

# Deep Imbalanced Regression

Yuzhe Yang<sup>1</sup> Kaiwen Zha<sup>1</sup> Ying-Cong Chen<sup>1</sup> Hao Wang<sup>2</sup>  
Dina Katabi<sup>1</sup>

<sup>1</sup>MIT Computer Science & Artificial Intelligence Laboratory

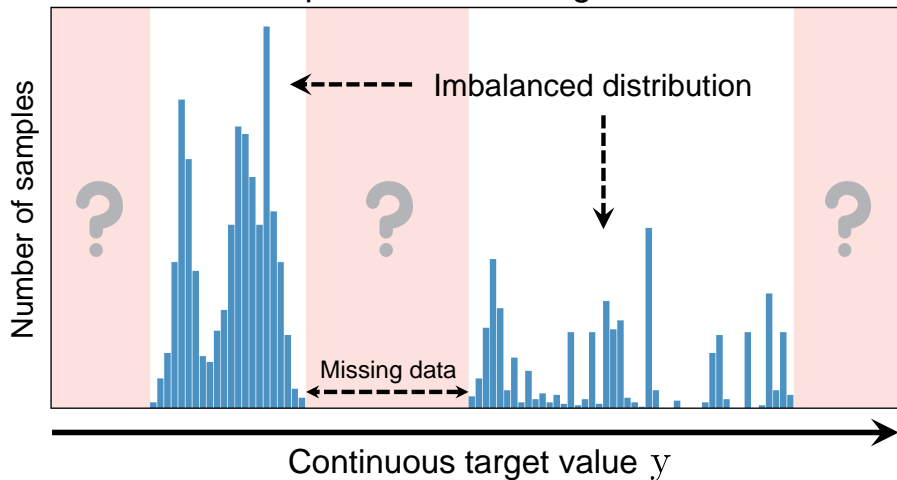
<sup>2</sup>Department of Computer Science, Rutgers University

ICML 2021

Presenter: Gianmarco Midena

26 November 2024

## Deep Imbalanced Regression



# 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
- $\mathcal{B} = \{1, \dots, M\} \subset \mathbb{Z}^+$ : index space
  - ▶ divides  $\mathcal{Y}$  into  $M$  groups (bins) with equal intervals  $[t_j, t_{j+1})$
  - ▶  $\{[t_0, t_1), \dots, [t_{M-1}, t_M)\}$ : 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
- $\theta$ : trainable model parameters

# Evaluation

- Divide target space into disjoint regions (bins)

- ▶ *Many-shot*: > 100 training examples
- ▶ *Medium-shot*: 20-100 training examples
- ▶ *Few-shot*: < 20 training examples
- ▶ *Zero-shot*: 0 training examples
  - Inspired by [Liu et al. 2019](#)

- Metrics

- ▶ Mean Absolute Error (MAE)
- ▶ Mean Squared Error (MSE)
- ▶ Pearson Correlation (PCC)
- ▶ Geometric Mean Error (GM)

$$GM = \sqrt[n]{\prod_{i=1}^n |y_i - \hat{y}_i|}$$

★ Pros: + fairness (uniformity) in prediction

# Datasets - Overview

Dataset	Target type	Target range	Bin size	Max bin density	Min bin density	# Training set	# Val. set	# Test set
IMDB-WIKI-DIR	Age	0 - 186	1	7,149	1	191,509	11,022	11,022
AgeDB-DIR	Age	0 - 101	1	353	1	12,208	2,140	2,140
STS-B-DIR	Text similarity score	0 - 5	0.1	428	1	5,249	1,000	1,000
NYUD2-DIR	Depth	0.7 - 10	0.1	$1.46 \times 10^8$	$1.13 \times 10^6$	50,688 ( $3.51 \times 10^9$ )	—	654 ( $8.70 \times 10^5$ )
SHHS-DIR	Health condition score	0 - 100	1	275	0	1,892	369	369

Table credit: [Yang et al. \(2021\)](#)

# (Training) Datasets - Label Distributions

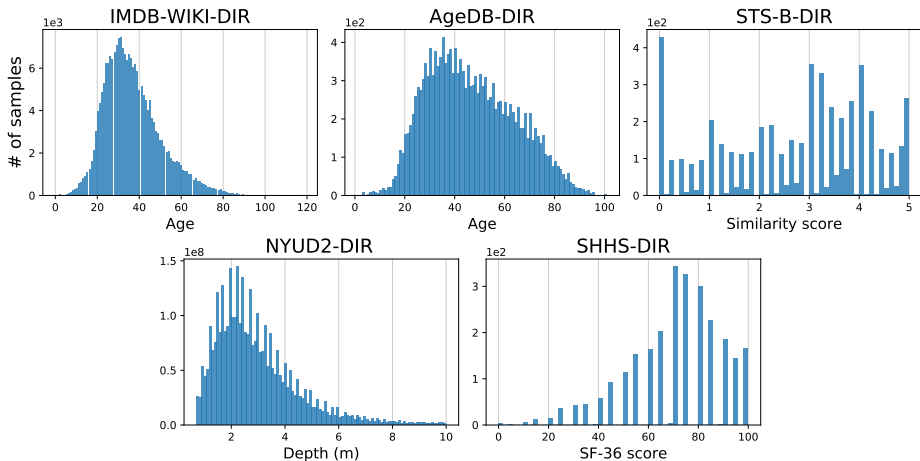
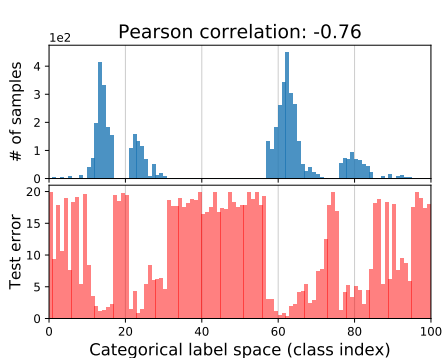


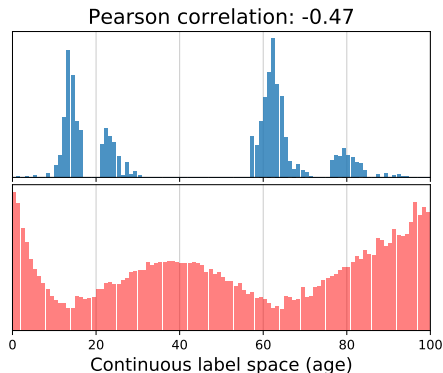
Image credit: [Yang et al. \(2021\)](#)

## Label Distribution Smoothing (LDS)

# Test Error on Categorical vs. Continuous Label Space



(a) CIFAR-100 (subsampled)

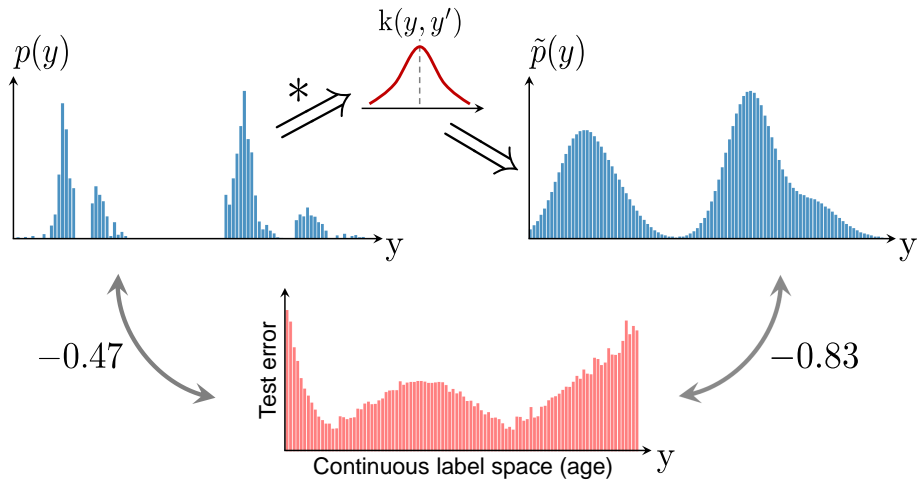


(b) IMDB-WIKI (subsampled)

Image credit: [Yang et al. \(2021\)](#)



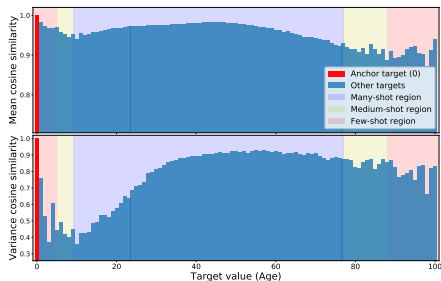
# The LDS Algorithm



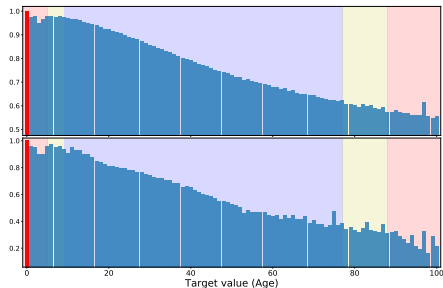
## Feature Distribution Smoothing (FDS)

# Feature statistics similarity (1/4)

Anchor age 0



(a) Baseline



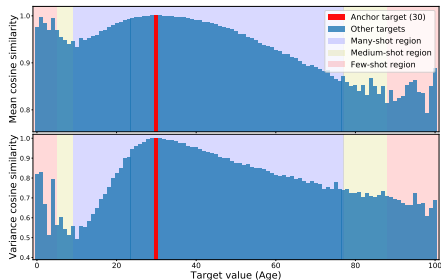
(b) FDS

- Improved feature statistics calibration:
  - ▶ High similarity only in neighbourhood
  - ▶ “The further the region the lower the similarity”
  - ▶ Gradual change of similarity

Image credit: [Yang et al. \(2021\)](#)

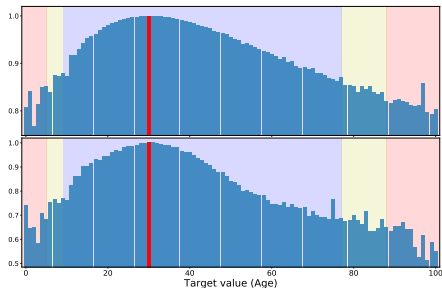
# Feature statistics similarity (2/4)

Anchor age 30



(a) Baseline

- High similarities with further regions
- Lower similarities with some closer regions



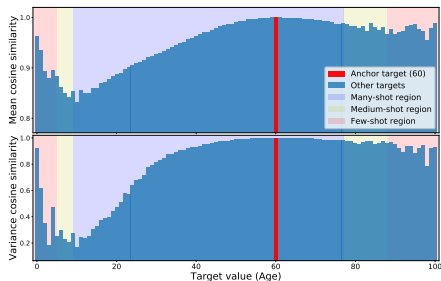
(b) FDS

- Improved feature statistics calibration:
  - ▶ High similarity only in neighbourhood
  - ▶ “The further the region the lower the similarity”
  - ▶ Gradual change of similarity

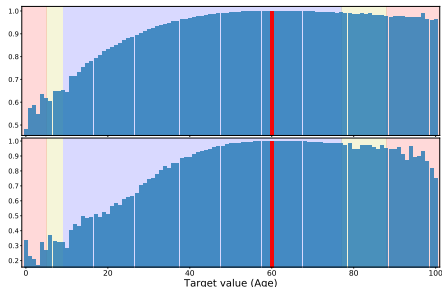
Image credit: Yang et al. (2021)

# Feature statistics similarity (3/4)

Anchor age 60



(a) Baseline



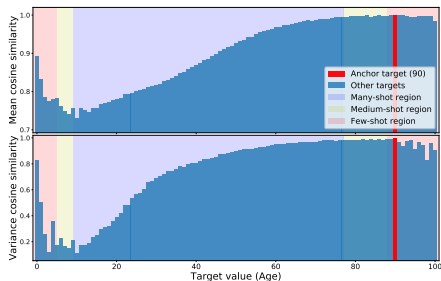
(b) FDS

- Improved feature statistics calibration:
  - ▶ High similarity only in neighbourhood
  - ▶ “The further the region the lower the similarity”
  - ▶ Gradual change of similarity

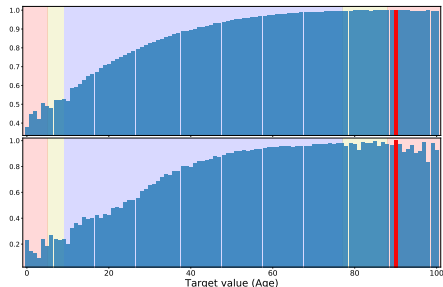
Image credit: [Yang et al. \(2021\)](#)

# Feature statistics similarity (4/4)

Anchor age 90



(a) Baseline

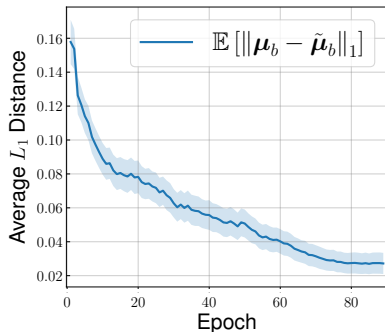


(b) FDS

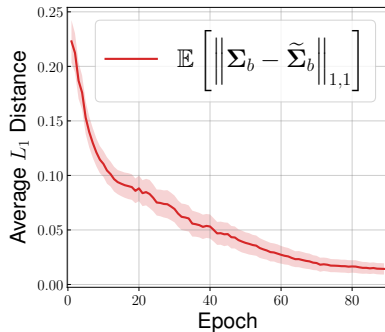
- Improved feature statistics calibration:
  - ▶ High similarity only in neighbourhood
  - ▶ “The further the region the lower the similarity”
  - ▶ Gradual change of similarity

Image credit: Yang et al. (2021)

# Change of feature statistics w.r.t. epoch



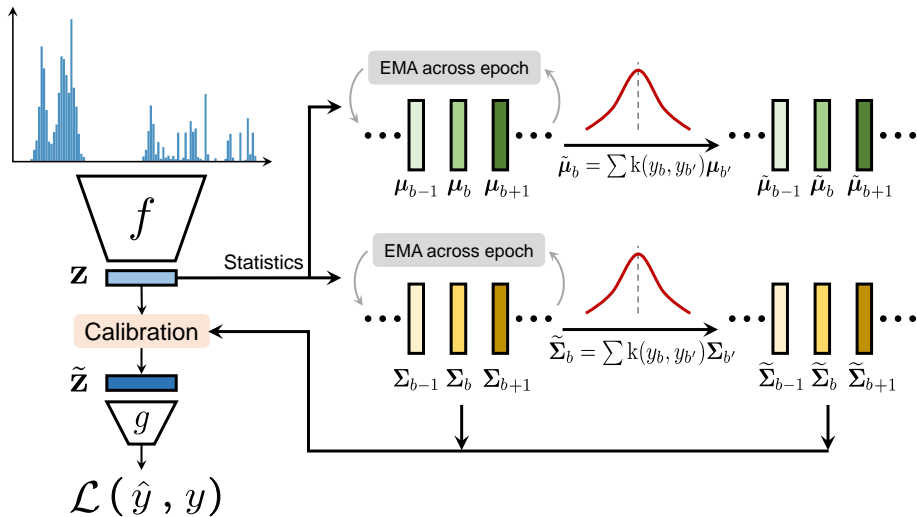
(a) Mean



(b) Variance

- $\mu, \Sigma$ : Running mean and variance
- $\tilde{\mu}, \tilde{\Sigma}$ : Smoothed mean and variance

# The FDS Algorithm





# 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
- ▶  $\beta, \gamma$ : 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)

## Results

# Could LDS + FDS help when the label distribution is skewed with one or more Gaussian peaks?

- Experimental setup

- ▶ Curated skewed label distributions with 1-4 Gaussian peaks on IMDB-WIKI-DIR
- ▶ Compared with the vanilla model

- Findings

- ▶ Robustness to distribution change
- ▶ Brings improvement

# Skewed label distribution with one Gaussian peak

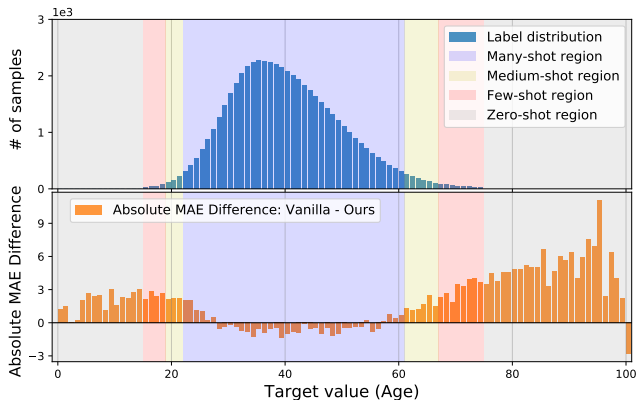


Figure: MAE gains of LDS + FDS over the vanilla model.

- Performance gains, esp. for extrapolation & interpolation

# Skewed label distribution with two Gaussian peaks

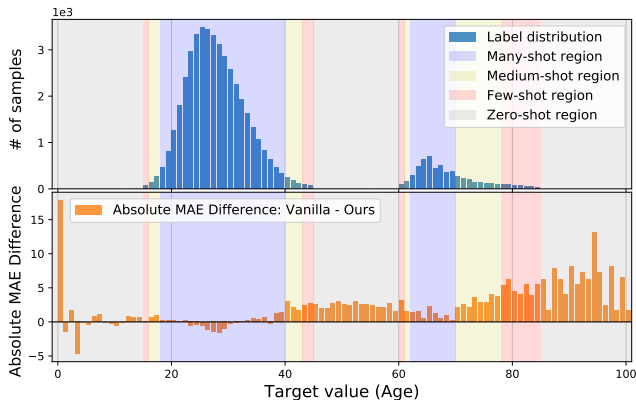


Figure: MAE gains of LDS + FDS over the vanilla model.

- Performance gains, esp. for extrapolation & interpolation

# Skewed label distribution with three Gaussian peaks

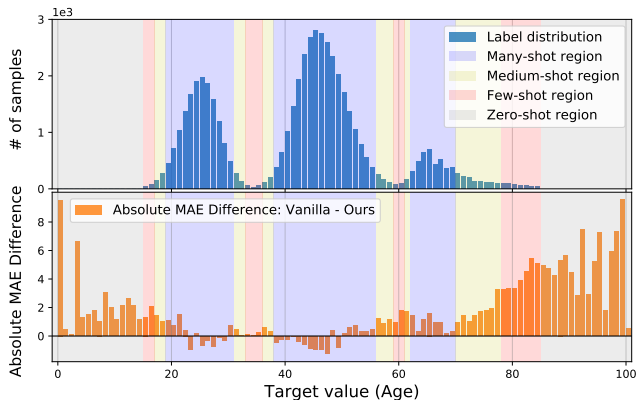


Figure: MAE gains of LDS + FDS over the vanilla model.

- Performance gains, esp. for extrapolation & interpolation

# Skewed label distribution with four Gaussian peaks

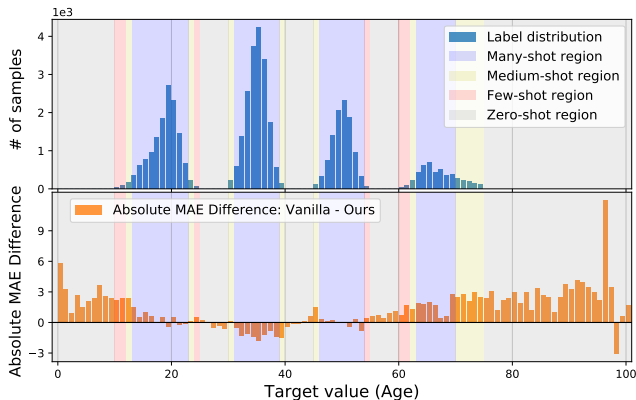


Figure: MAE gains of LDS + FDS over the vanilla model.

- Performance gains, esp. for extrapolation & interpolation



# Skewed label distribution with two Gaussian peaks on IMDB-WIKI-DIR

Metrics	MAE ↓				GM ↓			
Shot	All	w/ data	Interp.	Extrap.	All	w/ data	Interp.	Extrap.
VANILLA	11.72	9.32	16.13	18.19	7.44	5.33	14.41	16.74
VANILLA + <b>LDS</b>	10.54	8.31	14.14	17.38	6.50	4.67	12.13	15.36
VANILLA + <b>FDS</b>	11.40	8.97	15.83	18.01	7.18	5.12	14.02	16.48
VANILLA + <b>LDS</b> + <b>FDS</b>	<b>10.27</b>	<b>8.11</b>	<b>13.71</b>	<b>17.02</b>	<b>6.33</b>	<b>4.55</b>	<b>11.71</b>	<b>15.13</b>
<b>Ours (best)</b> vs. VANILLA	<b>+1.45</b>	<b>+1.21</b>	<b>+2.42</b>	<b>+1.17</b>	<b>+1.11</b>	<b>+0.78</b>	<b>+2.70</b>	<b>+1.61</b>

Table: Interpolation & extrapolation results

- Best results by smoothing both label & feature distributions

# Different skewed label distributions on IMDB-WIKI-DIR

Metrics	MAE ↓							GM ↓						
Shot	All	Many	Med.	Few	Zero	Interp.	Extrap.	All	Many	Med.	Few	Zero	Interp.	Extrap.
<b>1 peak:</b>														
VANILLA	11.20	6.05	11.43	14.76	22.67	—	22.67	7.02	<b>3.84</b>	8.67	12.26	21.07	—	21.07
VANILLA + LDS	10.09	6.26	9.91	12.12	19.37	—	19.37	6.14	3.92	6.50	8.30	16.35	—	16.35
VANILLA + FDS	11.04	<b>5.97</b>	11.19	14.54	22.35	—	22.35	6.96	<b>3.84</b>	8.54	12.08	20.71	—	20.71
VANILLA + LDS + FDS	<b>10.00</b>	6.28	<b>9.66</b>	<b>11.83</b>	<b>19.21</b>	—	<b>19.21</b>	<b>6.09</b>	3.96	<b>6.26</b>	<b>8.14</b>	<b>15.89</b>	—	<b>15.89</b>
<b>2 peaks:</b>														
VANILLA	11.72	6.83	11.78	15.35	16.86	16.13	18.19	7.44	3.61	8.06	12.94	15.21	14.41	16.74
VANILLA + LDS	10.54	6.72	9.65	12.60	15.30	14.14	17.38	6.50	3.65	<b>5.65</b>	9.30	13.20	12.13	15.36
VANILLA + FDS	11.40	6.69	11.02	14.85	16.61	15.83	18.01	7.18	<b>3.50</b>	7.49	12.73	14.86	14.02	16.48
VANILLA + LDS + FDS	<b>10.27</b>	<b>6.61</b>	<b>9.46</b>	<b>11.96</b>	<b>14.89</b>	<b>13.71</b>	<b>17.02</b>	<b>6.33</b>	3.54	5.68	<b>8.80</b>	<b>12.83</b>	<b>11.71</b>	<b>15.13</b>
<b>3 peaks:</b>														
VANILLA	9.83	7.01	9.81	11.93	20.11	—	20.11	6.04	3.93	6.94	9.84	17.77	—	17.77
VANILLA + LDS	9.08	<b>6.77</b>	8.82	10.48	18.43	—	18.43	<b>5.35</b>	<b>3.78</b>	5.63	7.49	15.46	—	15.46
VANILLA + FDS	9.65	6.88	9.58	11.75	19.80	—	19.80	5.86	3.83	6.68	9.48	17.43	—	17.43
VANILLA + LDS + FDS	<b>8.96</b>	6.88	<b>8.62</b>	<b>10.08</b>	<b>17.76</b>	—	<b>17.76</b>	5.38	3.90	<b>5.61</b>	<b>7.36</b>	<b>14.65</b>	—	<b>14.65</b>
<b>4 peaks:</b>														
VANILLA	9.49	7.23	9.73	10.85	12.16	8.23	18.78	5.68	3.45	6.95	8.20	9.43	6.89	16.02
VANILLA + LDS	8.80	<b>6.98</b>	8.26	10.07	11.26	8.31	<b>16.22</b>	5.10	<b>3.33</b>	<b>5.07</b>	7.08	8.47	6.66	<b>12.74</b>
VANILLA + FDS	9.28	7.11	9.16	10.88	11.95	8.30	18.11	5.49	3.36	6.35	8.15	9.21	6.82	15.30
VANILLA + LDS + FDS	<b>8.76</b>	7.07	<b>8.23</b>	<b>9.54</b>	<b>11.13</b>	<b>8.05</b>	16.32	<b>5.05</b>	3.36	<b>5.07</b>	<b>6.56</b>	<b>8.30</b>	<b>6.34</b>	13.10

Table credit: Yang et al. (2021)

# References

- Branco, Paula, Luís Torgo, and Rita P Ribeiro (2017). "SMOGLN: 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*.
- Liu, Ziwei et al. (2019). "Large-scale long-tailed recognition in an open world". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2537–2546.
- 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.