

SUPPORT VECTOR MACHINE FOR DATA ON MANIFOLDS: AN APPLICATION TO IMAGE ANALYSIS

Suman K. Sen, Mark Foskey, James S. Marron, Martin A. Styner

Medical Image Display & Analysis Group, University of North Carolina, Chapel Hill

ABSTRACT

The Support Vector Machine (SVM) is a powerful tool for classification. We generalize SVM to work with data objects that are naturally understood to be lying on curved manifolds, and not in the usual d -dimensional Euclidean space. Such data arise from medial representations (m-reps) in medical images, Diffusion Tensor-MRI (DT-MRI), diffeomorphisms, etc. Considering such data objects to be embedded in higher dimensional Euclidean space results in invalid projections (on the separating direction) while Kernel Embedding does not provide a natural separating direction. We use geodesic distances, defined on the manifold to formulate our methodology. This approach addresses the important issue of analyzing the change that accompanies the difference between groups by implicitly defining the notions of separating surface and separating direction on the manifold. The methods are applied in shape analysis with target data being m-reps of 3 dimensional medical images.

Index Terms— Image classification, Image shape analysis.

1. INTRODUCTION

Classification plays an important role in statistical shape analysis [1, 2, 3] of medical images. For example, brain disorders like Alzheimer's and schizophrenia are often accompanied by structural changes. By detecting these changes, classification help in understanding differences between populations.

Classification methods like Fisher Linear Discrimination (FLD) [4], Support Vector Machines [5, 6] and Distance Weighted Discrimination (DWD) [7] were designed for data which are vectors in Euclidean space and do not deal extensively with data which are parameterized by elements in curved manifolds. See [8] and [9] for an overview of common existing classification methods. Examples of data on manifold include m-reps [10, 11] and DT-MRIs [12].

SVM has been widely used in image analysis since it handles the issue of High Dimension Low Sample Size (HDLSS) reasonably well. Kernel Embedding [13] is another approach where the data are mapped to a higher dimensional feature space and the above mentioned Euclidean methods are applied. SVM is known to be widely used along with Kernel

Embedding. The drawback of this method is that the results are not interpretable since it does not provide a natural direction of separation. This issue is addressed here by generalizing SVM which naturally handles data on curved manifolds.

The notion of separating hyperplane, fundamental to many Euclidean classifiers, is intractable to explicitly compute when generalized to arbitrary manifolds. The approach adopted here is to find *control points* on the manifold, which represent the different classes of data, and then define the classifier as a function of the distances (geodesic distances on the manifold) of individual points from the control points (see section 2). We thus bypass the problem of explicitly finding separating boundaries on the manifold. In 2.3.1, a choice of the control points is proposed which generalizes the SVM criterion for manifolds.

This approach will enable us not only to use the method on our motivating example of m-rep data, but it is also applicable for DT-MRI, and several other sciences like human movement, mechanical engineering, robotics, computer vision and molecular biology where non-Euclidean data often appear.

1.1. M-reps as Shape Descriptor

The medial representation is based on the medial axis of Blum [14]. There is a medial manifold sampled over an approximately regular lattice, the elements of which are called *medial atoms*. When a medial figure consists of n medial atoms, its parameters are naturally understood to be lying in the cartesian product manifold $\{\mathbb{R}^3 \times \mathbb{R}^+ \times S^2 \times S^2\}^n$ [15, 16].

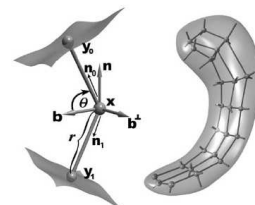


Fig. 1. Medial atom with a cross section of the boundary surface it implies (left). An m-rep model of a hippocampus and its boundary surface (right).

2. METHODS

Here we extend the ideas of separating hyperplane and separating direction (which are the foundations of many Euclidean classification methods such as Mean Difference, FLD, SVM and DWD) to data lying on a manifold. Our solution is based on the idea of *control points* (and the geodesic distance of data from these control points), as described in the following section.

2.1. Control Points and The General Classification Rule

We think about *control points* as being representatives of the two classes. If we name the control points as c_1 and c_2 , then the proposed classification rule $f_{c_1, c_2}(x)$ is given by

$$f_{c_1, c_2}(x) = d^2(c_2, x) - d^2(c_1, x), \quad (1)$$

where c_1, c_2 , and $x \in M$ and $d(\cdot, \cdot)$ is the geodesic distance metric defined on the manifold M . This rule assigns a new point x to class 1 if it is closer to c_1 than c_2 , and to class 2 otherwise.

2.2. The Implied Separating Surface and Direction of Separation

The zero level set of $f_{c_1, c_2}(\cdot)$ is the analog of the separating hyperplane, while the geodesic joining c_1 and c_2 is the analog of the direction of separation. Thus, the separating surface is the set of points which is equidistant from c_1 and c_2 . If we denote the separating surface by $H(c_1, c_2)$, we can write,

$$H(c_1, c_2) = \{x \in M : f_{c_1, c_2}(x) = 0\} \quad (2)$$

$$= \{x \in M : d^2(c_1, x) = d^2(c_2, x)\} \quad (3)$$

In d dimensional Euclidean space, $H(c_1, c_2)$ is a hyperplane of dimension $d - 1$ that is the perpendicular bisector of the line segment joining c_1 and c_2 . Note that the Mean Difference method is a particular case of this rule, where the control points are the means of the respective classes.

2.3. Choice of Control Points

Having set the framework for the general decision rule for manifolds the critical issue now is the choice of control points. For example, Fig. 2 shows that for the given set of data, the control points corresponding to the red solid separating boundary do a better job of classification than the pair corresponding to the black dotted boundary. So, the key to the construction of a good classification rule is to find the right pair of control points.

2.3.1. The Manifold SVM (MSVM) method

MSVM determines a pair of control points that maximizes the minimum distance to the separating boundary. While

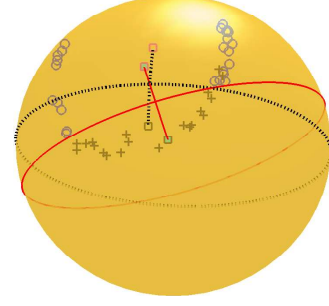


Fig. 2. Two pairs of control points showing their respective separating boundary and separating direction on the surface of the sphere. Different colors (with symbols) represent classes. The solid red surface (great circle) separates the data better than the dotted black surface.

the SVM criterion has many interpretations, it is the maximum margin idea that generalizes most naturally to manifolds where some Euclidean notions such as distance are much more readily available than others (e.g., inner product). MSVM appears to be the first approach where all calculations are done on the manifold. As in the classical SVM literature, y_i denotes the class label taking values -1 and 1.

The zero level set of the function $f_{c_1, c_2}(\cdot)$ (defined in (1)) defines the separating boundary $H(c_1, c_2)$ for a given pair (c_1, c_2) . Also, let $\hat{X}_{(c_1, c_2)}$ denote the set (to handle possible ties) of training points which are nearest to $H(c_1, c_2)$.

We would like to solve for some \tilde{c}_1 and \tilde{c}_2 such that

$$(\tilde{c}_1, \tilde{c}_2) = \arg \max_{c_1, c_2} \min_{i=1 \dots n} d(x_i, H(c_1, c_2)) \quad (4)$$

In other words, we want to maximize the minimum distance of the training points from the separating boundary.

It is important to note that the solution of $(\tilde{c}_1, \tilde{c}_2)$ in (4) is not unique. In fact, in the d dimensional Euclidean case there is a $d - 1$ dimensional space of solutions. Therefore, in order to make the search space for $(\tilde{c}_1, \tilde{c}_2)$ smaller we propose to find $(\tilde{c}_1, \tilde{c}_2)$ as

$$(\tilde{c}_1, \tilde{c}_2) = \arg \max_{(c_1, c_2) \in C_k} \min_{i=1 \dots n} d(x_i, H(c_1, c_2)) \quad (5)$$

where, for a given $k > 0$,

$$C_k = \{(c_1, c_2) : \hat{y}_{c_1, c_2} f_{c_1, c_2}(\hat{x}_{(c_1, c_2)}) = k\} \quad (6)$$

and,

$$\hat{x}_{(c_1, c_2)} = \arg \min_{x \in \hat{X}_{(c_1, c_2)}} f_{c_1, c_2}(x) \quad (7)$$

and \hat{y}_{c_1, c_2} is the class label of $\hat{x}_{(c_1, c_2)}$.

In Euclidean space, note that the distance of any point x from the separating boundary $H(c_1, c_2)$ is

$$d(x, H(c_1, c_2)) = \left| \frac{f_{c_1, c_2}(x)}{2d(c_1, c_2)} \right| \quad (8)$$

Therefore, using (5) - (8), and the fact $\hat{x}_{(c_1, c_2)}$ is one of the training points closest to $H(c_1, c_2)$, the optimization problem can be recast as the following penalized minimization problem:

$$\text{Minimize} \quad g_\lambda(c_1, c_2) = d^2(c_1, c_2) + \frac{\lambda}{n} \sum_{i=1}^n (h_i)_+ \quad (9)$$

where λ is the penalty for violating the constraints given by $h_i = k - y_i \{d^2(x_i, c_2) - d^2(x_i, c_1)\}$.

2.3.2. Tuning Parameter λ

Note that the second term in (9) not only penalizes misclassification, but also penalizes cases where training points come too close to the separating boundary. The parameter λ is called the tuning parameter. For large λ , the training error (proportion of misclassified training data) tends to decrease. But increasing λ indiscriminately tends to result in overfitting. This tradeoff is reflected by the cross-validation error (proportion of misclassified test data), which initially decreases, but increases when λ becomes large enough that the error is driven by overfitting. A sensible choice of λ is one which has low value of the cross-validation error.

3. RESULTS

In this section, we compare the performance of MSVM with two other methods. Method (a) is Euclidean SVM on a single tangent plane (with the overall geodesic mean as base point). We will call this method *TSVM*. Method (b) is called Geodesic Mean Difference (GMD), attained by choosing the control points as the geodesic means of the respective classes.

In our experiment, several values of the tuning parameter λ ($\lambda = 15^k, k = 0, \dots, 7$) are considered for each of MSVM and TSVM. The choice of the base 15 for λ is not set in stone: we choose it as a reasonable compromise between coverage of a large range for λ and the number of grid points.

3.1. Application To Hippocampi Data

This data consists of 82 m-rep models (of Hippocampi), 56 of which are from schizophrenic individuals and the remaining 26 are from healthy control individuals (see [17]). Each of these models has 24 medial atoms, placed in a 8×3 lattice (see Fig.1).

3.1.1. Training Error and Cross Validation Error

We conduct the simulation study in the following way. For each run we randomly remove five data points from the population of 82 and train our classifiers (for each λ) on the remaining 77 data points and test on the remaining five. Aggregating over several simulated replications, the training error and the cross-validation error are calculated.

Fig. 3 shows the performance (training error (left panel) and cross-validation error (right panel)) of the different methods.

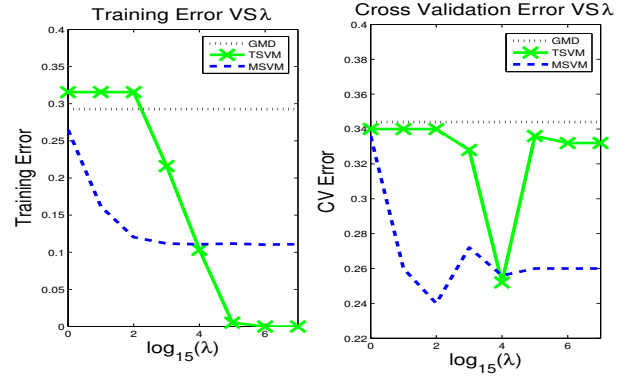


Fig. 3. Left: Training Errors against cost $\log_{15} \lambda$. Right: Cross-validation Errors against cost $\log_{15} \lambda$. Cross-validation error for MSVM is robust to the choice of λ .

From Fig. 3, we see that MSVM ($\lambda = 1$) has training error either very close to GMD or substantially smaller. On the other hand, for small values of λ the training error of TSVM is much higher. MSVM fails to attain a training error of zero while TSVM achieve zero training error (at $\lambda = 15^5$ or higher). But this could be due to overfitting by TSVM, and this idea is validated by the increased cross-validation error for high λ values.

We note that the cross-validation error TSVM is sensitive to the choice of λ , i.e., a good choice of λ appears to be critical. In contrast, MSVM is much more robust against the choice of λ . In particular, the fact that the cross-validation of MSVM is much more stable for high values of λ is promising. MSVM attains the minimum cross validation error of all the methods at $\lambda = 15^2$.

3.1.2. Shape Change Between The Two Groups

In our context, the rule that best shows structural change in the hippocampus is the most valuable. The structural change captured by each method is shown in Figure 4. For each classification rule (at the λ which has the least cross-validation error), we project the data points on to direction of separation. The mean of the projected data is calculated. The projected data points with the lowest and highest projection scores give the extent of structural change captured by the separating direction. The objects in the left are the projected shapes with the highest score, and on the right, with the lowest score. The colormap shows the surface distance maps of the mean (of projected data points) and projected shapes.

Fig 4 shows that GMD represents a large structural change. But its relevance is questionable because of its poor discriminating performance (Fig. 3). GMD shows a lot of structural change, but it fails to isolate the important

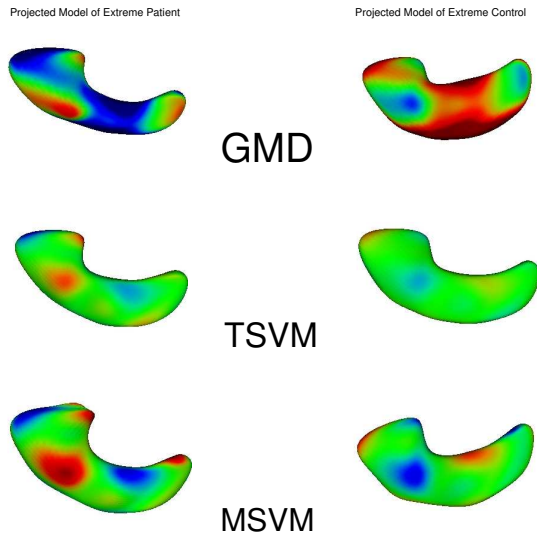


Fig. 4. Diagram showing the structural change captured by the different methods. Red, green, and blue are inward distance, zero distance, and outward distance respectively.

features which actually separate the two groups. Among the other methods, MSVM captures the change most strongly.

4. DISCUSSION

In this article we attempt to generalize SVM so that it can handle data on manifolds. The method was applied to m-rep models of hippocampi and compared with other approaches. The results were encouraging. It seems that by virtue of working on the manifold (and not a tangent plane, like TSVM), MSVM provides a good balance of classifying power and informative separating direction. TSVM hardly captures any change. This could be related to overfitting where the separating direction feels the microscopic noisy features of the training data and thus fails to capture the relevant structural changes.

One could argue for comparing the method with the “appropriate” Kernel Embedding approach. It should be recalled that such an approach will not produce interpretable projected m-reps (necessary to analyze the difference between the groups) and thus we are not interested in them. The possibility of doing Euclidean SVM on multiple tangent planes has not been explored here, but it can be an area of research. We believe this is one of the first attempts to do classification working on the manifold and preliminary results suggest that this approach identifies a meaningful separating direction.

5. ACKNOWLEDGEMENTS

The authors would like to thank Sarang Joshi for helpful insights and Ja-Yeon Jeong and Rohit Saboo for help with the

visualization tools and softwares. This research has been supported by NIH grant NCI/P01 CA47982.

6. REFERENCES

- [1] I. Dryden and K. Mardia, “Multivariate shape analysis,” *Sankhya*, vol. 55, pp. 460–480, 1993.
- [2] D. Thompson, *On Growth and Form*, Cambridge University Press, second edition, 1942.
- [3] C. G. Small, *The statistical theory of shape*, Springer, 1996.
- [4] R.A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [5] V. Vapnik, S. Golowich, and A Smola, “Support vector method for function approximation, regression estimation, and signal processing,” *Advances in Neural Information Processing Systems*, vol. 9, pp. 281–287, 1996.
- [6] C. J. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
- [7] J. S. Marron, M. Todd, and J. Ahn, “Distance weighted discrimination,” *Operations Research and Industrial Engineering, Cornell University, Technical Report 1339*, 2004.
- [8] R.O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, 2001.
- [9] T. Hastie, R. Tibshirani, and J.H. Friedman, *The Elements of Statistical Learning*, New York: Springer, 2001.
- [10] S. Pizer, D. Fritsch, P. Yushkevich, V. Johnson, and E. Chaney, “Segmentation, registration, and measurement of shape variation via image object shape,” *IEEE Transactions on Medical Imaging*, vol. 18, no. 851-865, 1999.
- [11] P. Yushkevich, P. T. Fletcher, S. Joshi, A. Thall, and S. M. Pizer., “Continuous medial representations for geometric object modeling in 2d and 3d,” *Image and Vision Computing*, vol. 21, no. 1, pp. 17–28, 2003.
- [12] P. J. Basser, J. Mattiello, and D. Le Bihan, “MR diffusion tensor spectroscopy and imaging,” *Biophysics Journal*, vol. 66, pp. 259–267, 1994.
- [13] B. Sch ö lkopf and A. Smola, *Learning with Kernels.*, Cambridge: MIT Press, 2002.
- [14] H. Blum, “A transformation for extracting new descriptors of shape, In W. Wathen-Dunn, editor,” *Models for the Perception of Speech and Visual Form. MIT Press, Cambridge MA*, pp. 363–380, 1967.
- [15] P. T. Fletcher, C. Lu, and S. Joshi., “Statistics of shape via principal geodesic analysis on lie groups,” *In Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 95101, 2003.
- [16] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi., “Principal geodesic analysis for the study of nonlinear statistics of shape,” *IEEE Transactions on Medical Imaging*, vol. 23, no. 8, pp. 995–1005, 2004.
- [17] M. Styner, J.A. Lieberman, D. Pantazis, and G. Gerig, “Boundary and medial shape analysis of the hippocampus in schizophrenia,” *MedIA*, vol. 197-203, 2004.