Proceedings of Machine Learning Research 260:-, 2024 ACML 2024

Towards Calibrated Losses for Adversarial Robust Reject Option Classification Supplementary Material

Vrund Shah VRUND.SHAH@RESEARCH.IIIT.AC.IN
Tejas Chaudhari TEJAS.CHAUDHARI@RESEARCH.IIIT.AC.IN
Naresh Manwani NARESH.MANWANI@IIIT.AC.IN
Machine Learning Lab, Kohli Research Block, IIIT Hyderabad, India - 500032

1. Proof of Proposition 8

Proposition *The Adversarial Robust Reject Option Loss ℓ_d^γ for the class of linear classifiers is γ-right shift of ℓ_d loss as follows.*

$$\ell_d^\gamma(yf(\mathbf{x}),\rho) = (1-d)\ \mathbb{1}_{\{yf(\mathbf{x})<-\rho+\gamma\}} +\ d\ \mathbb{1}_{\{\ yf(\mathbf{x})\leq\ \rho+\gamma\ \}} \tag{1}$$

Proof Let ℓ be a non-increasing function of $yf(\mathbf{x})$. The following property holds for ℓ.

$$\sup_{\mathbf{x}}\ \ell(yf(\mathbf{x})) = \ell(\inf_{\mathbf{x}}\ yf(\mathbf{x})) \tag{2}$$

Both indicator functions, $\mathbb{1}_{\{yf(\mathbf{x}')<-\rho\}}$ and $\mathbb{1}_{\{yf(\mathbf{x}')<\rho\}}$, are non-increasing with $yf(\mathbf{x}')$. Hence using (2) in the definition of the Adversarial Robust Reject Option Loss, we have

$$\ell_d^\gamma(yf(\mathbf{x}),\rho) = (1-d)\ \mathbb{1}_{\{\ \inf_{\mathbf{x}':\|\mathbf{x}-\mathbf{x}'\|\leq\gamma} yf(\mathbf{x}')<-\rho\ \}} +\ d\ \mathbb{1}_{\{\ \inf_{\mathbf{x}':\|\mathbf{x}-\mathbf{x}'\|\leq\gamma} yf(\mathbf{x}')\leq\rho\ \}}. \tag{3}$$

For $\mathcal{H}_{\text{lin}}$, $f(\mathbf{x}) = \mathbf{w}\cdot\mathbf{x}$ with $\|\mathbf{w}\| = 1$. The optimization problem formulated in eq. (3) is as follows :

$$\begin{aligned} \min_{\mathbf{x}'} \quad & y\ (\mathbf{w}\cdot\mathbf{x}') \\ \text{s.t.} \quad & \|\mathbf{x}-\mathbf{x}'\|\leq\gamma \end{aligned} \tag{4}$$

(4) is a convex optimization problem. The Lagrangian is given by

$$\mathcal{L}(\mathbf{x}',\lambda) = y\ \mathbf{w}\cdot\mathbf{x}' + \lambda\ (\|\mathbf{x}-\mathbf{x}'\|-\gamma)$$

where $\lambda\in\mathbb{R}$ is a Lagrangian multiplier. Applying KKT conditions, we get the following

1. $y\ \mathbf{w} - \lambda\ \frac{\mathbf{x}-\mathbf{x}'}{\|\mathbf{x}-\mathbf{x}'\|} = 0$
2. $\lambda\geq 0$
3. $\|\mathbf{x}-\mathbf{x}'\|\leq\gamma$
4. $\lambda\ (\|\mathbf{x}-\mathbf{x}'\|-\gamma) = 0$ (Complementary Slackness)

Using the condition from complementary slackness , we have the trivial case when $\lambda = 0$ as the objective function value is always 0. For $\lambda \neq 0$, it holds that $\|\mathbf{x} - \mathbf{x}'\| = \gamma$. Hence, the constraint $\|\mathbf{x} - \mathbf{x}'\| \leq \gamma$ is activated. From 1.) we have $\mathbf{x} - \mathbf{x}' = \frac{y\,\gamma}{\lambda}\mathbf{w}$. But, $\|\mathbf{x} - \mathbf{x}'\| = \gamma$, so $\|\frac{y\,\gamma}{\lambda}\mathbf{w}\| = \gamma$. Solving, we get $\lambda = \|\mathbf{w}\|$ and $\mathbf{x} - \mathbf{x}' = \frac{y\,\gamma}{\|\mathbf{w}\|}\mathbf{w}$.

The optimal solution to (4) is given by $(\mathbf{x}')^* = \mathbf{x} - \frac{y\,\gamma\,\mathbf{w}}{\|\mathbf{w}\|}$. Substituting this in (3), we get

$$\ell_d^\gamma(yf(\mathbf{x}), \rho) = (1-d)\ \mathbb{1}_{\{yf(\mathbf{x}) < -\rho+\gamma\ \}} +\ d\ \mathbb{1}_{\{\ yf(\mathbf{x}) \leq\ \rho+\gamma\ \}} \tag{5}$$

which is equivalent to γ-right shift of ℓ_d.

■

2. Proof of Lemma 9

Lemma *The excess-inner risk for target loss ℓ_d^γ is given by*

$$\Delta\mathcal{C}_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta) = \mathcal{C}_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta) - \mathcal{C}^*_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta) =$$

$$\begin{cases} (\eta-d)\ \mathbb{1}_{\min\{\eta,1-\eta\}-d\geq 0} + |2\eta-1|\ \mathbb{1}_{2\eta-1>0}\ \mathbb{1}_{\min\{\eta,1-\eta\}-d<0} & \text{if } \alpha < -\rho-\gamma \\ (1-d)\ \eta\ \mathbb{1}_{\min\{\eta,1-\eta\}-d\geq 0} + \\ \{(\eta-(1-\eta)(1-d))\ \mathbb{1}_{2\eta-1>0} + (1-\eta)d\ \mathbb{1}_{2\eta-1<0}\}\mathbb{1}_{\min\{\eta,1-\eta\}-d<0} & \text{if } -\rho-\gamma \leq \alpha < -\rho+\gamma \\ \{(d-(1-\eta))\ \mathbb{1}_{2\eta-1>0} + (d-\eta)\ \mathbb{1}_{2\eta-1<0}\}\mathbb{1}_{\min\{\eta,1-\eta\}-d<0} & \text{if } -\rho+\gamma \leq \alpha \leq \rho-\gamma \\ (1-d)\ (1-\eta)\ \mathbb{1}_{\min\{\eta,1-\eta\}-d\geq 0} + \\ \{\eta\ d\ \mathbb{1}_{2\eta-1>0} + ((1-\eta)-\eta\ (1-d))\ \mathbb{1}_{2\eta-1<0}\}\mathbb{1}_{\min\{\eta,1-\eta\}-d<0} & \text{if } \rho-\gamma < \alpha \leq \rho+\gamma \\ (1-\eta-d)\ \mathbb{1}_{\min\{\eta,1-\eta\}-d\geq 0} + |2\eta-1|\ \mathbb{1}_{2\eta-1<0}\ \mathbb{1}_{\min\{\eta,1-\eta\}-d<0} & \text{if } \rho+\gamma < \alpha \end{cases}$$

Proof A case by case breakdown of the definition of $\mathcal{C}_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta)$ is the first step. Depending upon prediction or rejection, the minimal inner-risk will be $\mathcal{C}^*_{\ell_d^\gamma,\mathcal{H}}(\eta) = \mathcal{C}^*_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta) = \min\{\eta, 1-\eta, d\}$. Additionally, we also assume that $\rho+\gamma < 1$. For each sub-case, a further splitting is done based on the minimal inner-risk value and using $\Delta\mathcal{C}_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta)$, we have the desired result. We prove it for one of the cases and for rest of the cases, proof follows the similar procedure. Consider the case of $\alpha < -\rho-\gamma$: when the minimal inner risk is d, then, $\Delta\mathcal{C}_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta) = \mathcal{C}_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta) - \Delta\mathcal{C}^*_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta) = \eta - d$. When the minimal inner risk is η, $\Delta\mathcal{C}_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta) = \eta-\eta = 0$ and for the sub-case when minimal inner risk is $1-\eta$, $\Delta\mathcal{C}_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta) = \eta-(1-\eta) = 2\eta-1$. Combining all the sub-cases using indicator functions, we have for

$$\Delta\mathcal{C}_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta) = (\eta-d)\ \mathbb{1}_{\min\{\eta,1-\eta\}-d\geq 0} + |2\eta-1|\ \mathbb{1}_{2\eta-1>0}\ \mathbb{1}_{\min\{\eta,1-\eta\}-d<0}$$

■

3. Proof of Theorem 10

Theorem *Any margin-based surrogate ℓ is $(\ell_d^\gamma, \mathcal{H})$-calibrated if and only if it satisfies the following :*

$$\inf_{\rho-\gamma<\alpha\leq\|\mathbf{x}\|} \mathcal{C}_{\ell,\mathcal{H}}(\alpha, \frac{1}{2}) > \inf_{0\leq\alpha\leq\|\mathbf{x}\|} \mathcal{C}_{\ell,\mathcal{H}}(\alpha, \frac{1}{2}) \tag{6}$$

$$\inf_{-\|\mathbf{x}\|\leq\alpha\leq\rho+\gamma} \mathcal{C}_{\ell,\mathcal{H}}(\alpha, \eta) > \inf_{-\|\mathbf{x}\|\alpha\leq\|\mathbf{x}\|} \mathcal{C}_{\ell,\mathcal{H}}(\alpha, \eta) \;\; \eta \in (\frac{1}{2}, 1] \tag{7}$$

Proof Let ℓ be a margin-based surrogate to ℓ_d^γ. Using Proposition 5, we have that ℓ is $(\ell_d^\gamma, \mathcal{H})$-calibrated if and only if its corresponding calibration function $\delta(\epsilon) > 0, \; \forall\epsilon$. The case for $\eta = 0.5$ is dealt separately. Based on range of η, two cases are made and for each one, a further split is made based on prediction or rejection, and then, the calibration function is computed. This further has 3 sub-cases - (based on the "Bayes classifier" and change in definition of $C_{\ell_d^\gamma,\mathcal{H}}$)

1. $1 - \eta < d$
2. $d \leq 1 - \eta$ and $\eta \geq \eta_{\text{right}}$
3. $d \leq 1 - \eta$ and $\eta < \eta_{\text{right}}$

$$\delta(\epsilon) = \delta_1 + \delta_2 \tag{8}$$

where

$$\delta_1 = \delta_r(\epsilon) \; \mathbb{1}_{\{\min(\eta,1-\eta)-d\geq 0\}} \tag{9}$$

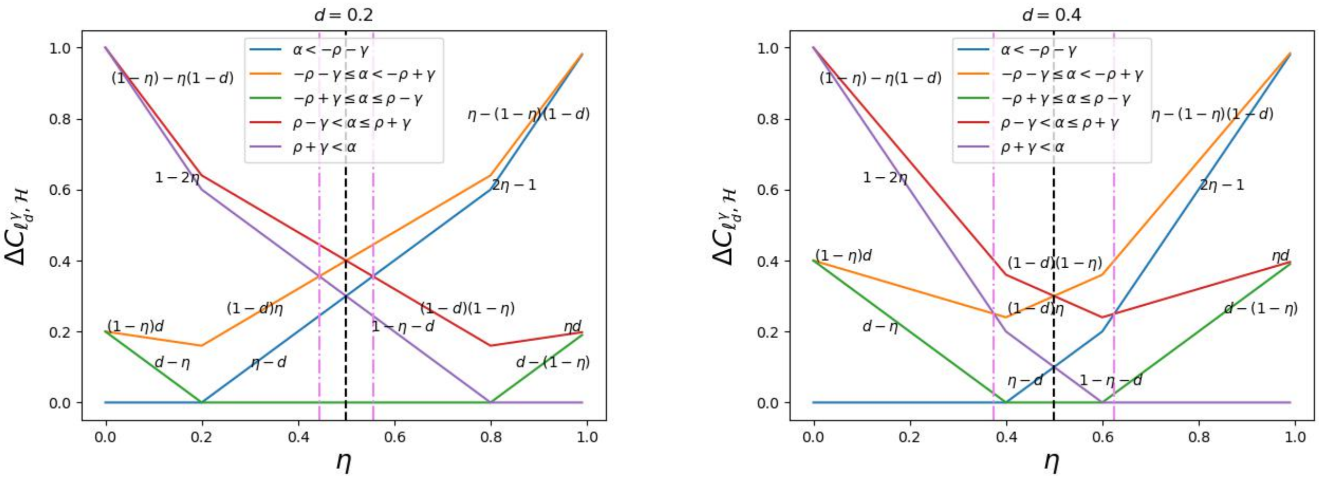
$$\delta_2 = \delta_p(\epsilon) \; \mathbb{1}_{\{\min(\eta,1-\eta)-d< 0\}} \tag{10}$$

For $\delta(\epsilon) > 0$ to hold $\forall\eta \in [0, 1]$, we need either $\delta_1(\epsilon) > 0$ or $\delta_2(\epsilon) > 0$ to hold $\forall\eta \in [0, 1]$. Note that one among δ_1 or δ_2 is always 0. We use Figure 1 to split the case of $\eta > 0.5$, into further sub-cases. First split is based on prediction or rejection, i.e minimizer is $\{\eta, 1 - \eta\}$ or d. In the prediction case, a definition change occurs around the points η_{left} (when $\eta < 0.5$) or η_{right} (when $\eta > 0.5$), as seen in Figure 1.

Case i) : $\eta > \frac{1}{2}$
Sub-case A) : $\min\{\eta, 1 - \eta\} < d$ (prediction)

$$\delta_p(\epsilon) = \begin{cases} \infty & \text{if } \epsilon > \eta - (1-\eta)(1-d) \\ \inf_{\{\alpha:-\rho-\gamma\leq\alpha<-\rho+\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } \eta - (1-\eta)(1-d) \geq \epsilon > 2\eta - 1 \\ \inf_{\{\alpha:\alpha<-\rho-\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } 2\eta - 1 \geq \epsilon > \eta\, d \\ \inf_{\{\alpha:\rho-\gamma<\alpha\leq\rho+\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } \eta\, d \geq \epsilon > d - (1-\eta) \\ \inf_{\{\alpha:-\rho+\gamma\leq\alpha\leq\rho-\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } d - (1-\eta) \geq \epsilon \end{cases} \tag{11}$$

Sub-case B) : $\min\{\eta, 1 - \eta\} \geq d$ (rejection)



(a) (b)

Figure 1: Graph of excess target risk vs η for two different d values.

I) $\eta \geq \eta_{\text{right}}$

$$\delta_r(\epsilon) = \begin{cases} \infty & \text{if } \epsilon > (1-d)\,\eta \\ \inf_{\{\alpha: -\rho-\gamma \leq \alpha < -\rho+\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } (1-d)\,\eta \geq \epsilon > \eta - d \\ \inf_{\{\alpha: \alpha < -\rho-\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } \eta - d \geq \epsilon > (1-\eta)(1-d) \\ \inf_{\{\alpha: \rho-\gamma < \alpha \leq \rho+\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } (1-\eta)(1-d) \geq \epsilon > 1-\eta-d \\ \inf_{\{\alpha: -\rho+\gamma \leq \alpha \leq \rho-\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } 1-\eta-d \geq \epsilon \end{cases} \tag{12}$$

II) $\eta < \eta_{\text{right}}$ [Narrow band]

$$\delta_r(\epsilon) = \begin{cases} \infty & \text{if } \epsilon > (1-d)\,\eta \\ \inf_{\{\alpha: -\rho-\gamma \leq \alpha < -\rho+\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } (1-d)\,\eta \geq \epsilon > (1-\eta)(1-d) \\ \inf_{\{\alpha: \alpha < -\rho-\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } (1-\eta)(1-d) \geq \epsilon > \eta - d \\ \inf_{\{\alpha: \rho-\gamma < \alpha \leq \rho+\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } \eta - d \geq \epsilon > 1-\eta-d \\ \inf_{\{\alpha: -\rho+\gamma \leq \alpha \leq \rho-\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } 1-\eta-d \geq \epsilon \end{cases} \tag{13}$$

NOTE: For margin-based surrogate, $\mathcal{C}_{\ell,\mathcal{H}}(f(\mathbf{x}),\eta)$ and $\Delta\mathcal{C}_{\ell,\mathcal{H}}(f(\mathbf{x}),\eta)$ are symmetrical about $\eta = \frac{1}{2}$. Hence, the definitions for Case ii) can be obtained by replacing η with $1-\eta$ from Case i).

Case ii) : $\eta < \frac{1}{2}$

Sub-case A) : $\min\{\eta, 1-\eta\} < d$ (prediction)

$$\delta_p(\epsilon) = \begin{cases} \infty & \text{if } \epsilon > (1-\eta) - \eta(1-d) \\ \inf_{\{\alpha:-\rho-\gamma\leq\alpha<-\rho+\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } (1-\eta)-\eta(1-d) \geq \epsilon > 1-2\eta \\ \inf_{\{\alpha:\alpha<-\rho-\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } 1-2\eta \geq \epsilon > (1-\eta)\, d \\ \inf_{\{\alpha:\rho-\gamma<\alpha\leq\rho+\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } (1-\eta)\, d \geq \epsilon > d-\eta \\ \inf_{\{\alpha:-\rho+\gamma\leq\alpha\leq\rho-\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } d-\eta \geq \epsilon \end{cases} \tag{14}$$

Sub-case B) : $\min\{\eta, 1-\eta\} \geq d$ (rejection)
I) $\eta \leq \eta_{\text{left}}$

$$\delta_r(\epsilon) = \begin{cases} \infty & \text{if } \epsilon > (1-d)\,(1-\eta) \\ \inf_{\{\alpha:-\rho-\gamma\leq\alpha<-\rho+\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } (1-d)\,(1-\eta) \geq \epsilon > 1-\eta-d \\ \inf_{\{\alpha:\alpha<-\rho-\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } 1-\eta-d \geq \epsilon > \eta\,(1-d) \\ \inf_{\{\alpha:\rho-\gamma<\alpha\leq\rho+\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } \eta\,(1-d) \geq \epsilon > \eta-d \\ \inf_{\{\alpha:-\rho+\gamma\leq\alpha\leq\rho-\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } \eta-d \geq \epsilon \end{cases} \tag{15}$$

II) $\eta > \eta_{\text{left}}$ [Narrow band]

$$\delta_r(\epsilon) = \begin{cases} \infty & \text{if } \epsilon > (1-d)\,(1-\eta) \\ \inf_{\{\alpha:-\rho-\gamma\leq\alpha<-\rho+\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } (1-d)\,(1-\eta) \geq \epsilon > (1-d)\,\eta \\ \inf_{\{\alpha:\alpha<-\rho-\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } (1-d)\,\eta \geq \epsilon > 1-\eta-d \\ \inf_{\{\alpha:\rho-\gamma<\alpha\leq\rho+\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } 1-\eta-d \geq \epsilon > \eta-d \\ \inf_{\{\alpha:-\rho+\gamma\leq\alpha\leq\rho-\gamma\}} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) & \text{if } \eta-d \geq \epsilon \end{cases} \tag{16}$$

For each of these sub-cases, we arrive at this definition using the graph below.

For each case, the corresponding calibration function definitions are (11) , (12) and (13) respectively. Using **Proposition 5**, it holds that ℓ is $(\ell_d^\gamma, \mathcal{H})$-calibrated if and only if its corresponding calibration function $\delta(\epsilon) > 0$. Applying this for each case , we get

1. $\inf_{\alpha>\rho+\gamma} \mathcal{C}_{\ell,\mathcal{H}}(\alpha, 1-\eta) > \inf_{\alpha\in\mathrm{Z}} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta)$

   $\equiv$

   $\inf_{\alpha<-\rho-\gamma} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) > \inf_{\alpha\in\mathrm{Z}} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta)$

2. $\inf_{\rho-\gamma<\alpha\leq\rho+\gamma} \mathcal{C}_{\ell,\mathcal{H}}(\alpha, 1-\eta) > \inf_{\alpha\in\mathrm{Z}} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta)$

   $\equiv$

   $\inf_{-\rho-\gamma\leq\alpha\leq-\rho+\gamma} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) > \inf_{\alpha\in\mathrm{Z}} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta)$

3. $\inf_{|\alpha|\leq\rho-\gamma} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) > \inf_{\alpha\in\mathrm{Z}} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta)$

   $\equiv$

   $\inf_{-\rho+\gamma\leq\alpha\leq\rho-\gamma} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) > \inf_{\alpha\in\mathrm{Z}} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta)$

4. $\inf_{\rho-\gamma<\alpha\le\rho+\gamma} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) > \inf_{\alpha\in \mathrm{Z}} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta)$

where $\mathrm{Z} = \big[-\|\mathbf{x}\| \ , \ \|\mathbf{x}\|\big]$.

By combining all 4 conditions mentioned above, we get

$$\inf_{-\|\mathbf{x}\|\le\alpha\le\rho+\gamma} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) > \inf_{-\|\mathbf{x}\|\alpha\le\|\mathbf{x}\|} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta) \tag{17}$$

Now, we compute $\delta(\epsilon)$ for $\eta = 0.5$

$$\Delta\mathcal{C}_{\ell_d^\gamma,\mathcal{H}}(\alpha,\eta) = \begin{cases} 0 & \text{if } \ |\alpha| \le \rho-\gamma \\ \frac{1-d}{2} & \text{if } \ \rho-\gamma < |\alpha| \le \rho+\gamma \\ \frac{1}{2}-d & \text{if } \ \rho+\gamma < |\alpha| \end{cases} \tag{18}$$

$$\delta(\epsilon) = \begin{cases} \infty & \text{if } \ |\alpha| \le \rho-\gamma \text{ or } \ \epsilon > \frac{1-d}{2} \\ \inf_{\rho-\gamma<|\alpha|\le\rho+\gamma} \Delta C_{\ell,\mathcal{H}}\ (\alpha,\frac{1}{2}) & \text{if } \ |\alpha| > \rho-\gamma \text{ and } \ \frac{1-d}{2} \ge \epsilon > \frac{1}{2}-d \\ \inf_{|\alpha|>\rho+\gamma} \Delta C_{\ell,\mathcal{H}}\ (\alpha,\frac{1}{2}) & \text{if } \ |\alpha| > \rho-\gamma \text{ and } \ \frac{1}{2}-d \ge \epsilon \end{cases} \tag{19}$$

$$\text{Bayes-inner risk : } \mathcal{C}^*_{\ell_d^\gamma,\mathcal{H}}(\alpha,\tfrac{1}{2}) = d$$

Using **Proposition 5** on (19), we get

A surrogate ℓ is $(\ell_d^\gamma,\mathcal{H})$ calibrated if and only if

$$\inf_{\rho-\gamma<|\alpha|\le\rho+\gamma} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\frac{1}{2}) > 0 \implies \inf_{\rho-\gamma<\alpha\le\rho+\gamma} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\frac{1}{2}) > \inf_{\alpha\in \mathrm{Z}} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\frac{1}{2}) \tag{20}$$

and

$$\inf_{|\alpha|>\rho+\gamma} \Delta\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\frac{1}{2}) > 0 \implies \inf_{\alpha>\rho+\gamma} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\frac{1}{2}) > \inf_{\alpha\in \mathrm{Z}} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\frac{1}{2}) \tag{21}$$

By combining (20) and (21) , we get

$$\inf_{\rho-\gamma<\alpha\le\|\mathbf{x}\|} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\frac{1}{2}) > \inf_{0\le\alpha\le\|\mathbf{x}\|} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\frac{1}{2}) \tag{22}$$

NOTE : $\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\frac{1}{2})$ is symmetric about 0. So, $\mathcal{C}^*_{\ell,\mathcal{H}}(\alpha,\frac{1}{2}) = \inf_{0\le\alpha\le\|\mathbf{x}\|} \mathcal{C}_{\ell,\mathcal{H}}(\alpha,\frac{1}{2})$. Thus, for any surrogate ℓ to be $(\ell_d^\gamma,\mathcal{H})$-calibrated if and only if it satisfies (6) and (7).

■

4. Proof of Theorem 11

Theorem *Let ℓ be a differentiable and convex margin based surrogate to ℓ_d^γ. Then, ℓ is not $(\ell_d^\gamma, \mathcal{H})$-calibrated.*

Proof Assume that ℓ, convex, differentiable surrogate to ℓ_d^γ is $\mathcal{H}$-calibrated. For $\eta = \frac{1}{2}$, the minimizer of the conditional risk lies at 0. As ℓ is convex, $\mathcal{C}_{\ell,\mathcal{H}}(\alpha, \eta)$ is also convex and $\mathcal{C}_{\ell,\mathcal{H}}(\alpha, \frac{1}{2}) = 0.5\ \bar{\ell}(\alpha)$. From convexity of $\bar{\ell}$, we have $\bar{\ell}(0) \leq \bar{\ell}(\alpha)\ \forall \alpha$. Thus, calibration condition (6) is satisfied. Next, we consider the case when $\eta > \frac{1}{2}$ and use proof by contradiction.

Any convex function on a compact set $[\theta_1, \theta_2] \subset \mathbb{R}$ can be characterised as :

1. Non-increasing
2. Non-decreasing
3. Non-increasing upto ω ($\in [\theta_1, \theta_2]$) and non-decreasing on $[\omega, \theta_2]$

Using the above characterization, for calibration condition (7) to hold, two cases are possible.

Let $\alpha^\star = \text{argmin}_\alpha\ \mathcal{C}_{\ell,\mathcal{H}}(\alpha, \eta)$.

Case i) : $\rho + \gamma < \alpha^\star \leq \|\mathbf{x}\|$ (when there exists a minimizer inside the compact set)

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} \mathcal{C}_{\ell,\mathcal{H}}(\alpha, \eta)\bigg|_{\alpha=\alpha^\star} = 0$$

$$\therefore\ \eta\ \ell'(\alpha^\star) = (1-\eta)\ \ell'(-\alpha^\star)$$

As $\eta \in (\frac{1}{2}, 1]$, we have that $\frac{\eta}{1-\eta} > 1$. Thus, $\ell'(-\alpha^\star) > \ell'(\alpha^\star)$. But as ℓ is convex, ℓ' is monotone. Hence, $[\ell'(\alpha^\star) - \ell'(-\alpha^\star)](2\ \alpha^\star) \geq 0$. This implies that $\ell'(\alpha^\star) \geq \ell'(-\alpha^\star)$. Hence, we have arrived at a contradiction.

Case ii) : $\alpha^\star > \|\mathbf{x}\|$ (non-increasing on the compact set)

Then, it holds that $\frac{\mathrm{d}}{\mathrm{d}\alpha} \mathcal{C}_{\ell,\mathcal{H}}(\alpha, \eta)\Big|_{\alpha=\|\mathbf{x}\|} < 0$ and since (7) requires the minima to lie in $(\rho+\gamma, \|\mathbf{x}\|\ ]$, the following must hold :

$$\mathcal{C}_{\ell,\mathcal{H}}(\rho+\gamma, \eta) > \mathcal{C}_{\ell,\mathcal{H}}(\|\mathbf{x}\|, \eta) \quad \forall \mathbf{x} \text{ such that } \|\mathbf{x}\| > \rho + \gamma$$

Using the definition of conditional risk and rearranging the terms, we get

$$\left(\frac{\eta}{1-\eta}\right)[\ \ell(\|\mathbf{x}\|) - \ell(\rho+\gamma)\ ] < [\ \ell(-\rho-\gamma) - \ell(-\|\mathbf{x}\|)\ ]$$

Since $\eta \in (\frac{1}{2}, 1]$, it holds that $\frac{\eta}{1-\eta} > 1$. Thus, $\ell(\|\mathbf{x}\|) - \ell(\rho+\gamma) < \ell(-\rho-\gamma) - \ell(-\|\mathbf{x}\|)$ which implies that $\bar{\ell}(\|\mathbf{x}\|) < \bar{\ell}(\rho+\gamma)$. But, $\bar{\ell}$ is a even, convex function. Hence, it holds that $\bar{\ell}(\|\mathbf{x}\|) - \bar{\ell}(\rho+\gamma) \geq \bar{\ell}'(\rho+\gamma)\ [\ \|\mathbf{x}\| - (\rho+\gamma)\ ]$ and $\bar{\ell}'(0) = 0$. So, $\bar{\ell}'(\rho+\gamma) > 0$ and we have that $\bar{\ell}(\|\mathbf{x}\|) \geq \bar{\ell}(\rho+\gamma)$ resulting in a contradiction.

Thus, no differentiable convex surrogate is $(\ell_d^\gamma, \mathcal{H})$-calibrated.

∎

5. Proof of Theorem 12

Theorem *No margin-based surrogate ℓ satisfying the property of Quasi-concavity of the conditional risk $\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta)$ in α, $\forall \eta \in [0,1]$ is $(\ell_d^\gamma, \mathcal{H})$-calibrated.*

Proof Let ℓ be a margin-based surrogate whose $\mathcal{C}_{\ell,\mathcal{H}}(\alpha,\eta)$ is quasi-concave in α , $\forall \eta \in [0,1]$ and assume that ℓ is $(\ell_d^\gamma, \mathcal{H})$-calibrated. Let $\bar{\ell}(\alpha) = \ell(\alpha) + \ell(-\alpha)$. Since it holds true $\forall \eta \in [0,1]$, it must hold for $\eta = \frac{1}{2}$. At $\eta = \frac{1}{2}$, Quasi-concavity is transferred onto $\bar{\ell}$. Also, every quasi-concave function on $\mathbb{R}$ can be characterized as following :

1. non-increasing on $\mathbb{R}$
2. non-decreasing on $\mathbb{R}$
3. non-decreasing up to a point of maxima θ i.e on $(-\infty, \theta]$, constant upto to ω $(\theta \leq \omega)$ and non-increasing on $[\omega, \infty)$.

Also, $\bar{\ell}$ is symmetric about 0. Hence, Quasi-concavity for even function would imply that first two cases essentially are reduced to constant functions. Else, third case prevails and we get two scenarios, maxima on either side of 0, both of which imply that $\bar{\ell}(\alpha)$ is non-increasing for $\alpha > 0$. For any surrogate ℓ to be $(\ell_d^\gamma, \mathcal{H})$-calibrated , it must satisfy (6) i.e

$$\inf_{\alpha \in (\rho-\gamma, \|x\|\ ]} \mathcal{C}_{\ell,\mathcal{H}}(\alpha, \frac{1}{2}) > \inf_{\alpha \in [0, \|x\|\ ]} \mathcal{C}_{\ell,\mathcal{H}}(\alpha, \frac{1}{2})$$

$$\therefore \inf_{\alpha \in (\rho-\gamma, \|x\|\ ]} \bar{\ell}(\alpha) > \inf_{\alpha \in [0, \|x\|\ ]} \bar{\ell}(\alpha)$$

This is in contradiction to $\bar{\ell}(\alpha)$ being non-increasing. Hence, our initial assumption was incorrect.

■

6. Reproducibility

Link to the repository containing the code files for reproducing the simulations is given here.