The Equivalence of Robustness and Regularization with Non-Perturbable Predictors

Sheng Yang[1]

December 9, 2022

[1]Harvard John Paul School of Engineering and Applied Science

Motivation

Previously, uncertainty set $\mathcal{U}_{(h,g)}$ controls the magnitude of perturbation

$$\mathcal{U}_{(h,g)} = \{\mathbf{\Delta} : ||\mathbf{\Delta}||_{(h,g)} \leqslant \lambda\}$$

This implicitly assumes that every feature is *continuously* perturbable.

However, this assumption may fail:

- Categorical (contrast coded) columns do not admit continuous perturbabtions
    - gender
- Some continuous columns are prespecified without uncertainty
    - medical research controled trials

Penalizing only Perturbable Features

Theorem (Partial Penalization in its General Form)

If $g : \mathbb{R}^n \mapsto \mathbb{R}$ is a non-identically 0 seminorm and $h : \mathbb{R}^n \mapsto \mathbb{R}$ is a norm, then for any $\boldsymbol{z} \in \mathbb{R}^n$ and $\boldsymbol{\beta} \in \mathbb{R}^p$,

$$\max_{\boldsymbol{\Delta} \in \mathcal{U}_{(h,g)}} g(\boldsymbol{z} + \boldsymbol{\Delta S}\beta) = g(\boldsymbol{z}) + \lambda h(\boldsymbol{S}\boldsymbol{\beta}) \tag{1}$$

where $\boldsymbol{S} = diag(s_j)$ and $s_j = 1_{[j\text{-th predictor is perturbable}]}$

Corollary (Partial Penalization in LASSO)

$$\min_{\boldsymbol{\beta}} \max_{||\boldsymbol{\Delta}_i||_2 \leqslant \lambda} ||y - (X + \boldsymbol{\Delta S})\beta||_2 = \min_{\boldsymbol{\beta}} ||y - X\boldsymbol{\beta}||_2 + \lambda \sum_{j: s_j = 1} |\beta_j| \tag{2}$$

Perturbing a subset of features is equivalent to regularizing parameters of the corresponding features!

Application in Robust Classification: SVM

Formulate classification problems to robusitfy against perturbable features:

$$\min_{\beta, b} \max_{\boldsymbol{\Delta} \in \mathcal{U}_q} \sum_{i=1}^{n} \max(1 - y_i(\beta^T(x_i + \boldsymbol{S\Delta}_i) - b), 0) \tag{3}$$

Theorem (Partially Robust SVM)

Equation 3 is equivalent to

$$\begin{aligned} &\min_{\beta, b} \sum_{i=1}^{n} \xi_i \\ s.t.\ \ & y_i(\beta^T x_i - b) - \lambda ||\boldsymbol{S}\beta||_{q^*} \geqslant 1 - \xi_i \quad \forall i \in \{1, ..., n\} \\ & \xi_i \geqslant 0, \quad \forall i \in \{1, ..., n\} \end{aligned} \tag{4}$$

where l_{q^*} is the dual-norm of l_q

Similar results are proven for Logistic Regression and OCT in report.

Empirical Study

Simulation: Partially penalizing Gaussian Mixture; does not do very well

data	terms	LR	LR Regularized	LR Robust	SVM	SVM Regularized	SVM Robust
continuous	both	0.9686 ± 0.015	0.9649 ± 0.023	**0.9705 ± 0.014**	0.9730 ± 0.014	**0.9755 ± 0.011**	0.9753 ± 0.012
	first	**0.9686 ± 0.015**	0.9657 ± 0.021	0.9642 ± 0.019	**0.9730 ± 0.014**	0.9693 ± 0.018	0.9692 ± 0.017
	second	**0.9686 ± 0.015**	0.9650 ± 0.023	0.9638 ± 0.019	**0.9730 ± 0.015**	0.9695 ± 0.018	0.9692 ± 0.018
categorical	first	**0.9425 ± 0.005**	0.9327 ± 0.018	0.9411 ± 0.006	**0.9348 ± 0.009**	0.9307 ± 0.013	0.9327 ± 0.010

Table 1: simulated accuracy across 2000 random initialization

UCI Dataset: 5 dataset with both continuous and categorical columns; marginal Improvement

dataset	LR	LR Regularized	LR Robust	SVM	SVM Regularized	SVM Robust
australian	0.8246 ± 0.03	0.8260 ± 0.04	**0.8304 ± 0.04**	0.7478 ± 0.04	0.7319 ± 0.02	**0.7478 ± 0.04**
bands	0.6894 ± 0.07	**0.6894 ± 0.07**	0.6847 ± 0.07	0.6706 ± 0.02	0.6659 ± 0.02	**0.6706 ± 0.02**
heart	0.8370 ± 0.07	0.8370 ± 0.07	**0.8370 ± 0.07**	0.8259 ± 0.06	**0.8333 ± 0.05**	0.8259 ± 0.06
hepatitis	1.0000 ± 0.00	1.0000 ± 0.00	**1.0000 ± 0.00**	0.9226 ± 0.04	0.9226 ± 0.03	**0.9226 ± 0.04**
horse	0.7100 ± 0.05	0.7167 ± 0.05	**0.7200 ± 0.06**	0.6967 ± 0.04	0.7100 ± 0.06	**0.7167 ± 0.06**

Table 2: 5-fold CV accuracy of UCI dataset

Conclusion

In this work, we have

- provided a theoretical justification of equivalence of robustness and classification with non-perturbable predictors
- demonstrated the empirical weak performance: not robust enough
- Outlined future work: support automatic perturbable set discovery

All codes are public in yangshengaa/robust_classification_partial !