CLASSIFICATION WITH MIXTURES OF CURVED MAHALANOBIS METRICS

Frank Nielsen[1,2*], *Boris Muzellec*[1] *Richard Nock*[3,4]

[1]École Polytechnique, France
[2]Sony Computer Science Laboratories, Japan

[3]Data61 & [4]ANU
Sydney, Australia

ABSTRACT

We study the classification with respect to the class of curved Mahalanobis metrics that extend the celebrated flat Mahalanobis distances to constant curvature spaces. We prove that these curved Mahalanobis k-NN classifiers define piecewise linear decision boundaries, and report the performance of learning those metrics within the framework of the Large Margin Nearest Neighbor (LMNN). Finally, we show experimentally that a mixture of curved Mahalanobis metrics define a composite metric distance that improves the classification performance.

Index Terms— Classification, Mahalanobis distance, metric learning, Large Margin Nearest Neighbor (LMNN), Cayley-Klein geometry

1. INTRODUCTION AND CONTRIBUTION

1.1. Introduction

In supervised classification [1, 2], one of the simplest classifier $\mathcal{M}$ is the k-NN classifier that classes an unlabeled observation x by taking the majority of the labels of the k nearest neighbors (NN) of x in the training set $\mathcal{T} = \{(x_i, y_i) : i \in [n]\}$ with $x_i = (x_i^{(1)}, ..., x_i^{(d)}) \in \mathbb{R}^d$, $y_i \in \{-1, 1\}$ (with $[n] = \{1, ..., n\}$). To avoid ties in binary classification, k is chosen odd. The notion of "nearest" neighbor depends on the selected distance function $D(\cdot, \cdot)$. The distance is often chosen to be the Euclidean distance: $D(p, q) = \sqrt{\sum_{i=1}^d (p^{(i)} - q^{(i)})^2} = \|p - q\|$ where $\|\cdot\|$ is the L_2 norm induced by the Euclidean inner product $\langle p, q\rangle = p^\top q$ (commonly called scalar or dot product in Euclidean geometry): $\|x\| = \sqrt{\langle x, x\rangle}$. Learning an appropriate distance from the training set allows one to improve the performance of the k-NN classifier over the ordinary Euclidean distance. A well-known generalization of the Euclidean distance is the Mahalanobis distance [1, 2] $D_\Sigma(p, q) = \sqrt{(p - q)^\top \Sigma (p - q)}$, where $\Sigma \succ 0$ is a symmetric $d \times d$ positive definite matrix. The Euclidean distance is a Mahalanobis distance obtained for the identity matrix I. The Mahalanobis distance is a *metric distance* that satisfies the three axioms of metrics: (i) reflexivity: $D_\Sigma(p, q) = 0 \Leftrightarrow p = q$, (ii) symmetry: $D_\Sigma(p, q) = D_\Sigma(q, p)$, and (iii) triangle inequality: $D_\Sigma(p, q) + D_\Sigma(q, r) \geq D_\Sigma(p, r)$. To learn an appropriate Mahalanobis distance (*i.e.*, matrix $\Sigma \succ 0$), various algorithms relying on *side information* have been proposed: For example, the Mahalanobis Metric Clustering [3] (MMC) for clustering and the Large Margin Nearest neighbor [4] (LMNN) for classification. In image retrieval systems by image query, the Mahalanobis k-NN classifier on image features allows one to return a ranked list of similar images to the query [5]. Notice that since the k nearest neighbors of a query point does not change by considering any monotonically increasing function of the selected distance (like a squaring operation), we may consider equivalently the squared Mahalanobis distance $D_\Sigma^2(p, q)$ (but doing so we loose the triangle inequality property, and it is not anymore a metric). By generalizing the Mahalanobis distance one may further hope to improve the Mahalanobis k-NN classifier performance. It turns out that the squared Mahalanobis distance is a particular case of a larger family of distortion measures, called Bregman divergence [6] $B_F(p, q)$ defined for a strictly convex and differentiable generator F by $B_F(p, q) = F(p) - F(q) - \langle p - q, \nabla F(q)\rangle$. For the generator $F_\Sigma(x) = x^\top \Sigma x$, we get $B_F(p, q) = D_\Sigma^2(p, q)$. However the cone space of such convex and differentiable functions is infinite-dimensional, and it is challenging to design methods for learning appropriate Bregman generators. In [5] (2015), a neat generalization of Mahalanobis distances has been proposed, called generalized hyperbolic and generalized elliptical Mahalanobis distances, and the classification using the generalized elliptical Mahalanobis distances has been proven superior compared to "Euclidean" Mahalanobis distances.

1.2. Contributions and outline

In this work, we refine and extend the framework of [5]. We summarize our contributions as follows: (i) We prove that curved Mahalanobis k-NN classifiers are always piecewise linear. (ii) We describe how to perform negatively-curved Mahalanobis metric learning using LMNN [4], extending the approach in [5] that considered only positively-curved Mahalanobis setting. (iii) We consider learning a mixture of curved

*Contact author: Frank.Nielsen@acm.org

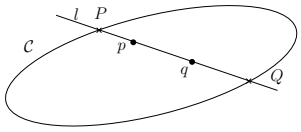

Fig. 1. Measuring length distances in Cayley-Klein geometries: $L(p,q) \propto \log(p,q;P,Q)$, where $(p,q;P,Q)$ denotes the cross-ratio of 4 collinear points.

Mahalanobis distances that induces a Riemannian geometry that is not anymore of constant curvature, and show experimentally that this mixed metric distance improves over the curved Mahalanobis distances for the classification task. Besides, we also show that curved Mahalanobis balls are equivalent to Euclidean Mahalanobis balls with shifted centers (and ellipsoid shapes) and report corresponding radii values.

The paper is organized as follows: Section 2 introduces the basic notions of Cayley-Klein geometries and explained the two hyperbolic/elliptical metric Cayley-Klein geometries that induce the negatively-curved and positively-curved Mahalanobis distances. Section 3 proves that the decision boundaries of k-NN for the curved Mahalanobis metric distances are piecewise-linear. Section 4 report the basic mapping transformations to transform any curved/flat Mahalanobis space into an equivalent canonical space of canonical curvature $\kappa \in \{-1, 0, +1\}$. Section 5 presents the curved Mahalanobis Large Margin Nearest Neighbor algorithm, further considers a mixture of curved Mahalanobis distances, and report experimentally on the accuracies of classification.

2. CURVED MAHALANOBIS GEOMETRIES

To define the curved Mahalanobis metric distances, we introduce their underlying Cayley-Klein geometries [7]. In brief, the Cayley-Klein geometries unify the metric Euclidean/elliptical/hyperbolic geometries with other spacetime geometries (Minkowskian, Galilean, de Sitter, etc.) from the viewpoint of *projective geometry* [7]. In a Cayley-Klein geometry [7], the *signed length* $L(p,q)$ between two points is defined according to a *fundamental conic* [7] $\mathcal{C}$ and a prescribed constant c as $L(p,q) = c \times \log(p,q;P,Q) = c \log \frac{pP \times qQ}{qP \times pQ}$, where P and Q are the two intersection points of the line l passing through p and q with $\mathcal{C}$ (Figure 1). The length measurements are signed: $L(q,p) = -L(p,q)$ (checked from the property of the cross ratio $(p,q;P,Q) = 1/(q,p;P,Q)$), and furthermore $L(p,p) = 0$ since $(p,p,P,Q) = 0$ (reflexivity). For three collinear points p, q and r we have $L(p,q) + L(q,r) = L(p,r)$. The proof follows easily from the properties of the cross-ratio: $(p,q;x,y)(q,r;x,y) = (p,r;x,y)$ (taking the logarithm yields the additive property). Let $D(p,q) = |L(p,q)|$, then the distance $D(\cdot,\cdot)$ satisfies the reflexivity/symmetry/triangle inequality axioms of a metric.

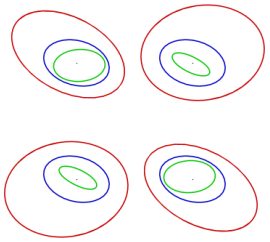

Fig. 2. Riemannian metric tensors induced by the flat Euclidean Mahalanobis distance (blue, constant), the negatively-curved hyperbolic Mahalanobis distance (green), and the positively-curved elliptical Mahalanobis distance (red).

To define the Cayley-Klein distance without the intersection points P and Q on the conic $\mathcal{C}$, let S be an invertible symmetric real-valued matrix of dimension $(d+1) \times (d+1)$, and consider the symmetric bilinear map (not necessarily an inner product since it may also yield negative values) on $p, q \in \mathbb{R}^d$ defined by:$S(p,q) = \tilde{p} S \tilde{q} = \begin{bmatrix} p^\top & 1 \end{bmatrix}^\top S \begin{bmatrix} q \\ 1 \end{bmatrix} = S(q,p)$. Here, we shall distinguish between two particular cases for an invertible matrix S (with $|S| = \det(S) \neq 0$): Case 1: $S \succ 0$ (S is positive definite): All eigenvalues[1] are positive, and the induced Cayley-Klein geometry is said *elliptical* (with the fundamental conic $\mathcal{C} = \{x : \tilde{x}^\top S \tilde{x} = 0\}$ purely complex, and the intersection points P and Q are conjugates) defined over the *full domain* $\mathbb{D}_S = \mathbb{R}^d$. The elliptical geometry [7] is not to be confused with the Riemannian spherical geometry since we identify antipodal points in the projective setting. Case 2: The last eigenvalue of S, λ_{d+1}, is negative and all the others are positive, we get the *Cayley-Klein hyperbolic geometry* (with real fundamental conic $\mathcal{C}$) defined over the *partial conic* domain $\mathbb{D}_S = \{x : \tilde{x}^\top S \tilde{x} < 0\} \subset \mathbb{R}^d$.

To incorporate these two cases, let us write $S = \begin{bmatrix} \Sigma & a \\ a^\top & b \end{bmatrix}$. Then the bilinear form becomes $S(p,q) = S_{\Sigma,a,b}(p,q) = S_{p,q} = \tilde{p}^\top S \tilde{q} = p^\top \Sigma q + p^\top a + a^\top q + b$. For notational convenience, further define $\mu \in \mathbb{R}^d$ and $\kappa \in \mathbb{R}$ so that $a = -\Sigma\mu$ (that is, $\mu = -\Sigma^{-1}a$) and $b = \mu^\top \Sigma \mu + \text{sign}(\kappa)\frac{1}{\kappa^2}$ (that is, $\kappa = \begin{cases} (b - \mu^\top \mu)^{-\frac{1}{2}} & b > \mu^\top \mu \\ -(\mu^\top \mu - b)^{-\frac{1}{2}} & b < \mu^\top \mu \end{cases}$), then $S(p,q)$ can be written as $S(p,q) = S_{\Sigma,\mu,\kappa}(p,q) = (p-\mu)^\top \Sigma (q-\mu) + \text{sign}(\kappa)\frac{1}{\kappa^2}$. Finally, by choosing the arbitrary but appropriate constants to get real (and not complex) distances [7], we get the *curved Mahalanobis distances* between two points $p, q \in \mathbb{D}_S$ as:

$$D_S(p,q) = D_{\Sigma,\mu,\kappa}(p,q) = \frac{1}{2|\kappa|} \text{arccosh}\left(\frac{|S(p,q)|}{\sqrt{S(p,p)S(q,q)}}\right)$$

where $\kappa \in \mathbb{R}\backslash\{0\}$ denotes the curvature, and

[1] Eigenvalues of symmetric real matrices are guaranteed reals and not complex values.

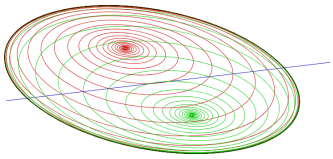

Fig. 3. Bisector for the negatively-curved Mahalanobis distance. The hyperbolic spheres are converted to equivalent flat Mahalanobis spheres for rasterization. The spheres become tangent to the fundamental conic as the radius tend to infinity.

$\mathrm{arccosh}(x) = \log(x + \sqrt{x^2-1})$ for $x \geq 1$ is a monotonically increasing function. We have [5]: $\lim_{\kappa\to 0^+} D_{\Sigma,\mu,\kappa}(p,q) = \lim_{\kappa\to 0^-} D_{\Sigma,\mu,\kappa}(p,q) = D_\Sigma(p,q)$. That is, the curved Mahalanobis distances generalize the Mahalanobis distance and $D_\Sigma(p,q) = D_{\Sigma,0,0}(p,q)$. By choosing $S = \mathrm{diag}(1,1,...,1,-1)$, we recover the usual hyperbolic geometry with distance [8] $D_h(p,q) = \mathrm{arccosh}\left(\frac{1-\langle p,q\rangle}{\sqrt{1-\langle p,p\rangle}\sqrt{1-\langle q,q\rangle}}\right)$ defined inside the interior of a unit ball, since we have $S(p,q) = \begin{bmatrix} p \\ 1 \end{bmatrix}^\top \begin{bmatrix} I & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} q \\ 1 \end{bmatrix} = p^\top Iq - 1 = p^\top q - 1$.

The Euclidean, hyperbolic and elliptical Cayley-Klein metric geometries can be interpreted as *Riemannian geometries* [9] with a corresponding metric tensor that yields Euclidean-straight geodesics. Figure 2 displays some Euclidean, hyperbolic and elliptical Cayley-Klein unit balls: We observe that the (flat) Mahalanobis balls have shapes independent of their center, but not the curved Mahalanobis balls with shapes varying according to their center position. In fact, it can be proved that *curved Mahalanobis balls* are equivalent to *flat Mahalanobis balls* with shifted centers (and corresponding radius values). We report the conversion formula (without proof for sake of conciseness): A Mahalanobis ball of center μ and covariance matrix Σ_M (and radius r_M) is defined by $(x-\mu_M)^\top \Sigma_M (x-\mu)_M = r_M^2$. That is, $x^\top \Sigma_M x - 2x^\top \Sigma_M \mu_M + \mu_M^\top \Sigma_M \mu_M = r_M^2$ By identifying the curved Mahalanobis ball of radius r and center c with the equation of the flat Mahalanobis ball, we find that: $\Sigma_M = r'^2\Sigma - a'a'^\top$, $\mu_M = \Sigma_M^{-1}(b'a' - r'^2 a)$, $r_M^2 = b'^2 - r'^2 b + c'^\top \Sigma_M c'$, with $a' = \Sigma c + a$, $b' = a^\top c + b$ and $r' = \sqrt{-S(c,c)}\cosh(r)$.

3. CURVED MAHALANOBIS K-NN CLASSIFIERS

The k-NN classifier associates for any point $x \in \mathbb{D}_S$ a label in $\{-1,1\}$ as the majority class of labels among the k-nearest neighbors (no ties for odd k). The bisector of two sites p and q is given by $\mathrm{Bi}(p,q) = \{x : D_S(p,x) = D_S(q,x)\}$. This yields the following bisector equation in the hyperbolic/elliptical case: $\left\langle x, \sqrt{|S(p,p)|}\Sigma q - \sqrt{|S(q,q)|}\Sigma p \right\rangle + \sqrt{|S(p,p)|}(a^\top(q+x)+b) - \sqrt{|S(q,q)|}(a^\top(p+x)+b) = 0$. Thus the bisector for the curved Mahalanobis distances are always Euclidean *hyperplanes*. Figure 3 illustrates such a bisector for a hyperbolic Cayley-Klein geometry.

It follows that the k-order Voronoi diagram [10] that partitions the space into elementary cells having the same equivalece class of the k-nearest neighbors is *piecewise linear*. Note that k-order Voronoi cells may be empty of generators when $k > 1$, see [10]. Since the decision boundary of the k-NN classifier is obtained from the boundaries of the union of those elementary k-order Voronoi cells after merging them by corresponding classes, we conclude that the curved Mahalanobis k-NN classifier is always piecewise linear. Furthermore, the VC dimension of those classifiers is $d+1$, see [1, 2]. Hence, curved Mahalanobis metrics boosts representation power but has no negative impact on generalisation compared to flat Mahalanobis metrics.

4. SPECTRAL DECOMPOSITION

It is well-known that one can apply the Cholesky decomposition $\Sigma = LL^\top$ (with L a lower triangular matrix) and transform the coordinate system x to $x' = L^\top x$ so that $D_\Sigma(p:q) = (p-q)^\top \Sigma (p-q) = (p-q)^\top LL^\top (p-q) = \|L^\top p - L^\top q\| = D_E(L^\top p, L^\top q)$. That is, the Mahalanobis distance amounts to compute an ordinary Euclidean distance on the affinely transformed space. Since the Euclidean geometry is flat and that an affine transformation yields an anisotropic stretching of space that is position independent, this motivates us again to use the term "flat Mahalanobis" distance. Now, consider the spectral decomposition of matrix $S = O\Lambda O^\top$ obtained by eigenvalue decomposition, and let us write canonically: $S = OD^{\frac{1}{2}} \begin{bmatrix} I & 0 \\ 0 & \lambda \end{bmatrix} D^{\frac{1}{2}} O^\top$, where $\lambda \in \{-1,1\}$ and O is an orthogonal matrix with $O^{-1} = O^\top$. Diagonal matrix D has all positive values, with $D_{i,i} = \Lambda_{i,i}$ and $D_{d+1,d+1} = |\Lambda_{d+1,d+1}|$ so that $D^{\frac{1}{2}}$ is defined as the diagonal matrix obtained by taking the square root values element-wise of the matrix. We rewrite the bilinear form into a canonical form by mapping the points x to $\tilde{x'} = D^{\frac{1}{2}} O^\top \begin{bmatrix} x \\ 1 \end{bmatrix} = \begin{bmatrix} x'' \\ w \end{bmatrix}$. Since $\tilde{x'} = \begin{bmatrix} x' \\ 1 \end{bmatrix}$, we can then find $x' = \frac{x''}{w}$. When $\lambda > 0$ (elliptical with $D_{d+1,d+1} > 0$), we have $S_S(p,q) = S_I(p',q')$. When $\lambda < 0$ (hyperbolic with $D_{d+1,d+1} < 0$), we have $S_S(p,q) = S_H(p',q')$, with $H = \mathrm{diag}(1,...,1,-1)$ the canonical matrix form for hyperbolic Cayley-Klein spaces. Notice that in the ordinary Mahalanobis case, instead of using the Cholesky decomposition, we may use the $L_1 D L_1^\top$ matrix decomposition where L_1 is a unit lower triangular matrix (with diagonal elements all 1), and D is a diagonal matrix of positive elements. The mapping is then $x' = D^{\frac{1}{2}} L_1^\top$ or $x' = (L_1 D^{\frac{1}{2}})^\top$ since $D = D^\top$. Thus by transformation the input space into one of the canonical Euclidean/elliptical/hyperbolic spaces, we avoid performing costly matrix multiplications in the bilinear form, and once the structure (say, a k-NN decision boundary) has been recov-

Dataset	d	n	k	Mahalanobis	Elliptical ($\kappa > 0$)	Hyperbolic ($\kappa < 0$)	Mixed	α	β
balance	4	625	3	0.846	0.910 (0.66)	0.904 (-0.15)	**0.920**	0.440	0.560
pima	8	768	2	0.709	0.712 (0.59)	0.699 (-0.04)	**0.720**	0.584	0.416
vowel	10	528	11	0.827	0.825 (1.16)	0.816 (-0.05)	**0.841**	0.407	0.593
sonar	60	208	2	0.733	0.788 (0.45)	0.640 (-0.01)	**0.802**	0.794	0.206

Table 1. LMNN classification accuracy: We observe experimentally on UCI datasets that positively-curved Mahalanobis distance (elliptical geometry) have better performance than negatively-curved Mahalanobis distance (hyperbolic geometry) that improves over the flat Mahalanobis distance (Euclidean geometry). Furthermore, a mixture of curved Mahalanobis distances (inducing a non-constant curvature space) improves the performance over a constant curvature space.

ered, we can map back to the original space (say, for classifying new observations using the original coordinate system).

5. CURVED METRIC LEARNING WITH LMNN

There are important differences between hyperbolic and elliptical spaces: While in the elliptical case, the domain is fully $\mathbb{R}^d$, the maximum distance is bounded by $\kappa\pi$. In the hyperbolic case, the distance is not bounded, and we need to ensure that the real conic matrix inducing the domain contains all points to classify. LMNN [4] aims to minimize the distances between points and their designated "target neighbors" (*i.e.*, points with the same label and likely to be close to each other) while keeping a distance margin with so-called "impostors" (*i.e.*, points with a different label but closer than a target neighbor). LMNN uses labeled triplet-wise constraints (x_i, x_j, x_l) as side information. We denote by $j \rightarrow i$ the fact that x_j is a neighbor of x_i. For a distance function D_S, LMNN minimizes the non-convex loss function l where γ is a trade-off between the two terms of the objective function: $l = \sum_{i,i\rightarrow j}(D_S(x_i,x_j) + \gamma\sum_l(1-y_{i,l})\zeta_{i,j,l})$, with $y_{i,j} \in \{0,1\}$, $\zeta_{i,j,l} = [1 + D_S(x_i,x_j) - D_S(x_i,x_l)]_+$ the hinge loss corresponding to impostors. This generic LMNN energy is minimized using gradient descent optimization.

We first review the adaptation to the elliptical case ($S \succ 0$, reported in [5]) before proposing an extension to hyperbolic metrics. For the elliptical case, we write $S = LL^\top$, and compute the gradients for minimizing $l(L)$ as follows: $\frac{\partial l(L)}{\partial L} = \sum_{i,i\rightarrow j}(\frac{\partial \rho_E(x_i,x_j)}{\partial L} + \gamma\sum_l(1-y_{i,l})\frac{\partial \zeta_{i,j,l}}{\partial L})$, $\frac{\partial D_S(x_i,x_j)}{\partial L} = \frac{\kappa}{\sqrt{S_{i,i}S_{j,j}-S_{i,j}^2}} L\left(\frac{S_{i,j}}{S_{i,i}}C_{i,i} + \frac{S_{i,j}}{S_{j,j}}C_{j,j} - (C_{i,j}+C_{j,i})\right)$, $\frac{\partial \zeta_{i,j,l}}{\partial L} = \frac{\partial D_S(x_i,x_j)}{\partial L} - \frac{\partial D_S(x_i,x_l)}{\partial L}$ when $\zeta_{i,j,l} \geq 0$, else 0, where $C_{i,j} = (x_i,1)^\top(x_j,1)$.[2]

Compared to the elliptical case, two difficulties arise in the hyperbolic case: First, we must ensure that matrix S has signature [7] $(d,1,0)$, and second, we must make sure that the input points remain at all time within the definition domain of $\mathbb{D}_S$ with $\mathbb{D}_S = \{x : S_{xx} < 0\}$. To address the first difficulty, we impose S to be of the form $L^\top DL$ where D is a symmetric matrix of signature $(d,1,0)$ and L is positive semi-definite. As in the elliptical case, L will be our learning parameter, whereas D will remain fixed. The gradient of the distance wrt. is expressed as follows: $\frac{\partial D_H(x_i,x_j)}{\partial L} = \frac{\kappa}{\sqrt{S_{i,j}^2 - S_{i,i}S_{j,j}}} DL\left(\frac{S_{i,j}}{S_{i,i}}C_{i,i} + \frac{S_{i,j}}{S_{j,j}}C_{j,j} - (C_{i,j}+C_{j,i})\right)$. Substituting this new expression in the former gradient, we obtain the gradient for the loss function in the hyperbolic case. Assuming that we have found a proper initialization (that is, two matrices L and D for which the data lie in the definition domain), we can now perform our gradient descent. Intuitively, since when a point comes closer to the boundary of the definition domain its distance to the others becomes infinite, with a good initialization our data should remain within the definition domain all throughout the algorithm. However in practice it may happen (because of numerical precision and choice of the gradient step) that some point gets out of the definition domain. To circumvent this, we allow the algorithm to backtrack when this happens, and reduce the gradient step for the next iteration (in practice, we divide it by two). Another difficulty raised by the definition domain is the initialization. We pick any symmetric PSD matrix for $L \succ 0$ which yields a good initialization, and then compute D as $D = \mathrm{diag}(1,...,1,-\min_x\{\|Lx\|^2\})$. It can be checked that with such an initialization, all points lie in the hyperbolic domain defined by $S = L^\top DL$. In practice, we have tried $L = I_{d+1}$ or $L = \mathrm{diag}(L',1)$ where $\Sigma = L'L'^\top$ is the precision matrix (inverse covariance) of the data. Finally, we considered learning a *mixture* of elliptical and hyperbolic curved Mahalanobis distances: $M(p,q) = \alpha D_E(p,q) + \beta D_H(p,q)$ (with $\beta = 1-\alpha$). Since a scaled metric distance and the sum of two metric distances is a metric distance, the mixture distance is a metric. However, the Riemannian metric tensor is not a composite metric tensors since the geodesics need to be solved by a non-trivial partial differential equation [9]. Table 1 displays our experimental results on UCI data-sets.[3] In all cases, the mixed elliptical/hyperbolic combination yields better results.

[2] There is a slight error in the expression of $\frac{\partial D_S(x_i,x_j)}{\partial L}$ in the original paper, as $C_{i,j} + C_{j,i}$ was replaced by $2C_{i,j}$, which is not the distance gradient since it must be symmetric with respect to both x_i and x_j.

[3] https://archive.ics.uci.edu/ml/datasets.html. The various matrices S defining the bilinear form learnt from the data-sets under the Euclidean, hyperbolic, elliptical and composite settings are available online at https://www.lix.polytechnique.fr/~nielsen/CayleyKlein/ for reproducible research.

6. REFERENCES

[1] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[2] Keinosuke Fukunaga, *Introduction to statistical pattern recognition*, Academic press, 2013.

[3] Eric P. Xing, Michael I. Jordan, Stuart Russell, and Andrew Y. Ng, "Distance metric learning with application to clustering with side-information," in *Advances in neural information processing systems*, 2002, pp. 505–512.

[4] Kilian Q Weinberger and Lawrence K Saul, "Distance metric learning for large margin nearest neighbor classification," *The Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.

[5] Yanhong Bi, Bin Fan, and Fuchao Wu, "Beyond Mahalanobis metric: Cayley-Klein metric learning," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 2339–2347.

[6] Arindam Banerjee, Srujana Merugu, Inderjit S Dhillon, and Joydeep Ghosh, "Clustering with Bregman divergences," *The Journal of Machine Learning Research*, vol. 6, pp. 1705–1749, 2005.

[7] Jürgen Richter-Gebert, "Perspectives on projective geometry: A guided tour through real and complex geometry," 2011.

[8] Frank Nielsen and Richard Nock, "Hyperbolic Voronoi diagrams made easy," in *IEEE International Conference on Computational Science and Its Applications (ICCSA)*, 2010, pp. 74–80.

[9] Jürgen Jost, *Riemannian geometry and geometric analysis*, Springer Science & Business Media, 2008.

[10] Jean-Daniel Boissonnat and Mariette Yvinec, *Algorithmic geometry*, Cambridge University Press, 1998.