# Learning Sign-Constrained Support Vector Machines





Kenya Tajima, Kohei Tsuchida, Esmeraldo Ronnie Rey Zara, Naoya Ohta and Tsuyoshi Kato Faculty of Science and Technology, Gunma University

#### **Experiments and Results**

The convergence of the proposed method was tested using the CORA and MNIST datasets.

From the results, the proposed method worked as expected, converging to the minimum objective error. Note that the duality gaps converged to zero implying that it can be utilized as a stopping criterion.

Next, to demonstrate our proposed method's prediction performance against the conventional SVM-pairwise on classification tasks, we use the 3,583 yeast proteins dataset with annotations under the Smith-Waterman score.

With the best ROC scores in bold-faced and the underlined scores do not have a significant statistical difference from the best score, the proposed method is a promising technique for SVM-pairwise framework.

## Solving the Sign-Constrained SVM

We developed a projected gradient algorithm on P(w) by using the **Pegasos** method (PG) which is already used in the conventional SVM.

Under Sign-Constrained SVM though, the optimal solution lies in a larger ball.



In search for a stopping criterion, we investigated the dual problem  $D(\alpha)$  and use the Frank-Wolfe (FW) framework to solve it.

We found out that the update rule is in closed form and the time complexity is small, signifying the number of iterations is bounded.



The method converges **but** there is no definite stopping criterion since  $P(\boldsymbol{w}_{\star})$  is usually unknown even for a small

$$P(\boldsymbol{w}^{(t)}) - P(\boldsymbol{w}_{\star}) \leq \epsilon.$$



Table 1		
ROC scores for amino acid sequence classification		
Class	Conventional SVM-pairwise	Sign Constrained SVM-pairwise
1	0.731 (0.009)	<b>0.749</b> (0.010)
2	$0.681 \ (0.013)$	<b>0.756</b> (0.012)
3	0.735(0.013)	<b>0.758</b> (0.012)
4	0.745(0.010)	<b>0.768</b> (0.009)
5	0.709(0.017)	<b>0.784</b> (0.008)
6	0.625(0.007)	<b>0.692</b> (0.012)
7	0.628(0.023)	<b>0.702</b> (0.030)
8	0.664(0.018)	<b>0.733</b> (0.018)
9	0.603(0.019)	<b>0.681</b> (0.022)
10	0.712(0.008)	<b>0.739</b> (0.010)
11	0.523(0.026)	<b>0.561</b> (0.029)
12	0.894(0.013)	<b>0.905</b> (0.012)

The proposed method is promising tool for SVM but more benchmarking is needed in other applications which are applicable to it.

[1] Nello Cristianini and John *ICML*, volume 28 of *Proceedings* Shawe-Taylor. Cambridge of Machine Learning Research, university press, 2000. pages 427–435, Atlanta, Georgia, [2] Tsuyoshi Kato, Ayano USA, 17–19 Jun 2013. PMLR. Kobayashi, Wakana Oishi, Syun [4] E.S. Levitin and B.T. Polyak. suke Kadoya, Satoshi Okabe, USSR Computational Naoya Ohta, Mohan Amarasiri, Mathematics and Mathematical and Daisuke Sano. *Physics*, 6(5):1–50, January 1966. Journal of Water and Health, [5] Shai Shalev-Shwartz, Yoram 17(3):404–415, April 2019. Singer, Nathan Srebro, and [3] Martin Jaggi. In Sanjoy Andrew Cotter. Math. Program., Dasgupta and David McAllester, 127(1):3–30, 2011. editors, Proceedings of the 30th

This work was supported by JSPS KAKENHI Grant Number 19K04661. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics.



(Simulated with  $d = 2; w_1 \ge 0, w_2 \ge 0$ .)

### Conclusion

#### References

## Acknowledgements