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

Problem Objective
For a small number of samples under a domain knowledge known in advance, how can we learn support vector machines (SVM) in which we can reduce the misclassification rate?

SVM vs Sign-Constrained SVM
Let \( n \) be the number of samples and \( d \) be the number of features of each sample \( x \in \mathbb{R}^d \).

Conventional SVM

- True Distribution of Positive Examples
- True Distribution of Negative Examples

Sign-Constrained SVM

- True Distribution of Positive Examples
- True Distribution of Negative Examples

Water Quality of Rivers and E. coli
Water coming from rain, runoff from industries and households can contribute to the growth of Escherichia coli (E. Coli) in rivers.

Knowing these information from domain knowledge can help us in designing features so that the selected features are possibly correlated to the classes.

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 5,853 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 \( P(\omega) \) 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.

The method converges but there is no definite stopping criterion since \( P(\omega_*) \) is usually unknown even for a small enough \( \epsilon \) satisfying

\[
P(\omega^{(t)}) - P(\omega_*) \leq \epsilon.
\]

Proposed Frank-Wolfe Algorithm

\[
\text{for } t = 1, 2, \ldots \text{ do}
\]

\[
\text{begin}
\]

\[
\text{if } P(\omega) \geq D(\alpha) \geq P(\omega^{(t)}) - P(\omega_*) \text{ then, this will serve as the stopping criterion for the proposed method.}
\]

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

<table>
<thead>
<tr>
<th>Class</th>
<th>Conventional SVM-pairwise</th>
<th>Sign Constrained SVM-pairwise</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.731 (0.009)</td>
<td>0.740 (0.010)</td>
</tr>
<tr>
<td>2</td>
<td>0.681 (0.013)</td>
<td>0.756 (0.012)</td>
</tr>
<tr>
<td>3</td>
<td>0.735 (0.013)</td>
<td>0.758 (0.012)</td>
</tr>
<tr>
<td>4</td>
<td>0.745 (0.010)</td>
<td>0.768 (0.009)</td>
</tr>
<tr>
<td>5</td>
<td>0.709 (0.017)</td>
<td>0.784 (0.008)</td>
</tr>
<tr>
<td>6</td>
<td>0.625 (0.007)</td>
<td>0.692 (0.012)</td>
</tr>
<tr>
<td>7</td>
<td>0.628 (0.023)</td>
<td>0.702 (0.030)</td>
</tr>
<tr>
<td>8</td>
<td>0.664 (0.018)</td>
<td>0.733 (0.018)</td>
</tr>
<tr>
<td>9</td>
<td>0.693 (0.019)</td>
<td>0.681 (0.022)</td>
</tr>
<tr>
<td>10</td>
<td>0.712 (0.006)</td>
<td>0.739 (0.010)</td>
</tr>
<tr>
<td>11</td>
<td>0.523 (0.026)</td>
<td>0.561 (0.029)</td>
</tr>
<tr>
<td>12</td>
<td>0.894 (0.013)</td>
<td>0.905 (0.012)</td>
</tr>
</tbody>
</table>

References

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