

# The Equivalence of Robustness and Regularization with Non-Perturbable Predictors

Sheng Yang<sup>1</sup>

December 9, 2022

---

<sup>1</sup>Harvard John Paul School of Engineering and Applied Science

# Motivation

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

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

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

However, this assumption may fail:

- Categorical (contrast coded) columns do not admit continuous perturbations
  - gender
- Some continuous columns are prespecified without uncertainty
  - medical research controlled 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  $\mathbf{z} \in \mathbb{R}^n$  and  $\boldsymbol{\beta} \in \mathbb{R}^p$ ,

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

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

## Corollary (Partial Penalization in LASSO)

$$\min_{\boldsymbol{\beta}} \max_{\|\Delta_i\|_2 \leq \lambda} \|y - (X + \Delta \mathbf{S})\boldsymbol{\beta}\|_2 = \min_{\boldsymbol{\beta}} \|y - X\boldsymbol{\beta}\|_2 + \lambda \sum_{j:s_j=1} |\beta_j| \quad (2)$$

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

# Application in Robust Classification: SVM

Formulate classification problems to robustify against perturbable features:

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

## Theorem (Partially Robust SVM)

*Equation 3 is equivalent to*

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

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	$0.9686 \pm 0.015$	$0.9649 \pm 0.023$	<b><math>0.9705 \pm 0.014</math></b>	$0.9730 \pm 0.014$	<b><math>0.9755 \pm 0.011</math></b>	$0.9753 \pm 0.012$
	first	<b><math>0.9686 \pm 0.015</math></b>	$0.9657 \pm 0.021$	$0.9642 \pm 0.019$	<b><math>0.9730 \pm 0.014</math></b>	$0.9693 \pm 0.018$	$0.9692 \pm 0.017$
	second	<b><math>0.9686 \pm 0.015</math></b>	$0.9650 \pm 0.023$	$0.9638 \pm 0.019$	<b><math>0.9730 \pm 0.015</math></b>	$0.9695 \pm 0.018$	$0.9692 \pm 0.018$
categorical	first	<b><math>0.9425 \pm 0.005</math></b>	$0.9327 \pm 0.018$	$0.9411 \pm 0.006$	<b><math>0.9348 \pm 0.009</math></b>	$0.9307 \pm 0.013$	$0.9327 \pm 0.010$

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	$0.8246 \pm 0.03$	$0.8260 \pm 0.04$	<b><math>0.8304 \pm 0.04</math></b>	$0.7478 \pm 0.04$	$0.7319 \pm 0.02$	<b><math>0.7478 \pm 0.04</math></b>
bands	$0.6894 \pm 0.07$	<b><math>0.6894 \pm 0.07</math></b>	$0.6847 \pm 0.07$	$0.6706 \pm 0.02$	$0.6659 \pm 0.02$	<b><math>0.6706 \pm 0.02</math></b>
heart	$0.8370 \pm 0.07$	$0.8370 \pm 0.07$	<b><math>0.8370 \pm 0.07</math></b>	$0.8259 \pm 0.06$	<b><math>0.8333 \pm 0.05</math></b>	$0.8259 \pm 0.06$
hepatitis	$1.0000 \pm 0.00$	$1.0000 \pm 0.00$	<b><math>1.0000 \pm 0.00</math></b>	$0.9226 \pm 0.04$	$0.9226 \pm 0.03$	<b><math>0.9226 \pm 0.04</math></b>
horse	$0.7100 \pm 0.05$	$0.7167 \pm 0.05$	<b><math>0.7200 \pm 0.06</math></b>	$0.6967 \pm 0.04$	$0.7100 \pm 0.06$	<b><math>0.7167 \pm 0.06</math></b>

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](https://github.com/yangshengaa/robust_classification_partial) !

yangshengaa / robust\_classification\_partial Public

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags Go to file Add file >> Code

yangshengaa update ignore and readme after real training	f288bee 5 days ago	10 commits
command	added real dataset and parsing; added real training and scripts	5 days ago
data	added real dataset and parsing; added real training and scripts	5 days ago
results	tested logistic regression; setup synthetic training run script	6 days ago
src	added real dataset and parsing; added real training and scripts	5 days ago
.gitignore	update ignore and readme after real training	5 days ago
LICENSE	Initial commit	9 days ago
README.md	update ignore and readme after real training	5 days ago

About MIT 15.095 ML Opt Final Presentation: Equivalence between Robustness and Regularization with Non-Perturbable Predictors

Readme MIT license 0 stars 1 watching 0 forks

Releases