STA3000F: Homework 2

Due: Nov 22 end of day on Quercus

Q1: VC dimension of two-layer neural networks

Given a pair of integers $m, d > 2$, consider the following class of two-layer neural networks that map $\mathbb{R}^d$ to binary labels

$$\mathcal{F} := \Big\{ x \mapsto \sigma\big(\sum_{i=1}^{m} a_i \sigma(w_i^\top x) \big) \ : \ w_i \in \mathbb{R}^d, a_i \in \mathbb{R}, \text{ for any } i = 1, 2, \cdots, m \Big\},$$

where we use a binary activation function $\sigma(x) := \mathbf{1}_{x \geq 0}$.

Prove that

$$\mathsf{VC}(\mathcal{F}) \leq cmd\big(\log m + \log d \big),$$

for some universal constant $c > 0$.

Q2: rate theorem with non-standard growth conditions

Following empirical process notations, we denote $Pf := \mathbb{E}_P[f(X)]$ and $P_n f := n^{-1} \sum_{i=1}^n f(X_i)$. Let us consider the M-estimator $\widehat{\theta}_n := \arg\min_{\theta \in K} P_n f_\theta$. We define the population-level minimizer $\theta^* := \arg\min_{\theta \in K} P f_\theta$. Suppose that

$$P f_\theta - P f_{\theta^*} \geq \|\theta - \theta^*\|_2^4, \quad \text{and} \quad \mathbb{E}\Big[\sup_{\substack{\theta \in K \\ \|\theta - \theta^*\|_2 \leq u}} |(P_n - P)(f_\theta - f_{\theta^*})| \Big] \leq \phi_n(u).$$

Find a high-probability bound on the convergence rate $\left\| \widehat{\theta}_n - \theta^* \right\|_2$, and prove your result. You may assume some growth condition on the function ϕ_n, and please specify such a condition used in your proof clearly.
[The answer should be expressed in terms of a fixed-point equation related to the function ϕ_n.]

Q3: classical Donsker's theorem

Let $(X_k : k \geq 0)$ be a one-dimensional simple random walk, i.e., $X_0 = 0$ and $X_{k+1} = X_k + \varepsilon_{k+1}$, where $(\varepsilon_k)_{k=1,2,\dots}$ are i.i.d. Rademacher random variables. Let $(B_t : t \geq 0)$ be a standard Brownian motion.[1] Prove that

$$\left(\frac{1}{\sqrt{n}} X_{\lfloor nt \rfloor} : 0 \leq t \leq T\right) \xrightarrow{d} (B_t : 0 \leq t \leq T),$$

for any $T > 0$.
[Hint: it suffices to verify finite-dimensional marginal convergence and stochastic equicontinuity. To verify the latter, we can use concentration inequalities discussed in previous lectures.]

[1] A standard Brownian motion is Gaussian process defined on $(t : t \geq 0)$ such that $\mathbb{E}[B_t] = 0$ and $\mathbb{E}[B_t B_s] = \min(t, s)$ for any $t, s \geq 0$.

Q4: fat-shattering dimension

Let $\mathcal{F}$ be the class of convex and 1-Lipschitz functions that maps $[0,1]$ to $[0,1]$.

1. Prove that $\mathsf{VC}(\mathcal{F}) = +\infty$.

2. For any $\varepsilon > 0$, prove an upper bound on $\mathsf{fat}_\varepsilon(\mathcal{F})$.

[Your grade will depend on how tight a bound you could get. For example, if the optimal bound is $\varepsilon^{-\alpha}$ but you get a correct proof of $\mathsf{fat}_\varepsilon(\mathcal{F}) \le \varepsilon^{-\beta}$ for some $\beta > \alpha$, you will get α/β fraction of the total marks. However, you do not need to justify the optimality of your bound.]