1. Proof of Theorem 2.1 in the paper "Classification algorithms using adaptive partitioning. For a given collection of sets $\mathcal{S}$ with VC-dimension $V = VC(\mathcal{S})$ let $f_S(x,y) := y(\chi_S(x) - \chi_{\Omega_\mathcal{S}}(x))$, $S \in \mathcal{S}$, where again $\Omega_\mathcal{S}$ is a best approximation to the Bayes set from $\mathcal{S}$. Since in these terms $\mathbb{E}(f_S) = \eta_S - \eta_{\Omega_\mathcal{S}}$, and $\frac{1}{n}\sum_{j=1}^n f_S(x_j, y_j) = \bar{\eta}_S - \bar{\eta}_{\Omega_\mathcal{S}}$, we need to estimate

$$\mathbb{P}\{\exists\, S \in \mathcal{S} : |\eta_S - \eta_{\Omega_\mathcal{S}} - (\bar{\eta}_S - \bar{\eta}_{\Omega_\mathcal{S}})| > e_n(S)\},$$

where

$$e_n(S) := e_n(S,r) := \sqrt{\rho_{S\Delta\Omega_\mathcal{S}}\varepsilon_n} + \varepsilon_n, \quad \varepsilon_n := \varepsilon_{n,r} := \frac{K\log n}{n}, \tag{1.1}$$

where $K := A\max\{r+1, V\}$. We want to show that the above probability is small provided A is chosen large enough.

We use the notation $\sigma_S^2 := \mathbb{E}_\rho(f_S^2)$, so that

$$\sigma_S^2 \le \rho_{S\Delta\Omega_\mathcal{S}}. \tag{1.2}$$

Rather than estimating the excess probability directly over all of $\mathcal{S}$ we first decompose the collection $\{f_S : S \in \mathcal{S}\}$ into the following slices. For any given $k = 1, \dots, n$, we define

$$\mathcal{S}_k := \{S \in \mathcal{S} : \varepsilon_n(k-1) \le \sigma_S^2 \le \varepsilon_n k\}, \tag{1.3}$$

Since $\varepsilon_n \ge \frac{1}{n}$, we have $\mathcal{S} = \mathcal{S}_1 \cup \dots \cup \mathcal{S}_n$. For later bounds (see (1.11) below) to remain well defined we remark that $\mathcal{S}_k = \emptyset$ for $k > n/(K\log n)$.

We now fix k and let

$$\mu := \varepsilon_n\sqrt{k}, \tag{1.4}$$

we observe that, by (1.2),

$$\begin{aligned} e_n(S) &= \sqrt{\rho_{S\Delta\Omega_\mathcal{S}}\varepsilon_n} + \varepsilon_n \ge \sigma_S\sqrt{\varepsilon_n} + \varepsilon_n \\ &\ge (\sqrt{(k-1)} + 1)\varepsilon_n \ge \mu, \quad S \in \mathcal{S}_k, \end{aligned} \tag{1.5}$$

which yields

$$\mathbb{P}\left\{\exists\, S \in \mathcal{S}_k : |\eta_S - \eta_{\Omega_\mathcal{S}} - (\bar{\eta}_S - \bar{\eta}_{\Omega_\mathcal{S}})| > e_n(S)\right\} \le \mathbb{P}\left\{\sup_{S\in\mathcal{S}_k} |\eta_S - \eta_{\Omega_\mathcal{S}} - (\bar{\eta}_S - \bar{\eta}_{\Omega_\mathcal{S}})| > \mu\right\}. \tag{1.6}$$

We define the random variable

$$\begin{aligned} Z(\mathbf{x}) &:= \sup_{S\in\mathcal{S}_k}\Big|\sum_{j=1}^n (f_S(x_j) - \mathbb{E}(f_S))\Big| \\ &= n\sup_{S\in\mathcal{S}_k}\big|\eta_S - \eta_{\Omega_\mathcal{S}} - (\bar{\eta}_S - \bar{\eta}_{\Omega_\mathcal{S}})\big|, \quad \mathbf{x} \in X^n, \end{aligned} \tag{1.7}$$

and note that

$$\sup_{S\in\mathcal{S}_k} \sigma_S \le \sqrt{\varepsilon_n k} = \sqrt{\frac{kK\log n}{n}} =: \sigma_k. \tag{1.8}$$

Since $\|f_S - \mathbb{E}(f_S)\|_{L_\infty} \le 2$, Talagrand's inequality as stated in Theorem 1.3 of [1], adapted to the present situation, asserts that

$$\mathbb{P}\{|Z - \mathbb{E}(Z)| > t\} \le C_0 \exp\Big\{ - c_0 t \log\Big(1 + \frac{2t}{nk\varepsilon_n + \mathbb{E}(Z)}\Big)\Big\}, \tag{1.9}$$

where c_0, C_0 are absolute constants. We next bound $\mathbb{E}(Z)$ by resorting to known bounds on expected sup-norms of empirical processes. Specifically, noting that

$$\|f_S\|_{L_\infty} \le 1, \quad |f_S(x,y)| \le \chi_{\Omega_{\mathcal{S}}}(x) \le 1, \quad \forall\ (x,y) \in X \times Y,\ S \in \mathcal{S}, \tag{1.10}$$

the bound from [2, (3.17),p. 46] yields

$$\mathbb{E}(Z) \le n\, C_1 \max\left\{ \sigma_k \sqrt{\frac{V\log(C_2\sigma_k^{-1})}{n}}, \frac{V\log(C_2\sigma_k^{-1})}{n} \right\}, \tag{1.11}$$

where C_1, C_2 are absolute constants. Observe that by (1.8) and (1.1), the first term on the right hand side of (1.11) exceeds the second one for each provided that $kK\log n/2V \ge \log\left(\frac{C_2^2 n}{kK\log n}\right)$, $k = 1,\dots,\lceil n/(K\log n)\rceil$. We now set $K = C_3 V$ and observe that by choosing A large we can attain any value of c_3. So the first term of the max in (1.11) is attained by the first term for all relevant k whenever

$$C_3 \ge C_2^2/V. \tag{1.12}$$

Under this condition we infer from (1.11) that

$$\mathbb{E}(Z) \le n\, C_1 \sqrt{k\varepsilon_n} \sqrt{\frac{V}{2n} \log\Big(\frac{C_2^2 n}{kC_3 V\log n}\Big)} =: nB_k \tag{1.13}$$

Therefore, returning to (1.9), we have for any $t \ge 2\mathbb{E}(Z)$

$$\begin{aligned} \mathbb{P}\{Z > t\} &\le \mathbb{P}\{|Z - \mathbb{E}(Z)| > t/2\} \\ &\le C_0 \exp\left\{ -c_0 \frac{t}{2} \log\left(1 + \frac{t}{nk\varepsilon_n + nB_k.}\right)\right\}. \end{aligned} \tag{1.14}$$

Recalling (1.4) and taking $t = n\mu = n\varepsilon_n\sqrt{k}$, we observe that $t \geq 2\mathbb{E}(Z)$ holds, by (1.13), whenever $\varepsilon_n\sqrt{k} \geq 2B_k$. In view of (1.13) and the definition of ε_n, this is indeed the case for all $k \leq \lceil n/(K\log n)\rceil$ whenever

$$\varepsilon_n = \frac{C_3 V\log n}{n} \geq \frac{2C_1^2 V}{n}\log\Big(\frac{C_2^2 n}{C_3 V\log n}\Big) \tag{1.15}$$

holds. This is certainly true when in addition to (1.12)

$$C_3 \geq 2C_1^2 \tag{1.16}$$

holds. Thus, (1.14) takes the form

$$\begin{aligned}\mathbb{P}\{Z > n\mu\} &\leq \mathbb{P}\{|Z-\mathbb{E}(Z)| > n\mu/2\}\\ &\leq C_0\exp\left\{-(c_0 n\mu/2)\log\left(1+\frac{\mu}{k\varepsilon_n+B_k}\right)\right\}.\end{aligned} \tag{1.17}$$

Since, as noted earlier, $\varepsilon_n\sqrt{k} \geq 2B_k$, the second term of the sum appearing in the denominator of the logarithm is smaller than the first one. Therefore, recalling (1.7),

$$\begin{aligned}\mathbb{P}\Big\{\sup_{S\in\mathcal{S}_k}\big|\eta_S-\eta_{\Omega_\mathcal{S}}-(\bar\eta_S-\bar\eta_{\Omega_\mathcal{S}})\big| > \mu\Big\} &\leq C_0\exp\Big\{-(c_0 n\mu/2)\log\Big(1+\frac{\mu}{2k\varepsilon_n}\Big)\Big\}\\ &\leq C_0\exp\Big\{-c_0\frac{n\mu^2}{4k\varepsilon_n}\Big\}\\ &\leq C_0\exp\Big\{-c_0\frac{n\varepsilon_n}{4}\Big\} \leq C_0 n^{-r-1},\end{aligned} \tag{1.18}$$

provided that

$$C_3 \geq \frac{4(r+1)}{c_0}\frac{4(r+1)}{c_0 V}. \tag{1.19}$$

The second inequality in (1.19) is obviously true since $V \geq 1$ and the first is true if C_3 (respectively A) is large enough. As we have already noted, every $S \in \mathcal{S}$ is in one of the $\mathcal{S}_k$. Therefore, using (1.6) and a union bound over $1 \leq k \leq \lceil n/(K\log n)\rceil \leq n$, collection the stipulations from (1.12), (1.16), (1.19), we arrive at the statement of the theorem with $\varepsilon_n = K\log n/n$ provided that for $K = C_3 V$

$$K \geq \max\Big\{\frac{4(r+1)}{c_0}, 2C_1^2 V, C_2^2\Big\}, \tag{1.20}$$

where c_0, C_1, C_2 are the constants from (1.11) and (1.9). □

References.

[1] P. Bartlett and S. Mendelson, *Empirical minimization*, Prob. Theory and Related Fields **135** (2003), 311-334.

[2] V. Kolchinskii, Oracle inequalities in empirical risk minimization and sparse recovery problems, Lecture Notes in Mathematics, **2033**, École d'Été de Probabilités de Saint Flour, Springer-Verlag, 2011.