A. Bayes near-optimality proof

Proof of Theorem 1. Since $\boldsymbol{\eta}$ is L-Lipschitz, given $x, x' \in \mathcal{X}$ we have

$$P(Y \neq Y' \mid x,x') = \sum_{j\in\mathcal{Y}} \boldsymbol{\eta}_j(x)(1-\boldsymbol{\eta}_j(x')) \tag{18}$$
$$\leq \sum_j \boldsymbol{\eta}_j(x)\left(1-\boldsymbol{\eta}_j(x)+Ld(x,x')\right)$$
$$= \sum_j \boldsymbol{\eta}_j(x)\left(1-\boldsymbol{\eta}_j(x)\right)+Ld(x,x').$$

By the definition of the nearest neighbor classifier g_{NN} in (6) we have $\mathbb{E}_S[P(g_{\mathrm{NN}}(X) \neq Y)] = \mathbb{E}_S[P(Y_{\pi_1(X)} \neq Y)]$, where the expectation is over the sample S determining g_{NN}. By (18) this error is bounded above by

$$\mathbb{E}_{S,X}[\sum_j \boldsymbol{\eta}_j(X)(1-\boldsymbol{\eta}_j(X))] + L\mathbb{E}_{S,X}[d(X, X_{\pi_1(X)})],$$

where now the expectation is over S and X. Denoting $k' = \mathrm{argmax}_j\, \boldsymbol{\eta}_j(X)$ and splitting the sum , the first term (which does not depend on S) satisfies

$$\mathbb{E}_X[\boldsymbol{\eta}_{k'}(X)(1-\boldsymbol{\eta}_{k'}(X))] + \mathbb{E}_X[\sum_{j\neq k'} \boldsymbol{\eta}_j(X)(1-\boldsymbol{\eta}_j(X))]$$
$$\leq \mathbb{E}_X[1-\boldsymbol{\eta}_{k'}(X)] + \mathbb{E}_X[\sum_{j\neq k'} \boldsymbol{\eta}_j(X)]$$
$$= 2\mathbb{E}_X[1-\boldsymbol{\eta}_{k'}(X)] = 2P(g^*(X) \neq Y).$$

It remains to bound $\mathbb{E}_{S,X}[d(X, X_{\pi_1(X)})]$ and we proceed exactly as in Ben-David & Shalev-Shwartz (2014). Let $\{C_1, \ldots, C_N\}$ be an ε-cover of $\mathcal{X}$ of cardinality $N = \mathcal{N}(\varepsilon, \mathcal{X}, d)$. Given a sample S, for $x \in C_i$ such that $S \cap C_i \neq \emptyset$ we have $d(x, X_{\pi_1(x)}) < \varepsilon$, while for $x \in C_i$ such that $S \cap C_i = \emptyset$ we have $d(x, X_{\pi_1(x)}) \leq \mathrm{diam}(\mathcal{X}) = 1$, thus $\mathbb{E}_{S,X}[d(X, X_{\pi_1(X)})]$ is bounded above by

$$\leq \mathbb{E}_S\left[\sum_{i=1}^N P(C_i)\left(\varepsilon\,\mathbb{1}_{\{S\cap C_i\neq\emptyset\}} + \mathbb{1}_{\{S\cap C_i=\emptyset\}}\right)\right]$$
$$= \sum_{i=1}^N P(C_i)\left(\varepsilon\,\mathbb{E}_S\left[\mathbb{1}_{\{S\cap C_i\neq\emptyset\}}\right] + \mathbb{E}_S\left[\mathbb{1}_{\{S\cap C_i=\emptyset\}}\right]\right).$$

Since $P(C_i)\mathbb{E}_S[\mathbb{1}_{S\cap C_i=\emptyset}] = P(C_i)(1-P(C_i))^n \leq 1/en$ and $N = \mathcal{N}(\varepsilon, \mathcal{X}, d)$ we get

$$\mathbb{E}_{S,X}[d(X, X_{\pi_1(X)})] \leq \varepsilon + \frac{\mathcal{N}(\varepsilon, \mathcal{X}, d)}{en}$$
$$\leq \varepsilon + \frac{1}{en}\left(\frac{2}{\varepsilon}\right)^D.$$

Setting $\varepsilon = 2n^{-\frac{1}{D+1}}$ concludes the proof. □

B. Rademacher analysis proofs

Proof of inequality (10). Dudley's chaining integral (Dudley, 1967) bounds from above the Rademacher complexity $\mathcal{R}_n(\mathcal{H}_L)$ by

$$\inf_{\alpha>0}\left(4\alpha + 12\int_\alpha^\infty \sqrt{\frac{\log \mathcal{N}(t, \mathcal{H}_L, \|\cdot\|_\infty)}{n}}dt\right).$$

By Lemma 2 the integral can be bounded as follows:

$$\int_\alpha^\infty \sqrt{\frac{\log \mathcal{N}(t, \mathcal{H}_L, \|\cdot\|_\infty)}{n}}dt$$
$$\leq \int_\alpha^\infty \sqrt{\frac{1}{n}\left(\frac{16L}{t}\right)^D \log\left(\frac{5k}{t}\right)}dt$$
$$\leq \int_\alpha^\infty \sqrt{\frac{\log 5k}{n}\left(\frac{16L}{t}\right)^D\left(\frac{1}{t}\right)}dt$$
$$= \sqrt{\frac{\log 5k}{n}}(16L)^{D/2}\int_\alpha^\infty\left(\frac{1}{t}\right)^{(D+1)/2}dt$$
$$= \sqrt{\frac{\log 5k}{n}}(16L)^{D/2}\left(\frac{2}{D-1}\right)\left(\frac{1}{\alpha^{(D-1)/2}}\right),$$

where in the second inequality we used the fact that for $x \in (0,1]$ and $c \geq e$ we have $\log(\frac{c}{x}) \leq \frac{\log c}{x}$. Choosing

$$\alpha^* = \left(9(16L)^D \frac{\log 5k}{n}\right)^{1/(D+1)}$$

yields the bound with

$$c_D = 16^{\frac{D}{D+1}}36^{\frac{1}{D+1}} + \frac{24}{(D-1)3^{\frac{D-1}{D+1}}}.$$

□

Proof of Theorem 4. An adaptation[4] of Mohri et al. (2012, Theorem 4.5) to $\mathcal{H}_L$ states that with probability $1-\delta$, for all $L > 0$, $h \in \mathcal{H}_L$,

$$\mathbb{E}[\mathcal{L}_{\text{margin}}(h)] \leq \widehat{\mathbb{E}}[\mathcal{L}_{\text{margin}}(h)] + 4\mathcal{R}_n(\mathcal{H}_L)$$
$$+ \sqrt{\left(\frac{\log\log_2 2L}{n}\right)_+} + \sqrt{\frac{\log\frac{2}{\delta}}{2n}}.$$

Since $\mathbb{1}_{\{u<0\}} \leq \mathcal{L}_{\text{margin}}(u)$ we have $P(g_h(X) \neq Y) \leq \mathbb{E}[\mathcal{L}_{\text{margin}}(h)]$. Since $\mathcal{L}_{\text{margin}}(u) \leq \mathcal{L}_{\text{cutoff}}(u)$ we can replace $\mathcal{L}_{\text{margin}}$ in the empirical loss by the loss function $\mathcal{L}_{\text{cutoff}}$. Bounding $\mathcal{R}_n(\mathcal{H}_L)$ using (10) concludes the proof. □

[4] essentially setting $\alpha = 1$ in Mohri et al. (2012) and doing the stratification on L instead

C. Scale sensitive analysis proof

Proof of Theorem 5. An application[5] of Guermeur (2010, Theorem 1) states that with probability $1-\delta$, for all $L>0$, $h \in \mathcal{H}_L$,

$$P(g_h(X) \neq Y) \leq \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{h(X_i,Y_i)<1\}}$$

$$+ \quad \sqrt{\frac{2}{n}\left(2\log\mathcal{N}\left(1/4, \mathcal{H}_L, \|\cdot\|_\infty\right) + \ln\left(\frac{2L}{\delta}\right)\right)} + \frac{1}{n}.$$

Applying the metric entropy bound in Lemma 2 proves the Theorem. □

D. Approximate NN proofs

First, we will show that $\tilde{h}$ is indeed a 2η additive perturbation of h, i.e.

$$\|h - \tilde{h}\|_\infty \leq 2\eta. \tag{19}$$

Instead of working directly with (16) we consider the following L-Lipschitz extension

$$\begin{aligned} h(x,y) &= \frac{1}{2}\mathrm{T}_{[-1,1]}\left(\min_{S_1}\{\xi(Y_i,y) + Ld(X_i,x)\}\right) \\ &+ \frac{1}{2}\mathrm{T}_{[-1,1]}\left(\max_{S_1}\{\xi(Y_i,y) - Ld(X_i,x)\}\right), \end{aligned}$$

easily seen to induce the same classifier g_h as (16). Consider the first term (the second term is treated similarly) and its approximate version:

$$\tilde{h}(x,y) = \mathrm{T}_{[-1,1]}\left(\min_{S_1}\left\{\xi(Y_i,y) + L\tilde{d}(X_i,x)\right\}\right),$$

where $d \leq \tilde{d} \leq (1+\eta)d$, given in (15), is the approximate "distance" as provided by the approximate nearest neighbor. For notational convenience, denote

$$\begin{aligned} h(x,y) &= \mathrm{T}_{[-1,1]}(\min_i q_i(x,y)) \\ \tilde{h}(x,y) &= \mathrm{T}_{[-1,1]}(\min_i \tilde{q}_i(x,y)) \\ q_i(x,y) &= h_i(y) + r_i(x) \\ \tilde{q}_i(x,y) &= \tilde{h}_i(y) + \tilde{r}_i(x), \end{aligned}$$

where $h_i(y) = \xi(Y_i,y)$, $r_i(x) = Ld(X_i,x)$, and $\tilde{h}_i$, $\tilde{r}_i$ defined analogously.

Observe that if $\tilde{r}_i(x) > 2$ then $r_i(x) > 2/(1+\eta) \geq 2(1-\eta)$. In this case, since h has range in $[-1,1]$, the eventual application of truncation operator $\mathrm{T}_{[-1,1]}$ will force $\tilde{h}(x,y) - h(x,y) \leq 2\eta$. Hence, we may assume that $\tilde{r}_i(x) \leq 2$ and so $r_i(x) \leq 2$. It is straightforward to verify that for $a,b \in \mathbb{R}^n$ with $\max_{i\in[n]}|a_i - b_i| \leq \eta$, we have

$$\left|\mathrm{T}_{[-1,1]}(\min_i a_i) - \mathrm{T}_{[-1,1]}(\min_i b_i)\right| \leq \eta.$$

Thus, establishing $|q_i(x,y) - \tilde{q}_i(x,y)| \leq 2\eta$ for all $i \in [|S_1|]$ and $y \in \mathcal{Y}$ with $\tilde{r}_i(x), r_i(x) \leq 2$ suffices to prove the claim. Indeed, by (15) we have

$$|r_i(x) - \tilde{r}_i(x)| \leq |r_i(x) - (1+\eta)r_i(x)| \leq 2\eta.$$

Proof of Lemma 6. Suppose $\tilde{h} \in \mathcal{H}_{L,\eta}$. By the definition of $\mathcal{H}_{L,\eta}$, there exists an $h \in \mathcal{H}_L$ such that $\|\tilde{h} - h\|_\infty \leq \eta$. Let h' be some element in a minimal ε-cover of $\mathcal{H}_L$ so that $\|h - h'\|_\infty \leq \varepsilon$. Then

$$\|\tilde{h} - h'\|_\infty \leq \|\tilde{h} - h\|_\infty + \|h - h'\|_\infty \leq \varepsilon + \eta.$$

Hence,

$$\mathcal{N}(\varepsilon+\eta, \mathcal{H}_{L,\eta}, \|\cdot\|_\infty) \leq \mathcal{N}(\varepsilon, \mathcal{H}_L, \|\cdot\|_\infty),$$

whence the claim follows. □

E. Dimensionality reduction proof

Proof of Theorem 7. Put $\tilde{S} = (\tilde{X}, Y)$. For $X_i \in X$ and $\tilde{X}_i \in \tilde{X}$, define $\delta_i(h) = h(X_i,Y_i) - h(\tilde{X}_i,Y_i)$. Then

$$\begin{aligned} \widehat{\mathcal{R}}_n(\mathcal{H}_L; S) &= \mathbb{E}\left[\sup_{h\in\mathcal{H}_L}\frac{1}{n}\sum_{i=1}^{n}\sigma_i h(X_i,Y_i)\,\middle|\, S\right] \\ &= \mathbb{E}\left[\sup_{h\in\mathcal{H}_L}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\left(h(\tilde{X}_i,Y_i) - \delta_i(h)\right)\,\middle|\, S\right] \\ &\leq \widehat{\mathcal{R}}_n(\mathcal{H}_L;\tilde{S}) + \mathbb{E}\left[\sup_{h\in\mathcal{H}_L}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\delta_i(h)\,\middle|\, S\right]. \end{aligned}$$

By (10), we have

$$\mathcal{R}_n(\mathcal{H}_L;\tilde{S}) \leq c_D L\left(\frac{\log 5k}{n}\right)^{1/(\beta+1)}. \tag{20}$$

Since by construction h is L-Lipschitz in its first argument, we have

$$\left|\sum_{i=1}^{n}\sigma_i\delta_i(h)\right| \leq \sum_{i=1}^{n}|\delta_i(h)| \leq L\sum_{i=1}^{n}d(X_i,\tilde{X}_i) \leq L\alpha. \tag{21}$$

Our claimed bound follows from (20) and (21). □

[5] setting $\gamma = 1$ in Guermeur (2010, Theorem 1) and doing the stratification on L instead