Submitted to the Annals of Statistics
arXiv: arXiv:1704.00642

SUPPLEMENTARY MATERIAL FOR 'LOCAL NEAREST NEIGHBOUR CLASSIFICATION WITH APPLICATIONS TO SEMI-SUPERVISED LEARNING'

By Timothy I. Cannings,*
Thomas B. Berrett† and Richard J. Samworth†,‡

University of Edinburgh and University of Cambridge†*

This is the supplementary material for Cannings, Berrett and Samworth (2019), hereafter referred to as the main text.

1. The relationship between our classes and the margin assumption. Recall from Mammen and Tsybakov (1999) that a distribution P on $\mathbb{R}^d \times \{0,1\}$ with marginal P_X on $\mathbb{R}^d$ and regression function η satisfies a *margin assumption* with parameter $\alpha > 0$ if there exists $C > 0$ such that

$$P_X\bigl(\{x : |\eta(x) - 1/2| \leq s\}\bigr) \leq Cs^\alpha$$

for all sufficiently small $s > 0$. The following lemma clarifies the relationship between our classes and the margin assumption.

Lemma 1. *Let $P \in \mathcal{P}_{d,\theta}$ for some $\theta = (\epsilon_0, M_0, \rho, \ell, g) \in \Theta$. Then P satisfies a margin assumption with parameter $\alpha = 1$.*

Proof. By the final part of **(A.3)**, we have

$$\begin{aligned} P_X\bigl(\{x : |\eta(x) - 1/2| \leq s\}\bigr) \leq P_X\bigl(&\{x : |\eta(x) - 1/2| \leq s\} \cap \mathcal{S}^{\epsilon_0}\bigr) \\ &+ P_X\bigl(\{x : \ell(\bar{f}(x)) \geq 1/s\}\bigr). \end{aligned} \tag{32}$$

Now, by Proposition 2 in Section 7.2, for $x \in \mathcal{S}^{\epsilon_0}$, there exists $x_0 \in \mathcal{S}$ and $t \in (-\epsilon_0, \epsilon_0)$ such that $x = x_0 + t\dot{\eta}(x_0)/\|\dot{\eta}(x_0)\|$. Thus, by a Taylor expansion,

$$|\eta(x) - 1/2| \geq |t|\epsilon_0 M_0 - \frac{1}{2}M_0 t^2 \geq \frac{1}{2}|t|\epsilon_0 M_0.$$

‡The research of the third author is supported by an Engineering and Physical Sciences Research Council Fellowship and a grant from the Leverhulme Trust.

MSC 2010 subject classifications: 62G20

Keywords and phrases: classification problems, nearest neighbours, nonparametric classification, semi-supervised learning

We deduce as in Step 5 of the proof of Theorem 5 that there exists $s_0 = s_0(d,\theta) > 0$ such that for all $s \in (0, s_0]$,

$$P_X\big(\{x : |\eta(x) - 1/2| \leq s\} \cap \mathcal{S}^{\epsilon_0}\big) \leq P_X\big(\mathcal{S}^{\frac{2s}{\epsilon_0 M_0}}\big) \leq \frac{8s}{\epsilon_0 M_0}\int_{\mathcal{S}} \bar{f}(x_0)\, d\mathrm{Vol}^{d-1}(x_0) \leq \frac{8sA_1'}{\epsilon_0 M_0}, \tag{33}$$

where the final bound follows from (28) in the main text. For the second term in (32), we exploit the fact that since $\ell \in \mathcal{L}$, there exists $A = A(d,\theta) > 0$ such that $\ell(\delta) \leq A\delta^{-\frac{\rho}{2(\rho+d)}}$ for all $\delta > 0$. Hence, arguing as in (30) in the main text, we find that

$$P_X\big(\{x : \ell(\bar{f}(x)) \geq 1/s\}\big) \leq P_X\Big(\Big\{x : \bar{f}(x) \leq (As)^{\frac{2(\rho+d)}{\rho}}\Big\}\Big) \leq As(1+M_0)^{\frac{\rho+2d}{2(\rho+d)}}\bigg\{\int_{\mathbb{R}^d} \frac{1}{(1+\|x\|^\rho)^{\frac{\rho+2d}{\rho}}}\, dx\bigg\}^{\frac{\rho}{2(\rho+d)}}. \tag{34}$$

The result follows from (32), (33) and (34). □

2. Example 1 from the main text. Recall that we consider the distribution P on $\mathbb{R}^d \times \{0,1\}$ for which $\bar{f}(x) = \frac{\Gamma(3+d/2)}{2\pi^{d/2}}(1-\|x\|^2)^2\mathbb{1}_{\{x \in B_1(0)\}}$ and $\eta(x) = \min(\|x\|^2, 1)$. Since $\bar{f}$ is continuous on all of $\mathbb{R}^d$, it is clear that **(A.1)** is satisfied.

Now, $\mathcal{S} = \{x \in \mathbb{R}^d : \|x\| = 2^{-1/2}\}$ and clearly $\mathcal{S} \cap \{x \in \mathbb{R}^d : \bar{f}(x) > 0\}$ is non-empty. For all $x_0 \in \mathcal{S}$ we have that $\bar{f}(x_0) = \frac{\Gamma(3+d/2)}{8\pi^{d/2}} \leq M_0$. Since $\epsilon_0 \leq 1/10$ we have that $\mathcal{S}^{\epsilon_0} \subseteq B_{9/10}(0) \setminus B_{3/5}(0)$ and thus $\bar{f}$ is twice continuously differentiable on $\mathcal{S}^{\epsilon_0}$. Differentiating $\bar{f}$ twice on $B_1(0)$, we have that $\dot{\bar{f}}(x) = -2\pi^{-d/2}\Gamma(3+d/2)(1-\|x\|^2)x$ and

$$\ddot{\bar{f}}(x) = 2\pi^{-d/2}\Gamma(3+d/2)\{2xx^T - (1-\|x\|^2)I\}.$$

Thus, for $x_0 \in \mathcal{S}$, we have $\|\dot{\bar{f}}(x_0)\|/\bar{f}(x_0) = 2^{5/2} \leq \ell(\bar{f}(x_0))$. We also have that, for any $x \in B_1(0)$,

$$\|\ddot{\bar{f}}(x)\|_{\mathrm{op}} = 2\pi^{-d/2}\Gamma(3+d/2)\|2xx^T - (1-\|x\|^2)I\|_{\mathrm{op}} \leq \frac{6\Gamma(3+d/2)}{\pi^{d/2}},$$

so that $\sup_{u \in B_{\epsilon_0}(0)} \|\ddot{\bar{f}}(x_0+u)\|_{\mathrm{op}}/\bar{f}(x_0) < 48 \leq \ell(\bar{f}(x_0))$ for any $x_0 \in \mathcal{S}$. Finally for **(A.2)** we consider the cases $x \in B_1(0) \setminus B_{\epsilon_0}(0)$ and $x \in B_{\epsilon_0}(0)$ separately. If $x \in B_1(0) \setminus B_{\epsilon_0}(0)$ then, for $r \in (0, \epsilon_0]$, at least a proportion 2^{-d} of the ball $B_r(x)$ is closer to the origin than x, and thus has larger density.

This gives us that, for such x and r, $p_r(x) \geq 2^{-d} a_d r^d \bar{f}(x) \geq \epsilon_0 a_d r^d \bar{f}(x)$. When $x \in B_{\epsilon_0}(0)$ and $r \in (0, \epsilon_0]$ we instead have that

$$p_r(x) \geq a_d r^d \frac{\Gamma(3+d/2)}{2\pi^{d/2}} (1-4\epsilon_0^2)^2 \geq a_d (1-4\epsilon_0^2)^2 r^d \bar{f}(x) \geq \epsilon_0 a_d r^d \bar{f}(x).$$

We now turn to condition **(A.3)**. First, for any $x_0 \in \mathcal{S}$ we have that $\|\dot{\eta}(x_0)\| = \|2x_0\| = 2^{1/2} \geq \epsilon_0 M_0$. For $x \in \mathcal{S}^{2\epsilon_0}$ we have that $\|\dot{\eta}(x)\| \leq 2(2^{-1/2} + 2\epsilon_0) \leq M_0$ and $\|\ddot{\eta}(x)\|_{\mathrm{op}} = \|2I\|_{\mathrm{op}} = 2 \leq M_0$. Since $\ddot{\eta}$ is constant on $\mathcal{S}^{2\epsilon_0}$ it is trivially true that

$$\sup_{x,z \in \mathcal{S}^{2\epsilon_0} : \|z-x\| \leq g(\epsilon)} \|\ddot{\eta}(z) - \ddot{\eta}(x)\|_{\mathrm{op}} \leq \epsilon$$

for any $g \in \mathcal{G}$. Now for $x \in \mathbb{R}^d \setminus \mathcal{S}^{\epsilon_0}$ we have that

$$|\eta(x) - 1/2| \geq 2^{1/2}\epsilon_0 - \epsilon_0^2 \geq \epsilon_0 \geq 1/\ell(\bar{f}(x)).$$

Since the support of $\bar{f}$ is equal to $B_1(0)$, we have that $\int_{\mathbb{R}^d} \|x\|^\rho \, dP_X(x) \leq 1 \leq M_0$, so **(A.4)** is satisfied.

We finally check **(A.5)** to show that $P \in \mathcal{Q}_{d,2,\lambda}$ for $\lambda \geq 6\pi^{-d/2}\Gamma(3+d/2)$. First, it is clear that $\|\bar{f}\|_\infty \leq \lambda$. Now, for any $x, y \in \mathbb{R}^d$ we have that

$$\|\dot{\bar{f}}(y) - \dot{\bar{f}}(x)\| \leq \|y-x\| \sup_{z \in B_1(0)} \|\ddot{\bar{f}}(z)\|_{\mathrm{op}} \leq 6\pi^{-d/2}\Gamma(3+d/2)\|y-x\|.$$

3. Example 2 from the main text.

Proof of claim in Example 2. Fix $\epsilon > 0$ and $k \in K_\beta$, let

$$\mathcal{T}_n := (0, 1/2) \times \big((1+\epsilon)\log(n/k), \infty\big),$$

and for $\gamma > 0$, let

$$B_{k,\gamma} := \bigcap_{x=(x_1,x_2) \in \mathcal{T}_n} \{\gamma < \|X_{(k+1)}(x) - x\| < x_2 - 1\}.$$

Now, for $\epsilon\beta \log n > 4$ and $\gamma \in [2, \epsilon\log(n/k)/2)$,

$$\mathbb{P}(B_{k,\gamma}^c) \leq \mathbb{P}(T \geq k+1) + \mathbb{P}(T' \leq k),$$

where $T \sim \mathrm{Bin}(n, p_\gamma^*)$, $T' \sim \mathrm{Bin}(n, p_*)$,

$$p_\gamma^* := \int_0^1 \int_{(1+\epsilon)\log(n/k)-\gamma}^\infty t_1 \exp(-t_2) \, dt_1 dt_2 \leq \frac{1}{2}\Big(\frac{k}{n}\Big)^{1+\epsilon} e^\gamma \leq \frac{1}{2}\Big(\frac{k}{n}\Big)^{1+\epsilon/2},$$

$$p_* := \int_0^1 \int_{3-3^{1/2}}^{3+3^{1/2}} t_1 \exp(-t_2) \, dt_1 dt_2 \geq \frac{1}{8}.$$

Therefore, there exists $n_0 \in \mathbb{N}$ such that $np_* - (k+1) \geq k/2$ and $k+1-np_\gamma^* \geq k/2$ for all $k \in K_\beta$, $\gamma \in [2, \epsilon\log(n/k)/2)$ and $n \geq n_0$. It follows by Bernstein's inequality that $\sup_{k\in K_\beta}\sup_{\gamma\in[2,\epsilon\log(n/k)/2)} \mathbb{P}(B_{k,\gamma}^c) = O(n^{-M})$ for every $M > 0$.

Now, for $x = (x_1, x_2) \in \mathcal{T}_n$, $\epsilon\beta\log n > 4$ and $\gamma \in [2, x_2 - 1)$, we have that

$$\begin{aligned}\frac{\int_{B_\gamma(x)} \eta(t)\bar{f}(t)\,dt}{\int_{B_\gamma(x)} \bar{f}(t)\,dt} &= \frac{\int_0^1 \int_{x_2-\{\gamma^2-(t_1-x_1)^2\}^{1/2}}^{x_2+\{\gamma^2-(t_1-x_1)^2\}^{1/2}} t_1^2 e^{-t_2}\,dt_2\,dt_1}{\int_0^1 \int_{x_2-\{\gamma^2-(t_1-x_1)^2\}^{1/2}}^{x_2+\{\gamma^2-(t_1-x_1)^2\}^{1/2}} t_1 e^{-t_2}\,dt_2\,dt_1} \\ &= \frac{\int_0^1 t_1^2 \sinh(\{\gamma^2-(t_1-x_1)^2\}^{1/2})\,dt_1}{\int_0^1 t_1 \sinh(\{\gamma^2-(t_1-x_1)^2\}^{1/2})\,dt_1} \\ &\geq \frac{2}{3}\frac{\sinh\bigl((\gamma^2-1)^{1/2}\bigr)}{\sinh(\gamma)} \geq \frac{2}{3}\frac{\sinh(3^{1/2})}{\sinh(2)} > \frac{1}{2}.\end{aligned}$$

Our next observation is that for $\gamma \in [0,\infty)$ and $x_{(k+1)} \in \mathbb{R}^d$ such that $\|x_{(k+1)} - x\| = \gamma$, we have that $(X_{(1)}, Y_{(1)}, \ldots, X_{(k)}, Y_{(k)})|(X_{(k+1)} = x_{(k+1)}) \overset{d}{=} (\tilde{X}_{(1)}, \tilde{Y}_{(1)}, \ldots, \tilde{X}_{(k)}, \tilde{Y}_{(k)})$, where the pairs $(\tilde{X}_1, \tilde{Y}_1), \ldots, (\tilde{X}_k, \tilde{Y}_k)$ are independent and identically distributed, and then $(\tilde{X}_{(1)}, \tilde{Y}_{(1)}), \ldots, (\tilde{X}_{(k)}, \tilde{Y}_{(k)})$ is a reordering such that $\|\tilde{X}_{(1)} - x\| \leq \ldots \leq \|\tilde{X}_{(k)} - x\|$. Here $\tilde{X}_1 \overset{d}{=} X|(\|X - x\| \leq \gamma)$ and $\mathbb{P}(\tilde{Y}_1 = 1|\tilde{X}_1 = x) = \eta(x)$. Writing $\tilde{S}_n(x) := \frac{1}{k}\sum_{i=1}^k \mathbb{1}_{\{\tilde{Y}_i=1\}}$ we therefore have by Hoeffding's inequality that, for $x \in \mathcal{T}_n$, $\epsilon\beta\log n > 4$ and $\|x_{(k+1)} - x\| \in [2, x_2 - 1)$,

$$\begin{aligned}\mathbb{P}\{\hat{S}_n(x) < 1/2\big|X_{(k+1)} = x_{(k+1)}\} &= \mathbb{P}\{\tilde{S}_n(x) < 1/2\} \\ &= \mathbb{P}\{\tilde{S}_n(x) - \mathbb{E}\tilde{S}_n(x) < -(\mathbb{E}\eta(\tilde{X}_1) - 1/2)\} \\ &\leq \exp\biggl(-2k\Bigl(\frac{2}{3}\frac{\sinh(3^{1/2})}{\sinh(2)} - \frac{1}{2}\Bigr)^2\biggr) = O(n^{-M})\end{aligned}$$

for all $M > 0$, uniformly for $k \in K_\beta$. Writing $P_{(k+1)}$ for the marginal distribution of $X_{(k+1)}$, we deduce that

$$\begin{aligned}&\mathbb{P}\{\hat{S}_n(x) < 1/2\} \\ &\quad\leq \mathbb{P}\{\hat{S}_n(x) < 1/2, \|X_{(k+1)} - x\| \in [2, x_2 - 1)\} + \mathbb{P}(B_{k,2}^c) \\ &\quad= \int_{B_{x_2-1}(x)\setminus B_2(x)} \mathbb{P}\{\hat{S}_n(x) < 1/2\big|X_{(k+1)} = x_{(k+1)}\}\,dP_{(k+1)}(x_{(k+1)}) + O(n^{-M}) \\ &\quad= O(n^{-M})\end{aligned}$$

for all $M > 0$, uniformly for $k \in K_\beta$. We conclude that for every $M > 0$,

$$
\begin{aligned}
&R_{\mathcal{T}_n}(\hat{C}_n^{k\mathrm{nn}}) - R_{\mathcal{T}_n}(C^{\mathrm{Bayes}}) \\
&= \int_{\mathcal{T}_n} \Big[\mathbb{P}\{\hat{S}_n(x) < 1/2\} - \mathbb{1}_{\{\eta(x)<1/2\}}\Big]\{2\eta(x) - 1\}\bar{f}(x)\,dx \\
&= \int_{(1+\epsilon)\log(n/k)}^{\infty} \int_0^{1/2} \mathbb{P}\{\hat{S}_n(x) \geq 1/2\}(1 - 2x_1)x_1 \exp(-x_2)\,dx_1\,dx_2 \\
&= \frac{1}{24}\Big(\frac{k}{n}\Big)^{1+\epsilon} + O(n^{-M}),
\end{aligned}
$$

uniformly for $k \in K_\beta$. The claim (4) follows from this together with Theorem 1(ii). □

4. Proof of Theorem 4.

Proof of Theorem 4. For an integer $q \geq 3$ and $\nu \geq 0$, define a grid on $\mathbb{R}^d$ by

$$
G_{q,\nu} := \Big\{\Big(\gamma_1, \gamma_2 + \frac{2\kappa_2 + 1}{2q}, \gamma_3 + \frac{2\kappa_3+1}{2q}, \ldots, \gamma_d + \frac{2\kappa_d + 1}{2q}\Big) : \gamma_1, \ldots, \gamma_d \in \{1, \ldots, \lceil q^\nu \rceil\}, \kappa_2, \ldots, \kappa_d \in \{0, 1, \ldots, q-1\}\Big\}.
$$

Now, for $x \in \mathbb{R}^d$, let $n_q(x)$ be the closest point to x among those in $G_{q,\nu}$ (if there are multiple points, pick the one that is smallest in the lexicographic ordering). Let $m := \lceil q^\nu \rceil^d q^{d-1}$ and define closed Euclidean balls $\mathcal{X}_1, \ldots, \mathcal{X}_m$ in $\mathbb{R}^d$ of radius $1/(2q)$, where the lth ball is centered at the lth grid point in the lexicographic ordering.

Writing $[z]$ for the closest integer to z (where we round half-integers to the nearest even integer), define the 'saw-tooth' function $\eta_0 : \mathbb{R}^d \to [3/8, 5/8]$, by $\eta_0(x) := 3/8 + \big|x_1 + 1/4 - [x_1 + 1/4]\big|/2$, for $x = (x_1, \ldots, x_d)^T$. Further, for $x \in \mathbb{R}^d$, set $u(x) := \frac{\alpha_0 g^{-1}(1/q)}{q^2}\big(1/4 - q^2\|x - n_q(x)\|^2\big)^4$, where $\alpha_0 := 1/27$.

For $\sigma := (\sigma_1, \ldots, \sigma_m)^T \in \{-1, 1\}^m$, we now define the distribution P_σ on $\mathbb{R}^d \times \{0, 1\}$ by setting the regression function to be $\eta_\sigma(x) := \eta_0(x) + \frac{1}{2}\sigma_l u(x)$, for $x \in \mathcal{X}_l$, $l = 1, \ldots, m$, and setting $\eta_\sigma(x) := \eta_0(x)$, otherwise. To define the marginal distribution on $\mathbb{R}^d$ induced by P_σ, which will be the same for each σ, we first define the boxes $B_0 := (0, \lceil q^\nu \rceil + 3/2)^d$ and $B_r := [-r/2 + 1/4 - a/16, -r/2 + 1/4 + a/16] \times [-a, a]^{d-1}$ for $r = 1, \ldots, 20$ and some $a > 0$ to be chosen later. We further define a modified bump function by

$$
h(x) := \begin{cases} 0 & \text{if } x \leq 0 \\ \Phi\big(\frac{2x-1}{x(1-x)}\big) & \text{if } x \in (0, 1) \\ 1 & \text{if } x \geq 1, \end{cases}
$$

where Φ denotes the standard normal distribution function. For $x \in \mathbb{R}^d$ we then set

$$\bar{f}(x) := w_0 h\bigl(1 - 4\mathrm{dist}(x, B_0)\bigr) + h\Bigl(1 - 16 \min_{r=1,\ldots,20} \mathrm{dist}(x, B_r)\Bigr)$$

for some $w_0 < 1/(\lceil q^\nu \rceil + 2)^d$ to be specified later. Here, a in the definition of B_r is chosen such that $\int_{\mathbb{R}^d} \bar{f} = 1$, and we note that

$$1 \geq 20\frac{a}{8}(2a)^{d-1} = \frac{5}{4}(2a)^d,$$

so $a \leq (4/5)^{1/d}/2$.

Let

$$\mathcal{P}_m := \Bigl\{P_\sigma : \sigma := (\sigma_1, \ldots, \sigma_m) \in \{-1, 1\}^m\Bigr\}.$$

We show below that $\mathcal{P}_m \subseteq \mathcal{P}_{d,\theta} \cap \mathcal{Q}_{d,2,\lambda}$ for all $\theta \in \Theta$ and $\lambda > 0$ satisfying the conditions of the theorem.

Letting $\mathbb{E}_\sigma$ denote expectation with respect to $P_\sigma^{\otimes n}$ and writing $[[x_1]] := x_1 - [x_1 + 1/4]$ for $x_1 \in \mathbb{R}$, we have that, for any classifier C_n,

$$\begin{aligned}
&\sup_{P \in \mathcal{P}_{d,\theta} \cap \mathcal{Q}_{d,2,\lambda}} \{R(C_n) - R(C^{\mathrm{Bayes}})\} \geq \max_{P \in \mathcal{P}_m} \{R(C_n) - R(C^{\mathrm{Bayes}})\} \\
&= \max_{\sigma \in \{-1,1\}^m} \int_{\mathbb{R}^d} \mathbb{E}_\sigma\{\mathbb{1}_{\{C_n(x)=0\}} - \mathbb{1}_{\{\eta_\sigma(x)<1/2\}}\}\{2\eta_\sigma(x) - 1\}\, dP_X(x) \\
&\geq \max_{\sigma \in \{-1,1\}^m} \sum_{l=1}^m \int_{\mathcal{X}_l} \mathbb{E}_\sigma\{\mathbb{1}_{\{C_n(x)=0\}} - \mathbb{1}_{\{\eta_\sigma(x)<1/2\}}\}\bigl\{[[x_1]] + \sigma_l u(x)\bigr\}\, dP_X(x) \\
&\geq \frac{1}{2^m} \sum_{\sigma \in \{-1,1\}^m} \sum_{l=1}^m \int_{\mathcal{X}_l} \mathbb{E}_\sigma\{\mathbb{1}_{\{C_n(x)=0\}} - \mathbb{1}_{\{\eta_\sigma(x)<1/2\}}\}\bigl\{[[x_1]] + \sigma_l u(x)\bigr\}\, dP_X(x).
\end{aligned}$$

Now let $\sigma_{l,r} := (\sigma_1, \ldots, \sigma_{l-1}, r, \sigma_{l+1}, \ldots, \sigma_m)$ for $l = 1, \ldots, m$, and $r \in \{-1, 0, 1\}$, and define the distribution $P_{l,r}$ on $\mathbb{R}^d \times \{0, 1\}$ by $\eta_{l,r}(x) := \eta_0(x) + (1/2)ru(x)$, for $x \in \mathcal{X}_l$ and $\eta_{l,r}(x) = \eta_{\sigma_{l,r}}(x) := \eta_\sigma(x)$ otherwise (the marginal distribution on $\mathbb{R}^d$ is again taken to be P_X). We write $\mathbb{E}_{l,r}$ to denote expectation with respect to $P_{l,r}^{\otimes n}$.

For $l = 1, \ldots, m$ and $r \in \{-1, 1\}$ define

$$L_{l,r} := \frac{\prod_{i=1}^n [Y_i \eta_{l,r}(X_i) + (1 - Y_i)\{1 - \eta_{l,r}(X_i)\}]}{\prod_{i=1}^n [Y_i \eta_{l,0}(X_i) + (1 - Y_i)\{1 - \eta_{l,0}(X_i)\}]}.$$

By the Radon–Nikodym theorem, we have that

$$
\begin{aligned}
&\frac{1}{2^m}\sum_{\sigma\in\{-1,1\}^m}\sum_{l=1}^m\int_{\mathcal{X}_l}\mathbb{E}_\sigma\{\mathbb{1}_{\{C_n(x)=0\}}-\mathbb{1}_{\{\eta_\sigma(x)<1/2\}}\}\bigl\{[[x_1]]+\sigma_l u(x)\bigr\}\,dP_X(x)\\
&=\frac{1}{2}\sum_{l=1}^m\mathbb{E}_{l,0}\biggl(\int_{\mathcal{X}_l}\bigl[L_{l,1}\{\mathbb{1}_{\{C_n(x)=0\}}-\mathbb{1}_{\{\eta_{l,1}(x)<1/2\}}\}\bigr]\bigl\{[[x_1]]+u(x)\bigr\}\,dP_X(x)\\
&\qquad\qquad+\int_{\mathcal{X}_l}\bigl[L_{l,-1}\{\mathbb{1}_{\{C_n(x)=0\}}-\mathbb{1}_{\{\eta_{l,-1}(x)<1/2\}}\}\bigr]\bigl\{[[x_1]]-u(x)\bigr\}\,dP_X(x)\biggr)\\
&\ge\frac{1}{2}\sum_{l=1}^m\mathbb{E}_{l,0}\biggl\{\biggl(\int_{\mathcal{X}_l}\{\mathbb{1}_{\{C_n(x)=0\}}-\mathbb{1}_{\{\eta_{l,1}(x)<1/2\}}\}\bigl\{[[x_1]]+u(x)\bigr\}\,dP_X(x)\\
&\quad+\int_{\mathcal{X}_l}\{\mathbb{1}_{\{C_n(x)=0\}}-\mathbb{1}_{\{\eta_{l,-1}(x)<1/2\}}\}\bigl\{[[x_1]]-u(x)\bigr\}\,dP_X(x)\biggr)\min(L_{l,1},L_{l,-1})\biggr\}.
\end{aligned}
$$

Now fix $x=(x_1,\ldots,x_d)^T\in\mathcal{X}_l$, and writing $C_n=C_n(x),\eta_{l,1}=\eta_{l,1}(x)$ and $\eta_{l,-1}=\eta_{l,-1}(x)$ as shorthand, observe that

$$
\begin{aligned}
&\{\mathbb{1}_{\{C_n=0\}}-\mathbb{1}_{\{\eta_{l,1}<1/2\}}\}\bigl\{[[x_1]]+u(x)\bigr\}\\
&\qquad\qquad\qquad\qquad+\{\mathbb{1}_{\{C_n=0\}}-\mathbb{1}_{\{\eta_{l,-1}<1/2\}}\}\bigl\{[[x_1]]-u(x)\bigr\}\\
&=2\Bigl\{\mathbb{1}_{\{C_n=0,\eta_{l,1}\ge1/2,\eta_{l,-1}\ge1/2\}}-\mathbb{1}_{\{C_n=1,\eta_{l,1}<1/2,\eta_{l,-1}<1/2\}}\Bigr\}[[x_1]]\\
&\qquad+\Bigl\{\mathbb{1}_{\{C_n=0,\eta_{l,1}\ge1/2,\eta_{l,-1}<1/2\}}-\mathbb{1}_{\{C_n=1,\eta_{l,1}<1/2,\eta_{l,-1}\ge1/2\}}\Bigr\}\{[[x_1]]+u(x)\}\\
&\qquad+\Bigl\{\mathbb{1}_{\{C_n=0,\eta_{l,1}<1/2,\eta_{l,-1}\ge1/2\}}-\mathbb{1}_{\{C_n=1,\eta_{l,1}\ge1/2,\eta_{l,-1}<1/2\}}\Bigr\}\{[[x_1]]-u(x)\}\\
&=2\Bigl\{\mathbb{1}_{\{C_n=0,\eta_{l,1}\ge1/2,\eta_{l,-1}\ge1/2\}}-\mathbb{1}_{\{C_n=1,\eta_{l,1}<1/2,\eta_{l,-1}<1/2\}}\Bigr\}[[x_1]]\\
&\qquad+\mathbb{1}_{\{\eta_{l,1}\ge1/2,\eta_{l,-1}<1/2\}}\bigl[\mathbb{1}_{\{C_n=0\}}\{[[x_1]]+u(x)\}-\mathbb{1}_{\{C_n=1\}}\{[[x_1]]-u(x)\}\bigr]\\
&\ge\mathbb{1}_{\{\eta_{l,1}\ge1/2,\eta_{l,-1}<1/2\}}\bigl\{u(x)-\bigl|[[x_1]]\bigr|\bigr\}.
\end{aligned}
$$

Here we used the fact that $\eta_{l,1}(x)\ge\eta_{l,-1}(x)$, so $\mathbb{1}_{\{\eta_{l,1}(x)<1/2,\eta_{l,-1}(x)\ge1/2\}}=0$, and that the minimum is attained by taking $C_n(x)=\mathbb{1}_{\{[[x_1]]\ge0\}}$ for $x\in\mathcal{X}_l$; it is interesting to note that this remains the optimal classifier even if $\bar{f}$ is known. Moreover, whenever $[[x_1]]\ge0$, we have $\eta_{l,1}(x)\ge1/2$, and when

$[[x_1]] < 0$, we have $\eta_{l,-1}(x) < 1/2$. It follows that

$$\sup_{P \in \mathcal{P}_{d,\theta}} \{R(C_n) - R(C^{\mathrm{Bayes}})\}$$
$$\geq \frac{1}{2}\sum_{l=1}^{m} \mathbb{E}_{l,0}\biggl\{\min(L_{l,1}, L_{l,-1})\int_{\mathcal{X}_l} \mathbb{1}_{\{\eta_{l,1}\geq 1/2, \eta_{l,-1}<1/2\}}\bigl\{u(x) - \bigl|[[x_1]]\bigr|\bigr\}\, dP_X(x)\biggr\}$$
$$= \sum_{l=1}^{m} \mathbb{E}_{l,0}\bigl\{\min(L_{l,1}, L_{l,-1})\bigr\} \int_{\mathcal{X}_l \cap \{[[x_1]]\geq 0\}} \mathbb{1}_{\{\eta_{l,-1}<1/2\}}\bigl\{u(x) - [[x_1]]\bigr\}\, dP_X(x)$$

(35)
$$= m w_0 \mathbb{E}_{1,0}\bigl\{\min(L_{1,1}, L_{1,-1})\bigr\} \int_{B_{1/(2q)}(0)\cap\{x_1\geq 0\}} \mathbb{1}_{\{\tilde{\eta}(x)<1/2\}}\bigl\{\tilde{u}(x) - x_1\bigr\}\, dx,$$

where $\tilde{u}(x) := \alpha_0 g^{-1}(1/q) q^6 (\frac{1}{4q^2} - \|x\|^2)^4$ and $\tilde{\eta}(x) := \frac{1}{2}\{1 + x_1 - \tilde{u}(x)\}$.

Now, observe that

$$\mathbb{E}_{1,0}\bigl\{\min(L_{1,1}, L_{1,-1})\bigr\} = 1 - d_{\mathrm{TV}}(P_{1,1}^{\otimes n}, P_{1,-1}^{\otimes n})$$

and

$$d_{\mathrm{TV}}^2(P_{1,1}^{\otimes n}, P_{1,-1}^{\otimes n}) \leq \frac{1}{2} d_{\mathrm{KL}}^2(P_{1,1}^{\otimes n}, P_{1,-1}^{\otimes n}) = \frac{n}{2} d_{\mathrm{KL}}^2(P_{1,1}, P_{1,-1}).$$

Moreover, using the fact that $\log(1+x) \leq x$ for $x \geq 0$, we have that

$$\begin{aligned} &d_{\mathrm{KL}}^2(P_{1,1}, P_{1,-1}) \\ &\quad= \int_{\mathbb{R}^d} \eta_{1,1}(x) \log\biggl(\frac{\eta_{1,1}(x)}{\eta_{1,-1}(x)}\biggr) + \{1 - \eta_{1,1}(x)\}\log\biggl(\frac{1-\eta_{1,1}(x)}{1-\eta_{1,-1}(x)}\biggr)\, dP_X(x) \\ &\quad\leq 24 \int_{\mathcal{X}_1} u^2(x)\, dP_X(x) \\ &\quad= 24\alpha_0^2 w_0 q^{12} g^{-1}(1/q)^2 \int_{B_{1/(2q)}(0)} \biggl(\frac{1}{4q^2} - \|x\|^2\biggr)^8 dx \\ &\quad= \frac{945\alpha_0^2 a_d w_0 g^{-1}(1/q)^2 \Gamma(1+d/2)}{2^{d+6}\Gamma(9+d/2)} q^{-(4+d)}. \end{aligned}$$

We now turn to finding a lower bound for the integral in (35). First, we observe that $\mathrm{sgn}\bigl(\tilde{u}(x) - x_1\bigr) = \mathrm{sgn}\bigl(1/2 - \tilde{\eta}(x)\bigr)$, and moreover for $d = 1$ and $0 \leq x_1 < \frac{\alpha_0 g^{-1}(1/q)}{2^{13}q^2}$, we have that

$$\begin{aligned} \tilde{u}(x_1) - x_1 = q^6 \alpha_0 g^{-1}(1/q)\biggl(\frac{1}{4q^2} - x_1^2\biggr)^4 - x_1 &> \frac{\alpha_0 g^{-1}(1/q)}{2^{12}q^2} - x_1 \\ &> \frac{\alpha_0 g^{-1}(1/q)}{2^{13}q^2} > 0. \end{aligned}$$

Thus

$$\int_0^{1/(2q)} \mathbb{1}_{\{\tilde{\eta}(x_1)<1/2\}}\{\tilde{u}(x_1)-x_1\}\,dx_1 \geq \frac{\alpha_0^2 g^{-1}(1/q)^2}{2^{26}q^4}.$$

Furthermore, for $d \geq 2$, writing $x_{-1} := (x_2,\ldots,x_d)^T$, we have that $\tilde{\eta}(x) < 1/2$ if and only if

$$0 > x_1 - q^6\alpha_0 g^{-1}(1/q)\Bigl(\frac{1}{4q^2}-\|x\|^2\Bigr)^4$$

which is satisfied if

$$\|x_{-1}\| < (1-2^{-1/4})^{1/2}\sqrt{\frac{1}{4q^2}-\Bigl(\frac{x_1}{q^6\alpha_0 g^{-1}(1/q)}\Bigr)^{1/4}-x_1^2} =: t(x_1).$$

Now $t(x_1)$ is real if $0 \leq x_1 \leq \frac{\alpha_0 g^{-1}(1/q)}{2^{14}q^2}$. Moreover, $t(x_1) > 1/(8q)$ for $x_1 \in \bigl[0, \frac{\alpha_0 g^{-1}(1/q)}{2^{14}q^2}\bigr]$. We also require the observation that $\tilde{u}(x) - x_1 \geq \frac{\alpha_0 g^{-1}(1/q)}{2^{14}q^2}$ when $x_1 \in \bigl[0, \frac{\alpha_0 g^{-1}(1/q)}{2^{14}q^2}\bigr]$ and $\|x_{-1}\| < t(x_1)$. Hence

$$\begin{aligned}
&\int_{B_{1/(2q)}(0)\cap\{x_1\geq 0\}} \mathbb{1}_{\{\tilde{\eta}(x)<1/2\}}\{\tilde{u}(x)-x_1\}\,dx \\
&\qquad\geq \int_0^{\frac{\alpha_0 g^{-1}(1/q)}{2^{14}q^2}} \int_{\|x_{-1}\|<t(x_1)} \{\tilde{u}(x)-x_1\}\,dx_{-1}\,dx_1 \\
&\qquad\geq \frac{\alpha_0 g^{-1}(1/q)}{2^{14}q^2} a_{d-1} \int_0^{\frac{\alpha_0 g^{-1}(1/q)}{2^{14}q^2}} t(x_1)^{d-1}\,dx_1 \\
&\qquad\geq \frac{\alpha_0^2 g^{-1}(1/q)^2}{2^{28}q^{3+d}} a_{d-1} 2^{-3(d-1)}.
\end{aligned}$$

We have therefore shown that, for $q \geq 3$,

$$\begin{aligned}
&\sup_{P\in\mathcal{P}_{d,\theta}} \{R(C_n) - R(C^{\mathrm{Bayes}})\} \\
&\quad\geq \frac{m w_0 a_{d-1}\alpha_0^2 g^{-1}(1/q)^2}{2^{28+3(d-1)}q^{3+d}}\biggl\{1-\sqrt{\frac{n945 a_d w_0 \alpha_0^2 g^{-1}(1/q)^2\Gamma(1+d/2)}{2^{d+6}\Gamma(9+d/2)q^{4+d}}}\biggr\},
\end{aligned}$$

where $a_0 := 1$. It follows that if we set

$$w_0 = \frac{q^{4+d}}{4^{d+1}ng^{-1}(1/q)^2},$$

and choose q to satisfy $\frac{q^{4+d+\nu(\rho+d)}}{g^{-1}(1/q)^2} = n$, then

$$\sup_{P \in \mathcal{P}_{d,\theta}} \{R(C_n) - R(C^{\mathrm{Bayes}})\} \geq \frac{q^{d+\nu d}}{n} \frac{a_{d-1}\alpha_0^2}{2^{28+5d}}$$
$$= g^{-1}(1/q)^{\frac{2d(1+\nu)}{4+d+\nu(\rho+d)}} n^{-\frac{4+\nu\rho}{4+d+\nu(\rho+d)}} \frac{a_{d-1}\alpha_0^2}{2^{28+5d}}.$$

It remains to show that P_σ belongs to the desired classes $\mathcal{P}_{d,\theta} \cap \mathcal{Q}_{d,2,\lambda}$ for each σ. First note that

$$w_0 = \frac{q^{4+d}}{4^{d+1} n g^{-1}(1/q)^2} = \frac{1}{4^{d+1}} q^{-\nu(\rho+d)} < \frac{1}{(\lceil q^\nu \rceil + 2)^d}.$$

Condition **(A.1)** is satisfied by $\bar{f}$ by construction. To verify the minimal mass assumption, we take $\epsilon_* < 2^{-\max(d,5)}$, and observe that when $\epsilon_0 \in (0, \epsilon_*]$,

$$\inf_{r_0 \in (0,\epsilon_0], x \in \mathbb{R}^d : \bar{f}(x) > 0} \frac{1}{a_d r_0^d \bar{f}(x)} \int_{B_{r_0}(x)} \bar{f}$$
$$\geq \inf_{r_0 \in (0,\epsilon_0]} \frac{1}{a_d r_0^d} \int_{B_{r_0}(0)} \Phi\biggl(\frac{1 - 32\|x\|}{16\|x\|(1 - 16\|x\|)}\biggr) \, dx \wedge 2^{-d} \geq 2^{-d},$$

as required. It follows that **(A.2)** is satisfied for such $\epsilon_0 \in (0, \epsilon_*]$ and for any $M_0 \geq 1$.

The main condition to check is **(A.3)**. For $x \in B_{1/(2q)}(0)$, consider

$$\tilde{\eta}_\pm(x) := (1/2)\biggl\{1 + x_1 \pm q^6 \alpha_0 g^{-1}(1/q)\biggl(\frac{1}{4q^2} - \|x\|^2\biggr)^4\biggr\}.$$

Then

$$\dot{\tilde{\eta}}_\pm(x) = (1/2, 0, \ldots, 0)^T \mp 8 q^6 \alpha_0 g^{-1}(1/q)\biggl(\frac{1}{4q^2} - \|x\|^2\biggr)^3 x,$$

and

$$\ddot{\tilde{\eta}}_\pm(x) = \mp 8 q^6 \alpha_0 g^{-1}(1/q)\biggl(\frac{1}{4q^2} - \|x\|^2\biggr)^3 I_{d\times d}$$
$$\pm 48 q^6 \alpha_0 g^{-1}(1/q)\biggl(\frac{1}{4q^2} - \|x\|^2\biggr)^2 x x^T.$$

From these calculations, we see that each η_σ is twice continuously differentiable on $\mathcal{S}^{2\epsilon_0}$, with $\|\dot{\eta}_\sigma(x)\| \in (1/4, 3/4)$ for all $x \in \mathcal{S}^{2\epsilon_0}$ and $\|\ddot{\eta}_\sigma(x)\|_{\mathrm{op}} \leq 1$.

We have that, when $n_q(z) = n_q(x)$,

$$
\begin{aligned}
&\|\ddot\eta_\sigma(z) - \ddot\eta_\sigma(x)\|_{\mathrm{op}} \\
&\le 8q^6\alpha_0 g^{-1}(1/q)\left|\left(\frac{1}{4q^2} - \|z\|^2\right)^3 - \left(\frac{1}{4q^2} - \|x\|^2\right)^3\right| \\
&\qquad + 48q^6\alpha_0 g^{-1}(1/q)\left\|\left(\frac{1}{4q^2} - \|z\|^2\right)^2 zz^T - \left(\frac{1}{4q^2} - \|x\|^2\right)^2 xx^T\right\| \\
&= 8q^6\alpha_0 g^{-1}(1/q)(\|z\| + \|x\|)\big|\|z\| - \|x\|\big|\biggl\{\Bigl(\frac{1}{4q^2} - \|z\|^2\Bigr)^2 \\
&\qquad\qquad + \Bigl(\frac{1}{4q^2} - \|z\|^2\Bigr)\Bigl(\frac{1}{4q^2} - \|x\|^2\Bigr) + \Bigl(\frac{1}{4q^2} - \|x\|^2\Bigr)^2\biggr\} \\
&\qquad + 48q^6\alpha_0 g^{-1}(1/q)\left\|\Bigl(\frac{1}{4q^2} - \|z\|^2\Bigr)^2\Bigl\{(z-x)(z-x)^T\right. \\
&\qquad \left. + x(z-x)^T + (z-x)x^T\Bigr\} + (\|x\|^2 - \|z\|^2)\Bigl\{\frac{1}{2q^2} - \|x\|^2 - \|z\|^2\Bigr\}xx^T\right\| \\
&\le \frac{1}{2}qg^{-1}(1/q)\|z - x\|. \qquad (36)
\end{aligned}
$$

Hence, using the fact that $r \mapsto r/g^{-1}(r)$ is increasing for sufficiently small $r > 0$, we have that for sufficiently large q,

$$
\sup_{\|x-z\| \le g(\epsilon), n_q(x) = n_q(z)} \|\ddot\eta_\sigma(x) - \ddot\eta_\sigma(z)\|_{\mathrm{op}} \le \epsilon.
$$

Now consider the case where $z \in \mathcal{X}_l$ and $x \in \mathcal{X}_{l'}$ with $l \neq l'$, so that $n_q(z) \neq n_q(x)$. Let z' denote the closest point in $\mathcal{X}_l$ to $\mathcal{X}_{l'}$ on the line segment joining x to z, and similarly let x' denote the closest point in $\mathcal{X}_{l'}$ to $\mathcal{X}_l$ on the same line segment. Then $\ddot\eta_\sigma(x') = \ddot\eta_\sigma(z') = 0$, so, by (36),

$$
\begin{aligned}
&\|\ddot\eta_\sigma(z) - \ddot\eta_\sigma(x)\|_{\mathrm{op}} \le \|\ddot\eta_\sigma(z) - \ddot\eta_\sigma(z')\|_{\mathrm{op}} + \|\ddot\eta_\sigma(x') - \ddot\eta_\sigma(x)\|_{\mathrm{op}} \\
&\quad \le \frac{1}{2}qg^{-1}(1/q)(\|z - z'\| + \|x - x'\|) \le \frac{1}{2}\bigl\{g^{-1}(\|z - z'\|) + g^{-1}(\|x - x'\|)\bigr\}.
\end{aligned}
$$

We therefore deduce that

$$
\sup_{\|x-z\| \le g(\epsilon)} \|\ddot\eta_\sigma(x) - \ddot\eta_\sigma(z)\|_{\mathrm{op}} \le \epsilon.
$$

For the final part of **(A.3)**, we note that

$$
\inf_{x \in (\epsilon_0 \pm \mathbb{Z}/2) \times \mathbb{R}^{d-1}} \Bigl|\eta_\sigma(x) - \frac{1}{2}\Bigr| \ge \frac{\epsilon_0}{2}.
$$

Finally, we check the moment condition in **(A.4)**. First,

$$\int_{\mathbb{R}^d} \|x\|^\rho \bar{f}(x)\,dx = w_0 \int_{x:\mathrm{dist}(x,B_0)\le 1/4} \|x\|^\rho h\Bigl(4\bigl(1-\mathrm{dist}(x,B_0)\bigr)\Bigr)\,dx$$
$$+ \int_{[-10,-1]\times[-a-1/16,a+1/16]^{d-1}} \|x\|^\rho h\biggl(16\Bigl(1-\min_{r=1,\ldots,20}\mathrm{dist}(x,B_r)\Bigr)\biggr)\,dx$$
$$\le w_0 d^{\frac{\rho}{2}}(\lceil q^\nu\rceil+2)^{d+\rho} + \max(1,2^{\frac{\rho-2}{2}})\{100^{\frac{\rho}{2}} + (d-1)^{\frac{\rho}{2}}(a+1/16)^{\frac{\rho}{2}}\}$$
$$\le \frac{3^{d+\rho}d^{\frac{\rho}{2}}}{4^{d+1}} + \max(1,2^{(\rho-2)/2})\{100^{\frac{\rho}{2}} + (d-1)^{\frac{\rho}{2}}(a+1/16)^{\frac{\rho}{2}}\} =: M_{01}(\rho),$$

say. We conclude that there exists $q_* = q_*(d)$ such that for $q \ge q_*$ and any $\nu \ge 0$, we have $P \in \mathcal{P}_{d,\theta}$ for $\theta = (\epsilon_0, M_0, \rho, \ell, g)$ with any $\rho > 0$, $M_0 \ge \max(M_{01}(\rho), 1)$, $\epsilon_0 \in (0, \min(2^{-\max(d,5)}, 1/(4M_0)))$, any $\ell \in \mathcal{L}$ with $\ell \ge 2/\epsilon_0$ and any $g \in \mathcal{G}$.

Finally, we note that $\|\bar{f}\|_\infty \le 1$ and

$$\|\ddot{\bar{f}}\|_\infty \le 2^8 \sup_{x\in(0,1)} \phi\biggl(\frac{2x-1}{x(1-x)}\biggr)\biggl\{\frac{2}{(1-x)^3} - \frac{2}{x^3} - \frac{2x-1}{x(1-x)}\biggl(\frac{1}{(1-x)^2} + \frac{1}{x^2}\biggr)\biggr\}$$
$$\le 2^{10}\times 5.$$

Hence $P \in \mathcal{Q}_{d,2,\lambda}$ for $\lambda \ge 2^{10}\times 5$. □

5. Proof of Theorem 5 (continued).

PROOF OF THEOREM 5 – **Step 7**. To complete the proof of Theorem 5, it remains to bound the error terms R_1, R_2, R_5 and R_6.

To bound R_1: We have

$$R_1 = \frac{1}{k_{\mathrm{L}}}\sum_{i=1}^{k_{\mathrm{L}}}\biggl(\mathbb{E}\eta(X_{(i)}) - \eta(x) - \mathbb{E}\{(X_{(i)}-x)^T\dot{\eta}(x)\}$$
$$- \frac{1}{2}\mathbb{E}\{(X_{(i)}-x)^T\ddot{\eta}(x)(X_{(i)}-x)\}\biggr).$$

By a Taylor expansion and **(A.3)**, for all $\epsilon \in (0,1)$, $x \in \mathcal{S}^{\epsilon_0}$ and $\|z-x\| < \min\{g(\epsilon),\epsilon_0\} =: r$,

$$\biggl|\eta(z) - \eta(x) - (z-x)^T\dot{\eta}(x) - \frac{1}{2}(z-x)^T\ddot{\eta}(x)(z-x)\biggr| \le \epsilon\|z-x\|^2.$$

Hence

$$|R_1| \leq \epsilon \frac{1}{k_{\mathrm{L}}} \sum_{i=1}^{k_{\mathrm{L}}} \mathbb{E}\{\|X_{(i)} - x\|^2 \mathbb{1}_{\{\|X_{(k_{\mathrm{L}})}-x\| \leq r\}}\} + 2\mathbb{P}\{\|X_{(k_{\mathrm{L}})} - x\| > r\}$$
$$+ \sup_{z \in \mathcal{S}^{\epsilon_0}} \|\dot{\eta}(z)\| \mathbb{E}\{\|X_{(k_{\mathrm{L}})} - x\| \mathbb{1}_{\{\|X_{(k_{\mathrm{L}})}-x\| > r\}}\}$$
$$+ \sup_{z \in \mathcal{S}^{\epsilon_0}} \|\ddot{\eta}(z)\|_{\mathrm{op}} \mathbb{E}\{\|X_{(k_{\mathrm{L}})} - x\|^2 \mathbb{1}_{\{\|X_{(k_{\mathrm{L}})}-x\| > r\}}\}. \tag{37}$$

Now, by similar arguments to those leading to (17), we have that

$$\frac{\epsilon}{k_{\mathrm{L}}} \sum_{i=1}^{k_{\mathrm{L}}} \mathbb{E}(\|X_{(i)} - x\|^2 \mathbb{1}_{\{\|X_{(k_{\mathrm{L}})}-x\| \leq r\}}) = \epsilon \Big(\frac{k_{\mathrm{L}}}{n a_d \bar{f}(x)}\Big)^{2/d} \frac{d}{d+2}\{1 + o(1)\}, \tag{38}$$

uniformly for $P \in \mathcal{P}_{d,\theta}$, $k_{\mathrm{L}} \in K_{\beta,\tau}$, $x_0 \in \mathcal{S}_n$ and $|t| < \epsilon_n$. Moreover, for every $M > 0$,

$$\mathbb{P}\{\|X_{(k_{\mathrm{L}})} - x\| > r\} = q_r^n(k_{\mathrm{L}}) = O(n^{-M}), \tag{39}$$

uniformly for $P \in \mathcal{P}_{d,\theta}$, $k_{\mathrm{L}} \in K_{\beta,\tau}$, $x_0 \in \mathcal{S}_n$ and $|t| < \epsilon_n$, by (16) in Step 1. For the remaining terms, note that

$$\mathbb{E}\{\|X_{(k_{\mathrm{L}})} - x\|^2 \mathbb{1}_{\{\|X_{(k_{\mathrm{L}})}-x\| > r\}}\}$$
$$= \mathbb{P}\{\|X_{(k_{\mathrm{L}})} - x\| > r\} + \int_{r^2}^{\infty} \mathbb{P}\{\|X_{(k_{\mathrm{L}})} - x\| > \sqrt{t}\}\, dt$$
$$= q_r^n(k_{\mathrm{L}}) + \int_{r^2}^{\infty} q_{\sqrt{t}}^n(k_{\mathrm{L}})\, dt. \tag{40}$$

Let $t_0 = t_0(x) := 5^{2/\rho}(1 + 2^{\rho-1})^{2/\rho}\big(M_0 + \|x\|^\rho\big)^{2/\rho}$. Then, for $t \geq t_0$, we have

$$1 - p_{\sqrt{t}} \leq (1 + 2^{\rho-1}) \frac{\mathbb{E}(\|X\|^\rho) + \|x\|^\rho}{t^{\rho/2}} \leq \frac{1}{5}.$$

It follows by Bennett's inequality that for $\rho\{n - (n-1)^{1-\beta}\} > 4$,

$$\int_{t_0}^{\infty} q_{\sqrt{t}}^n(k_{\mathrm{L}})\, dt$$
$$\leq e^{k_{\mathrm{L}}}(1 + 2^{\rho-1})^{(n-k_{\mathrm{L}})/2}\big\{M_0 + \|x\|^\rho\big\}^{(n-k_{\mathrm{L}})/2} \int_{t_0}^{\infty} t^{-\rho(n-k_{\mathrm{L}})/4}\, dt$$
$$= \frac{4e^{k_{\mathrm{L}}}5^{2/\rho}}{\rho(n-k_{\mathrm{L}}) - 4}(1 + 2^{\rho-1})^{2/\rho}\big\{M_0 + \|x\|^\rho\big\}^{2/\rho} 5^{-(n-k_{\mathrm{L}})/2}.$$

But, when $\beta\log(n-1) \geq (d+2)/d$ and $n \geq \max\{n_0, n_2\}$,

$$\sup_{x \in \mathcal{R}_n \cup \mathcal{S}_n^{\epsilon_n}} \|x\| \leq \epsilon_0 + \biggl\{\frac{(n-1)^{1-\beta} c_n^d M_0}{\mu_0 \beta^{d/2} \log^{d/2}(n-1)}\biggr\}^{1/\rho}.$$

We deduce that for every $M > 0$,

$$\sup_{P \in \mathcal{P}_{d,\theta}} \sup_{k \in K_{\beta,\tau}} \sup_{x \in \mathcal{R}_n \cup \mathcal{S}_n^{\epsilon_n}} \int_{t_0}^{\infty} q_{\sqrt{t}}^n(k_{\mathrm{L}})\,dt = O(n^{-M}). \tag{41}$$

Moreover, by Bernstein's inequality, for every $M > 0$,

$$\sup_{P \in \mathcal{P}_{d,\theta}} \sup_{k_{\mathrm{L}} \in K_{\beta,\tau}} \sup_{x \in \mathcal{R}_n \cup \mathcal{S}_n^{\epsilon_n}} \biggl\{q_r^n(k_{\mathrm{L}}) + \int_{r^2}^{t_0} q_{\sqrt{t}}^n(k_{\mathrm{L}})\,dt\biggr\} = O(n^{-M}). \tag{42}$$

We conclude from (14), (37), (38), (39), (40), (41) and (42), together with Jensen's inequality to deal with the third term on the right-hand side of (37), that (10) holds. With only simple modifications, we have also shown (13), which bounds R_2.

To bound R_5: Write

$$\begin{aligned} R_5 &:= \int_{\mathcal{S}_n} R_5(x_0)\,d\mathrm{Vol}^{d-1}(x_0) \\ &= \int_{\mathcal{S}_n} \int_{-\epsilon_n}^{\epsilon_n} t\|\dot{\psi}(x_0)\| \Bigl[\mathbb{P}\{\hat{S}_n(x_0^t) < 1/2\} - \mathbb{E}\Phi\bigl(\hat{\theta}(x_0^t)\bigr)\Bigr]\,dt\,d\mathrm{Vol}^{d-1}(x_0). \end{aligned}$$

Now by a non-uniform version of the Berry–Esseen theorem (Paditz, 1989, Theorem 1), for every $t \in (-\epsilon_n, \epsilon_n)$ and $x_0 \in \mathcal{S}_n$,

$$\bigl|\mathbb{P}\{\hat{S}_n(x_0^t) < 1/2 | X^n\} - \Phi\bigl(\hat{\theta}(x_0^t)\bigr)\bigr| \leq \frac{32}{k_{\mathrm{L}}(x_0^t)\hat{\sigma}_n(x_0^t, X^n)} \frac{1}{1 + |\hat{\theta}(x_0^t)|^3}. \tag{43}$$

Let

$$t_n = t_n(x_0) := C\max\biggl\{k_{\mathrm{L}}(x_0)^{-1/2}, \Bigl(\frac{k_{\mathrm{L}}(x_0)}{n\bar{f}(x_0)}\Bigr)^{2/d} \ell\bigl(\bar{f}(x_0)\bigr)\biggr\},$$

where

$$C := \frac{4}{a_d^{2/d}\epsilon_0}.$$

In the following we integrate the bound in (43) over the regions $|t| \leq t_n$ and $|t| \in (t_n, \epsilon_n)$ separately. Define the event

$$B_{k_{\mathrm{L}}} := \biggl\{\hat{\sigma}_n(x_0^t, X^n) \geq \frac{1}{3k_{\mathrm{L}}(x_0^t)^{1/2}} \text{ for all } x_0 \in \mathcal{S}_n,\, t \in (-\epsilon_n, \epsilon_n)\biggr\},$$

so that, by very similar arguments to those used to bound $\mathbb{P}(A_{k_{\mathrm{L}}}^c)$ in Step 2, we have $\mathbb{P}(B_{k_{\mathrm{L}}}^c) = O(n^{-M})$ for every $M > 0$, uniformly for $P \in \mathcal{P}_{d,\theta}$ and $k_{\mathrm{L}} \in K_{\beta,\tau}$. It follows by (43) and Step 2 that there exists $n_4 \in \mathbb{N}$ such that for all $n \geq n_4$, $k_{\mathrm{L}} \in K_{\beta,\tau}$ and $x_0 \in \mathcal{S}_n$,

$$\Bigl|\int_{-t_n}^{t_n} t\Bigl[\mathbb{P}\{\hat{S}_n(x_0^t) < 1/2\} - \mathbb{E}\Phi\bigl(\hat{\theta}(x_0^t)\bigr)\Bigr]\,dt\Bigr|$$

$$\leq \int_{-t_n}^{t_n} \mathbb{E}\biggl(\frac{32|t|\mathbb{1}_{B_{k_{\mathrm{L}}}}}{k_{\mathrm{L}}(x_0^t)\hat{\sigma}_n(x_0^t, X^n)}\biggr)\,dt + t_n^2\mathbb{P}(B_{k_{\mathrm{L}}}^c) \leq \frac{128t_n^2}{k_{\mathrm{L}}(x_0)^{1/2}}. \tag{44}$$

By Step 1, there exists $n_5 \in \mathbb{N}$ such that for $n \geq n_5$, $P \in \mathcal{P}_{d,\theta}$, $k_{\mathrm{L}} \in K_{\beta,\tau}$, $x_0 \in \mathcal{S}_n$ and $|t| \in (t_n, \epsilon_n)$,

$$\begin{aligned}|\mu_n(x_0^t) - 1/2| &\geq |\eta(x_0^t) - 1/2| - |\mu_n(x_0^t) - \eta(x_0^t)| \\ &\geq \frac{1}{2}\inf_{z \in \mathcal{S}}\|\dot{\eta}(z)\||t| - \frac{1}{4}C\epsilon_0 M_0\Bigl(\frac{k_{\mathrm{L}}(x_0)}{n\bar{f}(x_0)}\Bigr)^{2/d}\ell\bigl(\bar{f}(x_0)\bigr) \\ &> \frac{1}{4}\epsilon_0 M_0|t|. \end{aligned} \tag{45}$$

Thus for $n \geq n_5$, $P \in \mathcal{P}_{d,\theta}$, $k_{\mathrm{L}} \in K_{\beta,\tau}$, $x_0 \in \mathcal{S}_n$ and $|t| \in (t_n, \epsilon_n)$, we have that

$$\begin{aligned}&\mathbb{P}\Bigl\{|\hat{\theta}(x_0^t)| < \frac{1}{4}\epsilon_0 M_0 k_{\mathrm{L}}^{1/2}(x_0)|t|\Bigr\} \\ &\quad\leq \mathbb{P}\Bigl\{|\hat{\mu}_n(x_0^t, X^n) - \mu_n(x_0^t)| > |\mu_n(x_0^t) - 1/2| - \frac{1}{8}\epsilon_0 M_0|t|\Bigr\} \\ &\quad\leq \mathbb{P}\Bigl\{|\hat{\mu}_n(x_0^t, X^n) - \mu_n(x_0^t)| > \frac{1}{8}\epsilon_0 M_0|t|\Bigr\} \leq \frac{64\mathrm{Var}\{\hat{\mu}_n(x_0^t, X^n)\}}{\epsilon_0^2 M_0^2 t^2}. \end{aligned} \tag{46}$$

It follows by (43), (46) and Step 3 that, for $n \geq n_5$,

$$\begin{aligned}&\Bigl|\int_{|t| \in (t_n,\epsilon_n)} t\bigl[\mathbb{P}\{\hat{S}_n(x_0^t) < 1/2\} - \mathbb{E}\Phi\bigl(\hat{\theta}(x_0^t)\bigr)\bigr]\,dt\Bigr| \\ &\quad\leq \int_{|t| \in (t_n,\epsilon_n)} |t|\mathbb{E}\biggl(\frac{32\mathbb{1}_{B_{k_{\mathrm{L}}}}}{k_{\mathrm{L}}(x_0^t)\hat{\sigma}_n(x_0^t, X^n)}\frac{1}{1+\frac{1}{64}\epsilon_0^3 M_0^3 k_{\mathrm{L}}(x_0)^{3/2}|t|^3}\biggr)\,dt \\ &\qquad + \int_{|t| \in (t_n,\epsilon_n)} \frac{64\mathrm{Var}\{\hat{\mu}_n(x_0^t, X^n)\}}{\epsilon_0^2 M_0^2|t|}\,dt + \epsilon_n^2\mathbb{P}(B_{k_{\mathrm{L}}}^c) \\ &\quad\leq \frac{192}{k_{\mathrm{L}}(x_0)^{3/2}}\int_0^\infty \frac{u}{1 + \frac{1}{64}\epsilon_0^3 M_0^3 u^3}\,du \\ &\qquad + \frac{128}{\epsilon_0^2 M_0^2}\sup_{|t| \in (t_n,\epsilon_n)} \mathrm{Var}\{\hat{\mu}_n(x_0^t, X^n)\}\log\Bigl(\frac{\epsilon_n}{t_n}\Bigr) + \epsilon_n^2\mathbb{P}(B_{k_{\mathrm{L}}}^c) \\ &\quad= o\Bigl(\frac{1}{k_{\mathrm{L}}(x_0)}\Bigr) \end{aligned} \tag{47}$$

uniformly for $P \in \mathcal{P}_{d,\theta}$, $k_{\mathrm{L}} \in K_{\beta,\tau}$ and $x_0 \in \mathcal{S}_n$. We conclude from (44) and (47) that $|R_5| = o(\gamma_n(k_{\mathrm{L}}))$, uniformly for $P \in \mathcal{P}_{d,\theta}$ and $k_{\mathrm{L}} \in K_{\beta,\tau}$.

To bound R_6: Let $\theta(x_0^t) := -2k_{\mathrm{L}}(x_0^t)^{1/2}\{\mu_n(x_0^t) - 1/2\}$. Write

$$R_6 := \int_{\mathcal{S}_n} R_6(x_0)\, d\mathrm{Vol}^{d-1}(x_0) = R_{61} + R_{62},$$

where

$$R_{61} := \int_{\mathcal{S}_n} \int_{-\epsilon_n}^{\epsilon_n} t\|\dot{\psi}(x_0)\| \Big[\mathbb{E}\Phi\big(\hat{\theta}(x_0^t)\big) - \Phi\big(\theta(x_0^t)\big)\Big]\, dt\, d\mathrm{Vol}^{d-1}(x_0)$$

and

$$R_{62} := \int_{\mathcal{S}_n} \int_{-\epsilon_n}^{\epsilon_n} t\|\dot{\psi}(x_0)\| \Big[\Phi\big(\theta(x_0^t)\big) - \Phi\big(\bar{\theta}(x_0,t)\big)\Big]\, dt\, d\mathrm{Vol}^{d-1}(x_0).$$

To bound R_{61}: We again deal with the regions $|t| \leq t_n$ and $|t| \in (t_n, \epsilon_n)$ separately. First let $\tilde{\theta}(x_0^t) := -2k_{\mathrm{L}}(x_0^t)^{1/2}\{\hat{\mu}_n(x_0^t, X^n) - 1/2\}$. Writing ϕ for the standard normal density, and using the facts that $|\hat{\theta}(x_0^t)| \geq |\tilde{\theta}(x_0^t)|$, that $\hat{\theta}(x_0^t)$ and $\tilde{\theta}(x_0^t)$ have the same sign, and that $|x\phi(x)| \leq 1$, we have

$$\begin{aligned}
&\bigg|\int_{-t_n}^{t_n} t\big[\mathbb{E}\Phi\big(\hat{\theta}(x_0^t)\big) - \Phi\big(\theta(x_0^t)\big)\big]\, dt\bigg| \\
&\leq \int_{-t_n}^{t_n} |t|\mathbb{E}\Big\{|\hat{\theta}(x_0^t) - \tilde{\theta}(x_0^t)|\phi\big(\tilde{\theta}(x_0^t)\big)\mathbb{1}_{A_{k_{\mathrm{L}}}} + |\tilde{\theta}(x_0^t) - \theta(x_0^t)|\Big\}\, dt + t_n^2\mathbb{P}(A_{k_{\mathrm{L}}}^c) \\
&\leq \int_{-t_n}^{t_n} |t|\bigg[\mathbb{E}\bigg\{\mathbb{1}_{A_{k_{\mathrm{L}}}}\bigg|\frac{1}{2k_{\mathrm{L}}(x_0^t)^{1/2}\hat{\sigma}_n(x_0^t, X^n)} - 1\bigg|\bigg\} \\
&\qquad\qquad + 2k_{\mathrm{L}}(x_0^t)^{1/2}\mathrm{Var}^{1/2}\{\hat{\mu}_n(x_0^t, X^n)\}\bigg]\, dt + t_n^2\mathbb{P}(A_{k_{\mathrm{L}}}^c) = o(t_n^2)
\end{aligned}$$

uniformly for $P \in \mathcal{P}_{d,\theta}$, $k_{\mathrm{L}} \in K_{\beta,\tau}$ and $x_0 \in \mathcal{S}_n$. Note that for $|t| \in (t_n, \epsilon_n)$ and $x_0 \in \mathcal{S}_n$, we have when $\epsilon_n < \epsilon_0$ and $n \geq n_5$ that

$$\begin{aligned}
&\mathbb{E}\big\{\mathbb{1}_{A_{k_{\mathrm{L}}} \cap B_{k_{\mathrm{L}}}}\big|\hat{\theta}(x_0^t) - \theta(x_0^t)\big|\big\} \\
&\qquad \leq \mathbb{E}\bigg\{\frac{\mathbb{1}_{A_{k_{\mathrm{L}}} \cap B_{k_{\mathrm{L}}}}}{\hat{\sigma}_n(x_0^t, X^n)}|\hat{\mu}_n(x_0^t, X^n) - \mu_n(x_0^t)| \\
&\qquad\qquad + \mathbb{1}_{A_{k_{\mathrm{L}}} \cap B_{k_{\mathrm{L}}}}|\theta(x_0^t)|\bigg|\frac{1}{2k_{\mathrm{L}}(x_0^t)^{1/2}\hat{\sigma}_n(x_0^t, X^n)} - 1\bigg|\bigg\} \\
&\qquad \leq 3k_{\mathrm{L}}(x_0)^{1/2}\mathrm{Var}^{1/2}\{\hat{\mu}_n(x_0^t, X^n)\} \\
&\qquad\quad + \frac{5}{2}k_{\mathrm{L}}(x_0)^{1/2}M_0|t|\mathbb{E}\bigg\{\mathbb{1}_{A_{k_{\mathrm{L}}} \cap B_{k_{\mathrm{L}}}}\bigg|\frac{1}{2k_{\mathrm{L}}(x_0^t)^{1/2}\hat{\sigma}_n(x_0^t, X^n)} - 1\bigg|\bigg\}. \qquad (48)
\end{aligned}$$

Thus by (45), (46), (48) and Step 3, for $\epsilon_n < \epsilon_0$ and $n \geq n_5$,

$$\int_{|t|\in(t_n,\epsilon_n)} |t|\bigl|\mathbb{E}\Phi\bigl(\hat{\theta}(x_0^t)\bigr) - \Phi\bigl(\theta(x_0^t)\bigr)\bigr|\, dt$$

$$\leq \int_{|t|\in(t_n,\epsilon_n)} |t|\mathbb{E}\Bigl\{\mathbb{1}_{A_{k_{\mathrm{L}}}\cap B_{k_{\mathrm{L}}}}\bigl|\hat{\theta}(x_0^t) - \theta(x_0^t)\bigr|\Bigr\}\phi\Bigl(\frac{1}{4}\epsilon_0 M_0 k_{\mathrm{L}}^{1/2}(x_0)|t|\Bigr)\, dt$$

$$+ \mathbb{P}(A_{k_{\mathrm{L}}}^c \cup B_{k_{\mathrm{L}}}^c) + \frac{128}{\epsilon_0^2 M_0^2}\sup_{|t|\in(t_n,\epsilon_n)} \mathrm{Var}\{\hat{\mu}_n(x_0^t, X^n)\}\log\Bigl(\frac{\epsilon_n}{t_n}\Bigr)$$

$$= o\Bigl(\frac{1}{k_{\mathrm{L}}(x_0)}\Bigr) \tag{49}$$

uniformly for $P \in \mathcal{P}_{d,\theta}$, $k_{\mathrm{L}} \in K_{\beta,\tau}$ and $x_0 \in \mathcal{S}_n$.

To bound R_{62}: Let

$$u(x) \equiv u_n(x) := k_{\mathrm{L}}(x)^{1/2}\Bigl(\frac{k_{\mathrm{L}}(x)}{n\bar{f}(x)}\Bigr)^{2/d}.$$

Given $\epsilon > 0$ small enough that $\epsilon^2 + \frac{\epsilon}{2\epsilon_0} < 1/2$, by Step 1 there exists $n_6 \in \mathbb{N}$ such that for $n \geq n_6$, $P \in \mathcal{P}_{d,\theta}$, $k_{\mathrm{L}} \in K_{\beta,\tau}$, $x_0 \in \mathcal{S}_n$ and $|t| < \epsilon_n$,

$$\bigl|\theta(x_0^t) - \bar{\theta}(x_0,t)\bigr| \leq \epsilon^2\bigl\{|t|k_{\mathrm{L}}(x_0)^{1/2} + u(x_0)\ell\bigl(\bar{f}(x_0)\bigr)\bigr\}.$$

By decreasing ϵ and increasing n_6 if necessary, it follows that

$$\Bigl|\Phi\bigl(\theta(x_0^t)\bigr) - \Phi\bigl(\bar{\theta}(x_0,t)\bigr)\Bigr| \leq \epsilon^2\bigl\{|t|k_{\mathrm{L}}(x_0)^{1/2} + u(x_0)\ell\bigl(\bar{f}(x_0)\bigr)\bigr\}\phi\Bigl(\frac{1}{2}\bar{\theta}(x_0,t)\Bigr),$$

for all $n \geq n_6$, $P \in \mathcal{P}_{d,\theta}$, $k_{\mathrm{L}} \in K_{\beta,\tau}$, $x_0 \in \mathcal{S}_n$ and $t \in (-\epsilon_n, \epsilon_n)$ satisfying $2\epsilon u(x_0)\ell\bigl(\bar{f}(x_0)\bigr)\|\dot{\eta}(x_0)\| \leq |\bar{\theta}(x_0,t)|$. Substituting $u = \bar{\theta}(x_0,t)/2$, it follows that there exists $C^* > 0$ such that for all $n \geq n_6$, $P \in \mathcal{P}_{d,\theta}$ and $k_{\mathrm{L}} \in K_{\beta,\tau}$,

$$|R_{62}|$$

$$\leq \int_{\mathcal{S}_n}\int_{|u|\leq\epsilon u(x_0)\ell(\bar{f}(x_0))\|\dot{\eta}(x_0)\|} \frac{2\bar{f}(x_0)}{\|\dot{\eta}(x_0)\|k_{\mathrm{L}}(x_0)}|u + u(x_0)a(x_0)|\, du\, d\mathrm{Vol}^{d-1}(x_0)$$

$$+ \int_{\mathcal{S}_n}\int_{-\infty}^{\infty} \frac{2\bar{f}(x_0)|u + u(x_0)a(x_0)|}{\|\dot{\eta}(x_0)\|^2 k_{\mathrm{L}}(x_0)}\bigl\{\epsilon^2|u + u(x_0)a(x_0)|$$

$$+ \epsilon|u|\bigr\}\phi(u)\, du\, d\mathrm{Vol}^{d-1}(x_0) \leq C^*\epsilon\gamma_n(k_{\mathrm{L}}). \tag{50}$$

The combination of (49) and (50) yields the desired error bound on $|R_6|$ in (26), uniformly for $P \in \mathcal{P}_{d,\theta}$, $k_{\mathrm{L}} \in K_{\beta,\tau}$, and therefore completes the proof. □

6. Empirical analysis. In this section, we compare the k_{O}nn and k_{SS}nn classifiers, introduced in Section 4 of the main text, with the standard $k$nn classifier studied in Section 3 of the main text. We investigate three settings that reflect the differences between the main results in these sections.

- Setting 1: P_1 is the distribution of d independent $N(0,1)$ components; whereas P_0 is the distribution of d independent $N(1,1/4)$ components.
- Setting 2: P_1 is the distribution of d independent t_5 components; P_0 is the distribution of d independent components, the first $\lfloor d/2 \rfloor$ having a t_5 distribution and the remainder having a $N(1,1)$ distribution.
- Setting 3: P_1 is the distribution of d independent standard Cauchy components; P_0 is the distribution of d independent components, the first $\lfloor d/2 \rfloor$ being standard Cauchy and the remainder standard normal.

The corresponding marginal distribution P_X in Setting 1 satisfies **(A.4)** for every $\rho > 0$. Hence, for the standard k-nearest neighbour classifier when $d \geq 5$, we are in the setting of Theorem 1(i), while for $d \leq 4$, we can only appeal to Theorem 1(ii). On the other hand, for the local-k-nearest neighbour classifiers, the results of Theorems 2(i) and 3(i) apply for all dimensions, and we can expect the excess risk to converge to zero at rate $O(n^{-4/(d+4)})$. In Setting 2, **(A.4)** holds for $\rho < 5$, but not for $\rho \geq 5$. Thus, for the standard k-nearest neighbour classifier, we are in the setting of Theorem 1(ii) for $d < 20$, whereas Theorems 2(i) and 3(i) again apply for all dimensions for the local classifiers. Finally, in Setting 3, **(A.4)** does not hold for any $\rho \geq 1$, and only the conditions of Theorems 1(ii), 2(ii) and 3(ii) apply.

For the standard $k$nn classifier, we use 5-fold cross validation to choose k, based on a sequence of equally-spaced values between 1 and $\lfloor n/4 \rfloor$ of length at most 40. For the oracle classifier, we set

$$\hat{k}_{\mathrm{O}}(x) := \max\Big[1, \min\big[\lfloor \hat{B}_{\mathrm{O}}\{\bar{f}(x)n/\|\bar{f}\|_\infty\}^{4/(d+4)} \rfloor, n/2\big]\Big],$$

where $\hat{B}_{\mathrm{O}}$ was again chosen via 5-fold cross validation, but based on a sequence of 40 equally-spaced points between $n^{-4/(d+4)}$ (corresponding to the 1-nearest neighbour classifier) and $n^{d/(d+4)}$. Similarly, for the semi-supervised classifier, we set

$$\hat{k}_{\mathrm{SS}}(x) := \max\Big[1, \min\big[\lfloor \hat{B}_{\mathrm{SS}}\{\hat{f}_m(x)n/\|\hat{f}_m\|_\infty\}^{4/(d+4)} \rfloor, n/2\big]\Big],$$

where $\hat{B}_{\mathrm{SS}}$ was chosen analogously to $\hat{B}_{\mathrm{O}}$, and where $\hat{f}_m$ is the d-dimensional kernel density estimator constructed using a truncated normal kernel and bandwidths chosen via the default method in the `R` package `ks` (Duong,

TABLE 1
Misclassification rates for Settings 1, 2 and 3. In the final two columns we present the regret ratios given in (51) (with standard errors calculated via the delta method).

d	Bayes risk	n	$\hat{k}$nn risk	$\hat{k}_{\mathrm{O}}$nn risk	$\hat{k}_{\mathrm{SS}}$nn risk	O RR	SS RR
Setting 1							
1	22.67	50	$26.85_{0.13}$	$25.91_{0.12}$	$25.98_{0.13}$	$0.78_{0.022}$	$0.79_{0.023}$
		200	$24.07_{0.06}$	$23.52_{0.06}$	$23.48_{0.05}$	$0.61_{0.030}$	$0.58_{0.029}$
		1000	$23.20_{0.04}$	$22.93_{0.04}$	$22.94_{0.04}$	$0.48_{0.048}$	$0.50_{0.048}$
2	13.30	50	$17.70_{0.09}$	$16.96_{0.08}$	$16.95_{0.08}$	$0.83_{0.015}$	$0.83_{0.015}$
		200	$15.09_{0.05}$	$14.69_{0.04}$	$14.74_{0.05}$	$0.77_{0.018}$	$0.80_{0.019}$
		1000	$14.04_{0.04}$	$13.78_{0.03}$	$13.80_{0.03}$	$0.65_{0.025}$	$0.67_{0.025}$
5	3.53	50	$9.46_{0.07}$	$8.95_{0.06}$	$8.94_{0.06}$	$0.91_{0.006}$	$0.91_{0.006}$
		200	$6.94_{0.03}$	$6.67_{0.03}$	$6.70_{0.03}$	$0.92_{0.006}$	$0.93_{0.007}$
		1000	$5.49_{0.02}$	$5.18_{0.02}$	$5.23_{0.02}$	$0.84_{0.008}$	$0.87_{0.008}$
Setting 2							
1	31.16	50	$36.55_{0.14}$	$36.07_{0.14}$	$35.93_{0.14}$	$0.91_{0.020}$	$0.88_{0.020}$
		200	$32.93_{0.08}$	$32.38_{0.07}$	$32.42_{0.07}$	$0.69_{0.031}$	$0.71_{0.032}$
		1000	$31.62_{0.05}$	$31.37_{0.05}$	$31.37_{0.05}$	$0.46_{0.065}$	$0.47_{0.066}$
2	31.15	50	$37.79_{0.13}$	$38.02_{0.12}$	$37.90_{0.12}$	$1.02_{0.014}$	$1.01_{0.015}$
		200	$33.64_{0.08}$	$33.63_{0.07}$	$33.54_{0.07}$	$1.00_{0.028}$	$0.96_{0.026}$
		1000	$31.83_{0.05}$	$31.81_{0.05}$	$31.80_{0.05}$	$0.97_{0.039}$	$0.95_{0.038}$
5	20.10	50	$28.74_{0.12}$	$29.16_{0.12}$	$29.13_{0.11}$	$1.05_{0.011}$	$1.05_{0.011}$
		200	$23.60_{0.06}$	$23.75_{0.06}$	$23.93_{0.06}$	$1.04_{0.014}$	$1.09_{0.015}$
		1000	$21.86_{0.04}$	$21.71_{0.04}$	$21.77_{0.04}$	$0.91_{0.014}$	$0.95_{0.014}$
Setting 3							
1	37.44	50	$44.76_{0.10}$	$43.09_{0.12}$	$43.08_{0.12}$	$0.77_{0.013}$	$0.77_{0.013}$
		200	$41.86_{0.08}$	$40.18_{0.09}$	$40.23_{0.09}$	$0.62_{0.017}$	$0.63_{0.017}$
		1000	$38.68_{0.06}$	$37.85_{0.05}$	$37.89_{0.05}$	$0.33_{0.033}$	$0.36_{0.032}$
2	37.45	50	$46.20_{0.09}$	$44.81_{0.10}$	$45.24_{0.10}$	$0.84_{0.009}$	$0.89_{0.009}$
		200	$43.50_{0.07}$	$42.29_{0.08}$	$42.86_{0.08}$	$0.80_{0.011}$	$0.89_{0.011}$
		1000	$40.53_{0.06}$	$39.64_{0.06}$	$39.96_{0.06}$	$0.71_{0.013}$	$0.82_{0.014}$
5	23.23	50	$41.56_{0.11}$	$38.13_{0.11}$	$39.26_{0.12}$	$0.81_{0.005}$	$0.87_{0.005}$
		200	$36.02_{0.07}$	$33.34_{0.06}$	$34.68_{0.07}$	$0.79_{0.004}$	$0.90_{0.004}$
		1000	$31.46_{0.05}$	$29.91_{0.05}$	$30.58_{0.05}$	$0.81_{0.004}$	$0.89_{0.004}$

2015). In practice, we estimated $\|\hat{f}_m\|_\infty$ by the maximum value attained on the unlabelled training set.

In each of the three settings above, we generated a training set of size $n \in \{50, 200, 1000\}$ in dimensions $d \in \{1, 2, 5\}$, an unlabelled training set of size 1000, and a test set of size 1000. In Table 1, we present the sample mean and standard error (in subscript) of the risks computed from 1000 repetitions of each experiment. Further, we present estimates of the regret ratios, given by

$$\frac{R(\hat{C}_n^{\hat{k}_{\mathrm{O}}\mathrm{nn}}) - R(C^{\mathrm{Bayes}})}{R(\hat{C}_n^{\hat{k}\mathrm{nn}}) - R(C^{\mathrm{Bayes}})} \quad \text{and} \quad \frac{R(\hat{C}_n^{\hat{k}_{\mathrm{SS}}\mathrm{nn}}) - R(C^{\mathrm{Bayes}})}{R(\hat{C}_n^{\hat{k}\mathrm{nn}}) - R(C^{\mathrm{Bayes}})}, \tag{51}$$

for which the standard errors given are estimated via the delta method. From Table 1, we saw improvement in performance from the oracle and semi-supervised classifiers in 22 of the 27 experiments, comparable performance in three experiments, and there were two where the standard $k$nn classifier was the best of the three classifiers considered. In those latter two cases, the theoretical improvement expected for the local classifiers is small; for instance, when $d = 5$ in Setting 2, the excess risk for the local classifiers converges at rate $O(n^{-4/9})$, while the standard k-nearest neighbour classifier can attain a rate at least as fast as $o(n^{-1/3+\epsilon})$ for every $\epsilon > 0$. It is therefore perhaps unsurprising that we require the larger sample size of $n = 1000$ for the local classifiers to yield an improvement in this case. The semi-supervised classifier exhibits similar performance to the oracle classifier in all settings, though some deterioration is noticeable in higher dimensions, where it is harder to construct a good estimate of $\bar{f}$ from the unlabelled training data.

7. An introduction to differential geometry, tubular neighbourhoods and integration on manifolds. The purpose of this section is to give a brief introduction to the ideas from differential geometry, specifically tubular neighbourhoods and integration on manifolds, which play an important role in our analysis of misclassification error rates, but which we expect are unfamiliar to many statisticians. For further details and several of the proofs, we refer the reader to the many excellent texts on these topics, e.g. Guillemin and Pollack (1974), Gray (2004).

7.1. *Manifolds and regular values.* Recall that if $\mathcal{X}$ is an arbitrary subset of $\mathbb{R}^M$, we say $\phi : \mathcal{X} \to \mathbb{R}^N$ is *differentiable* if for each $x \in \mathcal{X}$, there exists an open subset $U \subseteq \mathbb{R}^M$ containing x and a differentiable function $F : U \to \mathbb{R}^N$ such that $F(z) = \phi(z)$ for $z \in U \cap \mathcal{X}$. If $\mathcal{Y}$ is also a subset of $\mathbb{R}^M$, we say $\phi : \mathcal{X} \to \mathcal{Y}$ is a *diffeomorphism* if ϕ is bijective and differentiable and if its inverse ϕ^{-1} is also differentiable. We then say $\mathcal{S} \subseteq \mathbb{R}^d$ is an *m-dimensional manifold* if for each $x \in \mathcal{S}$, there exist an open subset $U_x \subseteq \mathbb{R}^m$, a neighbourhood V_x of x in $\mathcal{S}$ and a diffeomorphism $\phi_x : U_x \to V_x$. Such a diffeomorphism ϕ_x is called a *local parametrisation* of $\mathcal{S}$ around x, and we sometimes suppress the dependence of ϕ_x, U_x and V_x on x. It turns out that the specific choice of local parametrisation is usually not important, and properties of the manifold are well-defined regardless of the choice made.

Let $\mathcal{S} \subseteq \mathbb{R}^d$ be an m-dimensional manifold and let $\phi : U \to \mathcal{S}$ be a local parametrisation of $\mathcal{S}$ around $x \in \mathcal{S}$, where U is an open subset of $\mathbb{R}^m$. Assume that $\phi(0) = x$ for convenience. The *tangent space* $T_x(\mathcal{S})$ to $\mathcal{S}$ at x is defined to be the image of the derivative $D\phi_0 : \mathbb{R}^m \to \mathbb{R}^d$ of ϕ at 0. Thus $T_x(\mathcal{S})$ is the m-dimensional subspace of $\mathbb{R}^d$ whose parallel translate

$x + T_x(\mathcal{S})$ is the best affine approximation to $\mathcal{S}$ through x, and $(D\phi_0)^{-1}$ is well-defined as a map from $T_x(\mathcal{S})$ to $\mathbb{R}^m$. If $f : \mathcal{S} \to \mathbb{R}$ is differentiable, we define the derivative $Df_x : T_x(\mathcal{S}) \to \mathbb{R}$ of f at x by $Df_x := Dh_0 \circ (D\phi_0)^{-1}$, where $h := f \circ \phi$.

In practice, it is usually rather inefficient to define manifolds through explicit diffeomorphisms. Instead, we can often obtain them as level sets of differentiable functions. Suppose that $\mathcal{R} \subseteq \mathbb{R}^d$ is a manifold and $\eta : \mathcal{R} \to \mathbb{R}$ is differentiable. We say $y \in \mathbb{R}$ is a *regular value* for η if $\mathrm{image}(D\eta_x) = \mathbb{R}$ for every $x \in \mathcal{R}$ for which $\eta(x) = y$. If $y \in \mathbb{R}$ is a regular value of η, then $\eta^{-1}(y)$ is a $(d-1)$-dimensional submanifold of $\mathcal{R}$ (Guillemin and Pollack, 1974, p. 21).

7.2. *Tubular neighbourhoods of level sets.* For any set $\mathcal{S} \subseteq \mathbb{R}^d$ and $\epsilon > 0$, we call $\mathcal{S} + \epsilon B_1(0)$ the *ϵ-neighbourhood* of $\mathcal{S}$. In circumstances where $\mathcal{S}$ is a $(d-1)$-dimensional manifold defined by the level set of a continuously differentiable function $\eta : \mathbb{R}^d \to \mathbb{R}$ with non-vanishing derivative on $\mathcal{S}$, the set $\mathcal{S}^\epsilon$ is often called a *tubular neighbourhood*, and $\dot\eta(x)^T v = 0$ for all $x \in \mathcal{S}$ and $v \in T_x(\mathcal{S})$. We therefore have the following useful representation of the ϵ-neighbourhood of $\mathcal{S}$ in terms of points on $\mathcal{S}$ and a perturbation in a normal direction.

PROPOSITION 2. *Let $\eta : \mathbb{R}^d \to [0,1]$, suppose that $\mathcal{S} := \{x \in \mathbb{R}^d : \eta(x) = 1/2\}$ is non-empty, and suppose further that η is continuously differentiable on $\mathcal{S} + \epsilon B_1(0)$ for some $\epsilon > 0$, with $\dot\eta(x) \neq 0$ for all $x \in \mathcal{S}$, so that $\mathcal{S}$ is a $(d-1)$-dimensional manifold. Then*

$$\mathcal{S} + \epsilon B_1(0) = \left\{ x_0 + \frac{t\dot\eta(x_0)}{\|\dot\eta(x_0)\|} : x_0 \in \mathcal{S}, |t| < \epsilon \right\} =: \mathcal{S}^\epsilon.$$

PROOF. For any $x_0 \in \mathcal{S}$ and $|t| < \epsilon$, we have $x_0 + t\dot\eta(x_0)/\|\dot\eta(x_0)\| \in \mathcal{S} + \epsilon B_1(0)$. On the other hand, suppose that $x \in \mathcal{S} + \epsilon B_1(0)$. Since $\mathcal{S}$ is closed, there exists $x_0 \in \mathcal{S}$ such that $\|x - x_0\| \leq \|x - y\|$ for all $y \in \mathcal{S}$. Rearranging this inequality yields that, for $y \neq x_0$,

$$2(x - x_0)^T \frac{(y - x_0)}{\|y - x_0\|} \leq \|y - x_0\|. \tag{52}$$

Let U be an open subset of $\mathbb{R}^{d-1}$ and $\phi : U \to \mathcal{S}$ be a local parametrisation of $\mathcal{S}$ around x_0, where without loss of generality we assume $\phi(0) = x_0$. Let $v \in T_{x_0}(\mathcal{S}) \setminus \{0\}$ be given and let $h \in \mathbb{R}^{d-1} \setminus \{0\}$ be such that $D\phi_0(h) = v$. Then for $t > 0$ sufficiently small we have $th \in U$, so by (52),

$$2(x - x_0)^T \frac{\{\phi(th) - \phi(0)\}}{\|\phi(th) - \phi(0)\|} \leq \|\phi(th) - \phi(0)\|.$$

Letting $t \searrow 0$ we see that $(x - x_0)^T v \leq 0$. Since $v \in T_{x_0}(\mathcal{S}) \setminus \{0\}$ was arbitrary and $-v \in T_{x_0}(\mathcal{S}) \setminus \{0\}$, we therefore have that $(x - x_0)^T v = 0$ for all $v \in T_{x_0}(\mathcal{S})$. Moreover, $\dot{\eta}(x_0)^T v = 0$ for all $v \in T_{x_0}(\mathcal{S})$, so $x - x_0 \propto \dot{\eta}(x_0)$, which yields the result. □

In fact, under a slightly stronger condition on η, we have the following useful result:

PROPOSITION 3. *Let $\mathcal{R}$ be a d-dimensional manifold in $\mathbb{R}^d$, suppose that $\eta : \mathcal{R} \to [0,1]$ satisfies the condition that $\mathcal{S} := \{x \in \mathcal{R} : \eta(x) = 1/2\}$ is non-empty. Suppose further that there exists $\epsilon > 0$ such that η is twice continuously differentiable on $\mathcal{S}^\epsilon$. Assume that $\dot{\eta}(x_0) \neq 0$ for all $x_0 \in \mathcal{S}$. Define $g : \mathcal{S} \times (-\epsilon, \epsilon) \to \mathcal{S}^\epsilon$ by*

$$g(x_0, t) := x_0 + \frac{t\dot{\eta}(x_0)}{\|\dot{\eta}(x_0)\|}.$$

If

$$\epsilon \leq \inf_{x_0 \in \mathcal{S}} \frac{\|\dot{\eta}(x_0)\|}{\sup_{z \in B_{2\epsilon}(x_0) \cap \mathcal{S}^\epsilon} \|\ddot{\eta}(z)\|_{\mathrm{op}}}, \tag{53}$$

then g is injective. In fact g is a diffeomorphism, with

$$Dg_{(x_0,t)}(v_1, v_2) = (I + tB)\biggl(v_1 + \frac{\dot{\eta}(x_0)}{\|\dot{\eta}(x_0)\|} v_2\biggr), \tag{54}$$

for $v_1 \in T_{x_0}(\mathcal{S})$ and $v_2 \in \mathbb{R}$, where

$$B := \frac{1}{\|\dot{\eta}(x_0)\|}\biggl(I - \frac{\dot{\eta}(x_0)\dot{\eta}(x_0)^T}{\|\dot{\eta}(x_0)\|^2}\biggr)\ddot{\eta}(x_0). \tag{55}$$

PROOF. Assume for a contradiction that there exist distinct points $x_1, x_2 \in \mathcal{S}$ and $t_1, t_2 \in (-\epsilon, \epsilon)$ with $|t_1| \geq |t_2|$ such that

$$x_1 + \frac{t_1\dot{\eta}(x_1)}{\|\dot{\eta}(x_1)\|} = x_2 + \frac{t_2\dot{\eta}(x_2)}{\|\dot{\eta}(x_2)\|}.$$

Then

$$0 < \|x_2 - x_1\|^2 = \frac{2t_1\dot{\eta}(x_1)^T(x_2 - x_1)}{\|\dot{\eta}(x_1)\|} + t_2^2 - t_1^2 \leq \frac{2t_1\dot{\eta}(x_1)^T(x_2 - x_1)}{\|\dot{\eta}(x_1)\|}. \tag{56}$$

By Taylor's theorem and (56),

$$\begin{aligned}|\dot\eta(x_1)^T(x_2-x_1)| &= |\eta(x_2)-\eta(x_1)-\dot\eta(x_1)^T(x_2-x_1)|\\ &\leq \frac{1}{2}\sup_{z\in B_{2\epsilon}(x_1)\cap\mathcal{S}^\epsilon}\|\ddot\eta(z)\|_{\mathrm{op}}\|x_2-x_1\|^2\\ &< \sup_{z\in B_{2\epsilon}(x_1)\cap\mathcal{S}^\epsilon}\|\ddot\eta(z)\|_{\mathrm{op}}\frac{\epsilon|\dot\eta(x_1)^T(x_2-x_1)|}{\|\dot\eta(x_1)\|},\end{aligned}$$

contradicting the hypothesis (53).

To show that g is a diffeomorphism, let $x_0\in\mathcal{S}$ be given and let $\phi: U\to\mathcal{S}$ be a local parametrisation around x_0 with $\phi(0)=x_0$. Define $\Phi: U\times(-\epsilon,\epsilon)\to\mathcal{S}\times(-\epsilon,\epsilon)$ by $\Phi(u,t):=(\phi(u),t)$, and $H:U\times(-\epsilon,\epsilon)\to\mathcal{S}^\epsilon$ by $H:=g\circ\Phi$. Finally, define the *Gauss map* $n:\mathcal{S}\to\mathbb{R}^d$ by $n(x_0):=\dot\eta(x_0)/\|\dot\eta(x_0)\|$. Then, for $h=(h_1^T,h_2)^T\in\mathbb{R}^{d-1}\times\mathbb{R}$ and $s\in\mathbb{R}\setminus\{0\}$,

$$\begin{aligned}&\lim_{s\to 0}\frac{H(sh_1,t+sh_2)-H(0,t)}{s}\\ &\quad=\lim_{s\to0}\biggl\{\frac{\phi(sh_1)-\phi(0)}{s}+\frac{t\{n(\phi(sh_1))-n(\phi(0))\}}{s}+h_2n\bigl(\phi(sh_1)\bigr)\biggr\}\\ &\quad= D\phi_0(h_1)+tDn_{x_0}\circ D\phi_0(h_1)+h_2n(x_0)\\ &\quad= Dg_{(x_0,t)}\circ D\Phi_{(0,t)}(h_1,h_2),\end{aligned}$$

where $Dg_{(x_0,t)}:T_{x_0}(\mathcal{S})\times\mathbb{R}\to\mathbb{R}^d$ is given in (54).

To show that $Dg_{(x_0,t)}$ is invertible, note that for $v_1\in T_{x_0}(\mathcal{S})$ and $|t|<\epsilon$,

$$\frac{|t|}{\|\dot\eta(x_0)\|}\biggl\|\biggl(I-\frac{\dot\eta(x_0)\dot\eta(x_0)^T}{\|\dot\eta(x_0)\|^2}\biggr)\ddot\eta(x_0)v_1\biggr\|\leq\frac{|t|\|\ddot\eta(x_0)\|_{\mathrm{op}}}{\|\dot\eta(x_0)\|}\|v_1\|<\|v_1\|,$$

where the final inequality follows from (53). Then, since $v_1+\frac{t}{\|\dot\eta(x_0)\|}\Bigl(I-\frac{\dot\eta(x_0)\dot\eta(x_0)^T}{\|\dot\eta(x_0)\|^2}\Bigr)\ddot\eta(x_0)v_1$ and $n(x_0)v_2$ are orthogonal, it follows that $Dg_{(x_0,t)}$ is indeed invertible. The inverse function theorem (e.g. Guillemin and Pollack, 1974, p. 13) then gives that g is a local diffeomorphism, and moreover, by Guillemin and Pollack (1974, Exercise 5, p. 18) and the fact that g is bijective, we can conclude that g is in fact a diffeomorphism. □

7.3. *Forms, pullbacks and integration on manifolds.* Let V be a (real) vector space of dimension m. We say $T:V^p\to\mathbb{R}$ is a *p-tensor* on V if it is p-linear, and write $\mathcal{F}^p(V^*)$ for the set of p-tensors on V. If $T\in\mathcal{F}^p(V^*)$ and $S\in\mathcal{F}^q(V^*)$, we define their *tensor product* $T\otimes S\in\mathcal{F}^{p+q}(V^*)$ by

$$T\otimes S(v_1,\ldots,v_p,v_{p+1},\ldots,v_{p+q}):=T(v_1,\ldots,v_p)S(v_{p+1},\ldots,v_{p+q}).$$

Let S_p denote the set of permutations of $\{1,\ldots,p\}$. If $\pi \in S_p$ and $T \in \mathcal{F}^p(V^*)$, we can define $T^\pi \in \mathcal{F}^p(V^*)$ by $T^\pi(v) := T(v_{\pi(1)},\ldots,v_{\pi(p)})$ for $v = (v_1,\ldots,v_p) \in V^p$. We say T is *alternating* if $T^\sigma = -T$ for all transpositions $\sigma : \{1,\ldots,p\} \to \{1,\ldots,p\}$. The set of alternating p-tensors on V, denoted $\Lambda^p(V^*)$, is a vector space of dimension $\binom{m}{p}$. The function $\mathrm{Alt} : \mathcal{F}^p(V^*) \to \Lambda^p(V^*)$ is defined by

$$\mathrm{Alt}(T) := \frac{1}{p!}\sum_{\pi \in S_p} (-1)^{\mathrm{sgn}(\pi)} T^\pi,$$

where $\mathrm{sgn}(\pi)$ denotes the sign of the permutation π. If $T \in \Lambda^p(V^*)$ and $S \in \Lambda^q(V^*)$, we define their *wedge product* $T \wedge S \in \Lambda^{p+q}(V^*)$ by

$$T \wedge S := \mathrm{Alt}(T \otimes S).$$

If W is another (real) vector space and $A : V \to W$ is a linear map, we define the *transpose* $A^* : \Lambda^p(W^*) \to \Lambda^p(V^*)$ of A by

$$A^*T(v_1,\ldots,v_p) := T(Av_1,\ldots,Av_p).$$

Let $\mathcal{S}$ be a manifold. A *p-form* ω on $\mathcal{S}$ is a function which assigns to each $x \in \mathcal{S}$ an element $\omega(x) \in \Lambda^p(T_x(\mathcal{S})^*)$. If ω is a p-form on $\mathcal{S}$ and θ is a q-form on $\mathcal{S}$, we can define their wedge product $\omega \wedge \theta$ by $(\omega \wedge \theta)(x) := \omega(x) \wedge \theta(x)$. For $j = 1,\ldots,m$, let $x_j : \mathbb{R}^m \to \mathbb{R}$ denote the coordinate function $x_j(y_1,\ldots,y_m) := y_j$. These functions induce 1-forms dx_j, given by $dx_j(x)(y_1,\ldots,y_m) = y_j$ (so $dx_j(x) = D(x_j)_x$ in our previous notation). Letting $\mathcal{I} := \{(i_1,\ldots,i_p) : 1 \leq i_1 < \ldots < i_p \leq m\}$, for $I = (i_1,\ldots,i_p) \in \mathcal{I}$, we write

$$dx_I := dx_{i_1} \wedge \ldots \wedge dx_{i_p}.$$

It turns out (Guillemin and Pollack, 1974, p. 163) that any p-form on an open subset U of $\mathbb{R}^m$ can be uniquely expressed as

$$\sum_{I \in \mathcal{I}} f_I \, dx_I, \tag{57}$$

where each f_I is a real-valued function on U.

Recall that the set of all ordered bases of a vector space V is partitioned into two equivalence classes, and an *orientation* of V is simply an assignment of a positive sign to one equivalence class and a negative sign to the other. If V and W are oriented vector spaces in the sense that an orientation has been specified for each of them, then an isomorphism $A : V \to W$ always either preserves orientation in the sense that for any ordered basis

β of V, the ordered basis $A\beta$ has the same sign as β, or it reverses it. We say an m-dimensional manifold $\mathcal{X}$ is *orientable* if for every $x \in \mathcal{X}$, there exist an open subset U of $\mathbb{R}^m$, a neighbourhood V of x in $\mathcal{X}$ and a diffeomorphism $\phi : U \to V$ such that $D\phi_u : \mathbb{R}^m \to T_x(\mathcal{X})$ preserves orientation for every $u \in U$. A map like ϕ above whose derivative at every point preserves orientation is called an *orientation-preserving* map.

If $\mathcal{X}$ and $\mathcal{Y}$ are manifolds, ω is a p-form on $\mathcal{Y}$ and $\psi : \mathcal{X} \to \mathcal{Y}$ is differentiable, we define the *pullback* $\psi^*\omega$ of ω by ψ to be the p-form on $\mathcal{X}$ given by

$$\psi^*\omega(x) := (D\psi_x)^*\omega\big(\psi(x)\big).$$

If V is an p-dimensional vector space and $A : V \to V$ is linear, then $A^*T = (\det A)T$ for all $T \in \Lambda^p(V)$ (Guillemin and Pollack, 1974, p. 160).

If ω is an m-form on an open subset U of $\mathbb{R}^m$, then by (57), we can write $\omega = f\,dx_1 \wedge \ldots \wedge dx_m$. If ω is an integrable form on U (i.e. f is an integrable function on U), we can define the integral of ω over U by

$$\int_U \omega := \int_U f(x_1, \ldots, x_m)\,dx_1 \ldots dx_m,$$

where the integral on the right-hand side is a usual Lebesgue integral. Now let $\mathcal{S}$ be an m-dimensional orientable manifold that can be parametrised with a single chart, in the sense that there exists an open subset U of $\mathbb{R}^m$ and an orientation-preserving diffeomorphism $\phi : U \to \mathcal{S}$. Define the *support* of an m-form ω on $\mathcal{S}$ to be the closure of $\{x \in \mathcal{S} : \omega(x) \neq 0\}$. If ω is compactly supported, then its pullback $\phi^*\omega$ is a compactly supported m-form on U; moreover $\phi^*\omega$ is integrable, and we can define the integral over $\mathcal{S}$ of ω by

$$\int_{\mathcal{S}} \omega := \int_U \phi^*\omega. \tag{58}$$

Alternatively, we can suppose that ω is non-negative and measurable in the sense that $\phi^*\omega = f\,dx_1 \wedge \ldots \wedge dx_m$, say, with f non-negative and measurable on U. In this case, we can also define the integral of ω over $\mathcal{S}$ via (58).

More generally, integrals of forms over more complicated manifolds can be defined via partitions of unity. Recall (Guillemin and Pollack, 1974, p. 52) that if $\mathcal{X}$ is an arbitrary subset of $\mathbb{R}^M$, and $\{V_\alpha : \alpha \in A\}$ is a (relatively) open cover of $\mathcal{X}$, then there exists a sequence of real-valued, differentiable functions (ρ_n) on $\mathcal{X}$, called a *partition of unity* with respect to $\{V_\alpha : \alpha \in A\}$, with the following properties:

1. $\rho_n(x) \in [0, 1]$ for all $n \in \mathbb{N}$;
2. Each $x \in \mathcal{X}$ has a neighbourhood on which all but finitely many functions ρ_n are identically zero;

3. Each ρ_n is identically zero except on some closed set contained in some V_α;
4. $\sum_{n=1}^\infty \rho_n(x) = 1$ for all $x \in \mathcal{X}$.

Now let $\mathcal{S} \subseteq \mathbb{R}^d$ be an m-dimensional, orientable manifold, so for each $x \in \mathcal{S}$, there exist an open subset U_x of $\mathbb{R}^m$, a neighbourhood V_x of x in $\mathcal{S}$ and an orientation-preserving diffeomorphism $\phi_x : U_x \to V_x$. If ω is a compactly supported m-form on $\mathcal{S}$ and (ρ_n) denotes a partition of unity on $\mathcal{S}$ with respect to $\{V_x : x \in \mathcal{S}\}$, we can define the integral of ω over $\mathcal{S}$ by

$$\int_{\mathcal{S}} \omega := \sum_{n=1}^\infty \int_{\mathcal{S}} \rho_n \omega. \tag{59}$$

In fact, writing Ω for the compact support of ω, we can find a neighbourhood W_x of $x \in \Omega$, $x_1, \ldots, x_N \in \Omega$ and a finite subset N_j of $\mathbb{N}$ such that $\{\rho_n : n \notin N_j\}$ are identically zero on W_{x_j}, and such that

$$\int_{\mathcal{S}} \omega = \sum_{j=1}^N \sum_{n \in N_j} \int_{\mathcal{S}} \rho_n \omega.$$

Thus the integral can be written as a finite sum. Similarly, if ω is a non-negative m-form on $\mathcal{S}$, we can again define the integral of ω over $\mathcal{S}$ via (59). Finally, if ω is an integrable m-form on $\mathcal{S}$, the integral can be defined by taking positive and negative parts in the usual way.

In our work, we are especially interested in integrals of a particular type of form. Given an m-dimensional, orientable manifold $\mathcal{S}$ in $\mathbb{R}^d$, the *volume form* $d\mathrm{Vol}^m$ is the unique m-form on $\mathcal{S}$ such that at each $x \in \mathcal{S}$, the alternating m-tensor $d\mathrm{Vol}^m(x)$ on $T_x(\mathcal{S})$ gives value $1/m!$ to each positively oriented orthonormal basis for $T_x(\mathcal{S})$. For example, when $\mathcal{S} = \mathbb{R}^m$, we have $d\mathrm{Vol}^m = dx_1 \wedge \ldots \wedge dx_m$, provided we consider the standard basis to be positively oriented. As another example, if $\mathcal{R} \subseteq \mathbb{R}^d$ is a d-dimensional manifold and $\eta : \mathcal{R} \to \mathbb{R}$ is continuously differentiable with $\mathcal{S} = \{x \in \mathcal{R} : \eta(x) = 1/2\}$ non-empty and $\dot{\eta}(x) \neq 0$ for $x \in \mathcal{S}$, then $\mathcal{S}$ is a $(d-1)$-dimensional, orientable manifold (Guillemin and Pollack, 1974, Exercise 18, p. 106). If we say that an ordered, orthonormal basis $e_1, \ldots, e_{d-1}$ for $T_{x_0}(\mathcal{S})$ is positively oriented whenever $\det(e_1, \ldots, e_{d-1}, \dot{\eta}(x_0)) > 0$, we have that

$$d\mathrm{Vol}^{d-1}(x_0) = \sum_{j=1}^d (-1)^{j+d} \frac{\eta_j(x_0)}{\|\dot{\eta}(x_0)\|} dx_1 \wedge \ldots \wedge dx_{j-1} \wedge dx_{j+1} \wedge \ldots \wedge dx_d(x_0),$$

where x_j denotes the jth coordinate function. We now define an ordered, orthonormal basis $(e_1, 0), \ldots, (e_{d-1}, 0), (0, 1)$ for $T_{x_0}(\mathcal{S}) \times \mathbb{R}$ to be positively oriented. Further, we define a $(d-1)$-form ω_1 and a 1-form ω_2 on $\mathcal{S}\times(-\epsilon,\epsilon)$ by

$$\omega_1(x_0,t)\big((v_1,w_1),\ldots,(v_{d-1},w_{d-1})\big) := d\mathrm{Vol}^{d-1}(x_0)(v_1,\ldots,v_{d-1})$$
$$\omega_2(x_0,t)(v_d,w_d) := dt(t)(w_d) = w_d.$$

Then, with g defined as in Proposition 3, and under the conditions of that proposition,

$$\begin{aligned}
&g^*(dx_1\wedge\ldots\wedge dx_d)(x_0,t)\big((e_1,0),\ldots,(e_{d-1},0),(0,1)\big)\\
&\quad= dx_1\wedge\ldots\wedge dx_d(x_0^t)\big(Dg_{(x_0,t)}(e_1,0),\ldots,Dg_{(x_0,t)}(e_{d-1},0),Dg_{(x_0,t)}(0,1)\big)\\
&\quad= \frac{1}{d!}\det(I+tB)\\
&\quad= \frac{1}{d}\det(I+tB)d\mathrm{Vol}^{d-1}(x_0)(e_1,\ldots,e_{d-1})dt(t)(1)\\
&\quad= \det(I+tB)\ (\omega_1\wedge\omega_2)(x_0,t)\big((e_1,0),\ldots,(e_{d-1},0),(0,1)\big),
\end{aligned}$$

so $g^*(dx_1\wedge\ldots\wedge dx_d)(x_0,t) = \det(I+tB)\ (\omega_1\wedge\omega_2)(x_0,t)$. It follows that if $h:\mathcal{S}\times(-\epsilon,\epsilon)\to\mathbb{R}$ is either compactly supported and integrable, or non-negative and measurable, then

$$\int_{\mathcal{S}\times(-\epsilon,\epsilon)} h\,\omega_1\wedge\omega_2 = \int_{\mathcal{S}}\int_{-\epsilon}^{\epsilon} h(x_0,t)\,dt\,d\mathrm{Vol}^{d-1}(x_0). \tag{60}$$

We also require the change of variables formula: if $\mathcal{X}$ and $\mathcal{Y}$ are orientable manifolds and are of dimension m, and if $\psi:\mathcal{X}\to\mathcal{Y}$ is an orientation-preserving diffeomorphism, then

$$\int_{\mathcal{X}}\psi^*\omega = \int_{\mathcal{Y}}\omega \tag{61}$$

for every compactly supported, integrable m-form on $\mathcal{Y}$ (Guillemin and Pollack, 1974, p. 168). In particular, if $f:\mathcal{S}^\epsilon\to\mathbb{R}$ is either compactly supported and integrable, or non-negative and measurable, then writing $x_0^t := x_0 + \frac{t\dot\eta(x_0)}{\|\dot\eta(x_0)\|}$, we have from (60) and (61) that

$$\begin{aligned}
\int_{\mathcal{S}^\epsilon} f(x)\,dx &= \int_{\mathcal{S}\times(-\epsilon,\epsilon)} \det(I+tB)f(x_0^t)\,(\omega_1\wedge\omega_2)(x_0,t)\\
&= \int_{\mathcal{S}}\int_{-\epsilon}^{\epsilon}\det(I+tB)f(x_0^t)\,dt\,d\mathrm{Vol}^{d-1}(x_0).
\end{aligned} \tag{62}$$

References.

Cannings, T. I., Berrett, T. B. and Samworth, R. J. (2019). Local nearest neighbour classification with applications to semi-supervised learning. *Ann. Statist.*, submitted.

Duong, T. (2015). `ks`: Kernel smoothing. `R` package version 1.9.4, `https://cran.r-project.org/web/packages/ks`.

Gray, A. (2004). *Tubes, 2nd ed.* Progress in Mathematics 221. Birkhäuser, Basel.

Guillemin, V. and Pollack, A. (1974). *Differential Geometry*. Prentice-Hall, New Jersey.

Mammen, E. and Tsybakov, A. B. (1999). Smooth discriminant analysis. *Ann. Statist.*, **27**, 1808–1829.

Paditz, L. (1989). On the analytical structure of the constant in the nonuniform version of the Esseen inequality. *Statistics*, **20**, 453–464.

School of Mathematics
James Clerk Maxwell Building
Peter Guthrie Tait Road
Edinburgh EH9 3FD
E-mail: timothy.cannings@ed.ac.uk
URL: http://www.maths.ed.ac.uk/%7Etcannings/

Statistical Laboratory
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WB
E-mail: t.berrett@statslab.cam.ac.uk
URL: www.statslab.cam.ac.uk/%7Etbb26/
E-mail: r.samworth@statslab.cam.ac.uk
URL: www.statslab.cam.ac.uk/%7Erjs57/