# Mathematical Foundations of Machine Learning, Spring 2022: Assignment I

1. [Minkowski integral inequality] Prove that for $1 \le p < \infty$ and a measurable function $F(x,t): \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$

$$\left( \int_{\mathbb{R}^m} \left( \int_{\mathbb{R}^n} |F(x,t)| dx \right)^p dt \right)^{1/p} \le \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^m} |F(x,t)|^p dt \right)^{1/p} dx .$$

**Hints:** for $1 < p < \infty$ $\left( \int_{\mathbb{R}^n} |F(x,t)| dx \right)^p = \left( \int_{\mathbb{R}^n} |F(x,t)| dx \right) \left( \int_{\mathbb{R}^n} |F(y,t)| dy \right)^{p-1}$, change order of integration of $t$ and $x$, use Hölder.

2. Let $f(x) := \sum_{m=1}^{M} c_m \mathbf{1}_{[2m,2m+1]}(x)$. Compute the modulus $\omega_1(f,t)_p$, for all $0 < t < 1/2$, and $0 < p \le \infty$.

3. Prove the following equality for any $N \ge 1$, $x, h \in \mathbb{R}^n$, $f: \mathbb{R}^n \to \mathbb{R}$,

$$\Delta_{Nh}^r(f,x) = \sum_{k_1=0}^{N-1} \cdots \sum_{k_r=0}^{N-1} \Delta_h^r(f, x + k_1 h + \ldots + k_r h).$$

**Hint:** recall we proved in class for $r = 1$. Now apply induction on $r$. <u>Make sure the notations are correct</u>.

4. Recall that a function $g \in L_1(\mathbb{R}^n)$ is the **distributional derivative** of $f \in L_1(\mathbb{R}^n)$, $g := \partial^\alpha f$, $\alpha \in \mathbb{Z}_+^n$, if

$$\int_{\mathbb{R}^n} g\phi = (-1)^{|\alpha|} \int_{\mathbb{R}^n} f \partial^\alpha \phi , \qquad \forall \phi \in C_0^\infty(\mathbb{R}^n) .$$

Prove $H'(x) = \begin{cases} 1, & -1 \le x < 0, \\ -1, & 0 \le x \le 1, \\ 0, & else. \end{cases}$ where $H(x) := \begin{cases} x+1, & -1 \le x < 0, \\ 1-x, & 0 \le x \le 1, \\ 0, & else. \end{cases}$

5. In class we proved that for $g \in W_p^r(\mathbb{R}) \cap C^r(\mathbb{R})$, $\omega_r(g,t)_{L_p(\mathbb{R})} \le t^r |g|_{W_p^r(\mathbb{R})}$. Use the density of $W_p^r(\mathbb{R}) \cap C^r(\mathbb{R})$ in $W_p^r(\mathbb{R})$ to prove this inequality for any $f \in W_p^r(\mathbb{R})$, $1 \le p < \infty$.

**Hint** By definition there exists a sequence $\{g_k\}$, $g_k \in W_p^r(\mathbb{R}) \cap C^r(\mathbb{R})$, $g_k \underset{W_p^r}{\to} f$.

6. [“Continuous” variance] Let $f: [0,1]^n \to \mathbb{R}^L$ and let $\Omega \subseteq [0,1]^n$. Prove that minimizing the variance over partitions $\Omega' \cup \Omega'' = \Omega$,

$$V_\Omega := \int_{\Omega'} \left| \vec{f}(x) - \vec{E}_{\Omega'} \right|_{l_2(\mathbb{R}^L)}^2 dx + \int_{\Omega''} \left| \vec{f}(x) - \vec{E}_{\Omega''} \right|_{l_2(\mathbb{R}^L)}^2 dx ,$$

is equivalent to maximizing the wavelet norms

$$\|\psi_{\Omega'}\|^2 + \|\psi_{\Omega''}\|^2 ,$$

where $\vec{E}_{\Omega'} = \frac{1}{|\Omega'|} \int_{\Omega'} \vec{f}(x) dx$, $\quad \|\psi_{\Omega'}\|_2 = |\Omega'|^{1/2} \left| \vec{E}_{\Omega'} - \vec{E}_\Omega \right|_{l_2(\mathbb{R}^L)}$.