

Rejoinder on: Support Vector Machines Maximizing Geometric Margins for Multi-Class Classification

Keiji Tatsumi · Tetsuzo Tanino

Published online: 15 August 2014
© Sociedad de Estadística e Investigación Operativa 2014

First of all, we thank the discussants for the time spent reading and commenting on our paper. Their comments raise interesting issues for further research. We next comment in turn on some of points raised by each discussant.

1 Discussion of Shigeo Abe

Abe raises some interesting questions and informative suggestions. The first question is a possibility of combination of MMSVM with OAO, DAG-SVM or the Mahalanobis metric. If the discriminant function is the same form (6) as the AT method, it is easy to derive an MMSVM in which the feasible region is restricted by the information based on OAO, while a different scalarization method and a different technique for its nonlinear extension are necessary in order to derive a single-objective SOCP. We are just now coming to grips with such extensions. Next, let us consider the Mahalanobis metric. Since the metric is basically measured as the distance from the center of each class, it may be difficult to use the metric directly for evaluating the geometric margin between a sample and the discriminant hyperplane. However, the covariance matrix of each class, which is effectively exploited in some references of Abe's Comment, will be useful in selecting a solution from many Pareto optimal ones of MMSVM.

This rejoinder refers to the comments available at doi:[10.1007/s11750-014-0339-7](https://doi.org/10.1007/s11750-014-0339-7),
doi:[10.1007/s11750-014-0340-1](https://doi.org/10.1007/s11750-014-0340-1), doi:[10.1007/s11750-014-0341-0](https://doi.org/10.1007/s11750-014-0341-0).

K. Tatsumi (✉) · T. Tanino
Graduate School of Engineering, Osaka University, Yamada-oka 2-1, Suita 565-0871 Osaka, Japan
e-mail: tatsumi@eei.eng.osaka-u.ac.jp

T. Tanino
e-mail: tanino@eei.eng.osaka-u.ac.jp

On the other hand, it may be difficult to incorporate MMSVM to DAG-SVMs straightforwardly because a unified discriminant function such as (6) is not used in DAG-SVMs, while the geometric margins might be useful for the classification based on DAG-SVMs.

The second question is about the difference between classification performances of MMSVM and MMSVM-OA shown in Table 7. It can be considered as follows. The numerical experiments demonstrated that the performance of discriminant functions obtained by the SOCPs for MMSVM and MMSVM-OA depend on the selection of a constant vector $\varepsilon_{-r,s}$. In addition, the vectors were selected by using the solutions of the existing AT and OAA, respectively, and the existing AT method is inferior to the OAA for many problems. Therefore, we can conclude that the superiority of MMSVM against MMSVM-OA results from the superiority of OAA. Moreover, the prominent difference for vehicle and glass can be also explained in a similar way. However, note that although solutions obtained by solving SOCPs of both models are remarkably different, Pareto optimal solutions of MMSVM-OA are feasible in (M2) or (NM) in MMSVM, and many of them can be close to Pareto optimal solutions in MMSVM. They show that MMSVM has many desirable solution, while it is critically important to select the desirable solution among many Pareto optimal solutions, and it is our future work.

The third question is about the relation between the sparseness of the solution and training time. In the numerical experiments, we verified that the reduction of the training time in MMSVM-OA is achieved by decreasing of the number of its decision variables and the size of matrices whose eigenvalues are required. However, since in the numerical experiments we did not evaluate the number of support vectors of solutions obtained by MMSVM or MMSVM-OA, it is difficult to discuss the sparseness of the solutions. The issue is worth analyzing in the future. Finally, Abe's suggestion about the fair comparison in numerical experiments are very informative. Thus, we will make it a reference in the future work.

2 Discussion of Yoonkyung Lee

Lee points out many important issues in the multi-class classification. In the following, we discuss each or some paragraphs of her comment.

2.1 Margin maximization as a form of regularization

In the paragraph "Margin Maximization as a Form of Regularization", Lee casts doubt on the significance of the distinction between the geometric and functional margins in the multi-category case, and its substantial differences in practice. In addition, she asserts that the original Vapnik's arguments rest on the functional margin rather than the geometric margin. Indeed, the functional margin $\|w\|$ works effectively as a measure of the complexity of a discriminant function and it is the most important element in binary classification. However, we cannot agree to her arguments. In our view, such a regularization term is effective only if $\|w\|$ can exactly measure the distance between samples and the discriminant hyperplane. It can be verified as follows:

Although in the existing AT model (O), the functional margins $\|w^p - w^q\|$ are minimized and its feasible region includes desirable solutions, its discriminant function is often inferior to other existing methods. They mean that the formulation is not suitable for multi-class classification, and the regularization term does not work effectively.

In addition, the proposed models maximizing the geometric margins, especially MMSVM-OA, can obtain better discriminant functions than other existing models in the numerical experiments. The results demonstrate the significance of maximizing the geometric margins. Of course, other factors also may contribute to the good performance of MMSVM-OA, as Lee mentioned in paragraph named “Differential Penalties”. For example, MMSVM-OA may find the solution by implicitly exploiting an appropriate balance between classes in a solution obtained by OAA. However, MMSVM-OA can improve the discriminant function obtained by the OAA, which confirms that maximizing the geometric margins is one of primary factors. At the same time, we do not claim that maximizing geometric margin is only the primary factor, we simply emphasize that geometric margins between samples and the discriminant hyperplanes are essential to construct a discriminant function.

Furthermore, in order to find a desirable discriminant function, it is not enough to maximize geometric margins. In the proposed multiobjective models, we have to select desirable solution from many Pareto optimal solutions. Therefore, a new regularization term or other criterion is required, as mentioned in the conclusion of our paper. Therefore, the latest results which Lee refers in paragraphs “Turning Focus from Margin (or Penalty) to Loss” and “Differential Penalties” provide beneficial information for further research.

2.2 Separable vs. non-separable case

In this paragraph, Lee comments that it is not clear how the notion of geometric margin is extended to the non-separable case. However, in our paper, soft-margin MMSVMs is shortly introduced in 4.4 with the references [31, 32]. In the references, the geometric margin and ξ_{ipq} are clearly defined. The variable ξ_{ipq} indicates how much the constraints for the correct classification is violated by sample i between classes p and q . Note that in the existing soft-margin AT, ξ_{ipq} is not exactly measured because of its inaccurate functional margins, and thus, the penalty function in the objective function does not work effectively, which is considerably different from soft-margin binary SVM (SP). The soft-margin MMSVM can overcome these drawbacks. On the other hand, since we can derive some kinds of variations of soft-margin MMSVMs [32], we are now investigating which models are suitable for multi-class classification.

2.3 Computation and empirical validation

Lee raises some questions about the selection of parameter values in MMSVM and MMSVM-OA. With respect to c_{rs} , we theoretically showed the upper bound of c_{rs} in the reference [27] of our paper. In addition, in the numerical experiments we easily verified that 10 is appropriate for c_{rs} for problems shown in Table 2.

With respect to $\varepsilon_{-r,s}$, as she pointed out, it is difficult to show a clear criterion for selecting $\varepsilon_{-r,s}$ theoretically, which is our future work as mentioned above. However, through numerical experiments, we observed that the selection based on the solution of the existing AT or OAA is effective. In addition, there are other possibilities of selecting $\varepsilon_{-r,s}$, and moreover, other scalarization methods can be applied to multiobjective (M2) or (NM). These properties indicate the extensibility of the proposed multiobjective models. At the same time, MMSVM-OA in which $\varepsilon_{-r,s}$ is selected by OAA can be interpreted as an improved method of OAA, and its additional training time is less than OAA. Then, MMSVM-OA can be regarded as an improved combination approach for multi-class classification.

3 Discussion of Yan Guermeur

Guermeur introduces the latest results of the statistical theory of large margin multi-category classifiers, and shows the connection between the geometric margins and the generalization performance. His comment includes highly suggestive and fruitful contents. Unfortunately, from the sense of the bound of margin, it seems to be difficult to justify our proposed models, MMSVM and MMSVM-OA, for the present. The result sounds a little disappointing. Nevertheless, we believe that our work derived from the viewpoint of optimization can contribute to the field of multi-class classification, where we theoretically pointed out the inaccurate formularization of the existing AT model whose error bound is statistically shown, and we proposed new models based on the exact geometric margin, which were empirically demonstrated to be effective and suitable for classification. Furthermore, in general terms, scientific progress cannot be achieved by verification of existing theories but rather by their critique and falsification from various kinds of viewpoints. Therefore, we hope that these discussions will be a trigger to generate a new statistical framework for multi-class classifiers which includes our work.

Acknowledgments Finally, we again thank the discussants for their critical and suggestive discussions which point out several important issues in our work and related research results. We believe that these discussions stimulate the interest of various researchers and will lead to producing new ideas.