Supplementary Material for: Voronoi Boundary Classification: A High-Dimensional Geometric Approach via Weighted Monte Carlo Integration

Vladislav Polianskii and Florian T. Pokorny

Lemma 1 (Change of Variables and Monte Carlo). *Let $P = V(t) \cap V(x_j)$. If $P \neq \emptyset$, P lies in some hyperplane $\Pi = \{x \in \mathbb{R}^d : \langle x, n \rangle = c\}$, where n is of unit norm. Denote by $\mathbb{S}^{d-1}$ the $d-1$ dimensional unit sphere centered at the origin and denote by f_j the map which maps $m \in \mathbb{S}^{d-1}$ to to the intersection between the ray starting at t in direction m with P if this intersection exists. Let $I_j = \int_{V(t)\cap V(x_j)} w(x)\mathrm{dVol}$, then*

$$I_j = \int_{f_j^{-1}(V(t)\cap V(x_j))} w(f_j(m)) \frac{\|f_j(m) - t\|^{d-1}}{|\langle m, n \rangle|} \mathrm{dVol}_{\mathbb{S}^{d-1}} \tag{1}$$

where $\mathrm{dVol}_{\mathbb{S}^{d-1}}$ denotes the standard induced volume form of the sphere in Euclidean space. Secondly, consider a sequence $\{m_i\}_1^T$ of uniform samples on $f_j^{-1}(V(t) \cap V(x_j)) \subset \mathbb{S}^{d-1}$. We have

$$I_j = V_j \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} w(f_j(m_i)) \frac{\|f_j(m_i) - t\|^{d-1}}{|\langle m_i, n \rangle|} \tag{2}$$

where V_j denotes the volume of the set $f_j^{-1}(V(t) \cap V(x_j))$ with respect to the standard induced metric on the unit sphere in Euclidean space.

Proof. Consider spherical coordinates $(\phi_1, \ldots, \phi_{d-1})$ on $\mathbb{S}^{d-1}$. Define $m = (m_1, \ldots, m_d)$ by $m_1 = \cos\phi_1$, $m_p = \cos\phi_p \prod_{j=1}^{p-1} \sin\phi_j$, for $p = 2, \ldots, d-1$ and $m_d = \prod_{j=1}^{d-1} \sin\phi_j$. One can verify that

$$B = \left\{ \frac{\frac{\partial m}{\partial \phi_1}}{\|\frac{\partial m}{\partial \phi_1}\|}, \ldots, \frac{\frac{\partial m}{\partial \phi_{d-1}}}{\|\frac{\partial m}{\partial \phi_{d-1}}\|}, m \right\}$$

forms an orthonormal basis of $\mathbb{R}^d$ and we note that the induced metric on $\mathbb{S}^{d-1}$ is given by $g_{ij} = \langle \frac{\partial m}{\partial \phi_i}, \frac{\partial m}{\partial \phi_j} \rangle$, which we note is a diagonal matrix in local ϕ_i

coordinates. We observe that $f_j(m) = l(m)m + t$, where $l(m) = \frac{c}{\langle m,n \rangle}$. We observe that the resulting induced metric is

$$\hat{g}_{ij} = \left\langle \frac{\partial f}{\partial \phi_i}, \frac{\partial f}{\partial \phi_j} \right\rangle = l^2 \left(g_{ij} + l^{-2} \frac{\partial l}{\partial \phi_i} \frac{\partial l}{\partial \phi_j} \right)$$

Using Sylvester's formula, we obtain

$$\begin{aligned} \det \hat{g} &= l^{2(d-1)} \det g \det(I + l^{-2} g^{-\frac{1}{2}} (\nabla l)^T \nabla l g^{-\frac{1}{2}}) \\ &= l^{2(d-1)} \det g (1 + l^{-2} \| g^{-\frac{1}{2}} \nabla l \|^2) \\ &= \frac{l^{2(d-1)}}{\langle m, n \rangle^2} \det g, \end{aligned}$$

where we observe in the last step that $1 + l^{-2} \| g^{-\frac{1}{2}} \nabla l \|^2$ is just $\frac{1}{\langle m,n \rangle^2}$ times the expression for $1 = \|n\|^2$ in the orthonormal basis B. The second part follows from the standard local coordinate Monte-Carlo integration theorems, see e.g. [1]. □

Lemma 2. *Let $t \in \mathbb{R}^d$ and $m \in \mathbb{S}^{d-1}$ and define*

$$l^*_m(x) = \frac{\|x - t\|^2}{2 \langle m, x - t \rangle}$$
$$\text{face}(m) = \arg \min_{i \in \{1,\dots,n\}} \left\{ l^*_m(x_i) \mid l^*_m(x_i) > 0 \right\}.$$

If there is no $i \in \{1, \dots, n\}$ such that $l^*_m(x_i) > 0$, then the ray R starting at t and in direction m is fully contained in $V(t)$ (and $V(t)$ is unbounded in direction m starting at t). Otherwise m lies in $f_j^{-1}(V(t) \cap V(x_{\text{face}(m)}))$.

Proof. Solving for $l \geq 0$, for an equidistant point between t and x along the ray $\{t + lm | t \geq 0\}$, we observe that $\|t + lm - t\|^2 = \|x - t - lm\|^2$ for $l = l^*_m(x)$. The expression for face(m) above hence selects the neighboring Voronoi cell to t along the direction m and $l^*_m(x_{\text{face}(m)})$ returns the distance to the Voronoi cell boundary along the ray. □

Theorem 1. *Consider a labeled dataset $D = \{(x_1, c_1), \dots, (x_n, c_n)\}$ of data points $x_i \in \mathbb{R}^d$ and corresponding class labels $c_i \in \{1, 2, \dots, k\}$ and a test point $t \in \mathbb{R}^d$, where $t \neq x_i$ for all $i \in \{1, \dots, n\}$. Assume that t only has a single nearest neighbor among $\{x_1, \dots, x_n\}$. Consider a sequence of weight functions $\{w_n\}_{n=1}^{\infty}$, $w_n : \mathbb{R}_+ \to \mathbb{R}_+$ which are each monotonically decreasing and where*

$$\lim_{n \to \infty} \int_{z_2}^{+\infty} \frac{w_n(z)}{w_n(z_1)} z^{d-2} \, dz = 0 \quad \textit{for all} \qquad 0 < z_1 < z_2 \tag{3}$$

Then, for sufficiently large n, the class $VBC(t|D, w_n)$ assigned by the Voronoi Boundary Classifier of t is equal to the class $1NN(t|D)$ assigned by the nearest neighbor classifier for t.

Proof. Let $\mathbb{B}_r(t)$ denote the ball of radius r around t in $\mathbb{R}^d$ and denote by Π_i the unique hyperplane containing $V(t) \cap V(x_i)$ whenever this intersection is non-empty and $d-1$ dimensional. We observe that, for all $r > 0$,

$$\begin{aligned}\int_{V(t)\cap V(x_i)} w_n(\|x-t\|)\mathrm{dVol} &\geq \int_{V(t)\cap V(x_i)\cap\mathbb{B}_r(t)} w_n(\|x-t\|)\mathrm{dVol} \\ &\geq \mathrm{Vol}(V(t)\cap V(x_i)\cap\mathbb{B}_r(t))\ w_n(r)\end{aligned}$$

since w decreases with distance from t.

Denote $\mathrm{Vol}(V(t)\cap V(x_i)\cap\mathbb{B}_r(t))$ as C_r. Let x_i be the nearest neighbor to t, and pick $r_1 = \frac{1}{2}\|x_i - t\| + \delta$ for some $\delta > 0$ so that $\mathrm{dist}(\Pi_j, t) > r_1$ for all $j \neq i$; such δ exists due to the assumed uniqueness of the nearest neighbor to t. Therefore, for any $j \neq i$ such that $V(t) \cap V(x_j) \neq \emptyset$,

$$\begin{aligned}\frac{\int_{V(t)\cap V(x_j)} w_n(\|x-t\|)\mathrm{dVol}}{\int_{V(t)\cap V(x_i)} w_n(\|x-t\|)\mathrm{dVol}} &\leq \frac{1}{C_{r_1}}\int_{V(t)\cap V(x_j)} \frac{w_n(\|x-t\|)}{w_n(r_1)}\mathrm{dVol} \\ &\leq \frac{1}{C_{r_1}}\int_{\Pi_j} \frac{w_n(\|x-t\|)}{w_n(r_1)}\mathrm{dVol}\end{aligned}$$

Let $r_2 = \mathrm{dist}(\Pi_j, t)$ so that $r_2 > r_1$. Pick polar coordinates on $\Pi_j = \mathbb{S}^{d-2}(p) \times \mathbb{R}_{\geq 0}$ centered at the closest point $p \in \Pi_j$ to t. Then

$$\begin{aligned}\frac{1}{C_{r_1}}\int_{\Pi_j} \frac{w_n(\|x-t\|)}{w_n(r_1)}\mathrm{dVol} &= \frac{1}{C_{r_1}}\int_0^\infty\int_{\mathbb{S}^{d-2}} \frac{w_n(\sqrt{r_2^2+r^2})}{w_n(r_1)} r^{d-2}\mathrm{dVol}_{\mathbb{S}^{d-2}}\mathrm{dr} \\ &= \frac{\mathrm{Vol}(\mathbb{S}^{d-2})}{C_{r_1}}\int_0^\infty \frac{w_n(\sqrt{r_2^2+r^2})}{w_n(r_1)} r^{d-2}\mathrm{dr} \\ &\leq \frac{\mathrm{Vol}(\mathbb{S}^{d-2})}{C_{r_1}}\int_0^\infty \frac{w_n(r_2+r)}{w_n(r_1)} r^{d-2}\mathrm{dr} \\ &\leq \frac{\mathrm{Vol}(\mathbb{S}^{d-2})}{C_{r_1}}\int_{r_2}^\infty \frac{w_n(r)}{w_n(r_1)} r^{d-2}\mathrm{dr},\end{aligned}$$

where the last term tends to zero. It follows that

$$\lim_{n\to\infty}\frac{\int_{V(t)\cap V(x_j)} w_n(\|x-t\|)\mathrm{dVol}}{\int_{V(t)\cap V(x_i)} w_n(\|x-t\|)\mathrm{dVol}} = 0.$$

We can replace the numerator with a finite sum over all x_k of the same class label as x_j, and the result still tends to zero. Hence, the Voronoi Boundary Rank for class label x_i dominates the Voronoi Boundary Rank for all other class labels for sufficiently large n and the result follows. □

References

[1] Philip J Davis and Philip Rabinowitz. *Methods of numerical integration.* Courier Corporation, 2007.