GEV-Canonical Regression for Accurate Binary Class Probability Estimation when One Class is Rare

Arpit Agarwal[1], Harikrishna Narasimhan[1], Shivaram Kalyanakrishnan[2] and Shivani Agarwal[1]

[1]Indian Institute of Science, Bangalore [2]Yahoo Labs, Bangalore

Abstract

We consider the problem of binary class probability estimation (CPE) when one class is rare compared to the other. It is well known that standard algorithms such as logistic regression do not perform well in this. Common fixes include under-sampling and weighting. Recently, Wang & Dey (2010) suggested the use of a parameterized family of asymmetric link functions based on the generalized extreme value (GEV) distribution. The approach showed promising initial results, but combined with the logarithmic CPE loss implicitly used in their work, it results in a non-convex loss that is difficult to optimize. In this paper, we use tools from the theory of proper composite losses (Buja et al., 2005; Reid & Williamson, 2010) to construct a canonical underlying CPE loss corresponding to the GEV link, which yields a convex proper composite loss that we call the GEV-canonical loss; this loss can be tailored for CPE when one class is rare, and can be easily minimized using an IRLS-type algorithm. Our experiments on both synthetic and real data demonstrate that the resulting algorithm outperforms common approaches such as under-sampling and weights-correction.

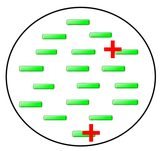
Binary problems where one class is rare

Fraud Detection

Medical Diagnosis

Web Advertising

Binary problems where one class is rare



Very few positive examples compared to negative examples

Problem Setup

Instance space $\mathcal{X}$, Label space $\mathcal{Y} = \{\pm 1\}$, Probability distribution D on

$\eta(x) = \mathbf{P}(Y = 1 \mid X = x)$ $\quad p = \mathbf{P}(Y = 1)$

We are interested in settings where $p \ll 0.5$

Given $S = ((x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)) \sim D^n$

GOAL? Learn a class probability estimation (CPE) model $\widehat{\eta}_S : \mathcal{X} \to [0, 1]$

Previous Approaches

- Logistic regression does not work well in this setting (Zhang, 2004)
- Previous approaches : Weighting, Undersampling, Asymmetric link functions
- Wang & Dey (2010) proposed GEV-log loss which is flexible but non-convex

Our Contribution

A new loss termed GEV-canonical loss which is convex, proper composite and flexible (can be tailored to different degrees of rarity)

Proper Loss function for CPE

A CPE loss function $c : \{\pm 1\} \times [0, 1] \to \overline{\mathbb{R}}_+$ is proper if

$$\eta \in \arg\min_{\widehat{\eta} \in [0,1]} \eta\, c_1(\widehat{\eta}) + (1 - \eta)\, c_{-1}(\widehat{\eta}) \qquad \forall \eta \in [0, 1]$$

and strictly proper if the minimizer is unique

Example : Log Loss

Link Function

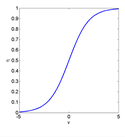
Let $\mathcal{V} \subseteq \mathbb{R}$, A link function

$$\psi : [0, 1] \to \mathcal{V}$$

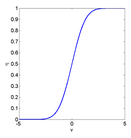
is any strictly increasing (and therefore invertible) function that maps probabilities in $[0, 1]$ to real-valued scores in $\mathcal{V}$

Examples :

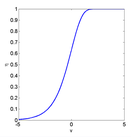
Logit Link



Probit Link



C-loglog link



Proper Composite Loss Function

A loss function $\ell : \{\pm 1\} \times \mathcal{V} \to \overline{\mathbb{R}}_+$ is said to be proper composite if $\exists$ a proper CPE loss $c : \{\pm 1\} \times [0, 1] \to \overline{\mathbb{R}}_+$ and a link $\psi : [0, 1] \to \mathcal{V}$ s.t.

$$\ell(y, v) = c(y, \psi^{-1}(v))$$

For every link function ψ there is a unique canonical proper loss function given by:

$$c_1(\widehat{\eta}) = \int_{\widehat{\eta}}^{1} (1 - q)\, \psi'(q)\, dq\,;$$

$$c_{-1}(\widehat{\eta}) = \int_{0}^{\widehat{\eta}} q\, \psi'(q)\, dq\,,$$

The resulting proper composite loss has some nice properties, including convexity.

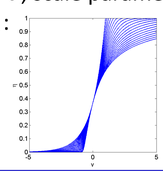
Examples : Logistic Loss = Log Loss + Logit Link — Canonical pair

GEV Link Family

CDF of GEV distribution with location parameter $\mu = 0$, scale parameter $\sigma = 1$ and parameterized by shape parameter $\xi \in \mathbb{R}$:

$$F_\xi(v) = \exp(-(1 + \xi v)_+^{-1/\xi})$$

ξ can be adjusted to fit different degrees of rarity in the data

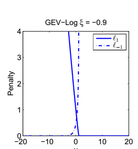


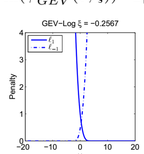
GEV-Log Loss (Wang & Dey, 2010)

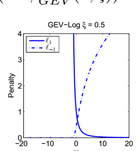
Log Loss + GEV Link = GEV-Log Loss

non-convex loss

$$\ell^{\text{GEV-log}(\xi)}(y, v) = -\mathbf{1}[y = 1]\, \ln(\psi^{-1}_{GEV}(v; \xi)) - \mathbf{1}[y = -1]\, \ln(1 - \psi^{-1}_{GEV}(v; \xi))$$







GEV-Canonical Loss & GEV-Canonical Regression (This paper)

Canonical Proper Loss for GEV Link

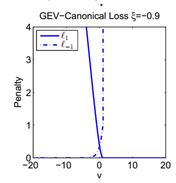
$$c_1^{\text{GEV-can}(\xi)}(\widehat{\eta}) = \int_{\widehat{\eta}}^{1} \frac{1 - q}{q(-\ln q)^{1+\xi}}\, dq$$

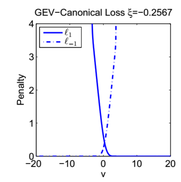
$$c_{-1}^{\text{GEV-can}(\xi)}(\widehat{\eta}) = \int_{0}^{\widehat{\eta}} \frac{1}{(-\ln q)^{1+\xi}}\, dq$$

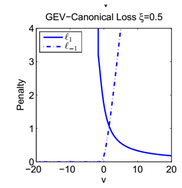
(Canonical Loss) + GEV Link = GEV-Canonical Loss

$$\ell^{\text{GEV-can}(\xi)}(y, v) = -\mathbf{1}[y = 1]\, c_1^{\text{GEV-can}(\xi)}(\psi^{-1}_{\text{GEV}}(v; \xi)) - \mathbf{1}[y = -1]\, c_{-1}^{\text{GEV-can}(\xi)}(1 - \psi^{-1}_{\text{GEV}}(v; \xi))$$

Canonical pair by construction; results in convex loss!





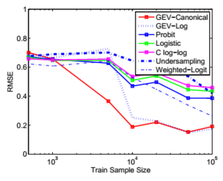


- Can be tailored for the problem of CPE for varying degrees of rarity
- Not available in closed form. But, the gradient and Hessian are available in closed form
- Can be efficiently minimized using IRLS type algorithm termed GEV-canonical regression

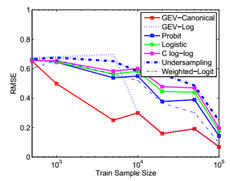
Experiments

- Parameter ξ selected using a validation set.

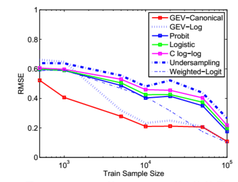
Synthetic Experiments — **Evaluation metric:** Root Mean Squared Error



Dataset 1 : p = 0.0158



Dataset 2 : p = 0.0312



Dataset 3 : p = 0.095

Real Experiments — **Evaluation metric:** Brier Score

Dataset	Logistic regression	Probit regression	Cloglog regression	Undersampling + King-Zeng correction	Weighted Logistic + correction	GEV-log regression	GEV-canonical regression
Nursery	**0.0084**	**0.0084**	0.0088 **	0.0124 **	0.0090 **	0.0172 **	**0.0084**
Letter-a	0.0079 **	0.0084 **	**0.0074**	0.0111 **	0.0112 **	0.0313 **	0.0080 **
Car	0.0266 **	0.0262	0.0267 **	0.0320 **	0.0271 *	0.0296 **	**0.0259**
Glass	0.0670	0.0671	0.0744 **	0.0623	0.0637	**0.0614**	0.0649
Ecoli	0.0646	0.0644	0.0689 **	0.0756 **	**0.0635**	0.0641	0.0641
Letter-vowel	0.1392 **	0.1392 **	0.1400 **	0.1416 **	0.1414 **	0.1405 **	**0.1367**
Cmc	0.1617	0.1617	0.1621	0.1642 **	**0.1615**	0.1626	0.1622
Vehicle	0.1399	0.1395	0.1422	0.1501 **	0.1408	0.1497 **	**0.1394**
Haberman	0.1828 *	0.1812	0.1907 **	0.1823 *	0.1814 **	**0.1761**	0.1769
Yeast	0.1634 **	0.1635 **	0.1666 **	0.1646 **	0.1635 **	0.1621	**0.1616**
German	0.1721	0.1731	0.1737	0.1754 **	**0.1714**	0.1787 **	0.1727
Pima	0.1617 **	0.1623 **	0.1652 **	0.1662 *	0.1626 **	0.1616	**0.1603**

Summary

	Flexible?	Convex?
Logistic	×	√
Probit	×	√
Cloglog	×	√
GEV-log(ξ)	√	×
GEV-canonical(ξ)	√	√

References

1. Wang, X. and Dey, D. K. Generalized extreme value regression for binary response data: an application to b2b electronic payments system adoption. The Annals of Applied Statistics, 4 (4):2000–2023, 2010.
2. Reid, M. D. and Williamson, R. C. Composite binary losses. Journal of Machine Learning Research, 11:2387–2422, 2010.
3. Buja, A., Stuetzle, W., and Shen, Y.. Loss functions for binary class probability estimation: Structure and applications. Technical report, University of Pennsylvania, November 2005.