Deep Learning Project Report

Title : Balanced MSE

Group-09
DEBAJYOTI MAZUMDER (2211002)
MAYUR MANKAR (20169)
MOHIT (20323)

Data Science And Engineering Dept.
Indian Institute of Science Education and Research Bhopal
Bhopal Bypass Road, Bhauri Bhopal 462 066 Madhya Pradesh, INDIA
Submitted on: 15th April, 2023

1 Discussion on application

Imbalanced regression occurs when the dataset used for regression analysis has an unequal distribution of target variables. In other words, there are significantly more data points for one class than the other. This can occur in many scenarios such as predicting the sale of luxury cars versus budget cars, predicting the failure of machines versus successful ones, and so on. Here are some applications of imbalanced regression:

1. Fraud Detection: When detecting fraud, there are typically many more legitimate transactions than fraudulent ones. Imbalanced regression can be used to identify fraudulent transactions, given that they are often rare events.

2. Medical Diagnosis: Medical diagnosis often involves detecting rare medical conditions. Imbalanced regression can be used to develop predictive models that can identify these rare medical conditions even when there are very few examples available in the dataset.

3. Credit Risk Assessment: When assessing credit risk, most people have good credit scores and a few have bad credit scores. Imbalanced regression can be used to identify individuals with bad credit scores, who are more likely to default on their loans.

4. Sales Forecasting: Sales forecasting can be used to predict the sales of a particular product in the future. Imbalanced regression can be used to predict sales of luxury items, where sales are often much lower than sales of everyday items.

Imbalanced regression can also be applied to age detection, where the distribution of age groups may not be balanced. Age detection involves predicting the age of an individual based on features such as facial characteristics, voice, or other biometric data.

In many cases, the distribution of age groups in the dataset may be imbalanced, with a large number of samples from certain age groups and a small number of samples from others. In such scenarios, imbalanced regression can be used to build accurate age detection models.

For example, in a facial recognition system, there may be many images of young people and a few images of older people. Imbalanced regression can be used to build a model that can accurately detect the age of older people, even with a small number of samples.

Similarly, in voice recognition systems, the distribution of voice samples may be imbalanced, with more samples from younger individuals than older individuals. Imbalanced regression can be used to build a model that can accurately detect the age of older individuals based on their voice, even with a small number of samples.

2 Literature survey

Imbalanced & Long-Tailed Classification For unbalanced & long-tailed classification, a variety of methods have been explored, including resampling [5, 4, 6, 8] and reweighting [3, 7] . Here, we concentrate on logit adjustment techniques because they are crucial to our job. Recent studies [9, 11, 16] show that modifying the logits in the mapping function, e.g., Softmax or Sigmoid, by an offset proportional to $logp_{train(y)}$ gives the Bayes optimal estimation of the $p_{bal(y|x)}$. The train-time loss function or the test-time adjustment can both be used with the logit adjustment procedures. [18] goes on to create an online version that collects label distribution statistics while training rather than collecting statistics for all training labels beforehand.

Imbalanced Regression Regression with imbalances is not often studied. The emphasis of earlier works [2, 17] has been on resampling and synthesising fresh samples for specialised labelling. Additional research uses ensemble regressors that were trained using various resampling strategies [1]. It is difficult to apply their strategy to high-dimensional observations like photographs. Recent study [15, 19] suggests using KDE to estimate the empirical training distribution before using the conventional reweighting method. A feature level smoothing is also suggested by [19], which is complementary to our study.

3 Methodology

Here, we are dealing with a regression based task where the labels(y) corresponding to the data has a highly skewed distribution, which can result into getting only the most frequent labels by the model. However, the error rate is still low, but issue of less frequent(minority) labels is not solved. To mitigate the problem of imbalance of label distribution, three kind of approaches are prevalent: first, balanced test set; second balanced evaluation metric; and third: balanced metric on arbitrary test set.

We assume that $p_{train}(x, y)$ and $p_{bal}(x, y)$ are drawn from different joint distributions, where the training set's label distribution $p_{train}(y)$ is skewed and the balanced test set's label distribution $p_{bal}(y)$ is uniform. The label-conditional probability $p(x, y)$ is assumed to be the same in both training and testing. Instead of learning $p_{train}(x, y)$, imbalanced regression's goal is to estimate $p_{bal}(x, y)$ to better perform on the balanced test set.

If the effective label density is available, techniques for addressing class imbalance problems can be directly adapted to the DIR context. For example, a straightforward adaptation can be the cost-sensitive re-weighting method, where we re-weight the loss function by multiplying it by the inverse of the Label Distribution Smoothing estimated label density for each target.

4 Your contributions

Table 1: Contributions

Tasks	Contributions
Paper Understanding	Debajyoti, Mayur, Mohit
Original results reproduced	Debajyoti, Mayur, Mohit(*)
Dataset introduced (ageDB)	Debajyoti, Mayur
Loss functions introduced (sMAPE, logcosh)	Mohit

* : Done after first phase of project

We have made several attempts at few things for novelty. Debajyoti and Mohit did few experiments with loss functions for regression. Mayur incorporated Pearson's correlation in NYUD2-DIR. Only the comparable results of experiments are being put on the above table.

5 Datasets

We have used three datasets from DIR benchmark of [19]. We use IMDB-WIKI-DIR dataset for age estimation from face images and NYUD2- DIR dataset to estimate depth maps from images of indoor scenes.

- IMDB-WIKI-DIR (age): We construct IMDB-WIKI-DIR using the IMDB-WIKI dataset [13], which contains 523.0K face images and the corresponding ages. We filter out unqualified images, and manually construct balanced validation and test set over the supported ages. The length of each bin is 1 year, with a minimum age of 0 and a maximum age of 186. The number of images per bin varies between 1 and 7149, exhibiting significant data imbalance. Overall, the curated dataset has 191.5K images for training, 11.0K images for validation and testing.
- NYUD2-DIR (depth): We create NYUD2-DIR based on the NYU Depth Dataset V2 [14], which provides images and depth maps for different indoor scenes. The depth maps have an upper bound of 10 meters and we set the bin length as 0.1 meter. Following standard practices, we use 50K images for training and 654 images for testing. We randomly select 9357 test pixels for each bin to make the test set balanced.
- AgeDB-DIR (age): AgeDB-DIR is constructed in a similar manner from the AgeDB dataset [10]. It contains 12.2K images for training, with a minimum age of 0 and a maximum age of 101, and maximum bin density of 353 images and minimum bin density of 1. The validation and test set are balanced with 2.1K images.

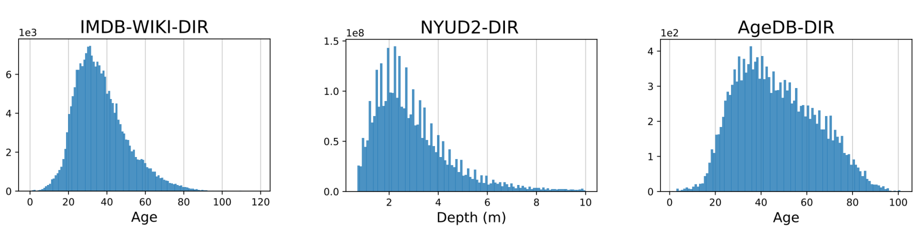
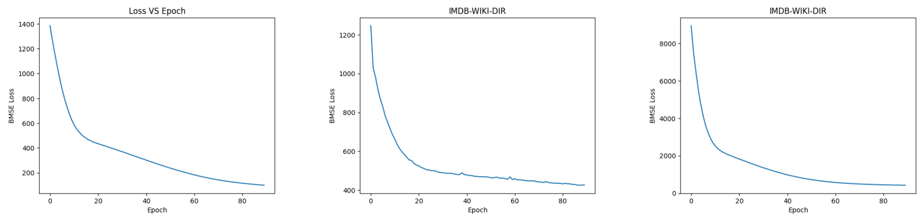


Figure 1: Overview of training set label distribution for three DIR datasets

6 Training plots



(a) ageDB Loss curve (b) IMDB on SMAPE (c) IMDB on logcosh

Figure 2: Loss curves

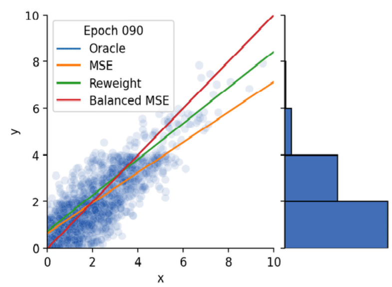


Figure 3: Comparison of Ablations

7 Evaluation metrics and comparisons

For metrics, we have used common regression based metrics, like Mean Absolute Error(MAE) and Mean-Squared Error(MSE). Since, the test set of AgeDB-DIR and IMDB-WIKI-DIR datasets is not balanced, therefore our assumption of balanced test set doesn't hold for them. So, to measure the model's performance uniformly, we use strategy of balanced metric and divide the label space into finite number of even sub-regions, compute the average average inside the sub-regions, and take the mean overall sub-regions. We name it balanced metric(bMAE). NYUD2-DIR's test set is balanced, we follow [19] and report RMSE. Following are the results which came out to be comparable to the original outcomes of the paper [12] and [19]:

Table 2: Reproduced results of IMDB-WIKI

Model	Overall	Many	Medium	Few
Vanilla	13.923	7.323	15.925	32.778
Vanilla (reproduced)	13.612	7.264	15.886	32.918
GAI	12.690	7.589	12.880	28.307
GAI (reproduced)	12.855	7.623	12.983	28.585
BMC	12.654	7.649	12.689	28.097
BMC (reproduced)	12.389	7.615	12.437	28.227
BNI	12.650	7.647	12.696	28.077
BNI (reproduced)	12.847	7.676	12.722	28.326

Table 3: Reproduced results of NYUD2

Model	Overall	Many	Medium	Few
GAI	1.279	0.819	0.917	1.705
GAI (reproduced)	1.293	0.884	0.925	1.792
BNI	1.281	0.833	0.856	1.714
BNI (reproduced)	1.294	0.835	0.877	1.732

Table 4: Logcosh on IMDB

Model	Overall	Many-Shot	Medium-Shot	Few-Shot
Vanilla	13.923	7.323	15.925	32.778
GAI	13.035	7.500	12.915	30.235
BMC	13.040	7.473	12.957	30.313
BNI	13.060	7.451	12.931	30.496

Table 5: SMAPE on IMDB

Model	Overall	Many-Shot	Medium-Shot	Few-Shot
Vanilla	13.923	7.323	15.925	32.778
GAI	13.064	7.391	12.943	30.695
BMC	13.007	7.549	12.917	29.948
BNI	13.036	7.476	12.930	30.301

Table 6: ageDB

Model	Overall	Many-Shot	Medium-Shot	Few-Shot
Vanilla	15.016	10.040	16.065	24.496
BMC	15.56	10.049	16.210	24.532
BNI	15.016	10.040	16.065	24.496
GAI	15.082	9.945	16.660	24.551

We believe few more modifications on the loss function and using it with the new dataset can bring out some novel results.

References

[1] Paula Branco, Luís Torgo, and Rita Ribeiro. “REBAGG: REsampled BAGGing for Imbalanced Regression”. In: Sept. 2018.

[2] Paula Branco, Luís Torgo, and Rita Ribeiro. “SMOGN: a Pre-processing Approach for Imbalanced Regression”. In: Sept. 2017.

[3] Kaidi Cao et al. “Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss”. In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc., 2019. URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/621461af90cadfdaf0e8d4cc25129f91-Paper.pdf.

[4] Peng Chu et al. “Feature Space Augmentation for Long-Tailed Data”. In: *CoRR* abs/2008.03673 (2020). arXiv: 2008.03673. URL: https://arxiv.org/abs/2008.03673.

[5] S T Chung. “Isolation and characterization of Streptomyces fradiae plasmids which are prophage of the actinophagephiSF1”. In: *Gene* 17 (1982), pp. 239–246. DOI: 10.1016/0378-1119(82)90139-1.

[6] Haibo He and Edwardo A. Garcia. “Learning from Imbalanced Data”. In: *IEEE Transactions on Knowledge and Data Engineering* 21.9 (2009), pp. 1263–1284. DOI: 10.1109/TKDE.2008.239.

[7] Muhammad Abdullah Jamal et al. “Rethinking Class-Balanced Methods for Long-Tailed Visual Recognition from a Domain Adaptation Perspective”. In: *CoRR* abs/2003.10780 (2020). arXiv: 2003.10780. URL: https://arxiv.org/abs/2003.10780.

[8] Jaehyung Kim, Jongheon Jeong, and Jinwoo Shin. “M2m: Imbalanced Classification via Major-to-minor Translation”. In: *CoRR* abs/2004.00431 (2020). arXiv: 2004.00431. URL: https://arxiv.org/abs/2004.00431.

[9] Aditya Krishna Menon et al. “Long-tail learning via logit adjustment”. In: *CoRR* abs/2007.07314 (2020). arXiv: 2007.07314. URL: https://arxiv.org/abs/2007.07314.

[10] Stylianos Moschoglou et al. "AgeDB: The First Manually Collected, In-the-Wild Age Database". In: July 2017, pp. 1997–2005. DOI: 10.1109/CVPRW.2017.250.

[11] Jiawei Ren et al. "Balanced Meta-Softmax for Long-Tailed Visual Recognition". In: *CoRR* abs/2007.10740 (2020). arXiv: 2007.10740. URL: https://arxiv.org/abs/2007.10740.

[12] Jiawei Ren et al. *Balanced MSE for Imbalanced Visual Regression*. 2022. arXiv: 2203.16427 [cs.CV].

[13] Rasmus Rothe, Radu Timofte, and Luc Van Gool. "Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks". In: *International Journal of Computer Vision* 126 (Apr. 2018). DOI: 10.1007/s11263-016-0940-3.

[14] Nathan Silberman et al. "Indoor Segmentation and Support Inference from RGBD Images". In: vol. 7576. Oct. 2012, pp. 746–760. ISBN: 978-3-642-33714-7. DOI: 10.1007/978-3-642-33715-4_54.

[15] Michael Steininger et al. "Density-based weighting for imbalanced regression". In: *Machine Learning* 110 (Aug. 2021), pp. 1–25. DOI: 10.1007/s10994-021-06023-5.

[16] Junjiao Tian et al. "Posterior Re-calibration for Imbalanced Datasets". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 8101–8113.

[17] Luís Torgo et al. "SMOTE for Regression". In: vol. 8154. Sept. 2013, pp. 378–389. ISBN: 978-3-642-40668-3. DOI: 10.1007/978-3-642-40669-0_33.

[18] Jiaqi Wang et al. "Seesaw Loss for Long-Tailed Instance Segmentation". In: (Aug. 2020).

[19] Yuzhe Yang et al. "Delving into Deep Imbalanced Regression". In: *CoRR* abs/2102.09554 (2021). arXiv: 2102.09554. URL: https://arxiv.org/abs/2102.09554.