Deep Imbalanced Regression

Yang Y, Zha K, Chen Y, Wang H, Katabi D

ICML 2021

Presenter: Gianmarco Midena

26 November 2024

Overview

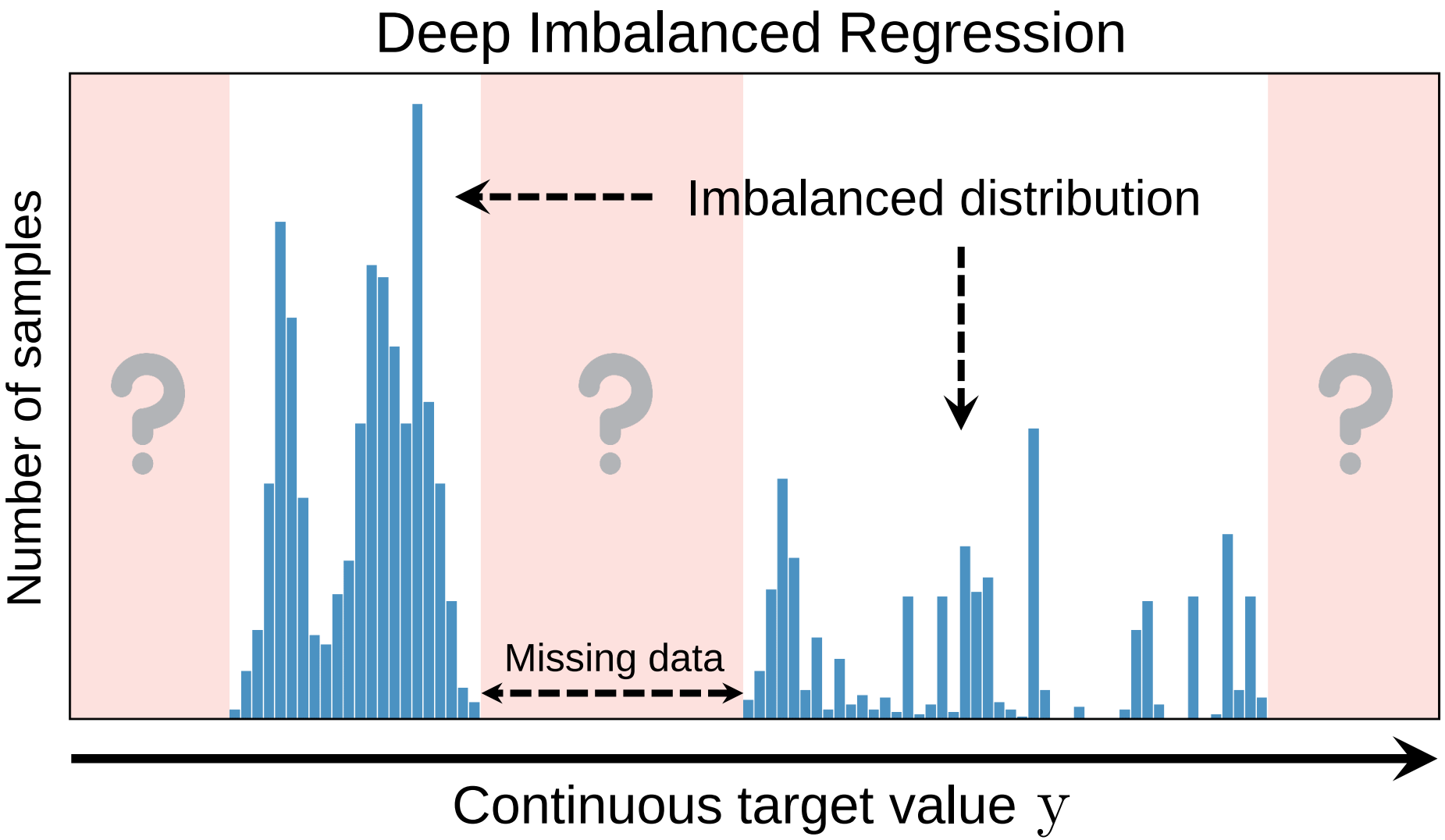
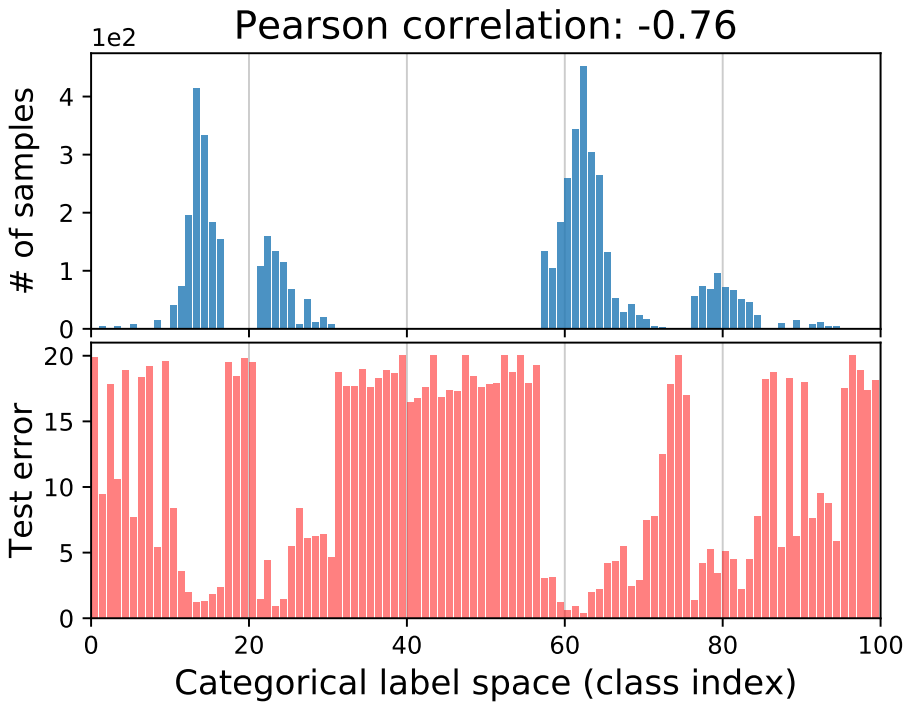
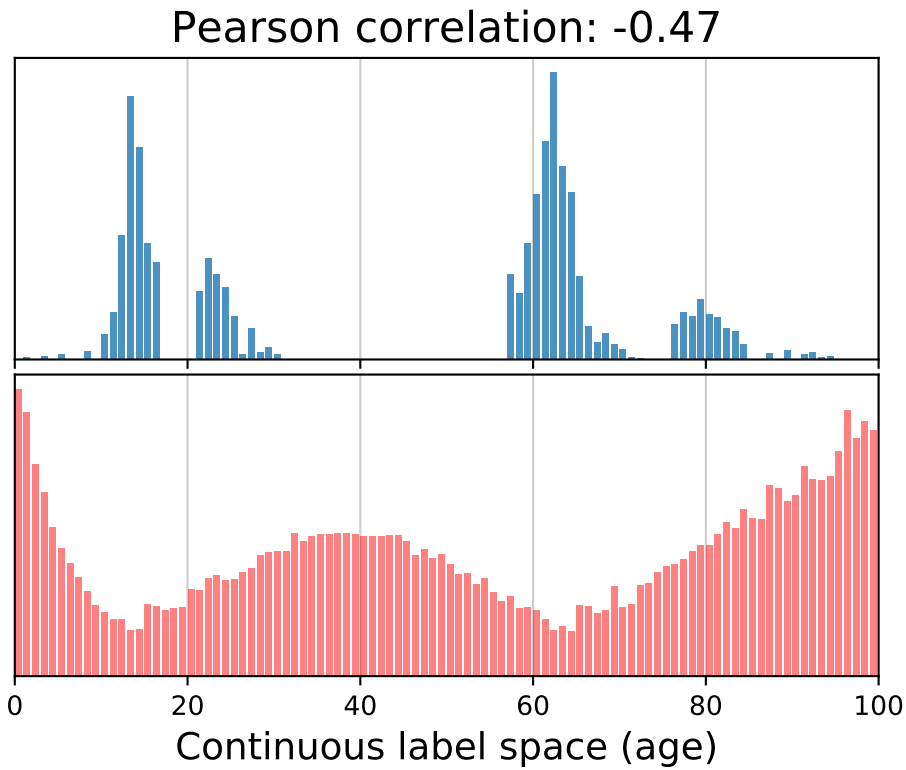


Image credit: *Yang et al. (2021)*

Test Error on Categorical vs. Continuous Label Space



(a) CIFAR-100 (subsampled)



(b) IMDB-WIKI (subsampled)

Image credit: *Yang et al. (2021)*

Problem Settings

- $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$: training set
- $\mathbf{x}_i \in \mathbb{R}^d$: input
- $y_i \in \mathcal{Y}$: continuous label or target
- $b_i \in \mathcal{B}$: discrete label or target
- $\mathcal{Y} \subset \mathbb{R}$: continuous label space
- $\mathcal{B} = \{1, \ldots, M\} \subset \mathbb{Z}^+$: index space
  - divides $\mathcal{Y}$ into M groups (bins) with equal intervals $[t_j, t_{j+1})$
  - $\{[t_0, t_1), \ldots, [t_{M-1}, t_M)\}$: discrete label space
  - $t_k \in \mathcal{Y}$
  - minimum resolution
    - ★ e.g., $\delta y \triangleq t_{j+1} - t_j = 1$ in age estimation
- $\hat{y}_i = g(\mathbf{z}_i) \in \mathbb{R}$: predicted continuous label
- $\mathbf{z}_i = f(\mathbf{x}_i; \theta) \in \mathbb{R}^{d'}$: learned representation
- θ: trainable model parameters

Label Distribution Smoothing

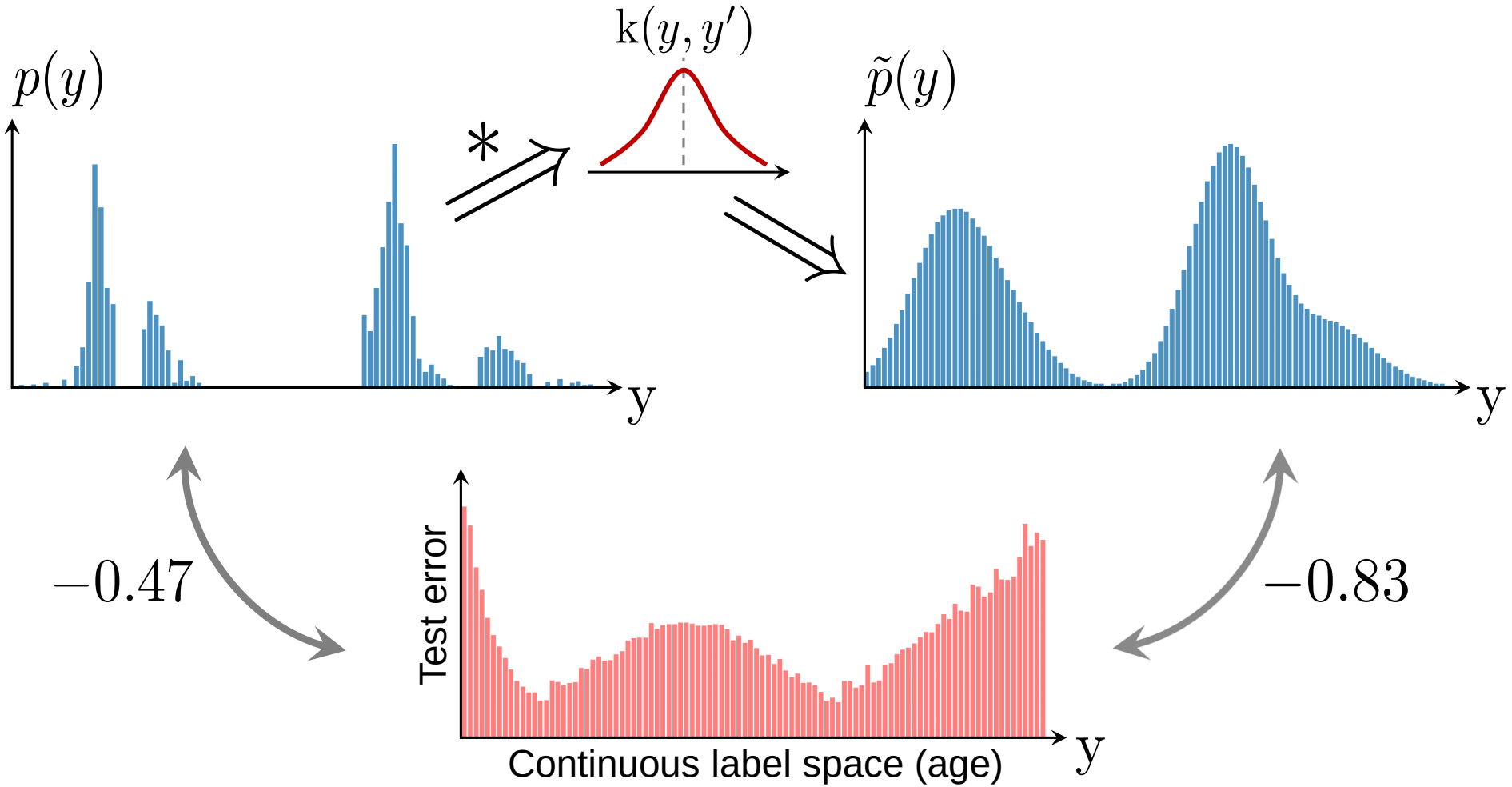


Image credit: *Yang et al. (2021)*

Feature Distribution Smoothing

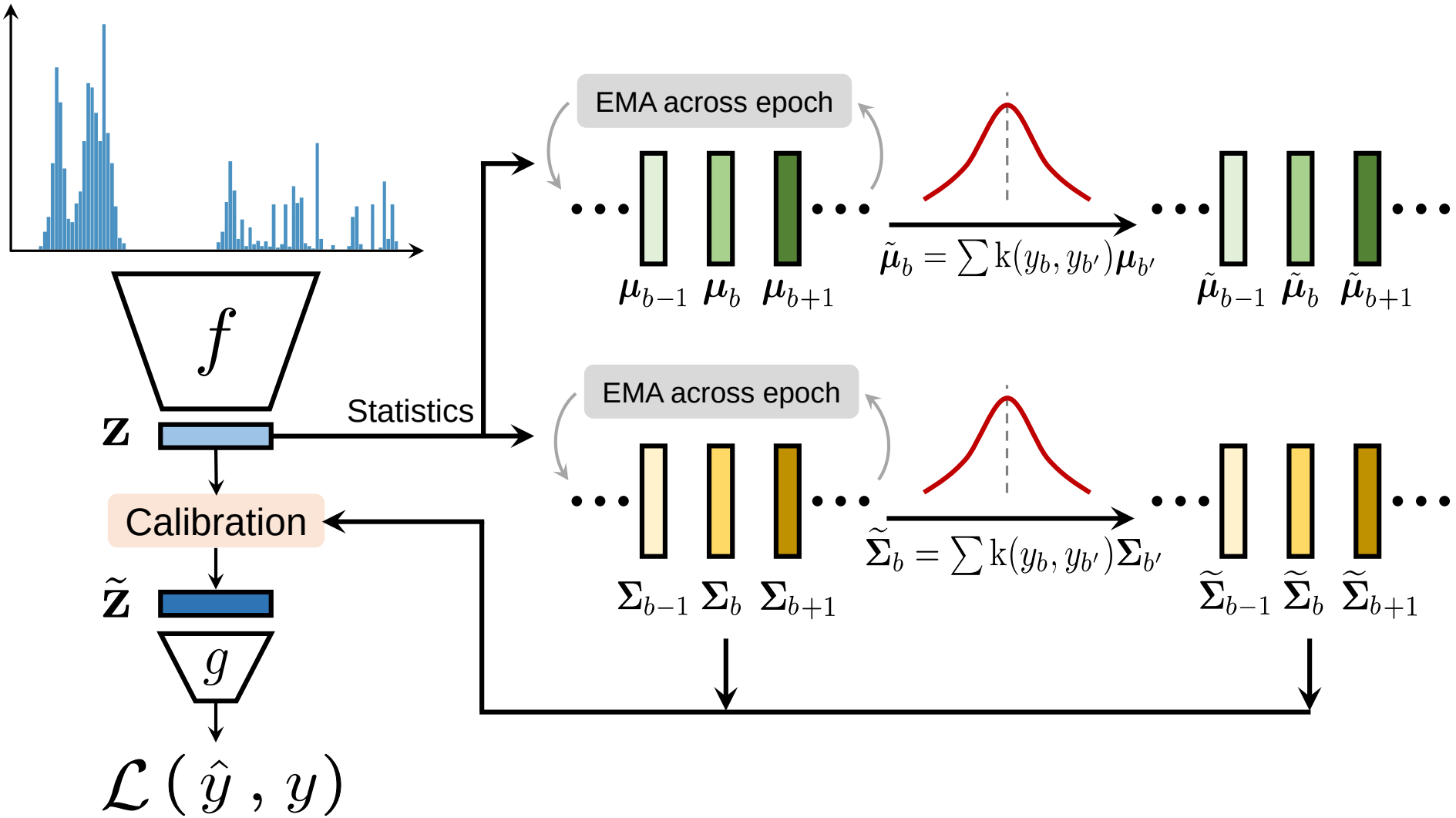


Image credit: *Yang et al. (2021)*

Baselines (1/2)

- Vanilla: neglects data imbalance

- Synthetic samples
  - SMOTER (Torgo et al. 2013)
    1. Defines frequent and rare regions using label density.
    2. Creates synthetic samples for pre-defined rare regions by linearly interpolating both inputs and labels.
  - SMOGN (Branco et al. 2017): augments SMOTER with Gaussian noise

- Focal-R

$$\frac{1}{n}\sum_{i=1}^{n}\sigma(|\beta e_i|)^{\gamma}e_i$$

  - Error-aware loss
  - Maps the absolute error into $[0, 1]$.
  - e_i: L_1 error for the i-th sample
  - β, γ: hyper-parameters
  - Inspired by Focal Loss (Lin 2017) for classification

Baselines (2/2)

- Regressor re-training (RRT)
  - Two-stage training
    1. Train encoder
    2. Re-train regressor with inverse re-weighting and frozen encoder.
  - Inspired by Kang et al. 2019

- Cost-sensitive re-weighting: re-weighting schemes based on label distribution
  - Inverse-frequency weighting (INV)
  - Square-root weighting variant (SQINV)

References

- Branco, Paula, Luís Torgo, and Rita P Ribeiro (2017). "SMOGN: a pre-processing approach for imbalanced regression". In: *First international workshop on learning with imbalanced domains: Theory and applications*. PMLR, pp. 36–50.
- Kang, Bingyi et al. (2019). "Decoupling representation and classifier for long-tailed recognition". In: *arXiv preprint arXiv:1910.09217*.
- Lin, T (2017). "Focal Loss for Dense Object Detection". In: *arXiv preprint arXiv:1708.02002*.
- Torgo, Luís et al. (2013). "Smote for regression". In: *Portuguese conference on artificial intelligence*. Springer, pp. 378–389.
- Yang, Yuzhe et al. (2021). "Delving into deep imbalanced regression". In: *International conference on machine learning*. PMLR, pp. 11842–11851.