Implicitly Constrained Least Squares Classification

Jesse H. Krijthe[1,2] & Marco Loog[1,3]

1 Pattern Recognition Laboratory, Delft University of Technology
2 Department of Molecular Epidemiology, Leiden University Medical Center
3 The Image Group, Department of Computer Science, University of Copenhagen

Background

In many machine learning applications, obtaining labeled objects for training is significantly more expensive than obtaining unlabeled objects. Some examples are document classification, protein function prediction or object recognition in images.

The goal of semi-supervised learning is to use these unlabeled objects to aid in the supervised prediction task. In practice, one at least wants to be sure using these unlabeled object will not degrade the performance of a supervised procedure trained using only the labeled data. Current semi-supervised approaches offer no such guarantees. Implicitly constrained semi-supervised learning aims at conservative semi-supervised learning, and attempts to offer guarantees on the non-degradation in performance.

Concept

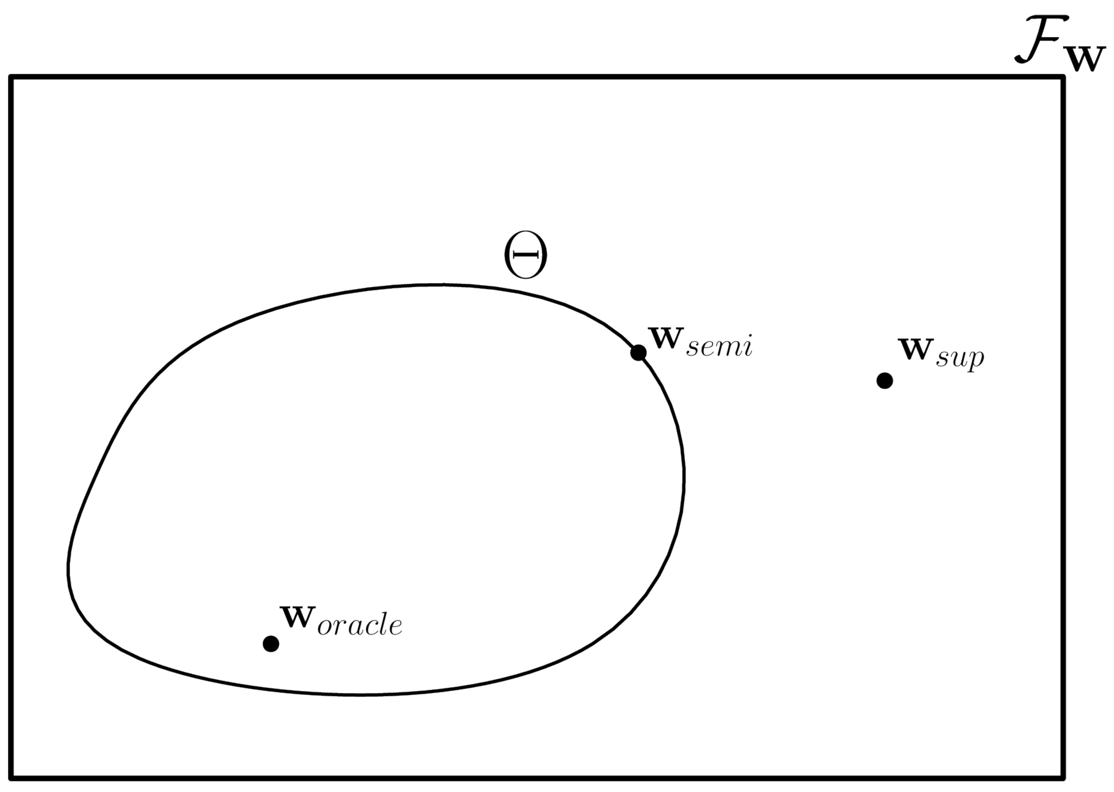
Let $\mathbf{X}$ be the design matrix containing the feature vectors of the labeled objects and the vector $\mathbf{y}$ the corresponding labels. The solution vector of the supervised least squares classifier is given by:

$$\mathbf{w}_{\text{sup}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

Additionally, let $\mathbf{X}_\text{e}$ be the design matrix of the labeled objects, extended with the unlabled objects. If we were to know the labels of the unlabeled objects $\mathbf{y}_\text{u}$, it would be obvious how to improve the supervised solution, by simply using this equation making use of all objects. Since we do not have these labels, the idea behind ICLS is to consider the solutions of the Least Squares Classifier for all possible labelings of the unlabeled objects. These form the set

$$\Theta = \left\{ \left(\mathbf{X}_\text{e}^\top \mathbf{X}_\text{e}\right)^{-1} \mathbf{X}_\text{e}^\top \begin{bmatrix} \mathbf{y} \\ \mathbf{y}_\text{u} \end{bmatrix} \mid \mathbf{y}_\text{u} \in [-1,1]^{N_u} \right\}$$

Note that this set is guaranteed to contain the oracle solution that we would obtain if we were to know the labels of all objects.



Conceptual representation of the ICLS classifier: we find the supervised solution within the constrained set that is closest to the supervised solution vector, measured in terms of the loss on the labeled objects only.

To choose a solution from this set conservatively, we propose to minimize the supervised empirical risk under the constraint that there has to be some labeling that leads to this solution

$$\mathbf{w}_{\text{semi}} = \underset{\mathbf{w} \in \Theta}{\operatorname{argmin}} \, ||\mathbf{X}\mathbf{w} - \mathbf{y}||^2$$

This can be formulated as a quadratic programming problem in terms of the unknown labels. Its solution gives a semi-supervised solution that almost never degrades performance while often increasing it.

Discussion

- Empirical results indicate the proposed procedure does lead to robust improvement over the supervised alternative.
- An alternative choice of objective function in the last step allows for a theoretical proof of non-degradation in some settings.
- Acts as a data-dependent regularization.
- How do we incorporate regularization or other loss functions?

Implicitly constrained least squares classification offers a semi-supervised version of the least squares classifier for which we can guarantee, in some settings, it will never deteriorate performance compared to the supervised alternative