

Volume Regularization for Binary Classification

Koby Crammer

Tal Wagner



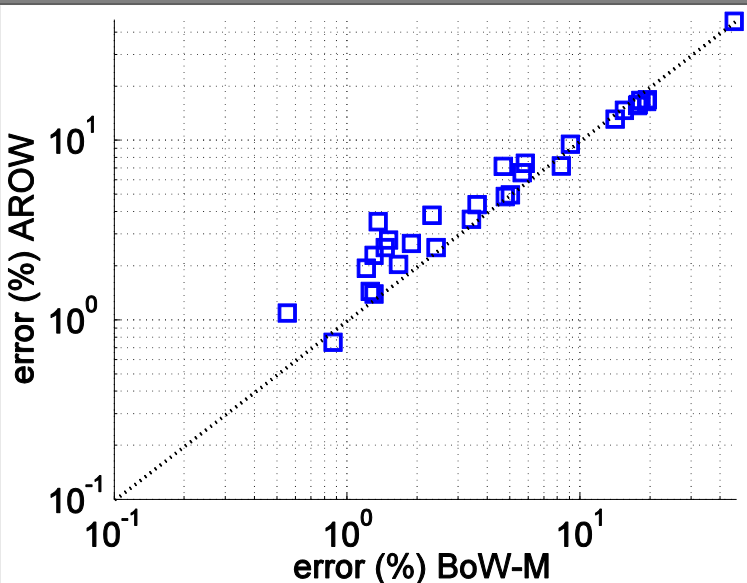
Volume Regularization for Binary Classification



Koby Crammer

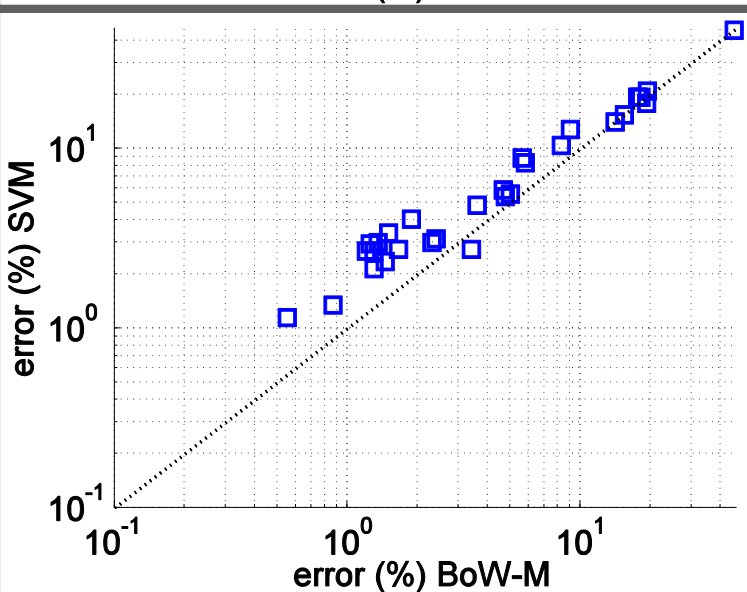
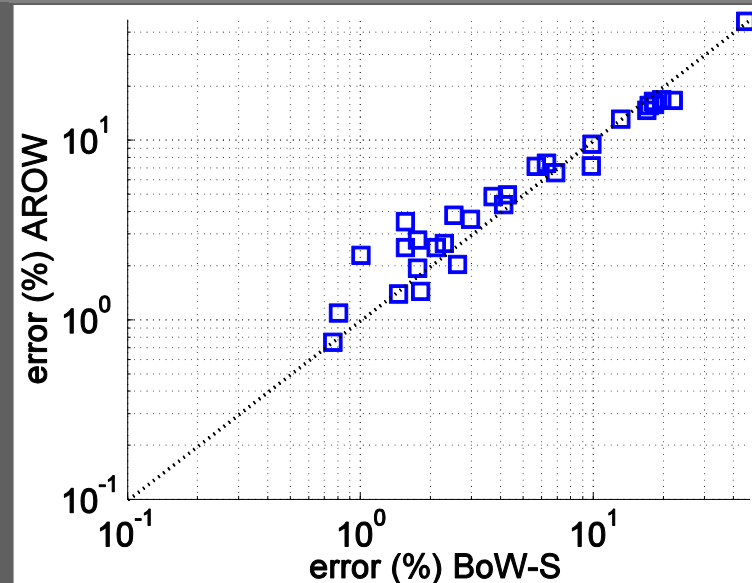
Tal Wagner

[Tu46]

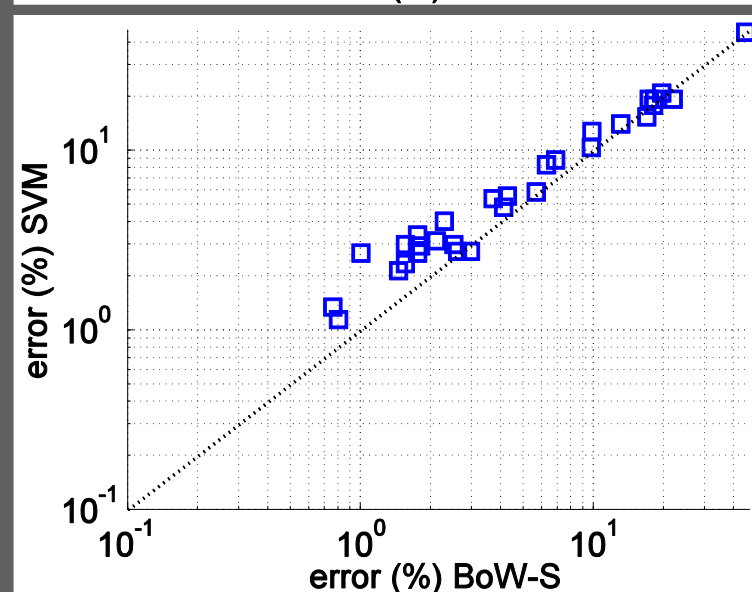


Results on binary 30 NLP datasets

Compared with - SVM
- AROW
(online learning with adaptive weights)

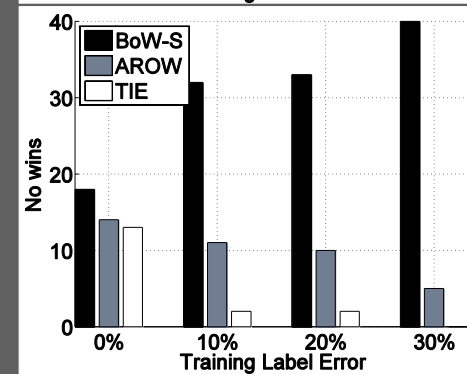
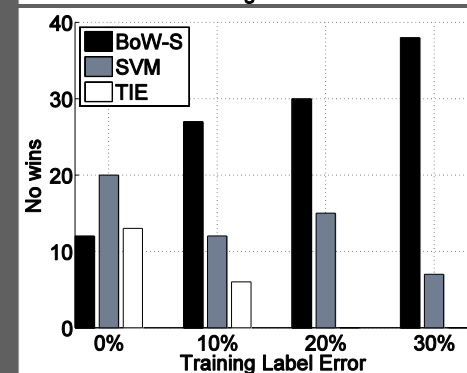
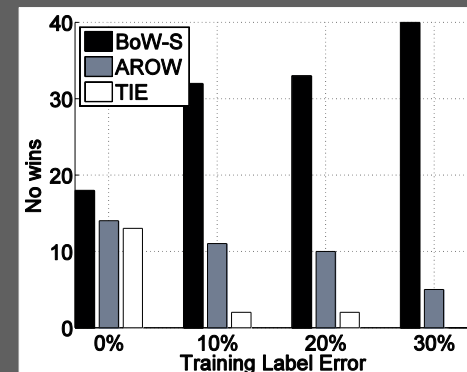
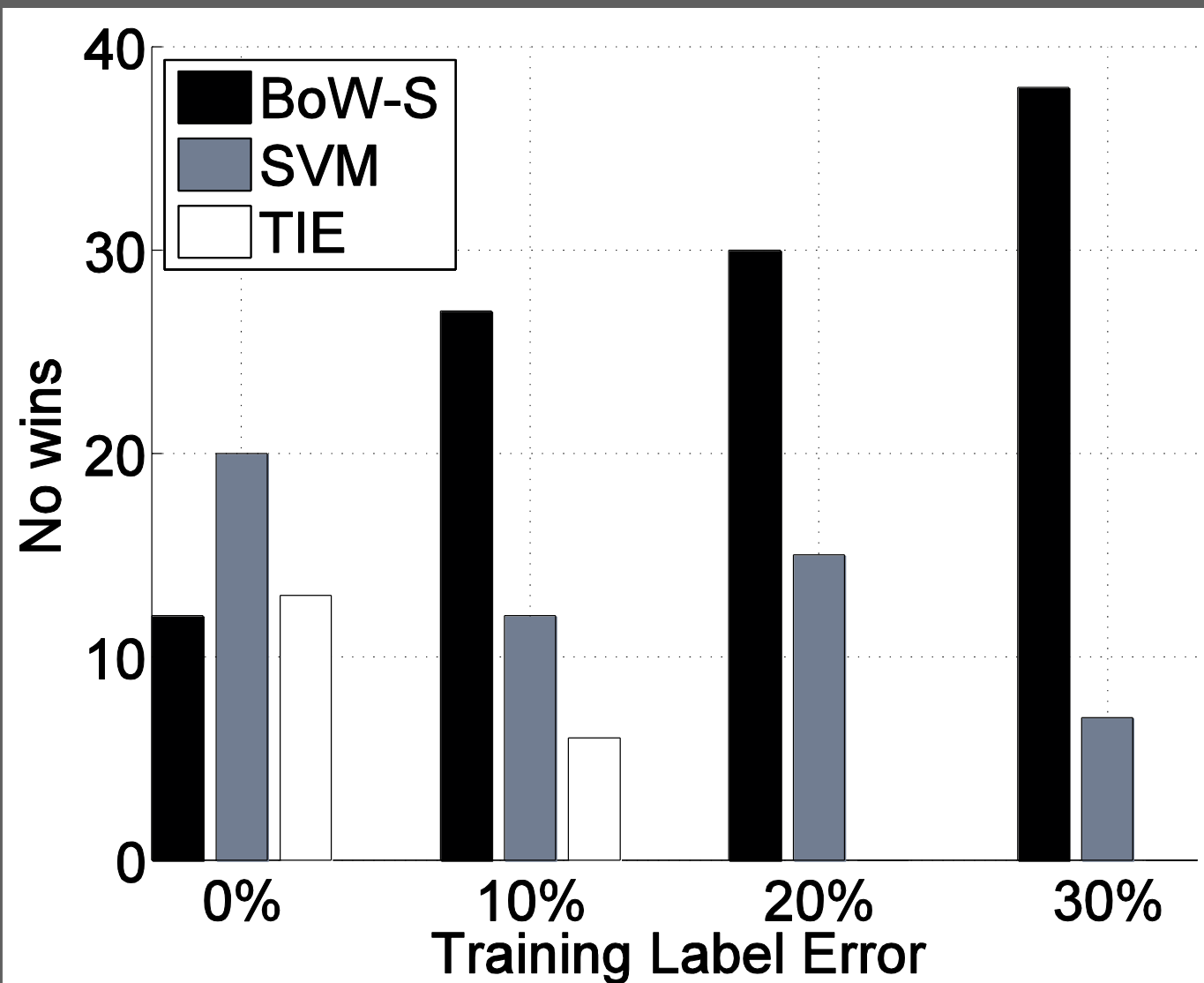


Points above diagonal line indicates superiority of BoW



Volume Regularization for Binary Classification

Number of wins of BoW vs SVM on all 45 pairs of USPS data-set



Volume Regularization for Binary Classification

- Empirical Risk Minimization with Regularization

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \left(\frac{1}{n} \sum_i \ell(\mathbf{w}, (\mathbf{x}_i, y_i)) + \text{const } \mathcal{R}(\mathbf{w}) \right)$$

- Same, over boxes of weights ...

$$B^* = \arg \min_B \sup_{\mathbf{w} \in B} \left(\frac{1}{n} \sum_i \ell(\mathbf{w}, (\mathbf{x}_i, y_i)) + \text{const } \mathcal{R}(\mathbf{w}) \right)$$

Some box
of minimal
size

Worst weight in box

- Two batch formulations
- Generalization bound based on PAC-Bayes
- Optimization algorithm guaranteed to converge
- Experiments ...