Mathematical Foundation of Machine Learning Spring 2023: Assignment III

1. [40%] Prove that for any $\alpha < 1/\tau$, there exists a constant $c(\alpha,\tau) > 0$, such that

$$\left|\Delta_j\right|_{B_\tau^\alpha([0,1])} \le c 2^{j\alpha},$$

where Δ_j is the sawtooth function with 2^{j-1} teeth.

Hints/comments

a. For simplicity, you can prove the case $0 < \alpha < 2$, which allows you to use

$$\left|\Delta_j\right|_{B_\tau^\alpha([0,1])} = \left(\int_0^\infty \left(t^{-\alpha} \omega_2\left(\Delta_j, t\right)_\tau \right)^\tau \frac{dt}{t} \right)^{1/\tau}.$$

For higher values of α, the definition calls for higher orders of the modulus $r \ge \lfloor \alpha \rfloor + 1 > 2$.

b. It is convenient to split the integration in t to: $\int_0^\infty = \int_0^{2^{-(j+1)}} + \int_{2^{-(j+1)}}^\infty$.

c. One can also show the lower bound that gives the equivalence we stated in class $\left|\Delta_j\right|_{B_\tau^\alpha([0,1])} \sim 2^{j\alpha}$.

2. [10%] Try to generalize (as much as you can) the construction of the sawtooth example to functions that can be realized by a neural network with j blocks and have Besov smoothness $\sim 2^{j\alpha}$.

Comment There is no need for full proof of smoothness estimates for these functions, but ensure that your construction has certain aspects "under control" as $j \to \infty$, otherwise the smoothness may behave differently.

3. [20%] Let $\left\{x_i, f(x_i)\right\}_{i\in I}$ be a dataset with $x_i \in [0,1]^n$ and $f:[0,1]^n \to \mathbb{R}^L$. Let $\mathcal{F}$ be a forest constructed over this data. For any $m > 0$, let $\tilde{x}_i = (x_i, z_i) \in [0,1]^{n+m}$, $z_i \in \mathbb{R}^m$, $i \in I$ and $\tilde{f}:[0,1]^{n+m} \to \mathbb{R}^L$, defined by $\tilde{f}(\tilde{x}) := f(\tilde{x}_1,\ldots,\tilde{x}_n)$. Let $\tilde{\mathcal{F}}$ be the natural extension of $\mathcal{F}$ over $[0,1]^{n+m}$ using the same trees with the same subdivisions over the first n dimensions. Prove that $N_\tau\left(\tilde{f},\tilde{\mathcal{F}}\right) = N_\tau\left(f,\mathcal{F}\right)$, for any $\tau > 0$.

Hint/comment

a. The wavelets do change with the increase of the dimension. So, you need to show the invariance of the wavelet norms.
b. This shows some invariance of the sparsity/smoothness indicators under dimension embeddings. Moreover, if the impactful features are only a subset of lower dimension (not necessarily the first n features), then the adaptive(!) sparsity/smoothness indicators will only be determined by them.

4. [30%] Show that for any $0 < \varepsilon < 1$, there exists a neural network $\tilde{D}$ with $O\left(\log^2\left(\varepsilon^{-1}\right)\right)$ weights that approximates the function $D(x_1,x_2) = e^{x_1} x_2$, $\left\|D - \tilde{D}\right\|_{L_\infty\left([1,1]^2\right)} \le \varepsilon$.

Hint/comment

You may use the following: For any $n \geq 1$, there exist neural networks $D_{1,n}, D_{2,n}$, each with $O(n)$ weights, such that $\max_{-1 \leq x \leq 1} \left| e^x - D_{1,n}(x) \right| \leq c_1 e^{-n}$ and $\max_{-1 \leq x_1, x_2 \leq 1} \left| x_1 x_2 - D_{2,n}(x_1, x_2) \right| \leq c_2 e^{-\sqrt{n}}$.