{ \|\mu\|_{\mathrm{TV}} : \forall x \in \mathcal{X}_r \; f(x) = \int_{\mathcal{S}} \sigma(w^\top x + b) d\mu(w,b) \right\} \tag{1}$$

and

$$\|f\|_{\mathcal{F}_2}^2 = \inf_{\eta \in L_2(\tau)} \left\{ \int_{\mathcal{S}} \eta(w,b)^2 d\tau(w,b) : \forall x \in \mathcal{X}_r \; f(x) = \int_{\mathcal{S}} \eta(w,b)\sigma(w^\top x + b) d\tau(w,b) \right\}. \tag{2}$$

Exercise 1 (On the total variation norm)**.**

1. *Let $\mu \in \mathcal{M}(\mathcal{S})$ be given by the finite combination of atoms $\mu = \sum_{j=1}^n \eta_j \delta_{(w_j,b_j)}$, where $\delta_{(w_j,b_j)}$ denotes the Dirac measure. Prove that $\|\mu\|_{\mathrm{TV}} = \sum_{j=1}^n |\eta_j|$.*

2. *Suppose $\mu \in \mathcal{M}(S)$ has a density η with respect to the measure τ. Prove that $\|\mu\|_{\mathrm{TV}} = \int_{\mathcal{S}} |\eta(w,b)| d\tau(w,b)$.*

3. *Define the normed space $(\mathcal{F}_1', \|\cdot\|_{\mathcal{F}_1'})$ by*

$$\|f\|_{\mathcal{F}_1'} = \inf_{\eta \in L_1(\tau)} \left\{ \int_{\mathcal{S}} |\eta(w,b)| d\tau(w,b) : \forall x \in \mathcal{X}_r \; f(x) = \int_{\mathcal{S}} \eta(w,b)\sigma(w^\top x + b) d\tau(w,b) \right\}.$$

   Suppose that $f \in \mathcal{F}_1'$. Prove that $f \in \mathcal{F}_1$ and $\|f\|_{\mathcal{F}_1} \leq \|f\|_{\mathcal{F}_1'}$.

Exercise 2 (On spaces $\mathcal{F}_1$ and $\mathcal{F}_2$)**.** *In this exercise, we explore some basic differences between the spaces $\mathcal{F}_1$ and $\mathcal{F}_2$.*

1. *Explain briefly why $(\mathcal{F}_2, \|\cdot\|_{\mathcal{F}_2})$ is a Hilbert space.*

*Lénaïc Chizat EPFL lenaic.chizat@epfl.ch

2. *Let* $\mathcal{S} = [-1, 1] \times \{0\}$ *and let* $\sigma(x) = \max(0, x)$ *be the relu activation. Consider the functions* $f_1 = \sigma(x)$ *and* $f_2 = -\sigma(-x)$. *Prove that* $\|f_1 + f_2\|_{\mathcal{F}_1} \geq 2$ *and* $\|f_1 - f_2\|_{\mathcal{F}_1} \geq 2$. *Deduce that the space* $(\mathcal{F}_1, \|\cdot\|_{\mathcal{F}_1})$ *is not a Hilbert space.*

   Hint: *for* $(\mathcal{F}_1, \|\cdot\|_{\mathcal{F}_1})$ *to be a Hilbert space it is necessary that the parallelogram law holds for the norm* $\|\cdot\|_{\mathcal{F}_1}$.

3. *Prove that* $\mathcal{F}_2 \subseteq \mathcal{F}_1$.

4. *Let* m *be a finite natural number and for* $j = 1, \dots, m$ *let* $(w_j, b_j) \in \mathcal{S}$. *Prove that the function* f *defined by a finite combinations of neurons* $f(x) = \sum_{j=1}^{m} \eta_j \sigma(w_j^\top x + b_j)$ *belongs to the space* $\mathcal{F}_1$. *Give an upper bound on its* $\mathcal{F}_1$ *norm.*

   Remark: *typically, even a single neuron – a function of the form* $x \mapsto \sigma(w^\top x + b)$ *– is not a member of the space* $\mathcal{F}_2$. *However, this fact is non-trivial to prove.*