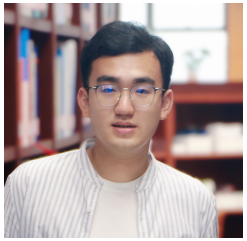
Delving into Deep Imbalanced Regression


Yuzhe Yang


Kaiwen Zha


Ying-Cong Chen


Hao Wang


Dina Katabi

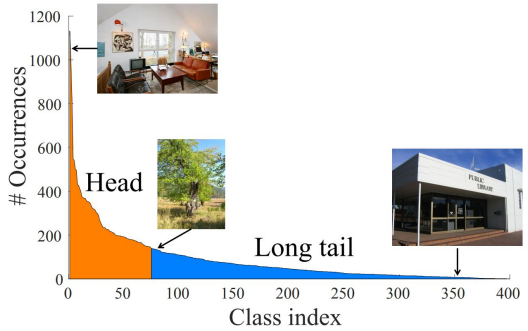
ICML 2021

Long Oral Presentation

http://dir.csail.mit.edu

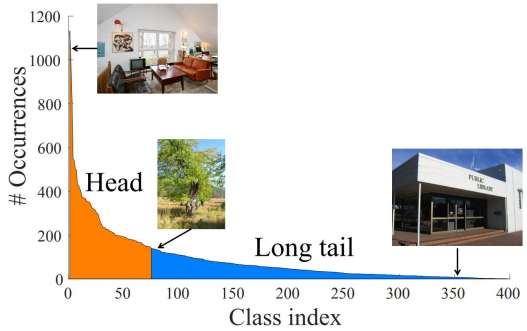
What is imbalanced data?

What is imbalanced data?

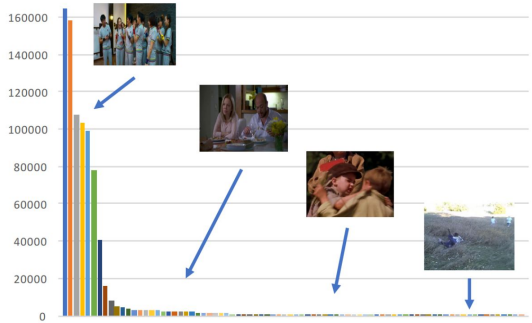


Places

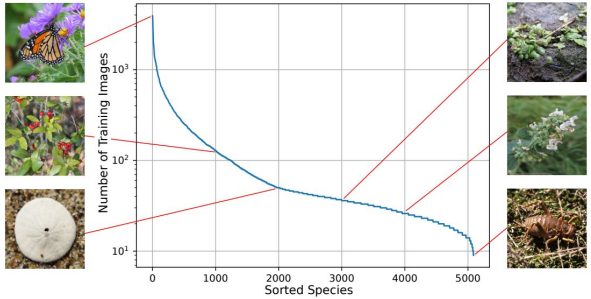
What is imbalanced data?



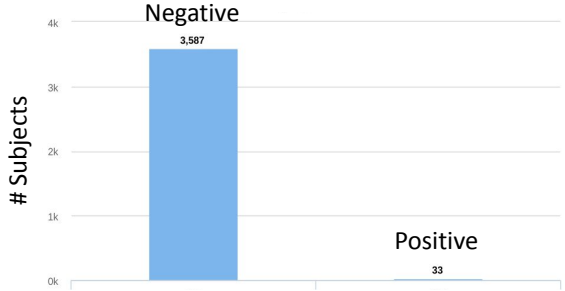
Places



Actions



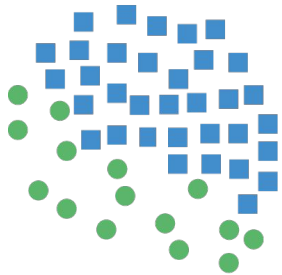
Species



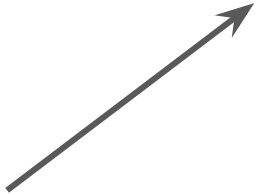
Medical diagnosis

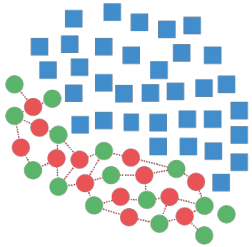
Solutions for learning from imbalanced data

Solutions for learning from imbalanced data



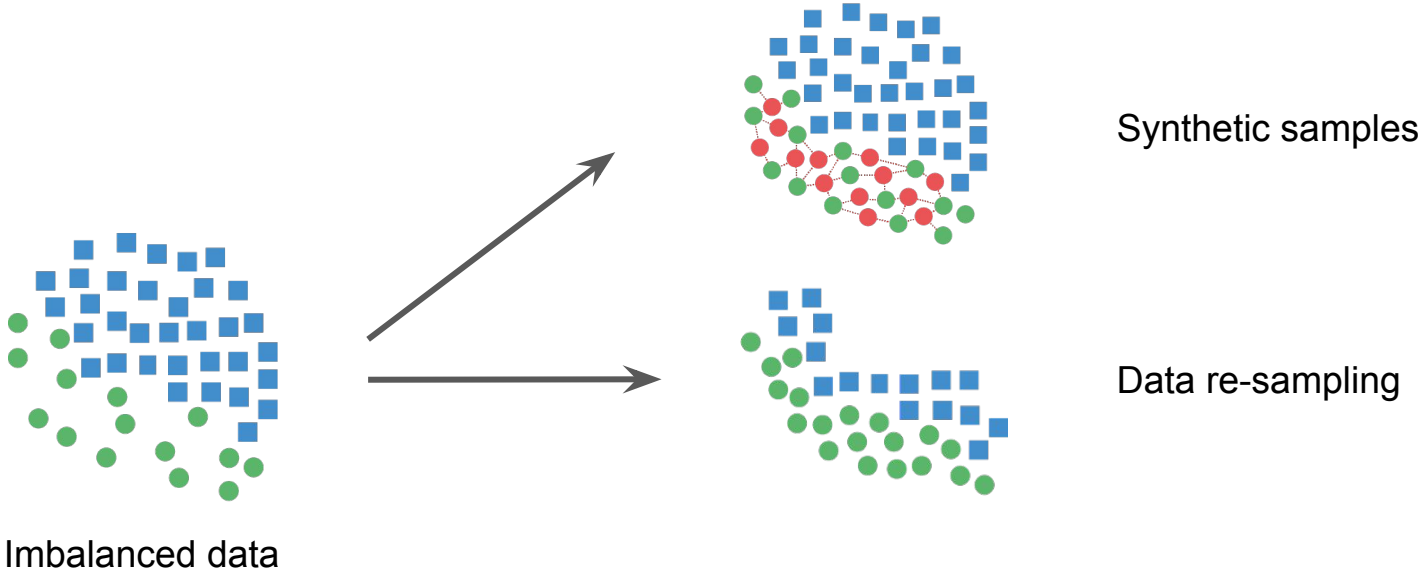
Imbalanced data



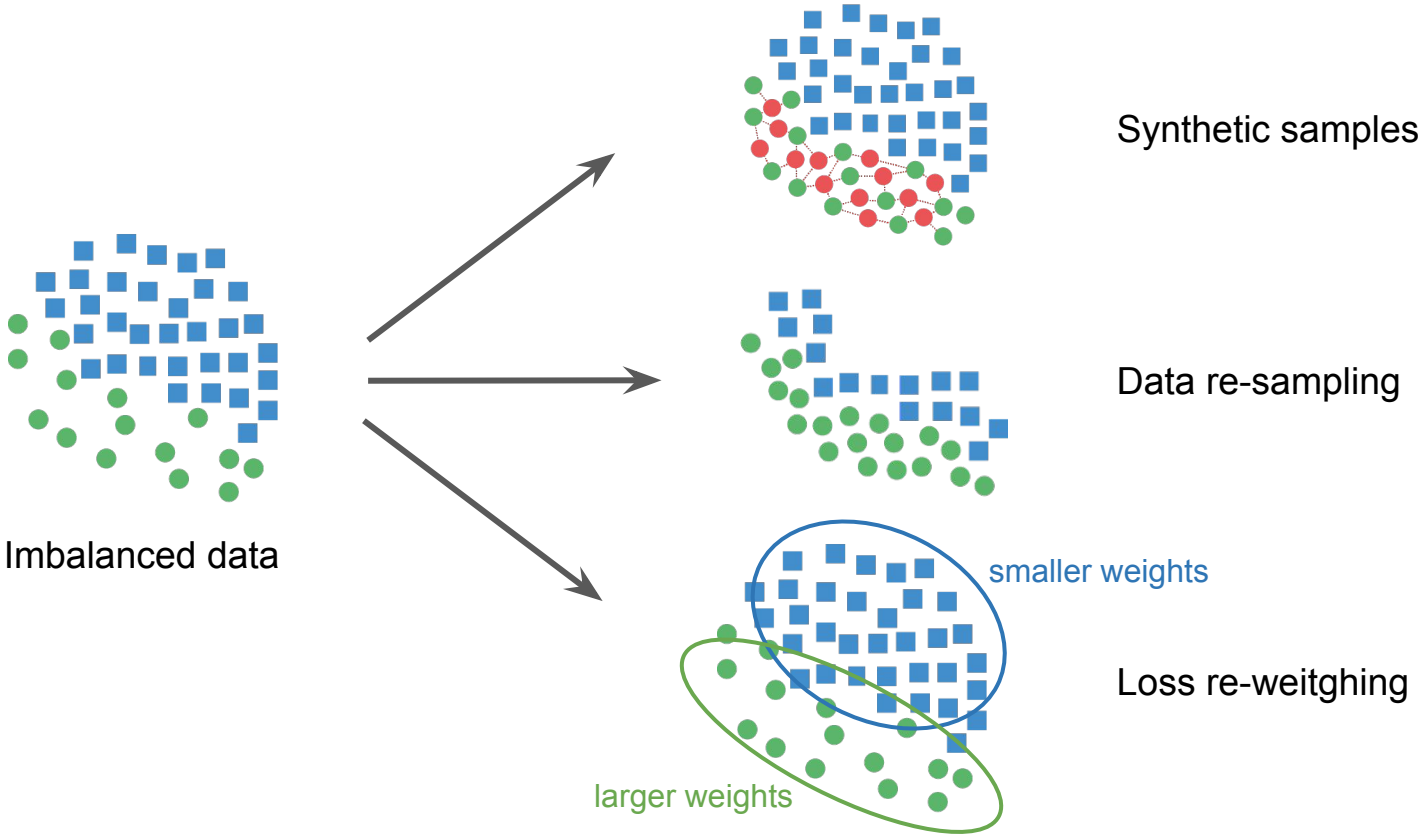


Synthetic samples

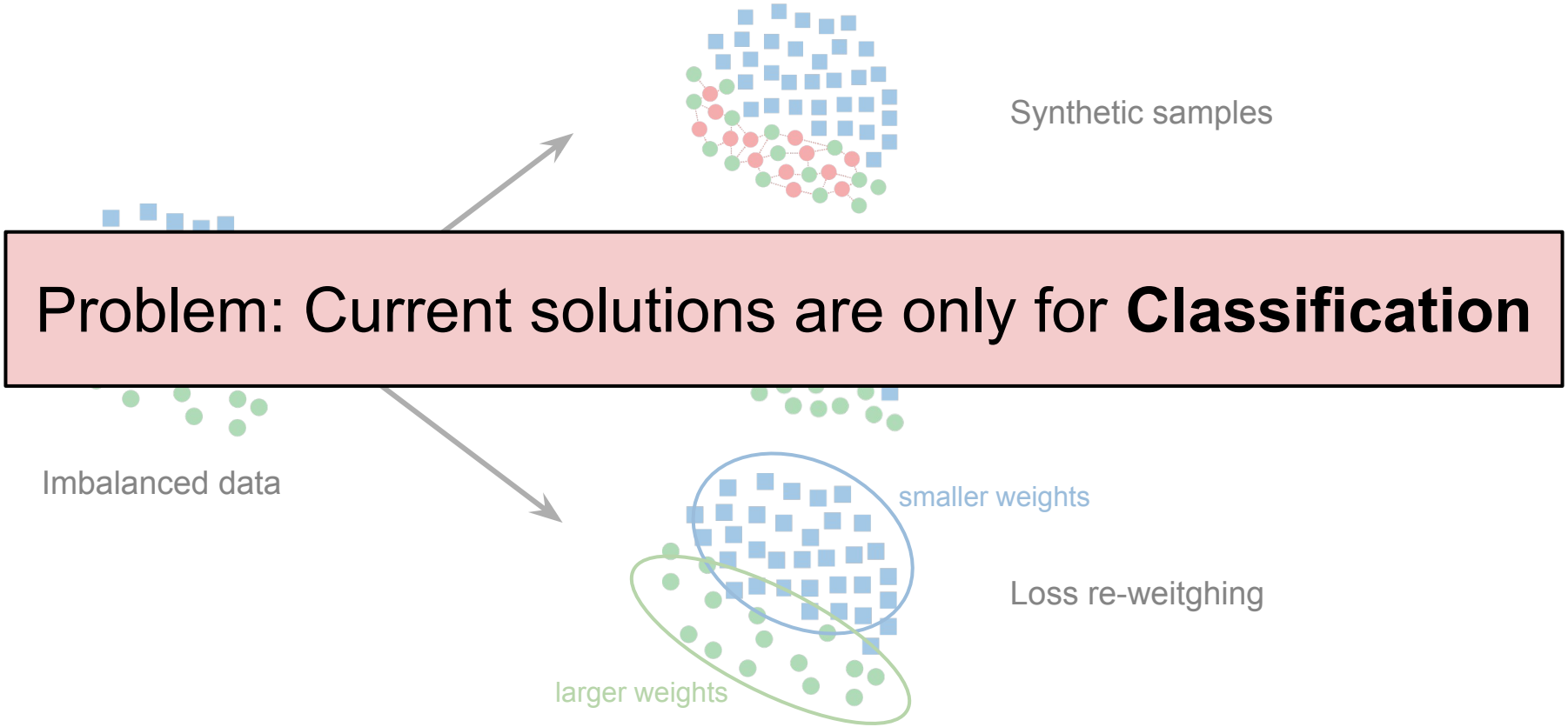
Solutions for learning from imbalanced data



Solutions for learning from imbalanced data



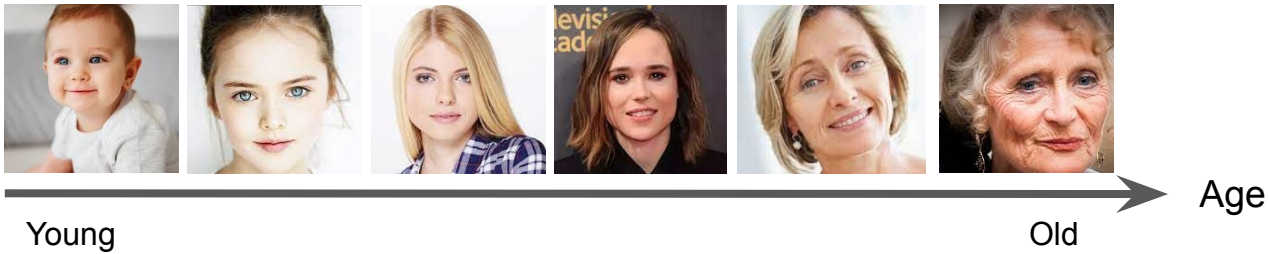
Solutions for learning from imbalanced data



Many tasks have Continuous targets: Regression

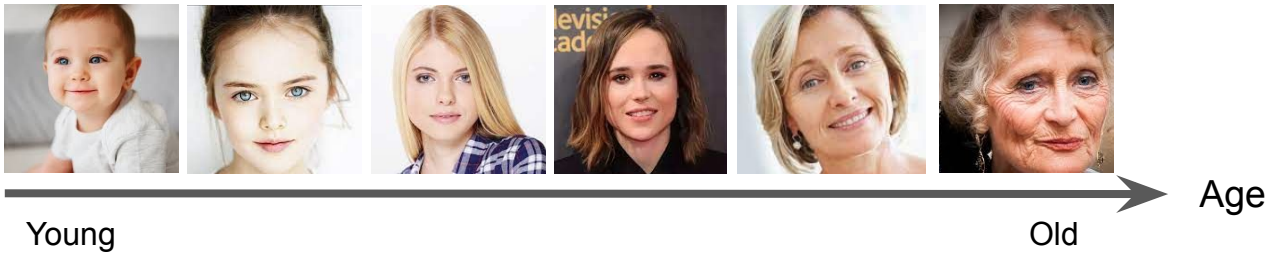
Many tasks have Continuous targets: Regression

- Vision application: Inferring age from visual appearance

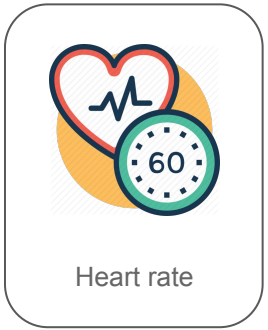


Many tasks have Continuous targets: Regression

- Vision application: Inferring age from visual appearance



- Medical application: Physiological signals that are continuous



Many tasks have Continuous targets: Regression

- Vision application: Inferring age from visual appearance

Why is regression different for imbalanced data?

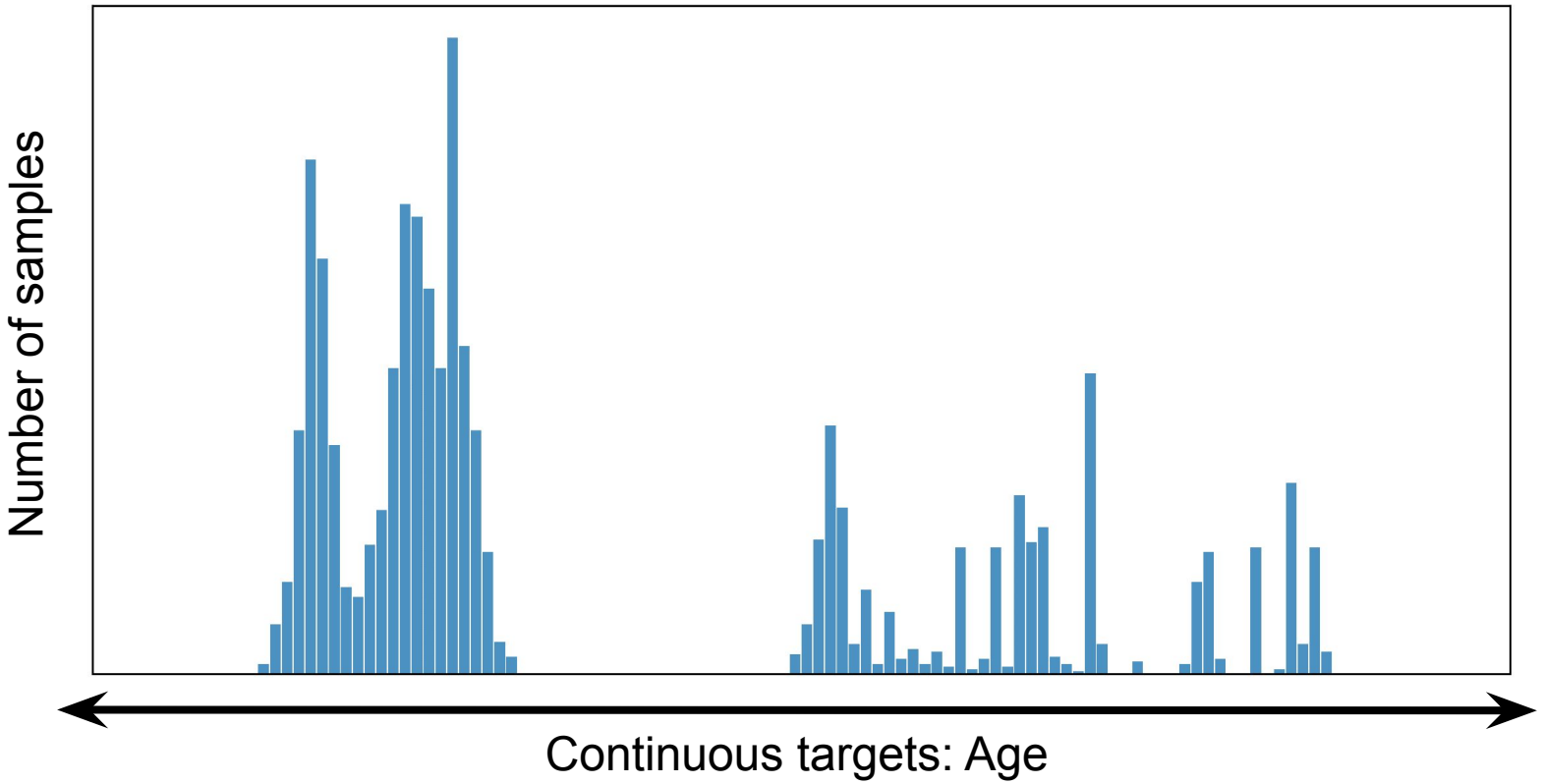
Heart rate

Blood pressure

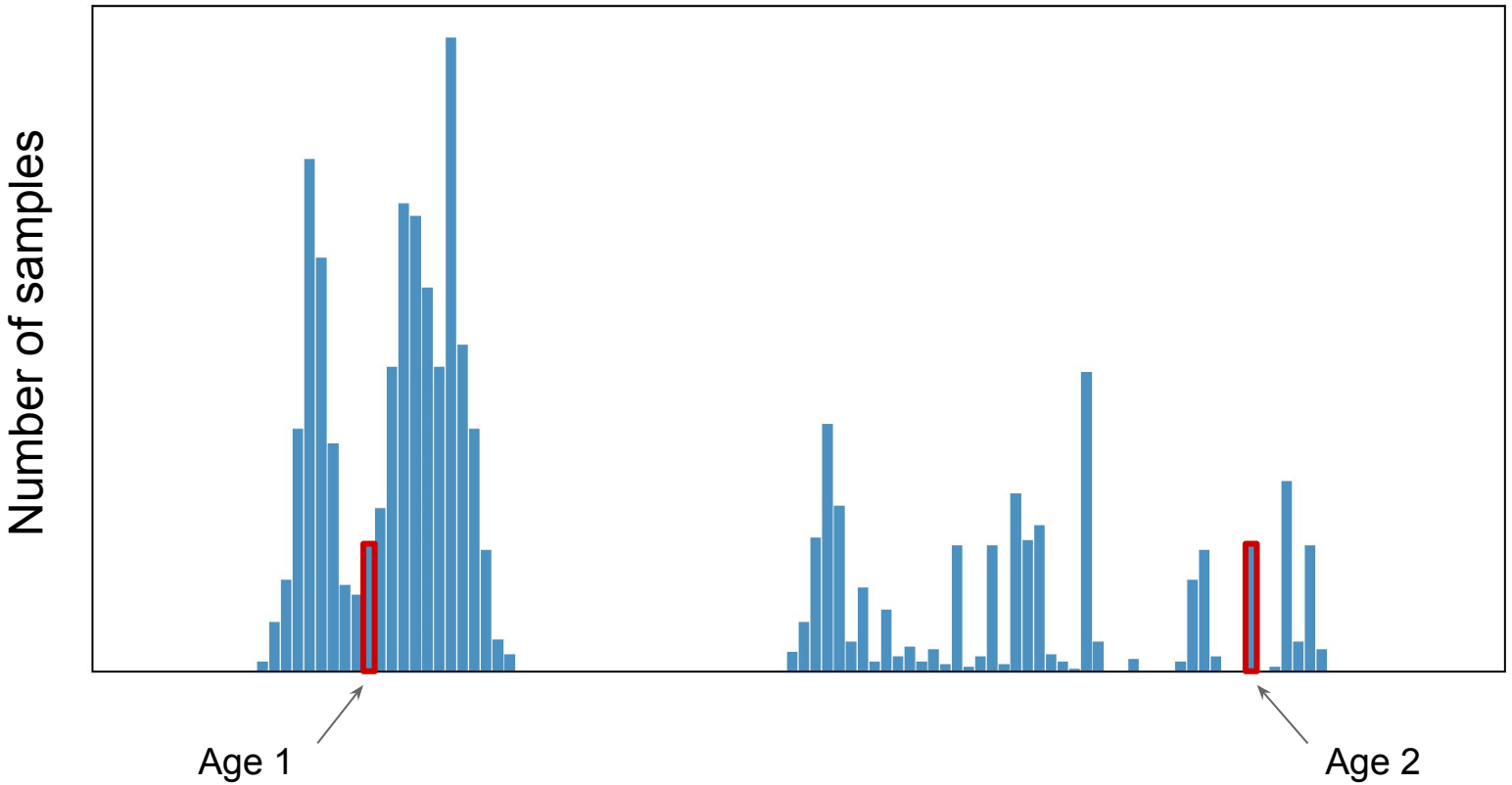
Oxygen saturation

Difference #1

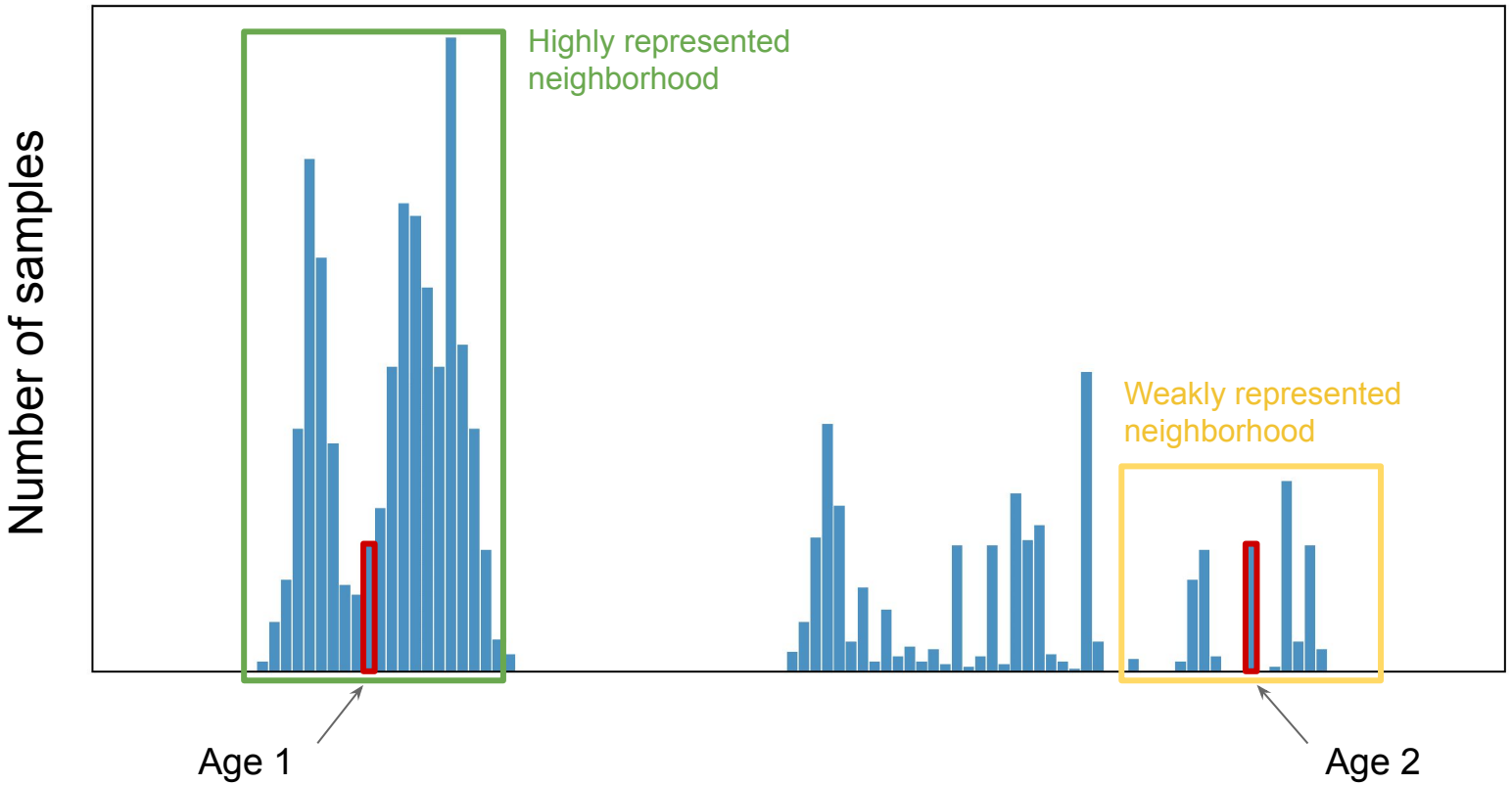
Difference #1



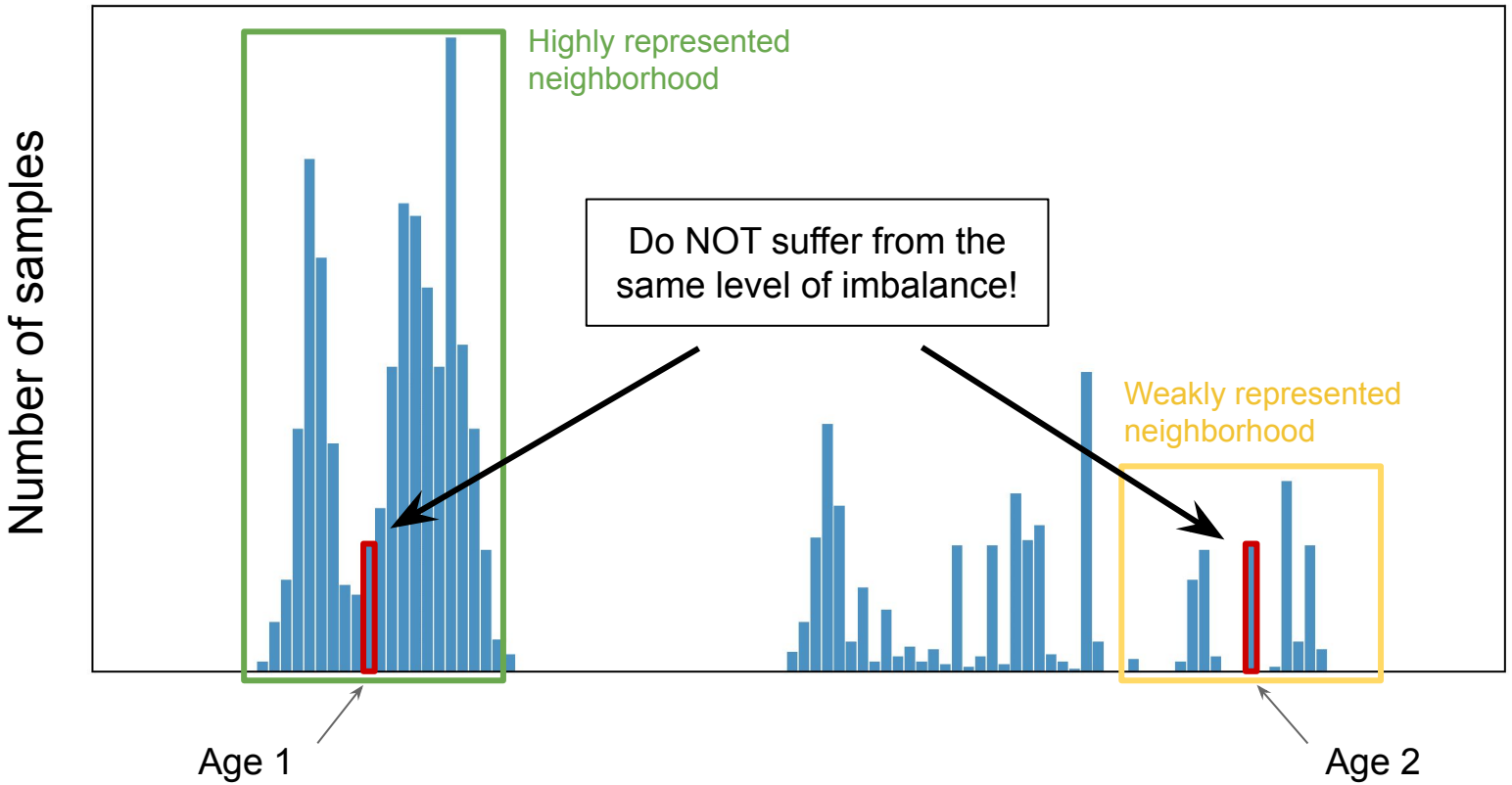
Difference #1



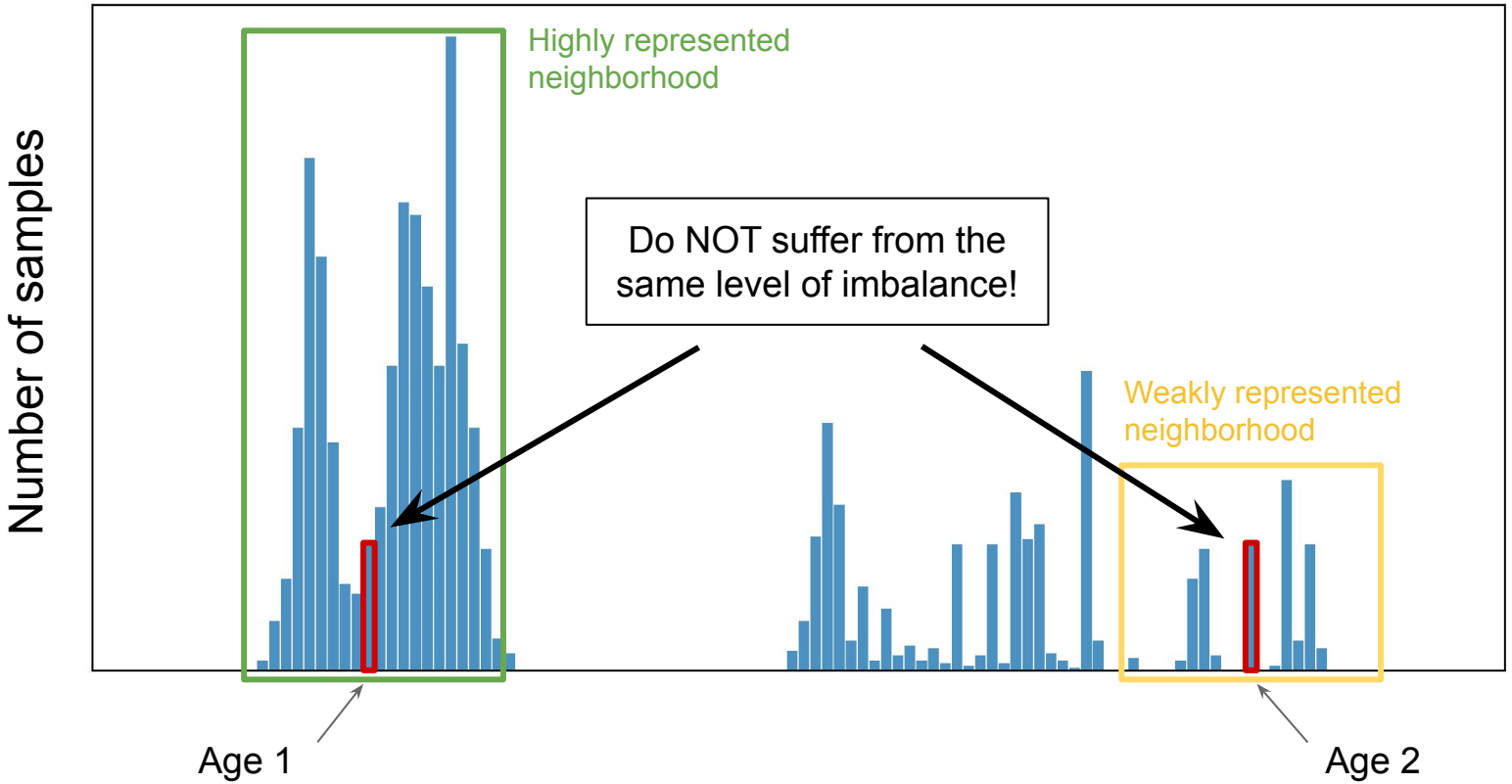
Difference #1



Difference #1

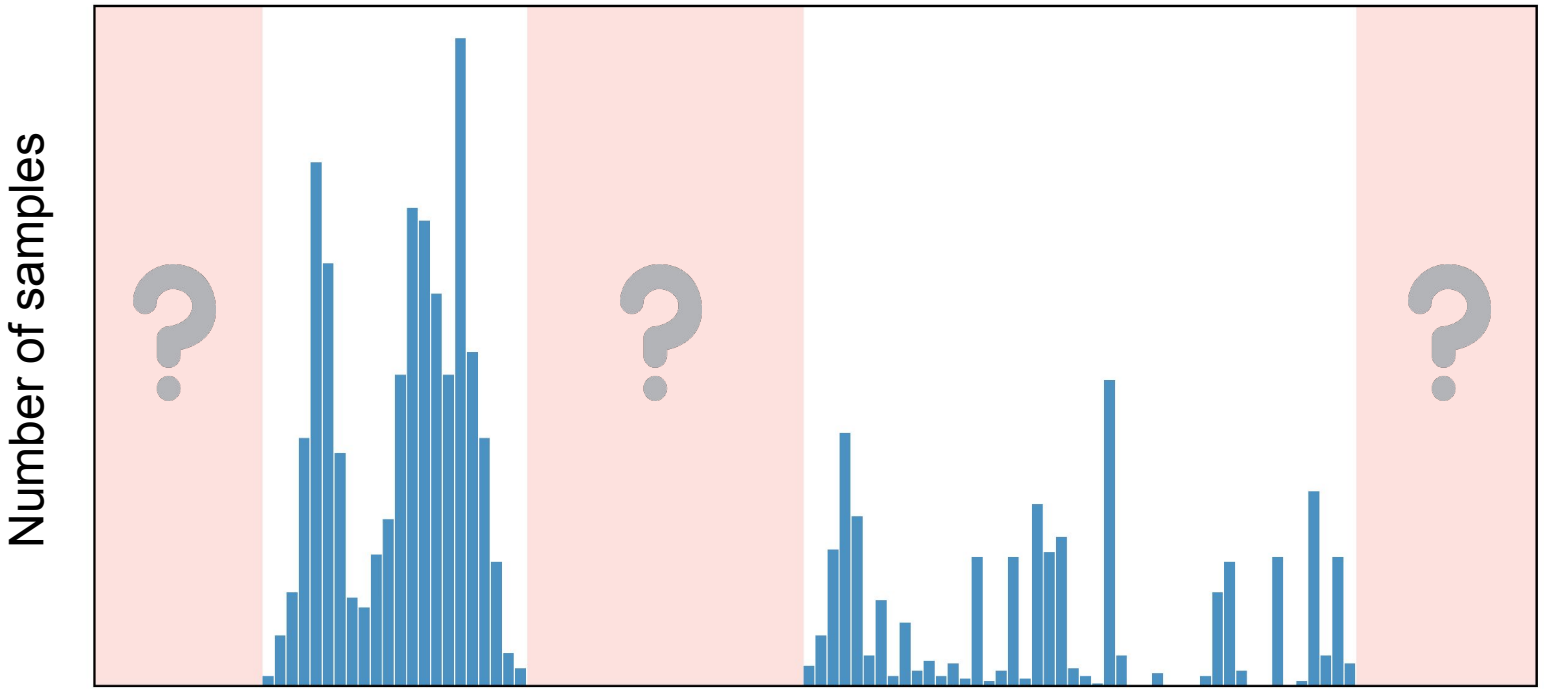


Difference #1: Equal number of examples does not mean equal balanceness

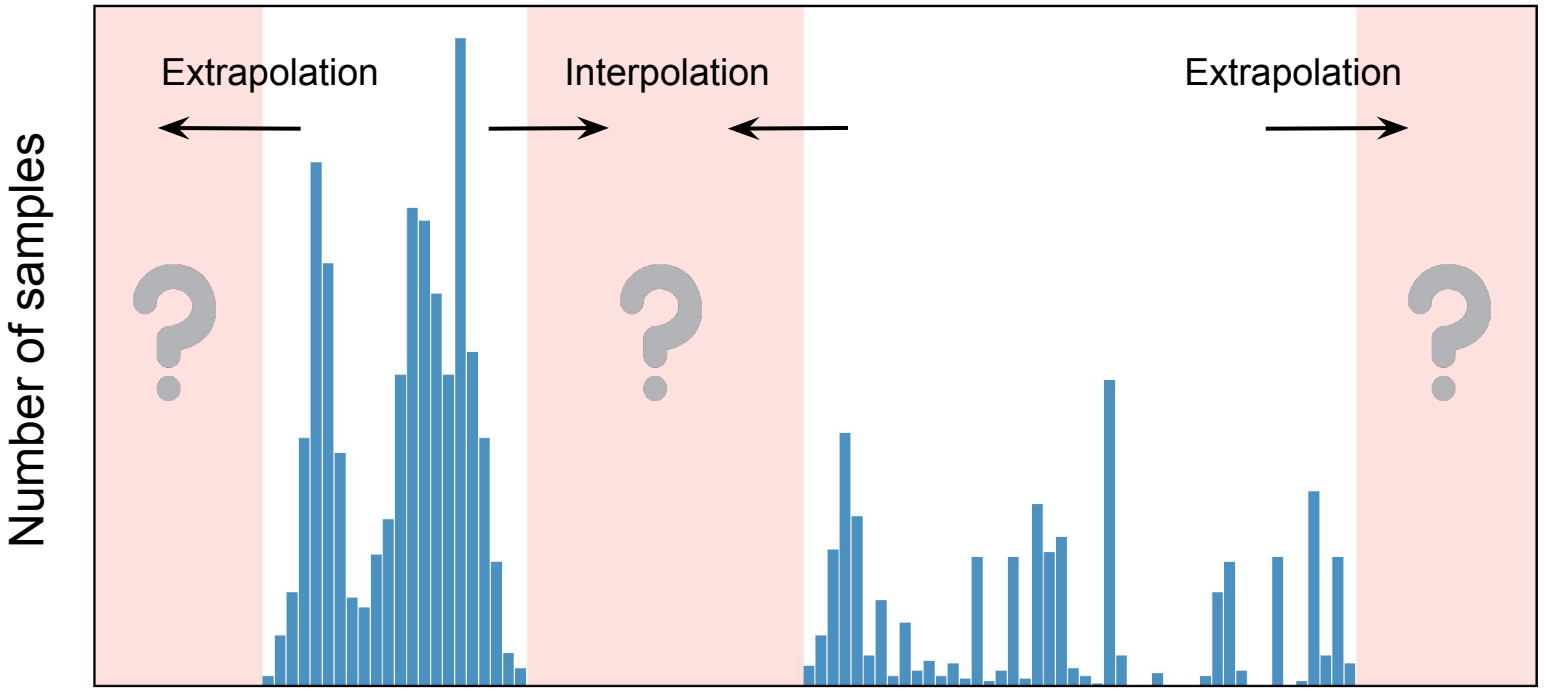


Difference #2

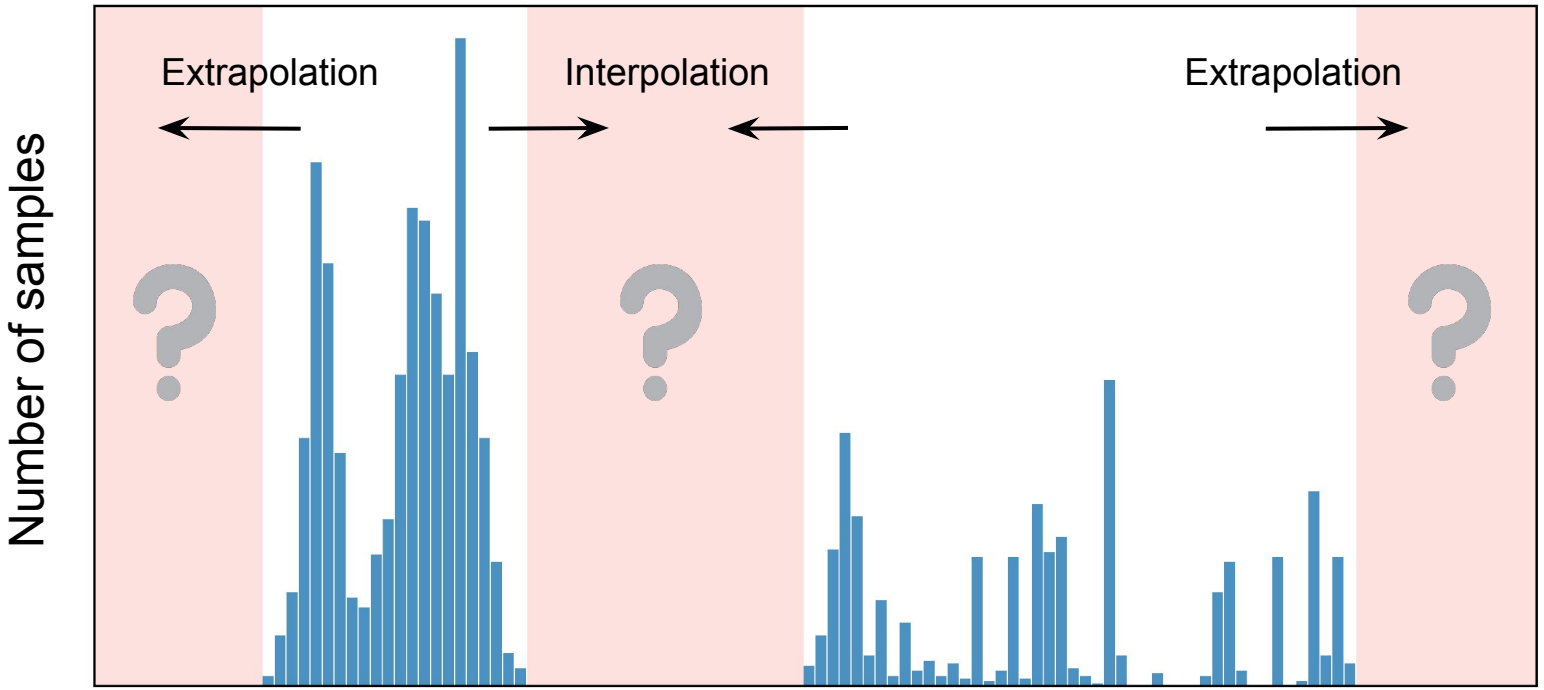
Difference #2



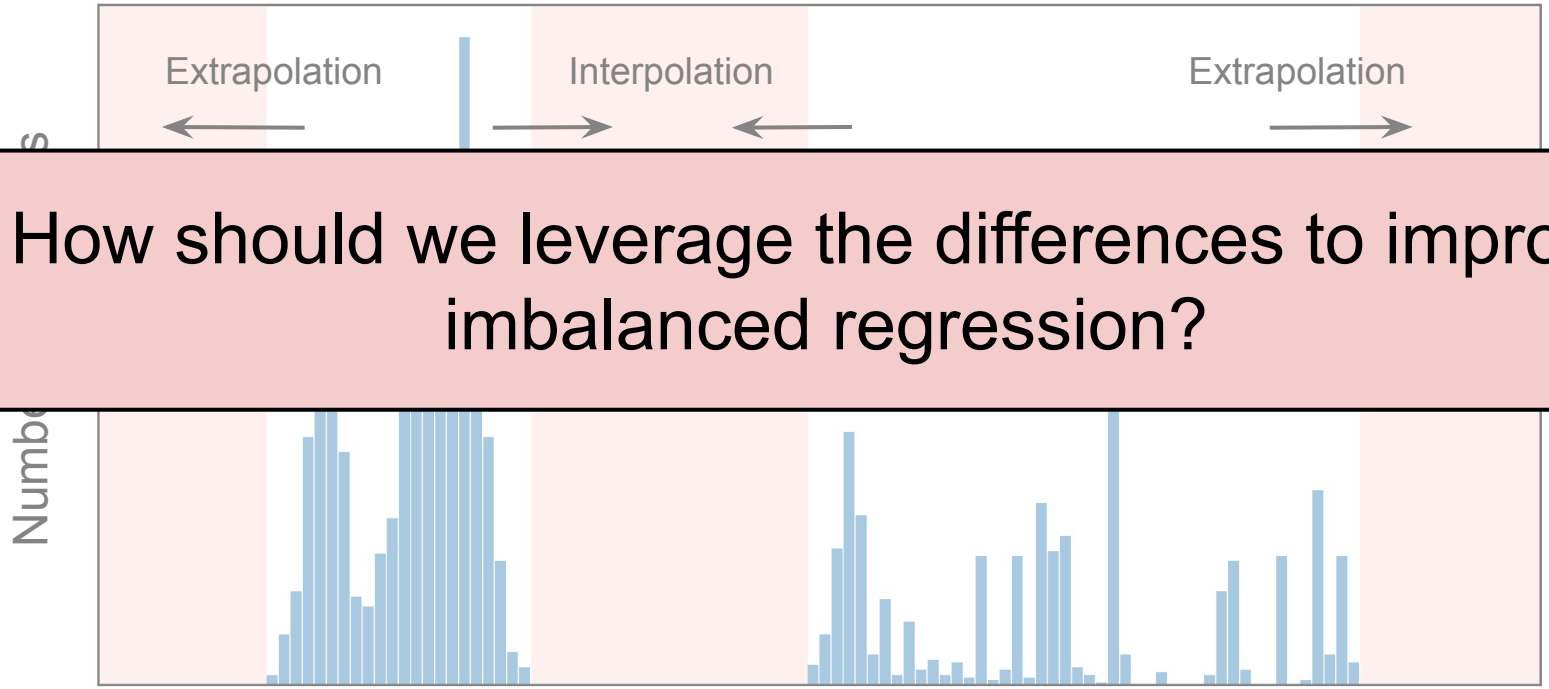
Difference #2



Difference #2: Continuity implies interpolation and extrapolation



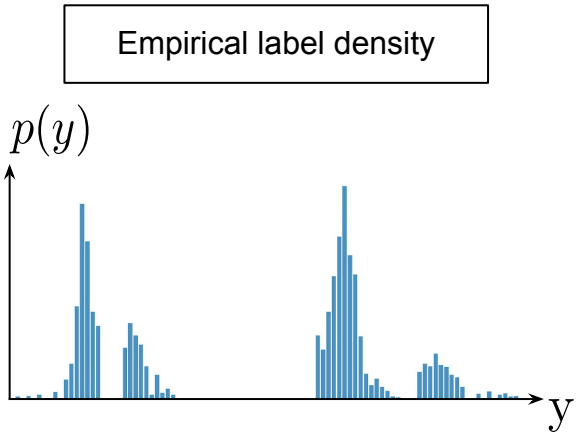
Difference #2: Continuity implies interpolation and extrapolation



How should we leverage the differences to improve imbalanced regression?

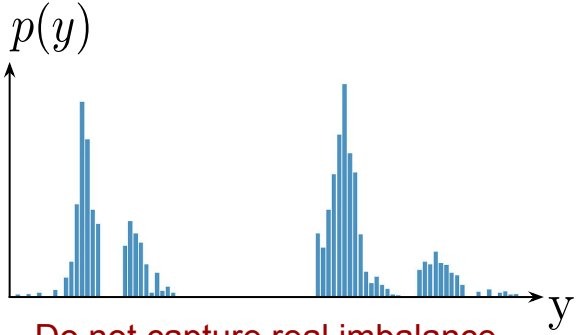
Solution #1: Label Distribution Smoothing (LDS)

Solution #1: Label Distribution Smoothing (LDS)



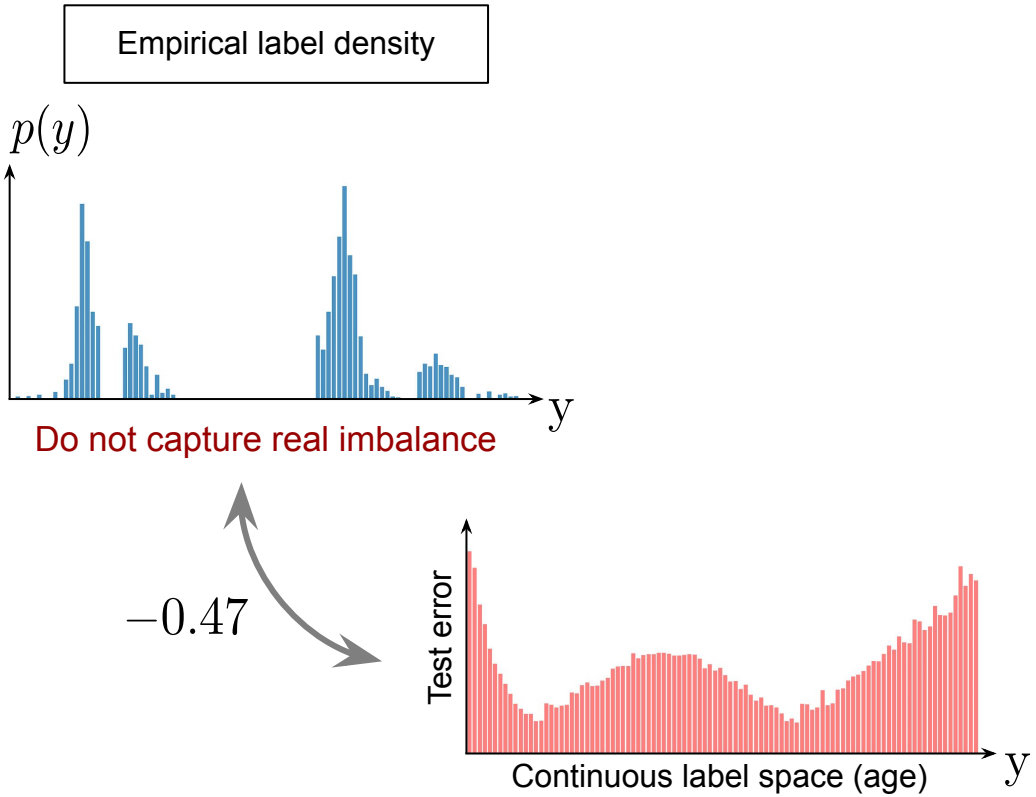
Solution #1: Label Distribution Smoothing (LDS)

Empirical label density

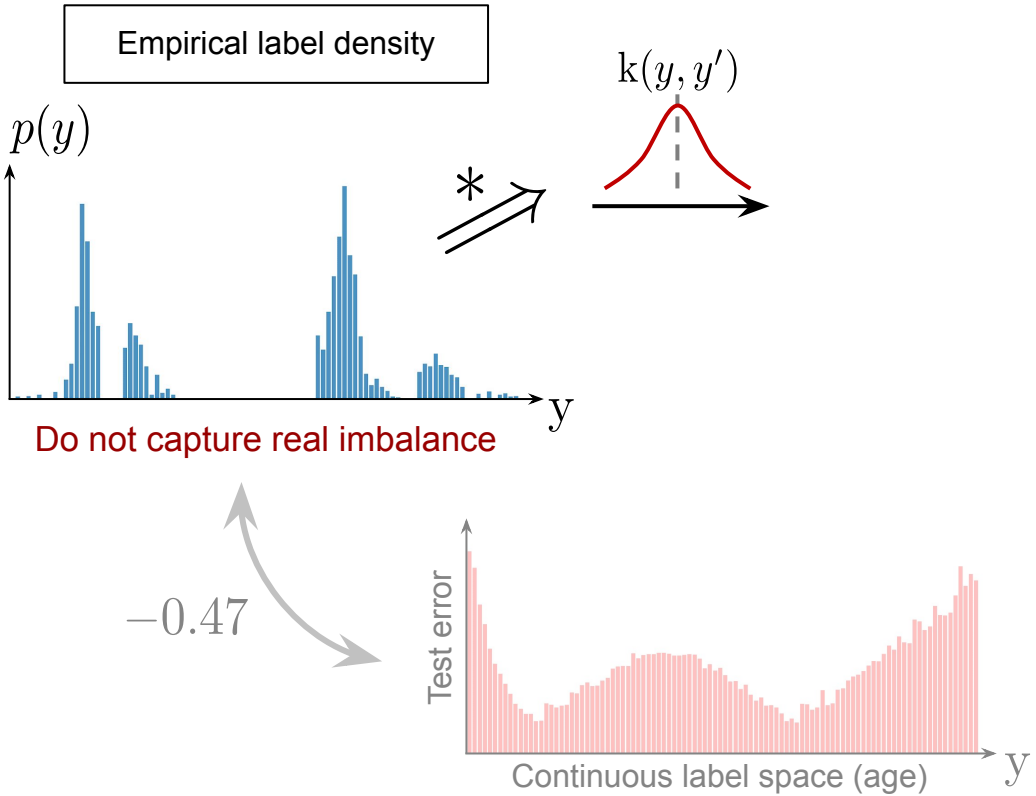


Do not capture real imbalance

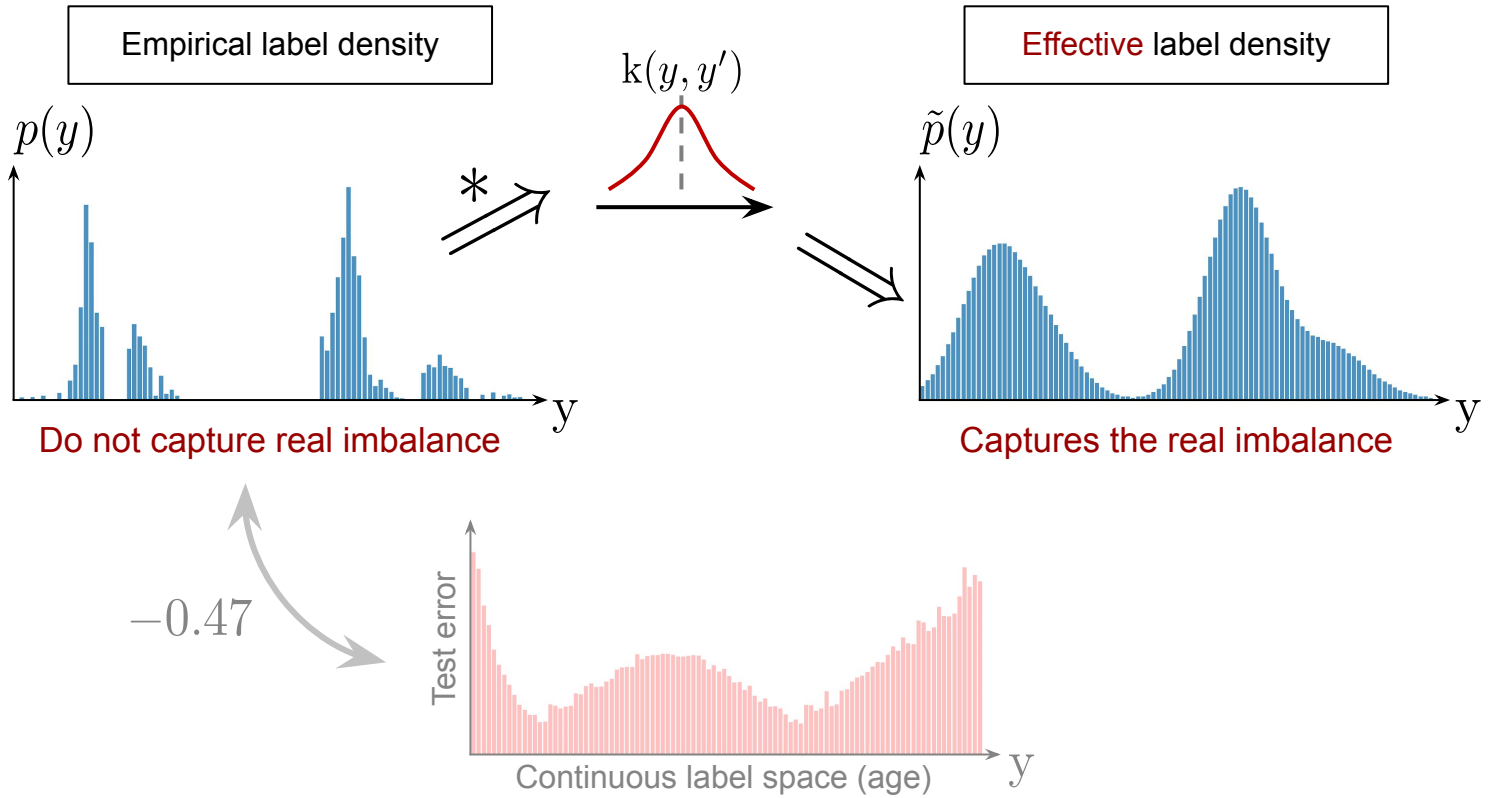
Solution #1: Label Distribution Smoothing (LDS)



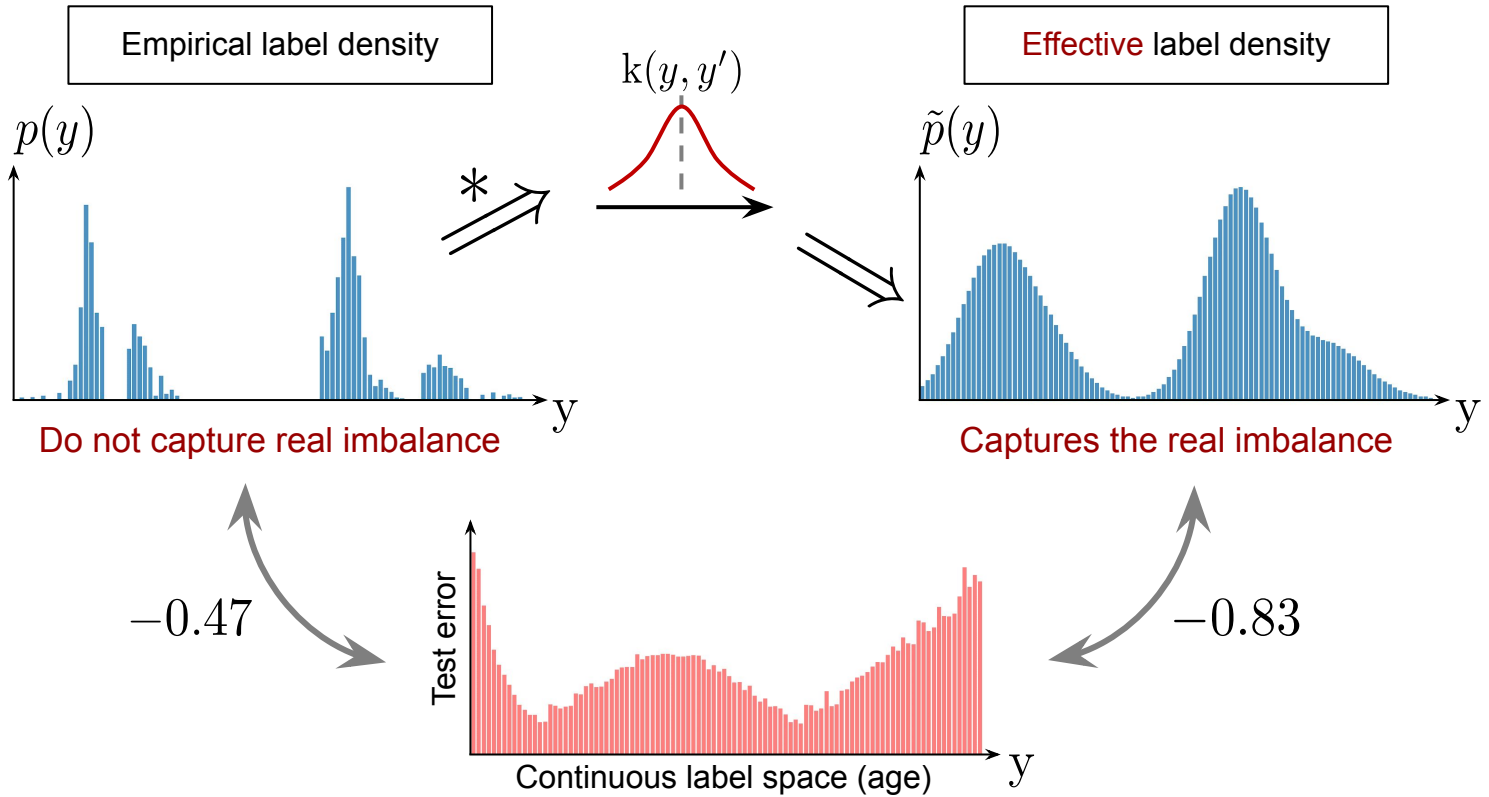
Solution #1: Label Distribution Smoothing (LDS)



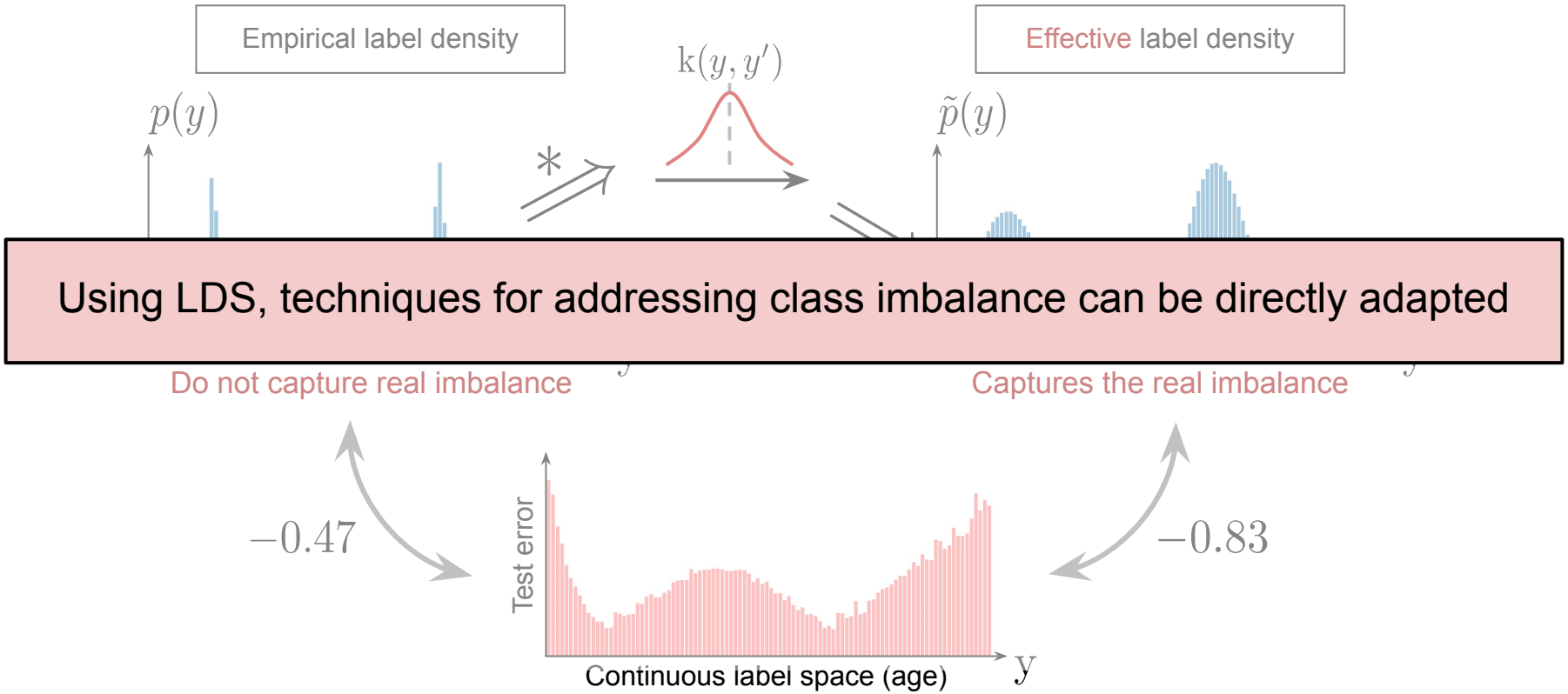
Solution #1: Label Distribution Smoothing (LDS)



Solution #1: Label Distribution Smoothing (LDS)

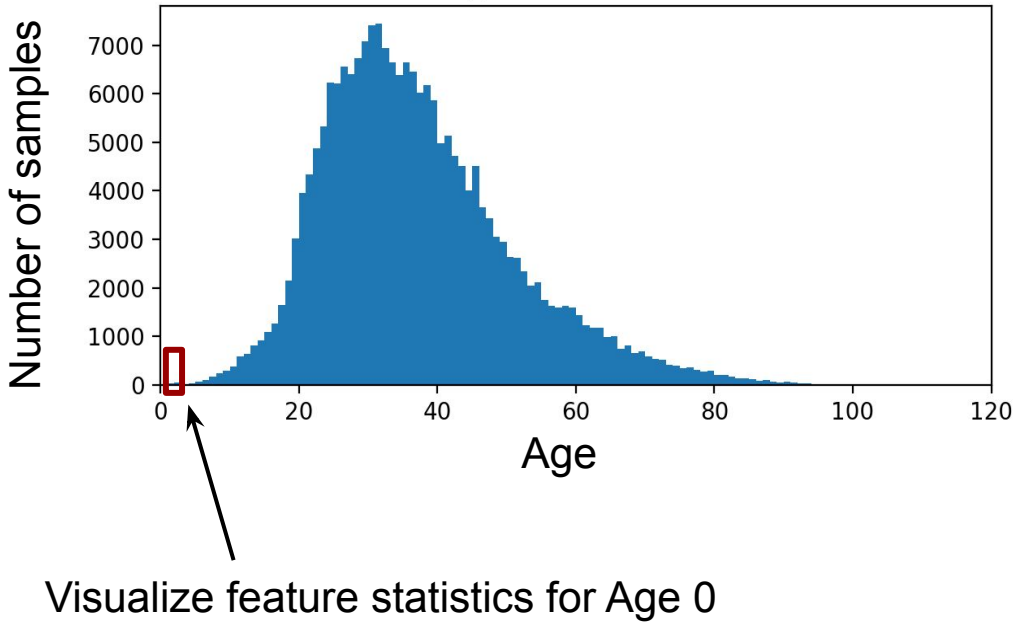


Solution #1: Label Distribution Smoothing (LDS)



Solution #2: Feature Distribution Smoothing (FDS)

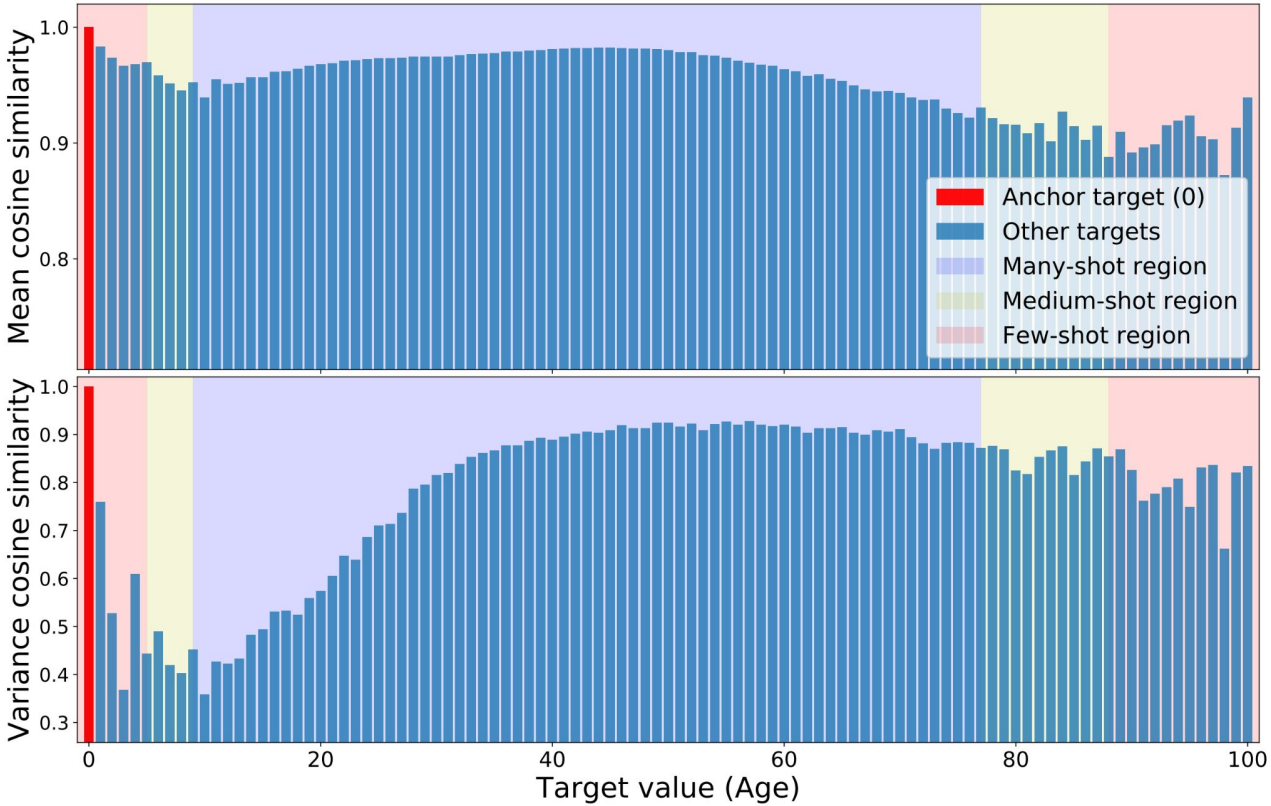
Solution #2: Feature Distribution Smoothing (FDS)



Solution #2: Feature Distribution Smoothing (FDS)

$\cos(\boldsymbol{\mu}_b, \boldsymbol{\mu}_{b_0})$

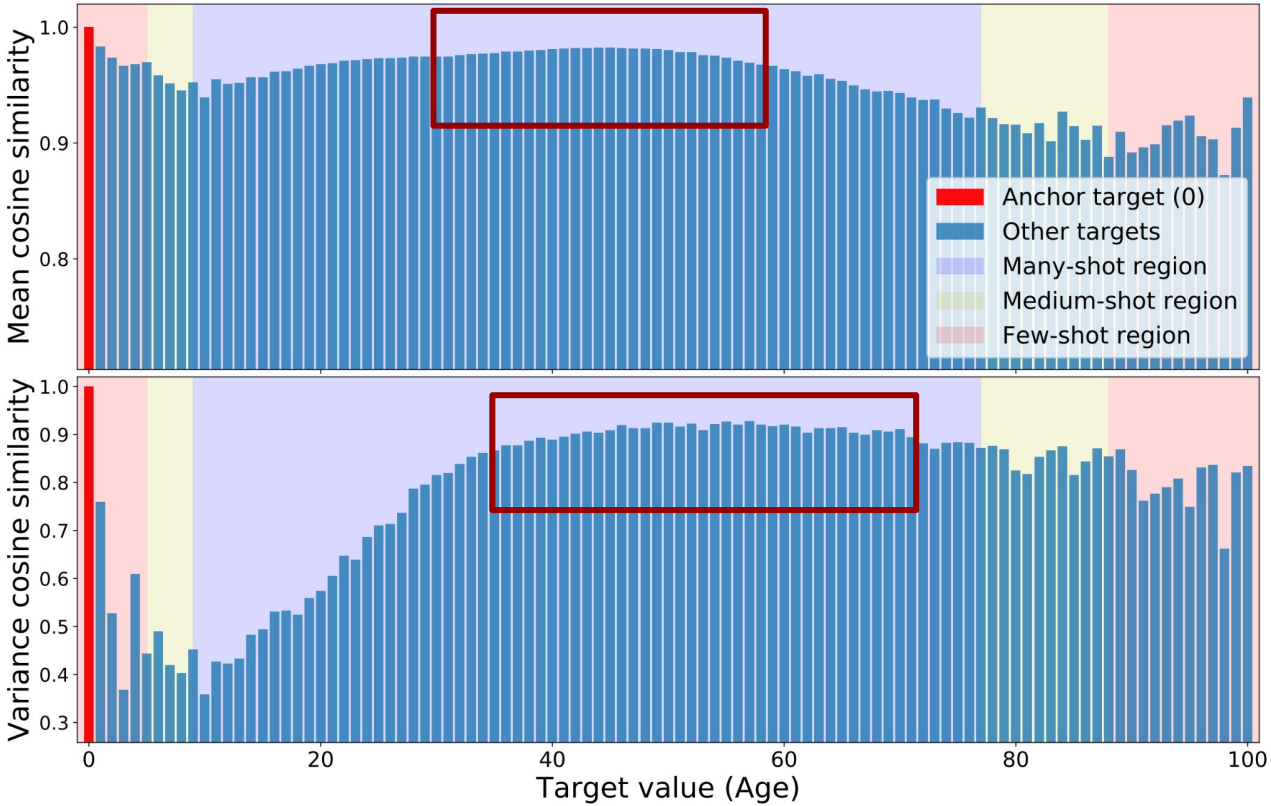
$\cos(\boldsymbol{\sigma}_b, \boldsymbol{\sigma}_{b_0})$



Solution #2: Feature Distribution Smoothing (FDS)

$\cos(\boldsymbol{\mu}_b, \boldsymbol{\mu}_{b_0})$

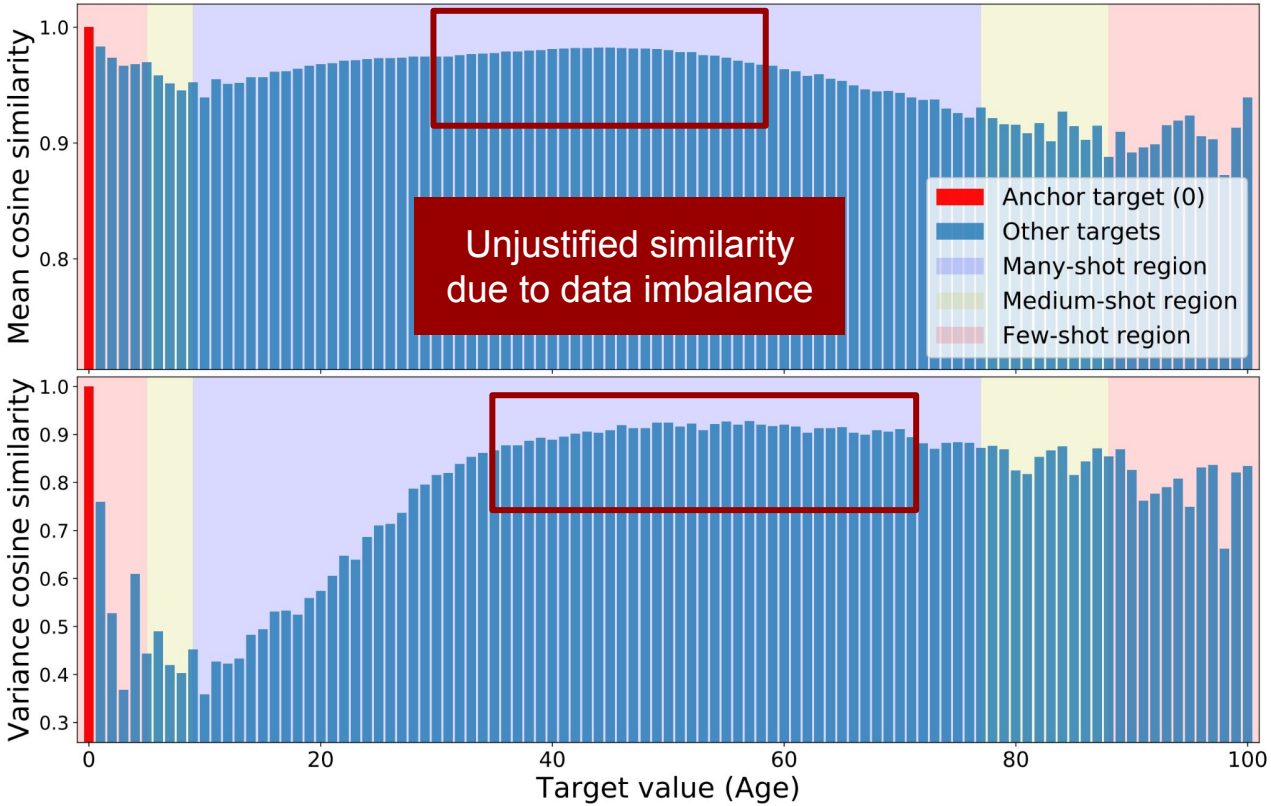
$\cos(\boldsymbol{\sigma}_b, \boldsymbol{\sigma}_{b_0})$



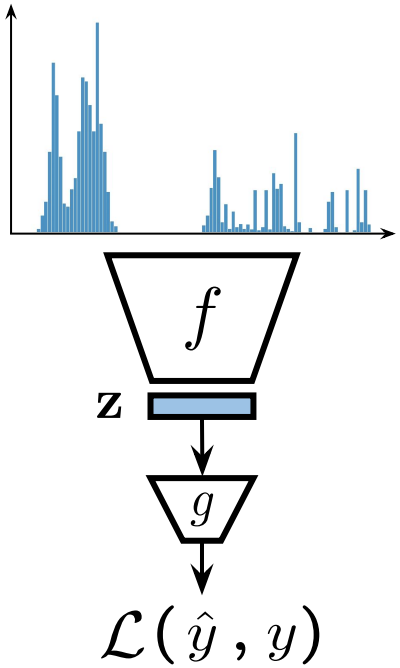
Solution #2: Feature Distribution Smoothing (FDS)

$\cos(\boldsymbol{\mu}_b, \boldsymbol{\mu}_{b_0})$

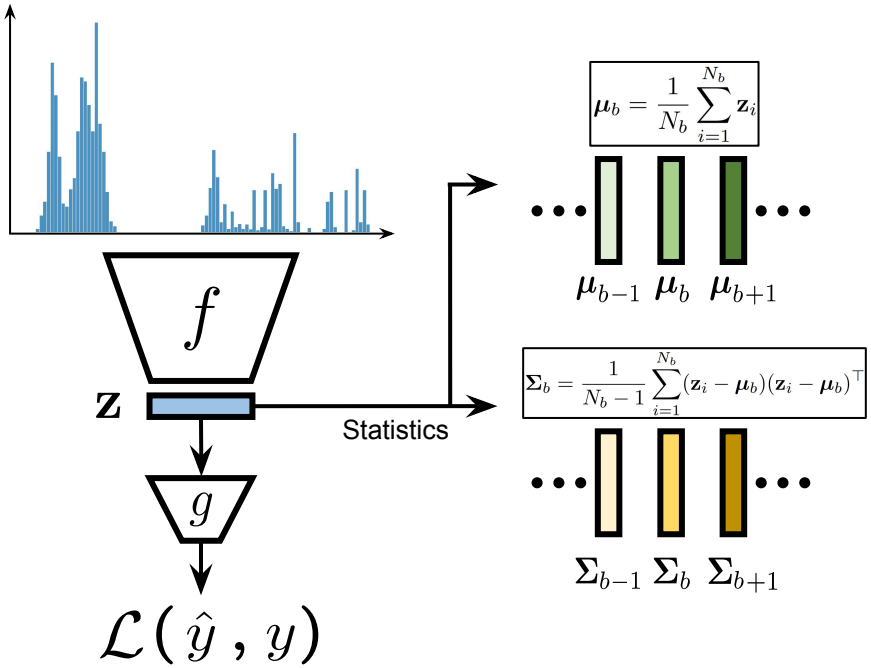
$\cos(\boldsymbol{\sigma}_b, \boldsymbol{\sigma}_{b_0})$



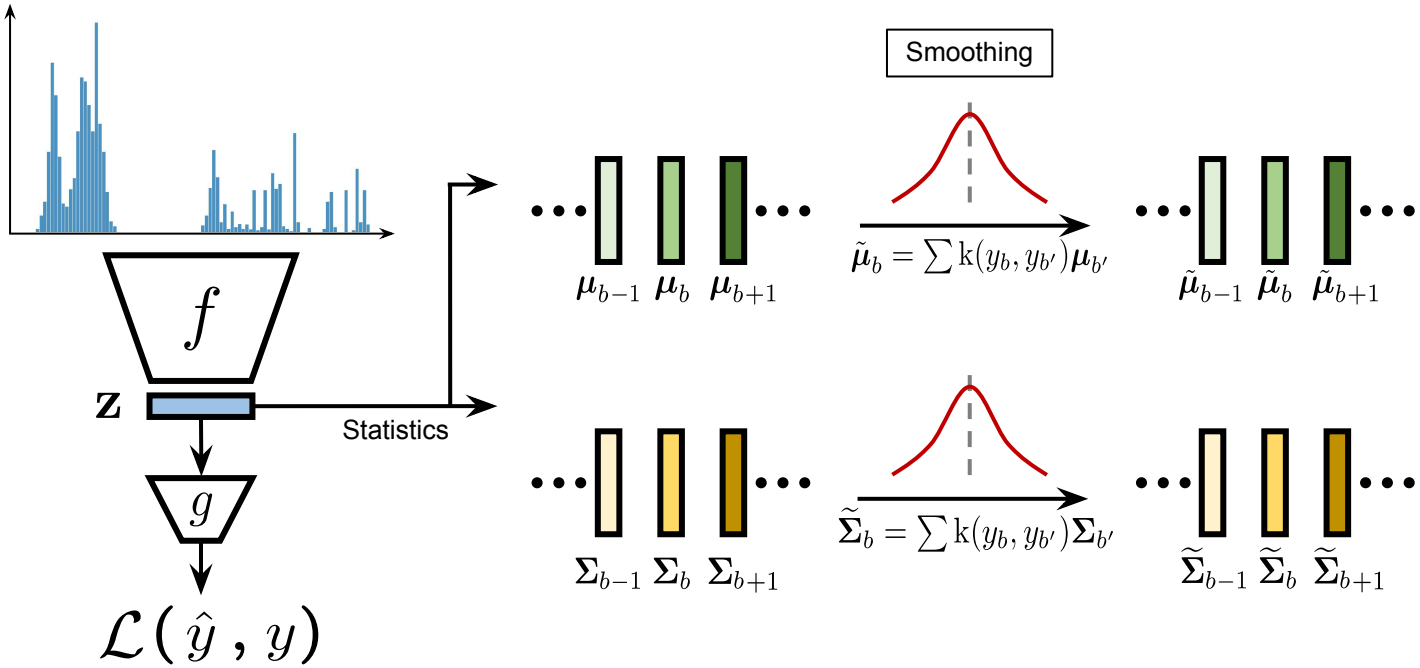
Solution #2: Feature Distribution Smoothing (FDS)



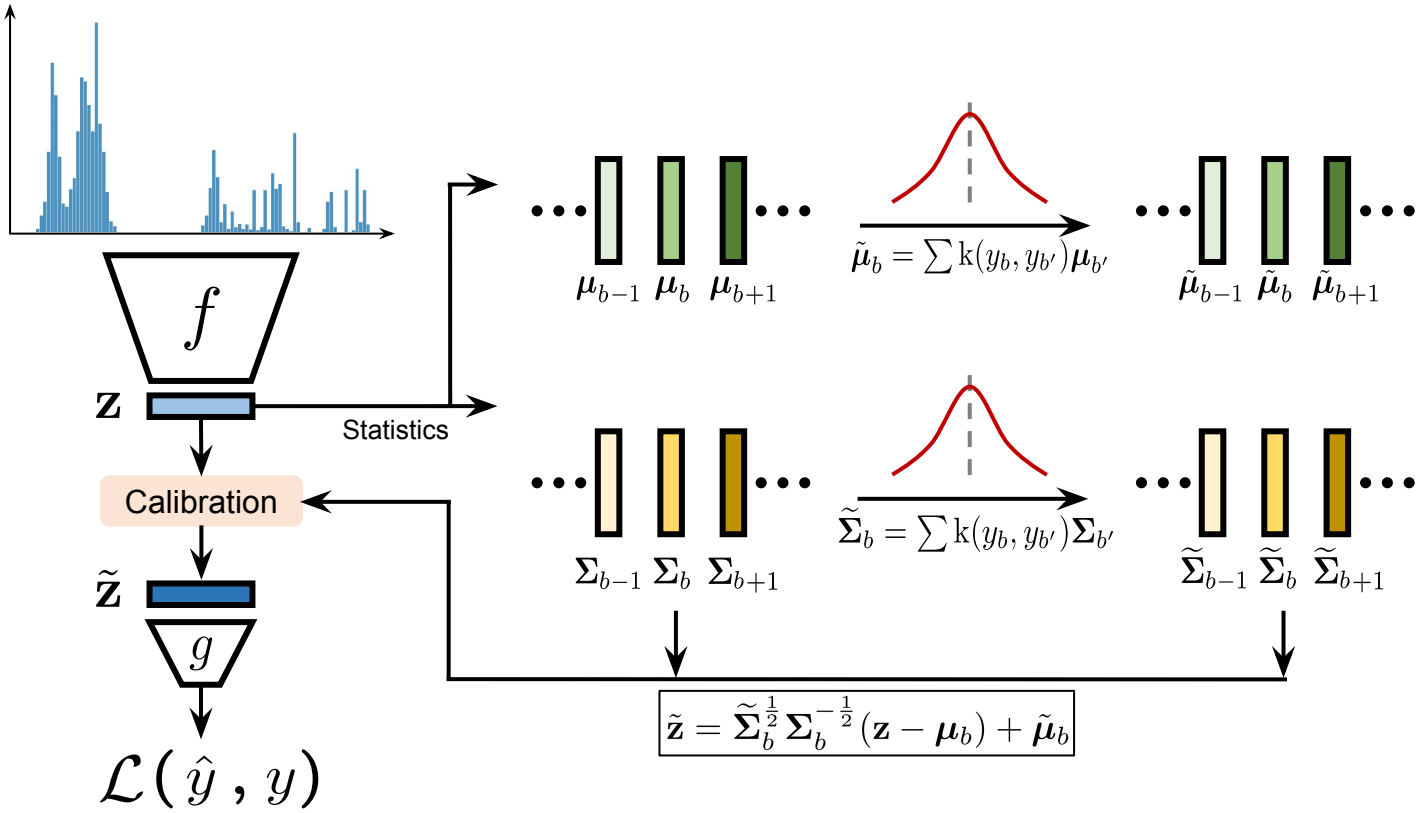
Solution #2: Feature Distribution Smoothing (FDS)



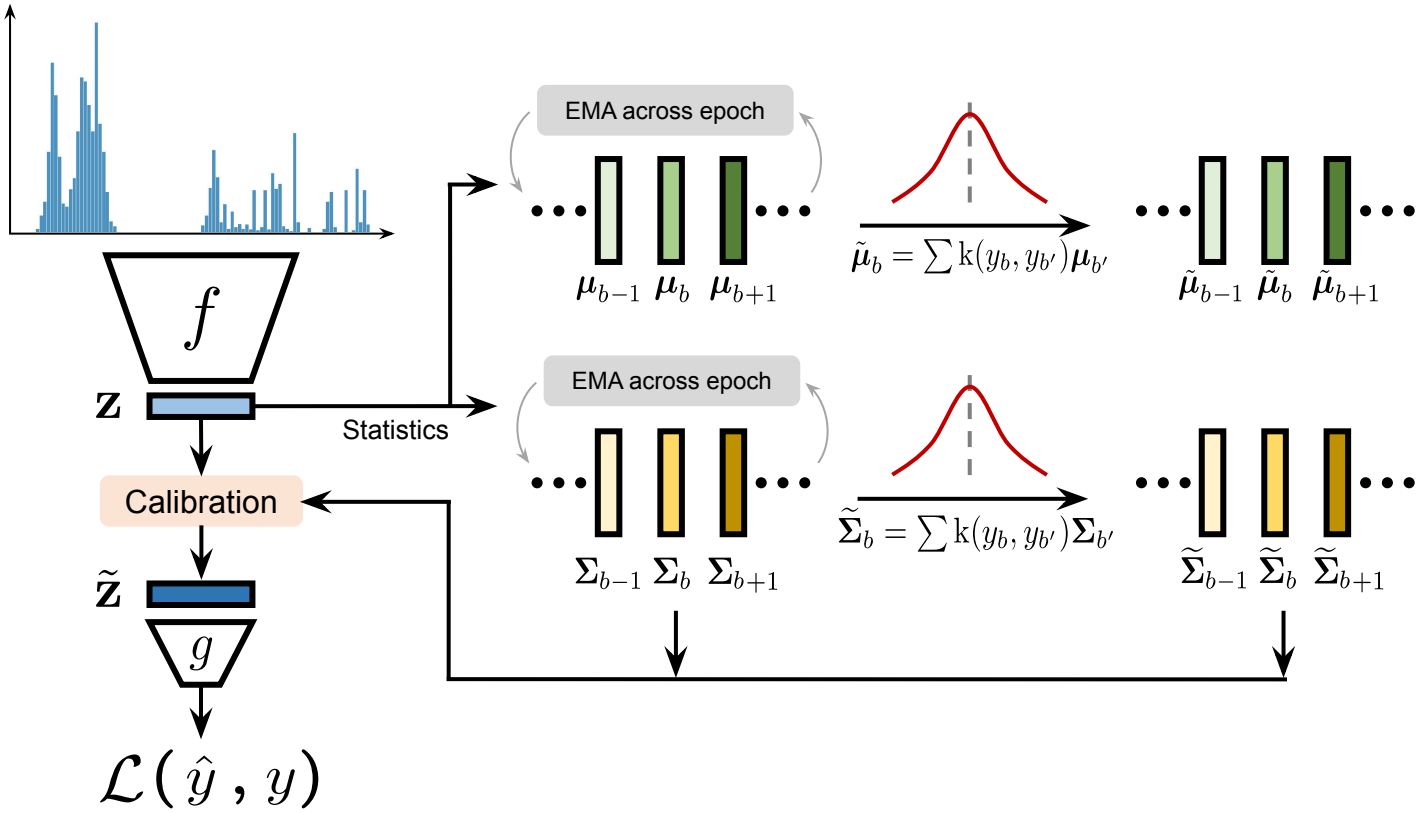
Solution #2: Feature Distribution Smoothing (FDS)



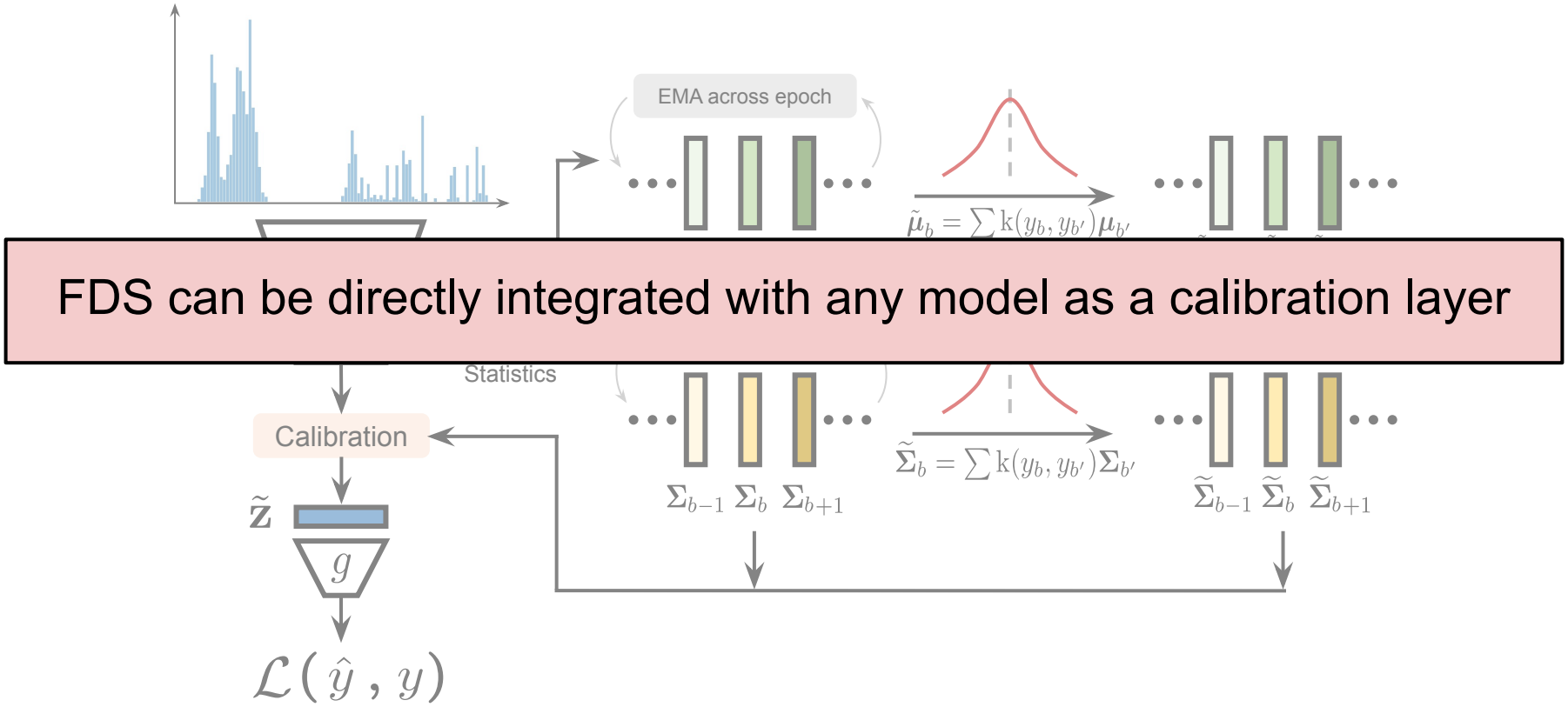
Solution #2: Feature Distribution Smoothing (FDS)



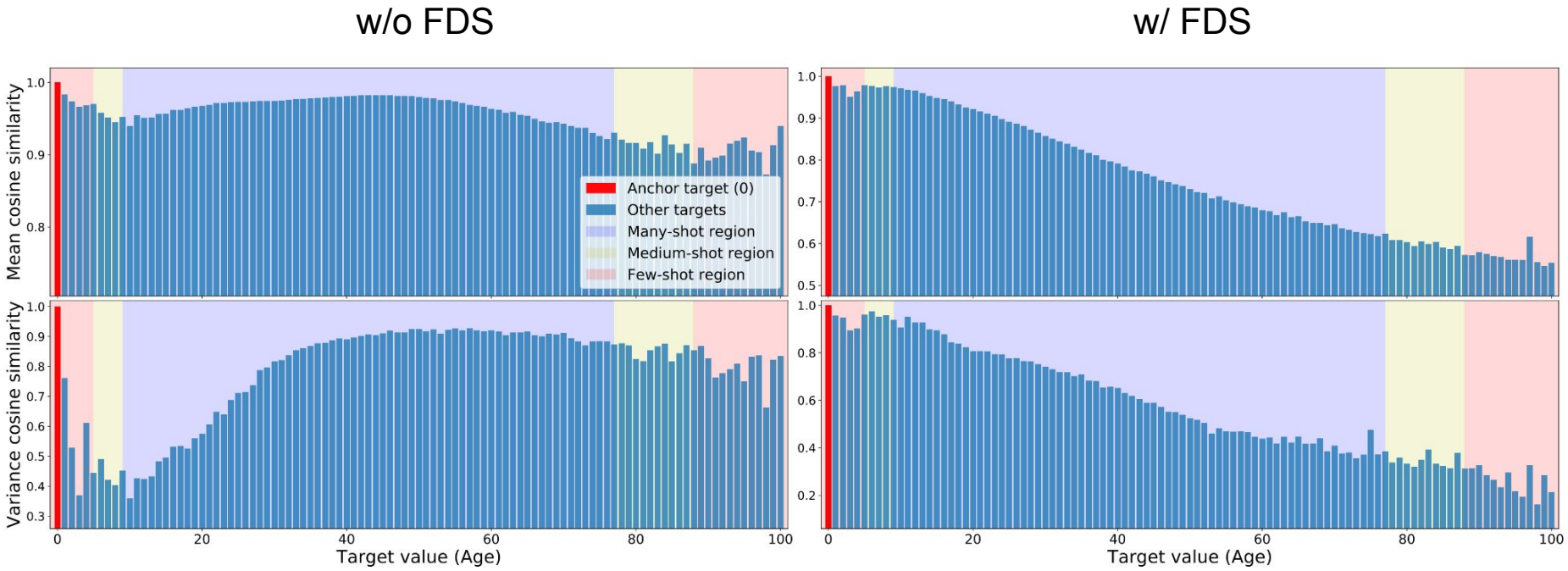
Solution #2: Feature Distribution Smoothing (FDS)



Solution #2: Feature Distribution Smoothing (FDS)



Solution #2: Feature Distribution Smoothing (FDS)



Benchmark imbalanced regression datasets

Computer vision

Natural language processing

Healthcare

Benchmark imbalanced regression datasets

Computer vision

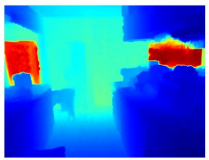
Natural language processing

Healthcare

age

age

depth



Ground truth

IMDB-WIKI-DIR

AgeDB-DIR

NYUD2-DIR

Benchmark imbalanced regression datasets

Computer vision

Natural language processing

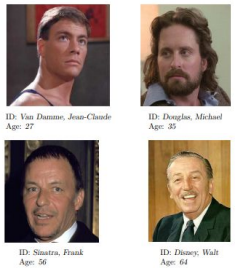
Healthcare

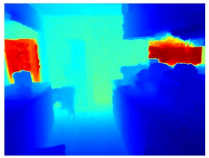
age

age

depth

text similarity





IMDB-WIKI-DIR

AgeDB-DIR

NYUD2-DIR

STS-B-DIR

Benchmark imbalanced regression datasets

Computer vision

Natural language processing

Healthcare

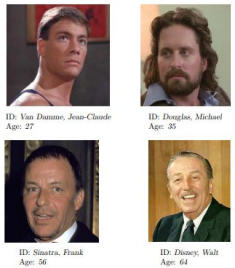
age

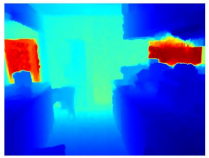
age

depth

text similarity

SF-36 (health score)





Ground truth

IMDB-WIKI-DIR

AgeDB-DIR

NYUD2-DIR

STS-B-DIR

SHHS-DIR

Benchmark imbalanced regression datasets

Computer vision

Natural language processing

Healthcare

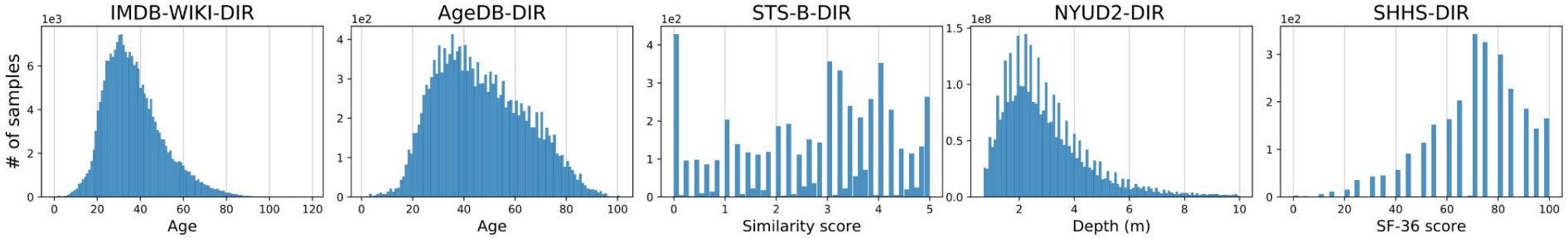
age

age

depth

text similarity

SF-36 (health score)



All datasets exhibit imbalanced distribution

Evaluation results

- Baselines: adapt from imbalanced classification
  - Synthetic samples: (1) SmoteR (2) SMOGN
  - Error-aware loss: (3) Focal-R ($\frac{1}{n}\sum_{i=1}^{n}\sigma(|\beta e_i|)^{\gamma}e_i$)
  - Two-stage training: (4) regressor re-training (RRT)
  - Cost-sensitive re-weighting: (5) naive inverse (INV) (6) square-root inverse (SQINV)

Evaluation results

- Baselines: adapt from imbalanced classification
  - Synthetic samples: (1) SmoteR (2) SMOGN
  - Error-aware loss: (3) Focal-R ($\frac{1}{n}\sum_{i=1}^{n}\sigma(|\beta e_i|)^{\gamma}e_i$)
  - Two-stage training: (4) regressor re-training (RRT)
  - Cost-sensitive re-weighting: (5) naive inverse (INV) (6) square-root inverse (SQINV)

All compatible with our solutions

- LDS
- FDS
- LDS + FDS

Evaluation results

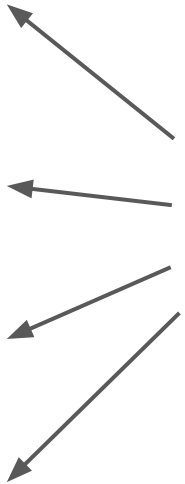
- IMDB-WIKI-DIR

Metrics
Shot
VANILLA
SMOTER (Torgo et al., 2013)
SMOGN (Branco et al., 2017)
SMOGN + **LDS**
SMOGN + **FDS**
SMOGN + **LDS** + **FDS**
FOCAL-R
FOCAL-R + **LDS**
FOCAL-R + **FDS**
FOCAL-R + **LDS** + **FDS**
RRT
RRT + **LDS**
RRT + **FDS**
RRT + **LDS** + **FDS**
SQINV
SQINV + **LDS**
SQINV + **FDS**
SQINV + **LDS** + **FDS**
OURS (BEST) VS. VANILLA

Evaluation results

- IMDB-WIKI-DIR

Metrics
Shot
VANILLA
SMOTER (Torgo et al., 2013)
SMOGN (Branco et al., 2017)
SMOGN + **LDS**
SMOGN + **FDS**
SMOGN + **LDS** + **FDS**
FOCAL-R
FOCAL-R + **LDS**
FOCAL-R + **FDS**
FOCAL-R + **LDS** + **FDS**
RRT
RRT + **LDS**
RRT + **FDS**
RRT + **LDS** + **FDS**
SQINV
SQINV + **LDS**
SQINV + **FDS**
SQINV + **LDS** + **FDS**
OURS (BEST) VS. VANILLA



4 group sections
according to basic strategies

Evaluation results

- IMDB-WIKI-DIR

Metrics	MAE ↓				GM ↓			
Shot	All	Many	Med.	Few	All	Many	Med.	Few
VANILLA								
SMOTER (Torgo et al., 2013)								
SMOGN (Branco et al., 2017)								
SMOGN + **LDS**								
SMOGN + **FDS**								
SMOGN + **LDS** + **FDS**								
FOCAL-R								
FOCAL-R + **LDS**								
FOCAL-R + **FDS**								
FOCAL-R + **LDS** + **FDS**								
RRT								
RRT + **LDS**								
RRT + **FDS**								
RRT + **LDS** + **FDS**								
SQINV								
SQINV + **LDS**								
SQINV + **FDS**								
SQINV + **LDS** + **FDS**								
OURS (BEST) VS. VANILLA								

Evaluation results

- IMDB-WIKI-DIR

Metrics	MAE ↓				GM ↓			
Shot	All	Many	Med.	Few	All	Many	Med.	Few
VANILLA	8.06	7.23	15.12	26.33	4.57	4.17	10.59	20.46
SMOTER (Torgo et al., 2013)								
SMOGN (Branco et al., 2017)								
SMOGN + **LDS**								
SMOGN + **FDS**								
SMOGN + **LDS** + **FDS**								
FOCAL-R								
FOCAL-R + **LDS**								
FOCAL-R + **FDS**								
FOCAL-R + **LDS** + **FDS**								
RRT								
RRT + **LDS**								
RRT + **FDS**								
RRT + **LDS** + **FDS**								
SQINV								
SQINV + **LDS**								
SQINV + **FDS**								
SQINV + **LDS** + **FDS**								
OURS (BEST) VS. VANILLA								

Evaluation results

- IMDB-WIKI-DIR

Metrics	MAE ↓				GM ↓			
Shot	All	Many	Med.	Few	All	Many	Med.	Few
VANILLA	8.06	7.23	15.12	26.33	4.57	4.17	10.59	20.46
SMOTER (Torgo et al., 2013)	8.14	7.42	14.15	25.28	4.64	**4.30**	9.05	19.46
SMOGN (Branco et al., 2017)	8.03	**7.30**	14.02	25.93	4.63	**4.30**	8.74	20.12
SMOGN + **LDS**	8.02	7.39	13.71	23.22	4.63	4.39	8.71	15.80
SMOGN + **FDS**	8.03	7.35	14.06	23.44	4.65	4.33	8.87	16.00
SMOGN + **LDS** + **FDS**	**7.97**	7.38	**13.22**	**22.95**	**4.59**	4.39	**7.84**	**14.94**
FOCAL-R	7.97	7.12	15.14	26.96	4.49	4.10	10.37	21.20
FOCAL-R + **LDS**	7.90	**7.10**	14.72	25.84	**4.47**	**4.09**	10.11	19.14
FOCAL-R + **FDS**	7.96	7.14	14.71	26.06	4.51	4.12	10.16	19.56
FOCAL-R + **LDS** + **FDS**	**7.88**	**7.10**	**14.08**	**25.75**	**4.47**	4.11	**9.32**	**18.67**
RRT	7.81	7.07	14.06	25.13	4.35	4.03	8.91	16.96
RRT + **LDS**	7.79	7.08	13.76	24.64	4.34	**4.02**	8.72	16.92
RRT + **FDS**	**7.65**	**7.02**	12.68	23.85	**4.31**	4.03	7.58	16.28
RRT + **LDS** + **FDS**	**7.65**	7.06	**12.41**	**23.51**	**4.31**	4.07	**7.17**	**15.44**
SQINV	7.87	7.24	12.44	22.76	4.47	4.22	7.25	15.10
SQINV + **LDS**	7.83	7.31	**12.43**	22.51	4.42	4.19	7.00	13.94
SQINV + **FDS**	7.83	7.23	12.60	22.37	4.42	4.20	**6.93**	13.48
SQINV + **LDS** + **FDS**	**7.78**	**7.20**	12.61	**22.19**	**4.37**	**4.12**	7.39	**12.61**
OURS (BEST) VS. VANILLA	**+0.41**	**+0.21**	**+2.71**	**+4.14**	**+0.26**	**+0.15**	**+3.66**	**+7.85**

Evaluation results

- IMDB-WIKI-DIR

LDS / FDS boosts results consistently!

Metrics	MAE ↓				GM ↓			
Shot	All	Many	Med.	Few	All	Many	Med.	Few
VANILLA	8.06	7.23	15.12	26.33	4.57	4.17	10.59	20.46
SMOTER (Torgo et al., 2013)	8.14	7.42	14.15	25.28	4.64	**4.30**	9.05	19.46
SMOGN (Branco et al., 2017)	8.03	**7.30**	14.02	25.93	4.63	**4.30**	8.74	20.12
SMOGN + **LDS**	8.02	7.39	13.71	23.22	4.63	4.39	8.71	15.80
SMOGN + **FDS**	8.03	7.35	14.06	23.44	4.65	4.33	8.87	16.00
SMOGN + **LDS** + **FDS**	**7.97**	7.38	**13.22**	**22.95**	**4.59**	4.39	**7.84**	**14.94**
FOCAL-R	7.97	7.12	15.14	26.96	4.49	4.10	10.37	21.20
FOCAL-R + **LDS**	7.90	**7.10**	14.72	25.84	**4.47**	**4.09**	10.11	19.14
FOCAL-R + **FDS**	7.96	7.14	14.71	26.06	4.51	4.12	10.16	19.56
FOCAL-R + **LDS** + **FDS**	**7.88**	**7.10**	**14.08**	**25.75**	**4.47**	4.11	**9.32**	**18.67**
RRT	7.81	7.07	14.06	25.13	4.35	4.03	8.91	16.96
RRT + **LDS**	7.79	7.08	13.76	24.64	4.34	**4.02**	8.72	16.92
RRT + **FDS**	**7.65**	**7.02**	12.68	23.85	**4.31**	4.03	7.58	16.28
RRT + **LDS** + **FDS**	**7.65**	7.06	**12.41**	**23.51**	**4.31**	4.07	**7.17**	**15.44**
SQINV	7.87	7.24	12.44	22.76	4.47	4.22	7.25	15.10
SQINV + **LDS**	7.83	7.31	**12.43**	22.51	4.42	4.19	7.00	13.94
SQINV + **FDS**	7.83	7.23	12.60	22.37	4.42	4.20	**6.93**	13.48
SQINV + **LDS** + **FDS**	**7.78**	**7.20**	12.61	**22.19**	**4.37**	**4.12**	7.39	**12.61**
OURS (BEST) VS. VANILLA	**+0.41**	**+0.21**	**+2.71**	**+4.14**	**+0.26**	**+0.15**	**+3.66**	**+7.85**

Evaluation results

- IMDB-WIKI-DIR

Metrics	MAE ↓				GM ↓			
Shot	All	Many	Med.	Few	All	Many	Med.	Few
VANILLA	8.06	7.23	15.12	26.33	4.57	4.17	10.59	20.46
SMOTER (Torgo et al., 2013)	8.14	7.42	14.15	25.28	4.64	**4.30**	9.05	19.46
SMOGN (Branco et al., 2017)	8.03	**7.30**	14.02	25.93	4.63	**4.30**	8.74	20.12
SMOGN + **LDS**	8.02	7.39	13.71	23.22	4.63	4.39	8.71	15.80
SMOGN + **FDS**	8.03	7.35	14.06	23.44	4.65	4.33	8.87	16.00
SMOGN + **LDS** + **FDS**	**7.97**	7.38	**13.22**	**22.95**	**4.59**	4.39	**7.84**	**14.94**
FOCAL-R	7.97	7.12	15.14	26.96	4.49	4.10	10.37	21.20
FOCAL-R + **LDS**	7.90	**7.10**	14.72	25.84	**4.47**	**4.09**	10.11	19.14
FOCAL-R + **FDS**	7.96	7.14	14.71	26.06	4.51	4.12	10.16	19.56
FOCAL-R + **LDS** + **FDS**	**7.88**	**7.10**	**14.08**	**25.75**	**4.47**	4.11	**9.32**	**18.67**
RRT	7.81	7.07	14.06	25.13	4.35	4.03	8.91	16.96
RRT + **LDS**	7.79	7.08	13.76	24.64	4.34	**4.02**	8.72	16.92
RRT + **FDS**	**7.65**	**7.02**	12.68	23.85	**4.31**	4.03	7.58	16.28
RRT + **LDS** + **FDS**	**7.65**	7.06	**12.41**	**23.51**	**4.31**	4.07	**7.17**	**15.44**
SQINV	7.87	7.24	12.44	22.76	4.47	4.22	7.25	15.10
SQINV + **LDS**	7.83	7.31	**12.43**	22.51	4.42	4.19	7.00	13.94
SQINV + **FDS**	7.83	7.23	12.60	22.37	4.42	4.20	**6.93**	13.48
SQINV + **LDS** + **FDS**	**7.78**	**7.20**	12.61	**22.19**	**4.37**	**4.12**	7.39	**12.61**
OURS (BEST) VS. VANILLA	**+0.41**	**+0.21**	**+2.71**	**+4.14**	**+0.26**	**+0.15**	**+3.66**	**+7.85**

Large improvements over Vanilla

Evaluation results

- Other datasets

Table 2. Benchmarking results on AgeDB-DIR.

Metrics	MAE ↓				GM ↓			
Shot	All	Many	Med.	Few	All	Many	Med.	Few
VANILLA	7.77	6.62	9.55	13.67	5.05	4.23	7.01	10.75
SMOTER (Torgo et al., 2013)	8.16	7.39	8.65	12.28	5.21	4.65	5.69	8.49
SMOGN (Branco et al., 2017)	8.26	7.64	9.01	12.09	5.36	4.90	6.19	8.44
SMOGN + **LDS**	7.96	7.44	8.64	11.77	5.03	4.68	5.69	7.98
SMOGN + **FDS**	8.06	7.52	8.75	11.89	5.02	4.66	5.63	8.02
SMOGN + **LDS** + **FDS**	**7.90**	**7.32**	**8.51**	**11.19**	**4.98**	**4.64**	**5.41**	**7.35**
FOCAL-R	7.64	6.68	9.22	13.00	4.90	4.26	6.39	9.52
FOCAL-R + **LDS**	7.56	**6.67**	8.82	12.40	4.82	4.27	5.87	8.83
FOCAL-R + **FDS**	7.65	6.89	8.70	**11.92**	4.83	4.32	5.89	**8.04**
FOCAL-R + **LDS** + **FDS**	**7.47**	6.69	**8.30**	12.55	**4.71**	**4.25**	**5.36**	8.59
RRT	7.74	6.98	8.79	11.99	5.00	4.50	5.88	8.63
RRT + **LDS**	7.72	7.00	8.75	11.62	4.98	4.54	5.71	8.27
RRT + **FDS**	7.70	**6.95**	8.76	11.86	4.82	**4.32**	5.83	8.08
RRT + **LDS** + **FDS**	**7.66**	6.99	**8.60**	**11.32**	**4.80**	4.42	**5.53**	**6.99**
SQINV	7.81	7.16	8.80	11.20	4.99	4.57	5.73	7.77
SQINV + **LDS**	7.67	**6.98**	8.86	10.89	4.85	4.39	5.80	7.45
SQINV + **FDS**	7.69	7.10	8.86	**9.98**	4.83	4.41	5.97	**6.29**
SQINV + **LDS** + **FDS**	**7.55**	7.01	**8.24**	10.79	**4.72**	**4.36**	**5.45**	6.79
OURS (BEST) VS. VANILLA	**+0.30**	**-0.05**	**+1.31**	**+3.69**	**+0.34**	**-0.02**	**+1.65**	**+4.46**

Table 3. Benchmarking results on STS-B-DIR.

Metrics	MSE ↓				Pearson correlation (%) ↑			
Shot	All	Many	Med.	Few	All	Many	Med.	Few
VANILLA	0.974	0.851	1.520	0.984	74.2	72.0	62.7	75.2
SMOTER (Torgo et al., 2013)	1.046	0.924	1.542	1.154	72.6	69.3	65.3	70.6
SMOGN (Branco et al., 2017)	0.990	0.896	1.327	1.175	73.2	70.4	65.5	69.2
SMOGN + **LDS**	0.962	0.880	1.242	1.155	74.0	71.5	65.2	69.8
SMOGN + **FDS**	0.987	0.945	**1.101**	1.153	73.0	69.6	**68.5**	69.9
SMOGN + **LDS** + **FDS**	**0.950**	**0.851**	1.327	**1.095**	**74.6**	**72.1**	65.9	**71.7**
FOCAL-R	0.951	0.843	1.425	0.957	74.6	72.3	61.8	76.4
FOCAL-R + **LDS**	0.930	**0.807**	1.449	0.993	**75.7**	**73.9**	62.4	75.4
FOCAL-R + **FDS**	**0.920**	0.855	**1.169**	1.008	75.1	72.6	**66.4**	74.7
FOCAL-R + **LDS** + **FDS**	0.940	0.849	1.358	**0.916**	74.9	72.2	66.3	**77.3**
RRT	0.964	0.842	1.503	0.978	74.5	72.4	62.3	75.4
RRT + **LDS**	0.916	0.817	1.344	0.945	75.7	73.5	64.1	76.6
RRT + **FDS**	0.929	0.857	**1.209**	1.025	74.9	72.1	**67.2**	74.0
RRT + **LDS** + **FDS**	**0.903**	**0.806**	1.323	**0.936**	**76.0**	**73.8**	65.2	**76.7**
INV	1.005	0.894	1.482	1.046	72.8	70.3	62.5	73.2
INV + **LDS**	0.914	0.819	1.319	0.955	75.6	73.4	63.8	76.2
INV + **FDS**	0.927	0.851	**1.225**	1.012	75.0	72.4	**66.6**	74.2
INV + **LDS** + **FDS**	**0.907**	**0.802**	1.363	**0.942**	**76.0**	**74.0**	65.2	**76.6**
OURS (BEST) VS. VANILLA	**+.071**	**+.049**	**+.419**	**+.068**	**+1.8**	**+2.0**	**+5.8**	**+2.1**

Table 4. Benchmarking results on NYUD2-DIR.

Metrics	RMSE ↓				δ_1 ↑			
Shot	All	Many	Med.	Few	All	Many	Med.	Few
VANILLA	1.477	0.591	0.952	2.123	0.677	0.777	0.693	0.570
VANILLA + **LDS**	1.387	0.671	0.913	1.954	0.672	0.701	0.706	0.630
VANILLA + **FDS**	1.442	**0.615**	0.940	2.059	0.681	**0.760**	0.695	0.596
VANILLA + **LDS** + **FDS**	**1.338**	0.670	**0.851**	**1.880**	**0.705**	0.730	**0.764**	**0.655**
OURS (BEST) VS. VANILLA	**+.139**	**-.024**	**+.101**	**+.243**	**+.028**	**-.017**	**+.071**	**+.085**

Table 5. Benchmarking results on SHHS-DIR.

Metrics	MAE ↓				GM ↓			
Shot	All	Many	Med.	Few	All	Many	Med.	Few
VANILLA	15.36	12.47	13.98	16.94	10.63	8.04	9.59	12.20
FOCAL-R	14.67	11.70	13.69	17.06	9.98	7.93	8.85	11.95
FOCAL-R + **LDS**	14.49	12.01	12.43	16.57	9.98	7.89	8.59	11.40
FOCAL-R + **FDS**	14.18	**11.06**	13.56	15.99	9.45	**6.95**	8.81	11.13
FOCAL-R + **LDS** + **FDS**	**14.02**	11.08	**12.24**	**15.49**	**9.32**	7.18	**8.10**	**10.39**
RRT	14.78	12.43	14.01	16.48	10.12	8.05	9.71	11.96
RRT + **LDS**	14.56	12.08	13.44	16.45	9.89	7.85	9.18	11.82
RRT + **FDS**	14.36	11.97	13.33	16.08	9.74	7.54	9.20	11.31
RRT + **LDS** + **FDS**	**14.33**	**11.96**	**12.47**	**15.92**	**9.63**	**7.35**	**8.74**	**11.17**
INV	14.39	11.84	13.12	16.02	9.34	7.73	8.49	11.20
INV + **LDS**	14.14	11.66	12.77	16.05	9.26	7.64	8.18	11.32
INV + **FDS**	13.91	**11.12**	12.29	15.53	8.94	**6.91**	7.79	10.65
INV + **LDS** + **FDS**	**13.76**	**11.12**	**12.18**	**15.07**	**8.70**	6.94	**7.60**	**10.18**
OURS (BEST) VS. VANILLA	**+1.60**	**+1.41**	**+1.80**	**+1.87**	**+1.93**	**+1.13**	**+1.99**	**+2.02**

Evaluation results

- Other datasets

Table 2. Benchmarking results on AgeDB-DIR.

Metrics	MAE ↓				GM ↓			
Shot	All	Many	Med.	Few	All	Many	Med.	Few
VANILLA	7.77	6.62	9.55	13.67	5.05	4.23	7.01	10.75
RRT + **LDS**	7.72	7.00	8.75	11.62	4.98	4.54	5.71	8.27
RRT + **FDS**	7.70	**6.95**	8.76	11.86	4.82	**4.32**	5.83	8.08
RRT + **LDS** + **FDS**	**7.66**	6.99	**8.60**	**11.32**	**4.80**	4.42	**5.53**	**6.99**
SQINV	7.81	7.16	8.80	11.20	4.99	4.57	5.73	7.77
SQINV + **LDS**	7.67	**6.98**	8.86	10.89	4.85	4.39	5.80	7.45
SQINV + **FDS**	7.69	7.10	8.86	**9.98**	4.83	4.41	5.97	**6.29**
SQINV + **LDS** + **FDS**	**7.55**	7.01	**8.24**	10.79	**4.72**	**4.36**	**5.45**	6.79
OURS (BEST) VS. VANILLA	**+0.30**	**-0.05**	**+1.31**	**+3.69**	**+0.34**	**-0.02**	**+1.65**	**+4.46**

Table 3. Benchmarking results on STS-B-DIR.

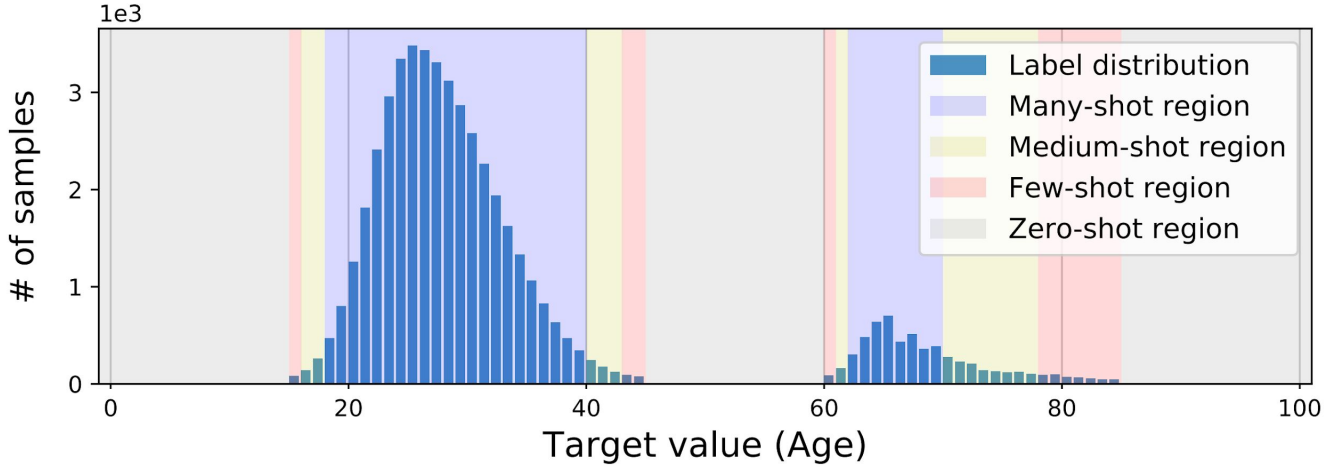
Metrics	MSE ↓				Pearson correlation (%) ↑			
Shot	All	Many	Med.	Few	All	Many	Med.	Few
VANILLA	0.974	0.851	1.520	0.984	74.2	72.0	62.7	75.2
RRT	0.964	0.842	1.503	0.978	74.5	72.4	62.5	75.4
RRT + **LDS**	0.916	0.817	1.344	0.945	75.7	73.5	64.1	76.6
RRT + **FDS**	0.929	0.857	**1.209**	1.025	74.9	72.1	**67.2**	74.0
RRT + **LDS** + **FDS**	**0.903**	**0.806**	1.323	**0.936**	**76.0**	**73.8**	65.2	**76.7**
INV	1.005	0.894	1.482	1.046	72.8	70.3	62.5	73.2
INV + **LDS**	0.914	0.819	1.319	0.955	75.6	73.4	63.8	76.2
INV + **FDS**	0.927	0.851	**1.225**	1.012	75.0	72.4	**66.6**	74.2
INV + **LDS** + **FDS**	**0.907**	**0.802**	1.363	**0.942**	**76.0**	**74.0**	65.2	**76.6**
OURS (BEST) VS. VANILLA	**+.071**	**+.049**	**+.419**	**+.068**	**+1.8**	**+2.0**	**+5.8**	**+2.1**

Table 4. Benchmarking results on NYUD2-DIR.

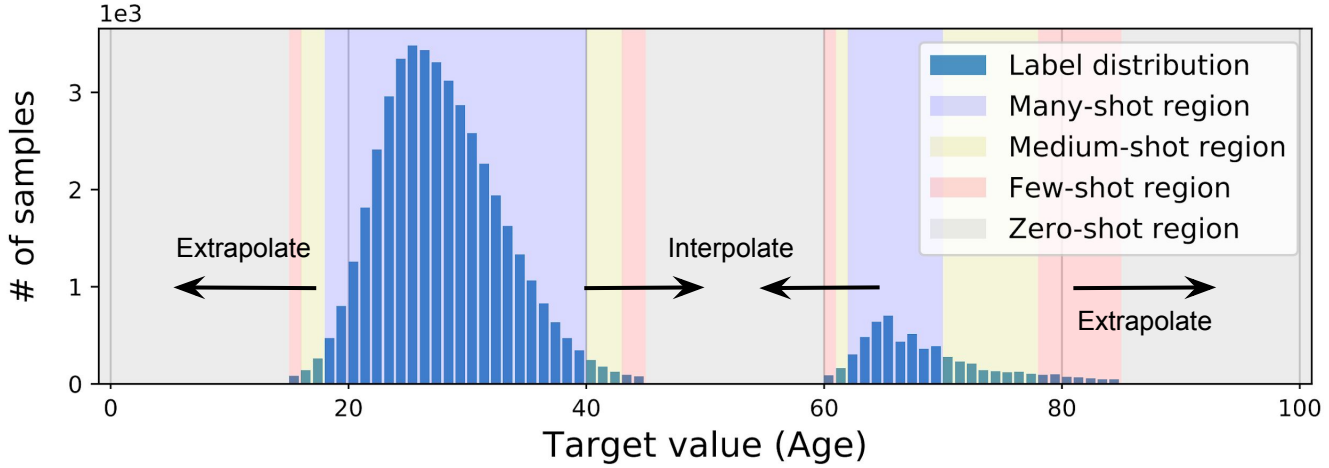
Metrics	RMSE ↓				δ_1 ↑			
Shot	All	Many	Med.	Few	All	Many	Med.	Few
VANILLA	1.477	0.591	0.952	2.123	0.677	0.777	0.693	0.570
VANILLA + **LDS**	1.387	0.671	0.913	1.954	0.672	0.701	0.706	0.630
VANILLA + **FDS**	1.442	**0.615**	0.940	2.059	0.681	**0.760**	0.695	0.596
VANILLA + **LDS** + **FDS**	**1.338**	0.670	**0.851**	**1.889**	**0.705**	0.730	**0.764**	**0.655**
FOCAL-R	14.67	11.70	13.69	17.06	9.98	7.93	8.83	11.93
FOCAL-R + **LDS**	14.49	12.01	12.43	16.57	9.98	7.89	8.59	11.40
FOCAL-R + **FDS**	14.18	**11.06**	13.56	15.99	9.45	**6.95**	8.81	11.13
FOCAL-R + **LDS** + **FDS**	**14.02**	11.08	**12.24**	**15.49**	**9.32**	7.18	**8.10**	**10.39**
RRT	14.78	12.43	14.01	16.48	10.12	8.05	9.71	11.96
RRT + **LDS**	14.56	12.08	13.44	16.45	9.89	7.85	9.18	11.82
RRT + **FDS**	14.36	11.97	13.33	16.08	9.74	7.54	9.20	11.31
RRT + **LDS** + **FDS**	**14.33**	**11.96**	**12.47**	**15.92**	**9.63**	**7.35**	**8.74**	**11.17**
INV	14.39	11.84	13.12	16.02	9.34	7.73	8.49	11.20
INV + **LDS**	14.14	11.66	12.77	16.05	9.26	7.64	8.18	11.32
INV + **FDS**	13.91	**11.12**	12.29	15.53	8.94	**6.91**	7.79	10.65
INV + **LDS** + **FDS**	**13.76**	**11.12**	**12.18**	**15.07**	**8.70**	6.94	**7.60**	**10.18**
OURS (BEST) VS. VANILLA	**+1.60**	**+1.41**	**+1.80**	**+1.87**	**+1.93**	**+1.13**	**+1.99**	**+2.02**

Superior performance across all DIR datasets!
(complete results in paper)

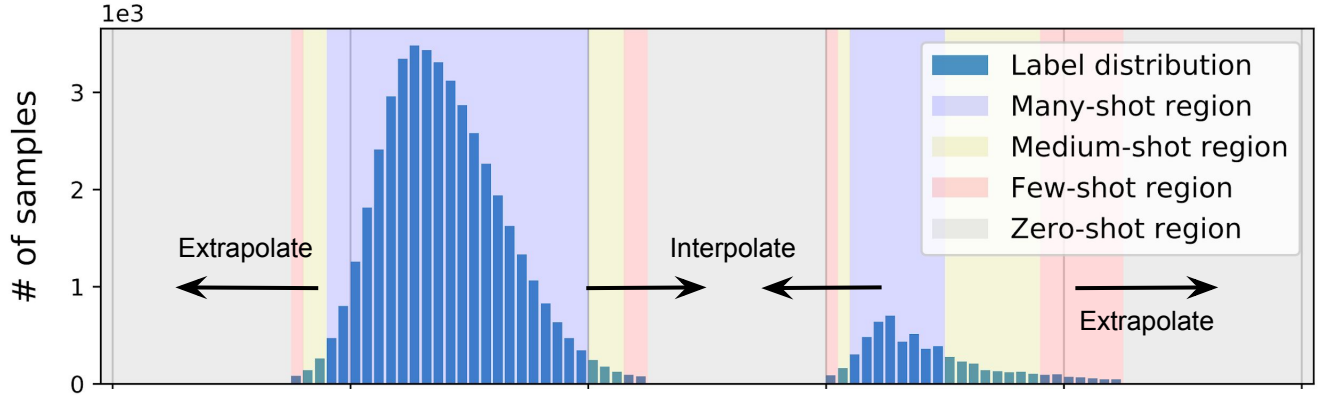
Analysis: Extrapolation & Interpolation



Analysis: Extrapolation & Interpolation

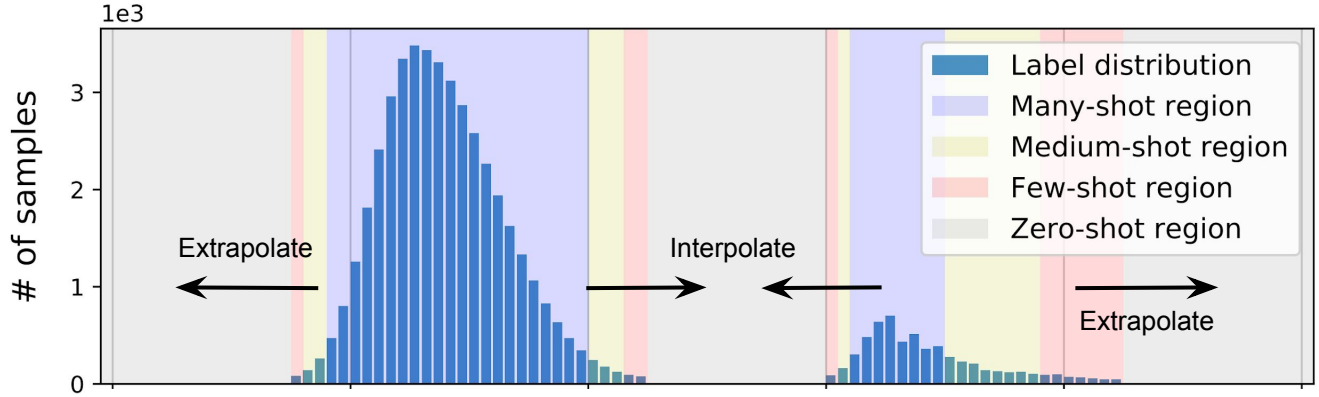


Analysis: Extrapolation & Interpolation



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 + **LDS**	10.54	8.31	14.14	17.38	6.50	4.67	12.13	15.36
VANILLA + **FDS**	11.40	8.97	15.83	18.01	7.18	5.12	14.02	16.48
VANILLA + **LDS** + **FDS**	**10.27**	**8.11**	**13.71**	**17.02**	**6.33**	**4.55**	**11.71**	**15.13**
OURS (BEST) VS. VANILLA	**+1.45**	**+1.21**	**+2.42**	**+1.17**	**+1.11**	**+0.78**	**+2.70**	**+1.61**

Analysis: Extrapolation & Interpolation



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 + **LDS**	10.54	8.31	14.14	17.38	6.50	4.67	12.13	15.36
VANILLA + **FDS**	11.40	8.97	15.83	18.01	7.18	5.12	14.02	16.48
VANILLA + **LDS** + **FDS**	**10.27**	**8.11**	**13.71**	**17.02**	**6.33**	**4.55**	**11.71**	**15.13**
OURS (BEST) VS. VANILLA	**+1.45**	**+1.21**	**+2.42**	**+1.17**	**+1.11**	**+0.78**	**+2.70**	**+1.61**

Summary

- New task: Deep Imbalanced Regression (DIR)
- New techniques: Label distribution smoothing (LDS) & Feature distribution smoothing (FDS)
- New benchmarks: IMDB-WIKI-DIR / AgeDB-DIR / STS-B-DIR / NYUD2-DIR / SHHS-DIR

Check out our paper and code at...

- Paper: https://arxiv.org/abs/2102.09554
- Code + data: https://github.com/YyzHarry/imbalanced-regression