An optimization problem on the sphere

Andreas Maurer
Adalbertstr 55
D80799 München

May 16, 2008

Abstract

We prove existence and uniqueness of the minimizer for the average geodesic distance to the points of a geodesically convex set on the sphere. This implies a corresponding existence and uniqueness result for an optimal algorithm for halfspace learning, when data and target functions are drawn from the uniform distribution.

1 Introduction

Let $\mathcal{S}^{n-1}$ be the unit sphere in $\mathbb{R}^n$ with normalized uniform measure σ and geodesic metric ρ, and let K be a proper convex cone with nonempty interior in $\mathbb{R}^n$. We will show that the function $\psi : \mathcal{S}^{n-1} \to \mathbb{R}$ defined by

$$\psi_K(w) = \int_{K \cap \mathcal{S}^{n-1}} \rho(w, y)\, d\sigma(y)$$

attains its global minimum at a unique point on $\mathcal{S}^{n-1}$. While existence of the minimum is straightforward, uniqueness seems surprisingly difficult to prove.

A similar problem has been considered in [2] and [1]. In these works the intention is to define a centroid, so integration is replaced by finite summation and $\rho(w, y)$ replaced by $\rho(w, y)^2$. Since the problem is rather obvious, it appears likely that a proof of the above result exists somewhere in the literature and we just haven't been able to find it.

2 Optimal halfspace learning

Our motivation to consider this problem arises in learning theory. Specifically we consider an experiment, where

1. A unit vector u is drawn at random from σ and kept concealed from the learner.

2. A sample $\mathbf{x} = (x_1, ..., x_m) \in \left(\mathcal{S}^{n-1}\right)^m$ is generated in m independent random trials of σ.

3. A label vector $\mathbf{y} = u(\mathbf{x}) \in \{-1, 1\}^m$ is generated according to the rule $y_i = sign(\langle u, x_i \rangle)$, where $\langle ., . \rangle$ is the euclidean inner product and $sign(t) = 1$ if $t > 0$ and -1 if $t < 0$. The *sign* of 0 is irrelevant, because it corresponds to events of probability zero.

4. The labeled sample $(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, u(\mathbf{x}))$ is supplied to the learner.

5. The learner produces a hypothesis $f(\mathbf{x}, \mathbf{y}) \in \mathcal{S}^{n-1}$ according to some learning rule $f : \left(\mathcal{S}^{n-1}\right)^m \times \{-1, 1\}^m \to \mathcal{S}^{n-1}$.

6. An unlabeled test point $x \in \mathcal{S}^{n-1}$ is drawn at random from σ and presented to the learner who produces the label $y = sign(\langle f(\mathbf{x}, \mathbf{y}), x \rangle)$.

7. If $sign(\langle u, x_i \rangle) = y$ the learner is rewarded one unit, otherwise a penalty of one unit is incurred.

We now ask the following question: Which learning rule f will give the highest average reward on a very large number of independent repetitions of this experiment?

Evidently the optimal learning rule has to minimize the following functional:

$$\Omega(f) = \mathbb{E}_{u \sim \sigma} \mathbb{E}_{\mathbf{x} \sim \sigma^m} \Pr_{x \sim \sigma} \{sign(\langle f(\mathbf{x}, u(\mathbf{x})), x \rangle) \neq sign(\langle u, x \rangle)\}.$$

Now a simple geometric argument shows that for any $v, u \in \mathcal{S}^{n-1}$ we have

$$\Pr_{x \sim \sigma} \{sign(\langle v, x \rangle) \neq sign(\langle u, x \rangle)\} = \rho(v, u) / \pi,$$

relating the misclassification probability to the geodesic distance. For a labeled sample $(\mathbf{x}, \mathbf{y}) \in \left(\mathcal{S}^{n-1}\right)^m \times \{-1, 1\}^m$ we denote

$$C(\mathbf{x}, \mathbf{y}) = \left\{u \in \mathcal{S}^{n-1} : u(\mathbf{x}) = \mathbf{y}\right\}.$$

$C(\mathbf{x}, \mathbf{y})$ is thus the set of all hypotheses consistent with the labeled sample $(\mathbf{x}, \mathbf{y})$. Observe that, given $\mathbf{x}$ and u there is exactly one $\mathbf{y}$ such that $\mathbf{y} = u(\mathbf{x})$, that is $u \in C(\mathbf{x}, \mathbf{y})$. We also have $C(\mathbf{x}, \mathbf{y}) = K(\mathbf{x}, \mathbf{y}) \cap \mathcal{S}^{n-1}$ where $K(\mathbf{x}, \mathbf{y})$ is the closed convex cone

$$K(\mathbf{x}, \mathbf{y}) = \{v \in \mathbb{R}^n : \langle u, y_i x_i \rangle \geq 0, \forall 1 \leq i \leq m\}.$$

We therefore obtain

$$\begin{aligned} \Omega(f) &= \pi^{-1} \mathbb{E}_{u \sim \sigma} \mathbb{E}_{\mathbf{x} \sim \sigma^m} \rho(f(\mathbf{x}, u(\mathbf{x})), u) \\ &= \pi^{-1} \mathbb{E}_{\mathbf{x} \sim \sigma^m} \sum_{\mathbf{y} \in \{-1,1\}^m} \mathbb{E}_{u \sim \sigma} \rho(f(\mathbf{x}, u(\mathbf{x})), u) 1_{C(\mathbf{x}, \mathbf{y})}(u) \\ &= \pi^{-1} \mathbb{E}_{\mathbf{x} \sim \sigma^m} \sum_{\mathbf{y} \in \{-1,1\}^m} \mathbb{E}_{u \sim \sigma} \rho(f(\mathbf{x}, \mathbf{y}), u) 1_{C(\mathbf{x}, \mathbf{y})}(u) \\ &= \pi^{-1} \mathbb{E}_{\mathbf{x} \sim \sigma^m} \sum_{\mathbf{y} \in \{-1,1\}^m} \psi_{K(\mathbf{x}, \mathbf{y})}(f(\mathbf{x}, \mathbf{y})). \end{aligned}$$

If $K(\mathbf{x},\mathbf{y})$ has empty interior then the corresponding summand vanishes, so we can assume that $K(\mathbf{x},\mathbf{y})$ has nonempty interior. Clearly $-y_i x_i \notin K(\mathbf{x},\mathbf{y})$ for all example points, so $K(\mathbf{x},\mathbf{y})$ is a proper cone. Our result therefore applies and asserts the existence of a unique minimizer $f^*(\mathbf{x},\mathbf{y})$ of the function $\psi_{K(\mathbf{x},\mathbf{y})}$. The map $f^* : (\mathbf{x},\mathbf{y}) \mapsto f^*(\mathbf{x},\mathbf{y})$ is then the unique optimal learning algorithm.

The map f^* also has the symmetry property $f^*(V\mathbf{x},\mathbf{y}) = Vf^*(\mathbf{x},\mathbf{y})$ for any unitary V on $\mathbb{R}^n$. This is so, because

$$\psi_{K(V\mathbf{x},\mathbf{y})}(w) = \psi_{K(\mathbf{x},\mathbf{y})}\left(V^{-1}w\right),$$

as is easily verified. We will also show, that the solution $f^*(\mathbf{x},\mathbf{y})$ must lie in the cone

$$\left\{\sum_{i=1}^{m} \alpha_i y_i x_i : \alpha_i \geq 0\right\}$$

and that $\psi_{K(\mathbf{x},\mathbf{y})}$ has no other local minima.

3 Proof of the main result

Notation 1 *$\rho(.,.)$ is the geodesic distance and σ the Haar measure on $\mathcal{S}^{n-1}$. For $A \subseteq \mathbb{R}^n$ we denote $A_1 = \{x \in A : \|x\| = 1\} = A \cap \mathcal{S}^{n-1}$. 'Cone' will always mean 'convex cone'. For $A \subseteq \mathbb{R}^n$ we denote*

$$\hat{A} = \{x : \langle x, v\rangle \geq 0, \forall v \in A\}.$$

This is always a closed convex set. A proper cone K is one which is contained in some closed halfspace. For a set A the indicator function of A will be denoted by 1_A

Lemma 2 *Let K be a closed cone*

(i) If $w \notin K$ then there is a unit vector $z \in \mathbb{R}^n$ such that $\langle z, w\rangle < 0$ and $\langle z, y\rangle \geq 0$ for all $y \in K$.

(ii) $\widehat{\left(\hat{K}\right)} = K$.

(iii) Suppose that K is proper and has nonempty interior, $w \in \mathcal{S}^{n-1}$, $w \notin \hat{K} \cup \left(-\hat{K}\right)$ and $\epsilon > 0$. Then there exists z with $\|z\| = 1$ such that $-\epsilon < \langle z, w\rangle < 0$ and $\langle z, y\rangle > 0$ for all $y \in \hat{K} \setminus \{0\}$.

Proof. (i) Let B be an open ball containing w such that $K \cap B = \emptyset$. Define

$$O = \{\lambda x : x \in B, \lambda > 0\}.$$

Then K and O are nonempty disjoint convex sets and O is open. By the Hahn-Banach theorem ([4], Theorem 3.4) there is $\gamma \in \mathbb{R}$ and $z \in \mathbb{R}^n$ such that

$$\langle z, x\rangle < \gamma \leq \langle z, y\rangle, \forall x \in O, y \in K.$$

Choosing $y = 0 \in K$ gives $\gamma \leq 0$, letting $\lambda \to 0$ in $\langle z, \lambda w \rangle < \gamma$ shows $\gamma \geq 0$, so that $\gamma = 0$. The normalization is trivial.

(ii) Trivially $K \subseteq \left(\hat{K}\right)^{\hat{}}$. On the other hand, if $w \notin K$ let z be the vector from part (i). Then $z \in \hat{K}$ but $\langle w, z \rangle < 0$, so that $w \notin \left(\hat{K}\right)^{\hat{}}$.

(iii) Since $w \notin \hat{K}$ there exists $x_1 \in K$ s.t. $\langle w, x_1 \rangle < 0$. Since the interior of K is nonempty, K is the closure of its interior (Theorem 6.3 in [3]), so we can assume $x_1 \in \operatorname{int}(K)$. Similarly, since $w \notin \left(-\hat{K}\right)$ we have $-w \notin \hat{K}$, so there is $x_2 \in \operatorname{int}(K)$ with $\langle -w, x_2 \rangle < 0$, that is $\langle w, x_2 \rangle > 0$. Since the interior of K is convex it contains the segment $[x_1, x_2]$, so by continuity of $\langle w, \cdot \rangle$ there is some $x_0 \in \operatorname{int}(K)$ with $\langle w, x_0 \rangle = 0$. Since K is a proper cone $0 \notin \operatorname{int}(K)$ and we can assume that $\|x_0\| = 1$.

Let $c > 0$ be such that $x' \in K$ whenever $\|x_0 - x'\| \leq c$. We define

$$z = (1 - \eta)^{1/2} x_0 - \eta^{1/2} w, \text{ where } 0 < \eta < \min\left\{\frac{c^2}{1 + c^2}, \epsilon^2\right\}.$$

Since $\langle w, x_0 \rangle = 0$ it is clear that z is a unit vector. Also $\langle w, z \rangle = -\eta^{1/2} > -\epsilon$, and for any $y \in \hat{K}_1$ we have $x_0 - cy \in K$, so $\langle y, x_0 - cy \rangle \geq 0$ and

$$\begin{aligned} \langle y, z \rangle &= (1 - \eta)^{1/2} \left(\langle y, x_0 - cy \rangle + c \langle y, y \rangle \right) - \eta^{1/2} \langle y, w \rangle \\ &\geq (1 - \eta)^{1/2} c - \eta^{1/2} > 0. \end{aligned}$$

■

Theorem 3 *Let $K \subset \mathbb{R}^{n-1}$ be a closed proper cone with nonempty interior, $g : [0, \pi] \to \mathbb{R}$ continuous and the function $\psi : \mathcal{S}^{n-1} \to \mathbb{R}$ defined by*

$$\psi(w) = \int_{K_1} g(\rho(w, y)) \, d\sigma(y).$$

(i) ψ attains its global minimum on $\mathcal{S}^{n-1}$.

(ii) If g is increasing then every local minimum of ψ must lie in $\hat{K} \cup \left(-\hat{K}\right)$ and every global minimum of ψ must lie in $K \cap \hat{K}$.

(iii) If g is increasing and convex in $[0, \pi/2]$ then the global minimum of ψ is unique and corresponds to the only local minimum outside $-\hat{K}$.

(iv) If g is increasing, convex in $[0, \pi/2]$ and concave in $[\pi/2, \pi]$ then the global minimum of ψ is unique and corresponds to its only local minimum on $\mathcal{S}^{n-1}$.

Proof. (i) is immediate since $\mathcal{S}^{n-1}$ is compact and ψ is continuous.

(ii) Fix $w \in \mathcal{S}^{n-1}$, $w \notin \hat{K} \cup \left(-\hat{K}\right)$. We will first show that there can be no local minimum of ψ at w. Let $\epsilon > 0$ be arbitrary and choose z as in the lemma (iii). The functional z divides the sphere $\mathcal{S}^{n-1}$ into two open hemispheres

$$L = \{u : \langle z, u \rangle < 0\} \text{ and } R = \{u : \langle z, u \rangle > 0\},$$

and an equator of σ-measure zero. Note that $w \in L$ and $\hat{K}_1 \subseteq R$. We can write

$$c = \min_{y \in \hat{K}_1} \langle y, z \rangle > 0,$$

since $\hat{K}_1$ is compact and $y \mapsto \langle y, z \rangle$ is continuous. With V we denote the reflection operator which exchanges points of L and R

$$Vx = -\langle x, z \rangle z + (x - \langle x, z \rangle z).$$

V is easily verified to an isometry and $V^2 = I$.

Suppose now that $u \in R$ and $Vu \in K$. We claim that u is in the interior of K. Indeed, if $u' \in \mathbb{R}^n$ satisfies $\|u - u'\| < 2 \langle u, z \rangle c$, then for all $y \in \hat{K}_1$ we have

$$\begin{aligned} \langle u', y \rangle &= \langle u, y \rangle - \langle u - u', y \rangle \geq \langle u, y \rangle - 2 \langle u, z \rangle c \\ &\geq \langle u, y \rangle - 2 \langle u, z \rangle \langle z, y \rangle = \langle Vu, y \rangle \geq 0, \end{aligned}$$

so $u' \in \left(\hat{K}\right)^{\hat{}} = K$, by part (ii) of the lemma. This establishes the claim and shows that $V(K) \cap R$ is contained in the interior of K. It follows that

$$\forall u \in R, 1_K(u) \geq 1_K(Vu). \tag{1}$$

Also $V(K) \cap R$ is relatively closed in R while $\text{int}(K) \cap R$ is open in R. Since R is connected they can only coincide if $V(K) \cap R = R$. But this is impossible, since then

$$\begin{aligned} L \cup R &= V(V(K) \cap R) \cup (V(K) \cap R) \subseteq V(V(K \cap L)) \cup \text{int}(K) \\ &= (K \cap L) \cup \text{int}(K) \subseteq K, \end{aligned}$$

and K is assumed to be a proper cone. So $V(K) \cap R$ is a proper subset of $\text{int}(K) \cap R$. The inequality (1) is therefore strict on the nonempty open set $(\text{int}(K) \cap R) \setminus (V(K) \cap R)$.

Using isometry and unipotence of V we now obtain

$$\begin{aligned} \psi(w) - \psi(Vw) &= \int_R (g(\rho(w,u)) - g(\rho(Vw,u))) 1_K(u) \, d\sigma(u) + \\ &\quad + \int_L (g(\rho(w,u)) - g(\rho(Vw,u))) 1_K(u) \, d\sigma(u) \\ &= \int_R (g(\rho(w,u)) - g(\rho(Vw,u))) (1_K(u) - 1_K(Vu)) \, d\sigma(u) \\ &> 0. \end{aligned}$$

The inequality holds, because the first factor $(g(\rho(w,u)) - g(\rho(Vw,u)))$ in the last integral is always positive for $u \in R$, since g is increasing and ρ is increasing in the euclidean distance. The second is nonnegative and positive on a set of positive measure. Since $\|w - Vw\| = 2\epsilon$ and $\epsilon > 0$ was arbitrary, we see that every neighborhood of w contains a point where ψ has a smaller value than at

w. We conclude that w cannot be a local minimum of ψ, which proves the first assertion of (ii).

If $w \notin K$ choose z as in part (i) of the lemma and let W be the isometry $Wx = -\langle x, z\rangle z + (x - \langle x, z\rangle z)$. The $\forall u \in K$ we have $g(\rho(w, u)) > g(\rho(Ww, u))$, so $\psi(w) > \psi(Ww)$ and w cannot be a global minimizer of ψ. So every global minimizer must be in $K \cap \left(\hat{K} \cap \left(-\hat{K}\right)\right)$. Since $K_1 \cap \left(-\hat{K}_1\right)$ is obviously empty the second assertion of (ii) follows.

(iii) Now let $w_1, w_2 \in \hat{K}_1$ with $w_1 \neq w_2$. Connect them with a geodesic in $\hat{K}_1$ and let $w^* \in \hat{K}_1$ be the midpoint of this geodesic, such that $\rho(w_1, w^*) = \rho(w^*, w_2) = \rho(w_1, w_2)/2 \leq \pi/2$. We define a map U by

$$Ux = \langle x, w^*\rangle w^* - (x - \langle x, w^*\rangle w^*).$$

Geometrically U is reflection on the one-dimensional subspace generated by w^*. Note that $w_2 = Uw_1$ and that $\rho(u, Uu) = 2\rho(u, w^*)$ if $\rho(u, w^*) \leq \pi/2$ and that $\rho(u, Uu) = 2\pi - 2\rho(u, w^*)$ if $\rho(u, w^*) \geq \pi/2$.

Take any $u \in K_1$. Since $w^* \in \hat{K}_1$ we have $\rho(u, w^*) \leq 2\pi$, whence $\rho(u, Uu) = 2\rho(u, w^*)$. All the four points w_1, w_2, u and Uu have at most distance $\pi/2$ from w^* and lie therefore together with w^* on a common hemisphere. By the triangle inequality

$$\begin{aligned} 2\rho(u, w^*) &= \rho(u, Uu) \\ &\leq \rho(u, w_1) + \rho(w_1, Uu) = \rho(u, w_1) + \rho(Uw_1, UUu) \\ &= \rho(u, w_1) + \rho(w_2, u). \end{aligned}$$

If u does not lie on the geodesic through w_1 and w_2 and not at distance $\pi/2$ from w^* strict inequality holds, and since K_1 has nonempty interior strict inequality holds on an open subset of K_1. If g is increasing and convex in $[0, \pi/2]$ then dividing by 2, applying g and integrating over K_1 we get

$$\psi(w^*) < (1/2)(\psi(w_1) + \psi(w_2)).$$

It follows that there can be at most one point in $\hat{K}_1$ where the gradient of ψ vanishes, and this point, if it exists, must correspond to a local minimum. By (ii) this is the unique global minimum and the only local minimum outside $-\hat{K}$, which establishes (iii).

(iv) If $x_1, x_2 \in -\hat{K}_1$ and $x^* \in -\hat{K}_1$ is their midpoint, then for $u \in K$ we obtain, using $\rho(x_i, u) = \pi - \rho(-x_i, u)$ and a reasoning analogous to the above,

$$\rho(u, w^*) \geq (1/2)(\rho(u, w_1) + \rho(u, w_2)),$$

the inequality being again strict on a set of positive measure and preserved under application of a function g which is increasing and concave in $[\pi/2, \pi]$, so that

$$\psi(w^*) > (1/2)(\psi(w_1) + \psi(w_2)).$$

It again follows that there can be at most one point in $-\hat{K}_1$ where the gradient of ψ vanishes, and this point must now correspond to a local maximum. We conclude that ψ has a unique local minimum which lies in $\hat{K}_1$. ■

Remark. An example of a function as in (iii) is $g(t) = t^2$, in which case the minimizer is the spherical mass centroid considered in [2] and [1]. Examples of functions as in (iv) are of course the identity function, in which case we obtain the result stated in the introduction. We could also set $g(t) = 2(1 - \cos t)$, in which case the function reads

$$\psi(w) = \int_{K_1} \|w - y\|^2 \, d\sigma(y).$$

In this case uniqueness of the minimum can be established with much simpler methods.

Acknowledgement. The author is grateful to Andreas Argyriou, Massimiliano Pontil and Erhard Seiler for many encouraging discussions.

References

[1] S. R. Buss and J. P. Fillmore. Spherical averages and applications to spherical splines and interpolation. *ACM Trans. Graph.* 20, 2: 95-126, 2001.

[2] G. A. Gal'perin. A concept of the mass center of a system of material points in the constant curvature spaces. *Comm. Math. Phys.* Volume 154, 1: 63-84, 1993.

[3] R. T. Rockafeller. *Convex Analysis.* Princeton University Press, 1970.

[4] W. Rudin. *Functional Analysis.* McGraw-Hill, 1974.