Mathematical Foundation of Machine Learning Spring 2025: Assignment II

1. [30%] Let $f:[0,1]^n \to \mathbb{R}^L$ and let $\Omega \subseteq [0,1]^n$. Prove that minimizing the variance over partitions $\Omega' \cup \Omega'' = \Omega$,

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

is equivalent to maximizing the wavelet norms

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

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

2. [40%] [Besov smoothness of indicator functions] Let $f(x) = 1_{\tilde{\Omega}}(x)$, where $\tilde{\Omega} \subset [0,1]^2$ is a domain with a smooth boundary.
   a. Prove that for an isotropic tree $\mathcal{T}_I$, that applies non-adaptive subdivisions at dyadic partitions of the two variables, we get that $f \in B^\alpha_\tau(\mathcal{T}_I)$, for $\alpha < 1/(2\tau)$.
   b. Assume that one can construct an anisotropic tree $\mathcal{T}_A$, such that from some level $k_0 \geq 1$, there are only $\leq c_1 2^m$ domains at the level $k_0 + 2m$ that intersect the boundary of $\tilde{\Omega}$, each of area $\leq c_2 2^{-3m}$. Prove that $f \in B^\alpha_\tau(\mathcal{T}_A)$, where $\alpha < 2/(3\tau)$. You can assume that there exists $0 < \rho < 1$, such that for any child Ω' of Ω, $|\Omega'| \leq \rho |\Omega|$.

Hints:

(i) For both cases you need to follow the proof of the case of the classic Besov smoothness with dyadic cubes. The case (a) is <u>very similar</u> to the classic case, because $B^{2\alpha}_\tau \sim B^\alpha_\tau(\mathcal{T}_I)$.

(ii) In both cases, you could first bound the contribution of the odd levels of the trees by the even levels. Then, it will be sufficient to only estimate the sum of contributions of even levels.

(iii) Use the ρ volume condition for (b) and the properties of the modulus. It is useful to observe that the condition $|\Omega'| \leq \rho |\Omega|$, also gives $|\Omega'| \geq (1-\rho)|\Omega|$.

3. [30%] [Convolutions] The convolution of $f, g \in L_1(\mathbb{R}^n)$ is defined by $f * g(x) := \int_{\mathbb{R}^n} f(x-y)g(y)dy$.

(i) Prove that $f * g \in L_1(\mathbb{R}^n)$,

(ii) Prove that $f * g = g * f$,

(iii) The Fourier Transform is defined by $\hat{f}(w) = \int_{\mathbb{R}^n} f(x) e^{-i\langle w,x \rangle} dx$, for $w \in \mathbb{R}^n$. Show that

$$(f * g)^\wedge(w) = \hat{f}(w)\hat{g}(w), \quad \forall w \in \mathbb{R}^n.$$

(iv) Let $f \in L_1(\mathbb{R}^2)$ be a piecewise constant function. Design a 'filter' $g \in L_\infty(\mathbb{R}^2)$, with support in $[-\varepsilon/2, \varepsilon/2]^2$, for some $\varepsilon > 0$, such that $f * g$ is 'significant' only in ε neighborhoods of points where f has 'almost' vertical edges.