# Margin Preserved Approximate Convex Hulls for Classification

Tetsuji Takahashi and Mineichi Kudo

Graduate school of Hokkaido University
Kita-14, Nishi-9, Kita-ku, Sapporo 060-0814, JAPAN
tetsuji@main.ist.hokudai.ac.jp

## Abstract

*The usage of convex hulls for classification is discussed with a practical algorithm, in which a sample is classified according to the distances to convex hulls. Sometimes convex hulls of classes are too close to keep a large margin. In this paper, we discuss a way to keep a margin larger than a specified value. To do this, we introduce a concept of "expanded convex hull" and confirm its effectiveness.*

***Keywords*-Pattern recognition, Margin, Convex hull**

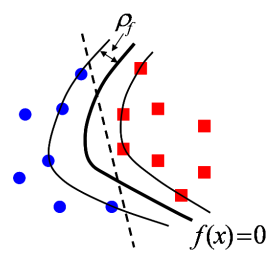



**Figure 1. Nonlinear margin $\rho_f$. The dashed line is the linear decision boundary with maximum linear margin.**

## I. Introduction

Use of convex hulls to approximate each class region has a large potential to realize a large margin classifier. Indeed, the closest point pair on the convex hulls each of which encloses all the samples of a single class gives the same liner classifier as SVM (Support Vector Machine) [1]. Even in a soft-margin linear SVM, a similar explanation is possible by reduced convex hulls [2]. A large difference between them is that SVM keeps the largest linear margin in the kernel space, while convex hulls keep a large nonlinear margin in the original feature space.

In linearly separable cases, SVM maximizes the linear margin between two classes [3], [4]. However, for nonlinearly separable cases, a large nonlinear margin is desirable to be kept. Fig. 1 gives an example. In general, a margin is defined by $\rho_f = \min_i y_i f(\boldsymbol{x}_i)$, and is usually positive, where $(\boldsymbol{x}_i, y_i)$, $\boldsymbol{x}_i \in R^p$, $y_i \in \{-1, +1\}$. In this paper, we consider a method achieving a large marign by using convex hulls.

We have already proposed a practical way to find a set of approximate convex hulls in each class as an estimate of the class region [5]. In the algorithm, training samples are perfectly separated by those convex hulls. Therefore model selection must be made for avoiding overfitting. In this framework, a model is specified by the number of facets, the number of convex hulls, and the number of irregular (too thin or too small) convex hulls. In a previous work [6], we have already tried model selection in the number of facets, and succseeded to improve the performance to some extent.

In this work, from the viewpoint of margin maximization, we modify the algorithm [5] using a new concept of *expanded convex hull*.

## II. Preparations

### A. Approximate Convex Hull [7]

Given a set of points $S$, we define the convex hull of $S$ using *support functions* $H(S, \boldsymbol{w}) = \sup\{\langle \boldsymbol{x}, \boldsymbol{w} \rangle | \boldsymbol{x} \in S\}$ with a unit vector $\boldsymbol{w}$ ($||\boldsymbol{w}|| = 1$), where $\langle \cdot, \cdot \rangle$ denotes the inner product. With the set $W_0$ of all possible vectors $\boldsymbol{w}$ of norm one, one can define the

convex hull as

$$conv(S, W_0) = \bigcap_{\boldsymbol{w} \in W_0} \{\boldsymbol{y} | \langle \boldsymbol{y}, \boldsymbol{w} \rangle \leq H(S, \boldsymbol{w})\}.$$

That is, the convex hull is given by *support planes* $\{\boldsymbol{x} | \langle \boldsymbol{x}, \boldsymbol{w} \rangle = H(S, \boldsymbol{w})\}$ for all possible directions $\boldsymbol{w}$.

Furthermore, by limiting $W_0$ to a finite set $W$, we can obtain an approximate convex hull. Clearly $conv(S, W_0) \subseteq conv(S, W)$, that is, with a finite $W$ we have a larger convex body than the convex hull.

### B. Nonlinear margin

For separation of two sample sets $S$ and $T$ from different classes, it is efficient to consider only the support planes that contribute to the separation of $S$ and $T$ at least in part. By collecting $\boldsymbol{w}$ for such *reflective support planes* [5], we have a smaller subset $W_r$ of $W_0$ while keeping the separation boundary the same.

The *linear margin* in direction $\boldsymbol{w}$ between $S$ and $T$ is defined by

$$\rho_{S,T}(\boldsymbol{w}) = -H(T, -\boldsymbol{w}) - H(S, \boldsymbol{w}).$$

Then, the linear margin between $S$ and $T$ with $W$ is given by

$$\rho_{S,T} = \sup_{\boldsymbol{w} \in W} \rho_{S,T}(\boldsymbol{w}).$$

For the linear margin to separate $conv(S)$ and $conv(T)$, we can use $\rho_{S,T}$. However, we consider more than one convex hull in each class, so that the nonlinear margin has to be defined clearly.

In two-class cases, we consider two families $\mathcal{S} = \{U \subseteq S\}$ and $\mathcal{T} = \{V \subseteq T\}$ where $U$ is a subset of the positive sample set $S$ and $V$ is a subset of the negative sample set $T$, where $U$ and $V$ are given by the algorithm described in the following sections. An example is shown in Fig. 2. Then, the nonlinear margin is defined by

$$\rho_{\mathcal{S},\mathcal{T}} = \min_{(U,V) \in \mathcal{S} \times \mathcal{T}} \rho_{U,V} = \min_{(U,V)} \sup_{\boldsymbol{w} \in W} \rho_{U,V}(\boldsymbol{w}).$$

That is, $\rho_{\mathcal{S},\mathcal{T}}$ is the minimum margin between all possible pairs of convex hulls of different classes.

### C. Approximate nonlinear margin

In what follows we use a finite set $W$ of unit vectors, supposedly close to $W_r$, to construct approximate convex hulls $conv(S, W)$. The *approximate reflective convex hull*, shortly ARCH, has at most $|W|$ faces facing against the opposite class. Any related value is affected by this set $W$. For example, the nonlinear margin becomes a little larger than the case when we use the exact convex hull with $W_0$. Fig. 2 shows ARCHs and the nonlinear margin between those two ARCHs.

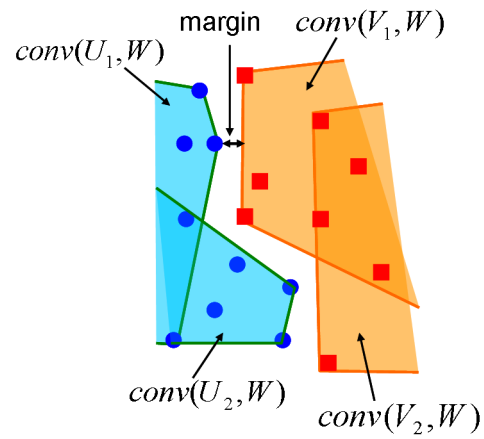



**Figure 2. Estimated class regions by ARCHs and the margin.**

For classification, we need to calculate the distance between a given point $\boldsymbol{x}$ and the boundary of a convex hull. It is easily calculated by $\max_{\boldsymbol{w} \in W} \rho_{S,\{\boldsymbol{x}\}}(\boldsymbol{w})$. This distance takes a negative value when $\boldsymbol{x}$ is inside of $conv(S, W)$. The cost is linear in the dimensionality.

## III. Expanded Convex Hull

It is often the case that two convex hulls are too close, so that the margin is small (Fig. 2). This is due to our strategy to find maximal subsets of samples in each class. However, there is room to widen the margin.

To improve our algorithm [5] so as to make the margin larger, we introduce the concept of an *expanded convex hull*. We define an expanded convex hull of a finite set $S$ ($|S| = n$) as follows:

$$Econv(S, \mu) = \{\boldsymbol{y} | \boldsymbol{y} = \textstyle\sum_{i=1}^n a_i \boldsymbol{x}_i, \boldsymbol{x}_i \in S, \sum_{i=1}^n a_i = 1, \; -\frac{\mu-1}{n} \leq a_i \leq \frac{(n-1)\mu+1}{n}, \mu \geq 1\}.$$

Where, $\mu$ is an expanding parameter. $Econv(S, \mu = 1.0)$ is reduced in the exact convex hull of $S$. That is, $Econv(S)$ is the convex hull of $S_\mu = \{\boldsymbol{x}'_i = \boldsymbol{w} + \mu(\boldsymbol{x}_i - \boldsymbol{w}) | \boldsymbol{w} = \frac{1}{n}\sum_i \boldsymbol{x}_i, \boldsymbol{x}_i \in S\}$. Fig. 3 shows an example of an expanded convex hull. The direction restricted version of an expanded convex hull is given by

$$Econv(S, W, \mu) = \bigcap_{\boldsymbol{w} \in W} \{\boldsymbol{y} | \langle \boldsymbol{y}, \boldsymbol{w} \rangle \leq H(S_\mu, \boldsymbol{w})\}.$$

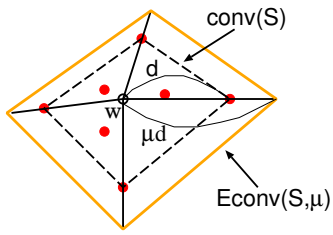



**Figure 3. An expanded convex hull of a finite set $S$ by $\mu$.**

Now we can state the algorithm. We construct a convex hull by adding training samples of the same class one by one starting from the empty set as long as the resultant convex hull does not include any other class samples. To keep the margin to a specific amount, we modify the condition evaluation from the convex hull, original setting in [5], to the expanded convex hull with a prespecified value of $\mu$.

The concrete algorithm is given as follows:

1) Let $S$ be the positive sample set of a target class and $T$ be the negative sample set of the other classes. Let $\mathcal{C} = W = \emptyset$. Let $L$ be an upper bound of the number of convex hulls and $K$ be the number of normal vectors.
2) Find at random $K$ pairs of $\boldsymbol{x}$ ($\in S$) and $\boldsymbol{y}$ ($\in T$) and put $\boldsymbol{w} = \frac{\boldsymbol{y}-\boldsymbol{x}}{\|\boldsymbol{y}-\boldsymbol{x}\|}$ in set $W$.
3) Repeat $L$ times the following Steps 4–5.
4) Let $U = \emptyset$. According to a random presentation order of positive samples, add a positive sample $\boldsymbol{x}$ to $U$ as long as $Econv(U \cup \{\boldsymbol{x}\}, W, \mu) \cap T = \emptyset$.
5) Add the obtained $conv(U, W)$ into $\mathcal{C}$, unless it is already in $\mathcal{C}$.
6) Select a minimal subset of $\mathcal{C}$ by a greedy set cover procedure for all positive samples.

In Step 2, we collect $\boldsymbol{x}$ and $\boldsymbol{y}$ randomly to have a set $W$. This way is valid because such a $\boldsymbol{w}$ is likely to be one for reflective support planes.

## IV. Experiments

We conducted experiments in order to confirm the effectiveness of our new algorithm.

To construct $W$ of normal vectors, we used randomly chosen $n_p$ positive samples and $n_p(c-1)$ negative samples, so that $K = n_p^2(c-1)$ unit vectors were chosen randomly, where $c$ is the number of classes. In following, we fixed the value of $n_p$ to 50, thus, $K$ to $2500(c-1)$. The recognition rate was estimated by 10-fold cross validation. We repeated the algorithm 10 times for reducing the effect of the other random factors, and selected the best one from their results. The loop number $L$ of the randomized subclass method was set to $L = 20$. That is, the number of convex hulls was limited to at most 20 in each class.

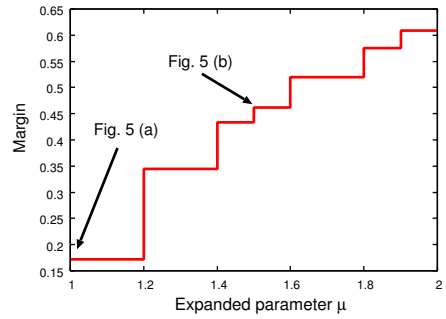


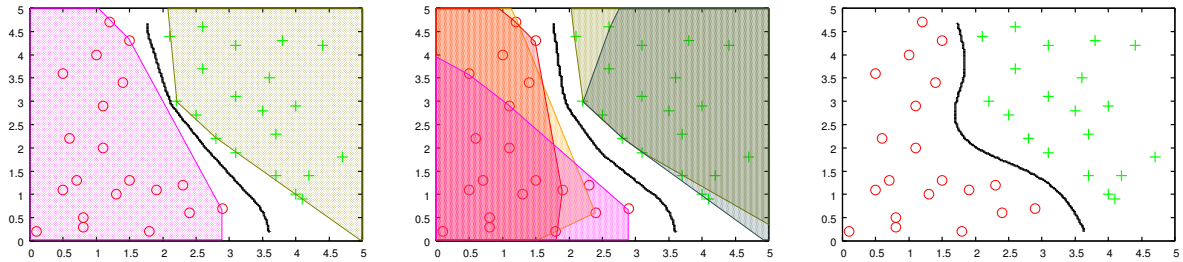

**Figure 4. The nonlinear margin $\rho_{\mathcal{S},\mathcal{T}}$ as the value of $\mu$ increases.**

**Table I. The recognition rate of $\mu$-ARCH.**

| Dataset | Recognition rate | | | |
|---|---|---|---|---|
| | $\mu = 1.0$ | $\mu = 1.1$ | $\mu = 1.2$ | SVM [6] |
| balance | **95.20** | 95.04 | 95.04 | 93.2 |
| diabetes | 77.99 | **78.52** | 77.47 | 64.1 |
| ecoli | 86.30 | **87.49** | 87.18 | 79.8 |
| glass | 54.13 | 51.84 | 49.98 | **66.3** |
| heart | 72.22 | 73.70 | **74.44** | 59.3 |
| ionos | 93.74 | **95.17** | 94.88 | 94.0 |
| iris | 96.67 | 96.67 | 96.67 | **98.0** |
| sonar | 90.86 | 90.38 | 90.86 | 77.4 |
| wine | 89.87 | **89.90** | 89.87 | 72.3 |
| average | 84.10 | **84.30** | 84.04 | 78.3 |

First we used an artificial dataset of two classes in 2-dimensional space. Fig. 4 shows the change of the margin when $\mu$ increases. The value of margin $\rho$ is 0.17 for $\mu = 1.0$ and 0.46 for $\mu = 1.5$. Fig. 5 shows the estimated ARCHs in $\mu = 1.0$ and 1.5. As seen in Fig. 5, we can see that the margin is widened by a larger value of $\mu$. The decision boundary does not differ between two cases of $\mu = 1.0$ and 1.5. They are similar even to that of SVM. However, it should be noted that the result of SVM is with the optimally chosen variance parameter $\sigma$. That is, $\mu$ in our algorithm and $\sigma$ in SVM have to be chosen optimally in some way.

Next, we used 9 datasets taken form UCI machine learning repository [8]. We set the value of $\mu$ to 1.0, 1.1 and 1.2. Table I shows the recognition rates in which the results taken from [6] about SVM with RBF kernel of the default value of $\sigma$ are also presented. In five datasets, an improvement is seen with $\mu > 1.0$. For the other two datasets (*balance* and *glass*), the rate is getting worse with $\mu > 1.0$. On average, the case of $\mu = 1.1$ shows the best performance. In many cases,



(a)$\mu = 1.0$ ($\#conv = 2$) (b)$\mu = 1.5$ ($\#conv = 5$) (b) SVM ($\sigma = 1.0$)

**Figure 5. Estimated regions and decision boundaries: (a)The case of $\mu = 1.0$, (b)The case of $\mu = 1.5$, and (c)The boundary of SVM with an RBF kernel.**

**Table II. Statistics of ARCH.**

| Dataset | #ARCHs | | #samples(%) | |
|---|---|---|---|---|
| | $\mu = 1.0$ | $\mu = 1.1$ | $\mu = 1.0$ | $\mu = 1.1$ |
| balance | 17.5 | 17.6 | 25.4 | 24.4 |
| diabetes | 17.3 | 17.4 | 8.0 | 4.8 |
| ecoli | 9.1 | 10.1 | 12.8 | 10.4 |
| glass | 17.3 | 18.1 | 7.2 | 2.2 |
| heart | 17.5 | 17.7 | 9.1 | 7.0 |
| ionos | 17.4 | 17.6 | 28.3 | 24.6 |
| iris | 6.2 | 7.9 | 28.5 | 27.5 |
| sonar | 17.3 | 17.7 | 38.8 | 32.6 |
| wine | 12.3 | 13.4 | 12.1 | 11.1 |
| average | 14.7 | 15.3 | 18.9 | 16.1 |

the performance is better than that of SVM.

Table II shows the average number of ARCHs per class and the average ratio of included samples in one ARCH. When the value of $\mu$ increases, the number of ARCHs increases and the number of included samples decreases. This is natural because the consistency condition of ARCH becomes stronger than before. In dataset *glass*, the expansion degrades the performance. It turned out that many training samples could not be included anymore in the convex hulls when $\mu > 1.0$, so that the training error was increased. We can see it from the average size of ARCHs in Table II.

## V. Conclusion

We have proposed an algorithm which produces a classifier with a larger margin compared with our previous algorithm. With a user-specified expansion coefficient $\mu$, we can preserve a large margin between approximate convex hulls of classes. To finish this study, we have to carry out model selection in $\mu$ and to give a good bound on the generalization error. There are already good bounds on the basis of the value of margin (e.g., [9] and [10](Chap. 1)), but analysis is not sufficient for nonlinear margin. Especially, a theoretical analysis is desirable for nonlinear margin taken in the original feature space.

## References


[1] D. Zhou, B. Xiao, and H. Zhou, "Global geometry of svm classifiers." *Technical report, AI Lab, Institute of Automation, Chinese Academy of Sciences*, 2002.

[2] M. E. Mavroforakis and S. Theodoridis, "A geometric approach to support vector machine (svm) classification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 671–682, 5 2006.

[3] B. Boser, I. Guyon, and V. Vapnic, "A training algorithm for optimal margin classifiers," *Anuual Workshop on Computational Leaning Theory*, pp. 144–152, 1992.

[4] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 9 1995.

[5] M. Kudo, I. Takigawa, and A. Nakamura, "Classification by reflective convex hulls," *Proceedings of 19th International Conference on Pattern Recognision(ICPR2008), Tampa, Florida, USA*, 2008.

[6] T. Takahashi, M. Kudo, and A. Nakamura, "Classifier selection in a family of polyhedron classifiers," *Progress in Pattern Recognition, Image Analysis, Computer Vision and Applications*, vol. LNCS 5856, pp. 441–448, 2009.

[7] P. Ghosh and K. Kumar, "Support function representation of convex bodies, its application in geometric computing, and some related representations." *Computer Vision and Image Understanding*, vol. 72, no. 3, pp. 379–403, 1998.

[8] A. Asuncion and D. Newman, "UCI machine learning repository," 2007.

[9] V. Vapnik, *Statistical Learning Theory*. John Wiley and Sons, 1998.

[10] A. J. Smola, P. L. Bartlett, B. Scholkopf, and D. Schurmans, Eds., *Advances in Large Margin Classifiers*. The MIT Press, 2000.