# IMPLICITLY CONSTRAINED SEMI-SUPERVISED LEARNING

Jesse H. Krijthe & Marco Loog











1.

#### *Robust* semi-supervised learning is important

#### 2.

#### ICSSL interesting direction towards this goal

## Supervised Learning Example



### Supervised Learning Example



#### Semi-Supervised Learning



### Applications

#### Image classification



#### Webpage classification



#### Protein function prediction



#### Part of speech tagging



#### Semi-Supervised Learning



#### Semi-Supervised Learning



### Good choice?



### Good choice?



How do we safely use unlabeled data to improve a classification rule?

## Safe/Robust =

semi-supervised classifier performs no worse than supervised counterpart

# Is robustness important?

# IMPLICITLY CONSTRAINED LEAST SQUARES CLASSIFICATION

#### Least Squares Classification

- Sometimes referred to as Fisher classifier
- Assume two classes labeled 1 and 2
- Simple idea : use ordinary least squares regression on these labels

$$\min_{\mathbf{w}} \|X\mathbf{w} - \mathbf{y}\|_2^2$$

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

## Implicitly constrained SSL

- Idea: If we would be able to generate all possible labelings of the unlabeled objects, the true labeling will be one of them.
- Calculate the least squares classifier for every labeling:

$$\Theta = \left\{ (\mathbf{X}_{e}^{\top} \mathbf{X}_{e})^{-1} \mathbf{X}_{e}^{\top} \begin{bmatrix} \mathbf{y} \\ \mathbf{y}_{u} \end{bmatrix} | \mathbf{y}_{u} \in [1, 2]^{N_{u}} \right\}$$

• And select the "best" classifier from this set

$$\hat{\mathbf{w}}_{semi} = \underset{\mathbf{w}\in\Theta}{\operatorname{argmin}} \quad d(\mathbf{w}, \hat{\mathbf{w}}_{sup})$$

#### Implicitly Constrained SSL





#### An interesting result

#### If we choose:

$$d(\mathbf{w}, \hat{\mathbf{w}}_{sup}) = \sqrt{(\mathbf{w} - \hat{\mathbf{w}}_{sup})^{\top} \mathbf{X}_{e}^{\top} \mathbf{X}_{e} (\mathbf{w} - \hat{\mathbf{w}}_{sup})}$$

Then:

$$L(\mathbf{w}_{semi}, \mathbf{X}_{e}, \mathbf{y}_{e}^{*}) \leq L(\mathbf{w}_{sup}, \mathbf{X}_{e}, \mathbf{y}_{e}^{*})$$

Always better in the *transductive* setting

#### Simulation Studies



### Conclusion

- Robust semi-supervised learning is important
- ICSSL
  - Strong theoretical guarantees in limited settings (currently)
  - Interesting empirical results
  - Very conservative
- Next steps
  - Other classifiers
  - Less conservatism?