LOSS FUNCTIONS IN IMBALANCED CLASSIFICATION

Janocha K., Pouplin A., Higgins C., Linton N., Bharath A.

Imperial College London

Abstract

One of the key characteristics of medical data is imbalance between classes, which can inappropriately bias deep learning models' predictions toward the majority class. Such a bias can reduce the effectiveness of a model in detecting rare pathologies. In this study, we aim to compare and analyse the effect of various loss objectives on models' behaviour with respect to the level of class imbalance.

Research objective

The problem of class imbalance is often approached by either class resampling, class-margin enlarging, or cost sensitive training [3]. (assuming prior knowledge of the between-class distribution). which assumes prior knowledge of the between-class distributions. However, the issue of how the choice of loss function in classification affects a model's generalisation abilities [2] remains underrepresented. By conducting empirical experiments, we aim to explore the impact that changing the loss function has on a model's behaviour in the specific case of imbalanced datasets.

Analysed loss functions

Table: List of losses analysed in this work. $\mathbf{y}$ is true label as one-hot encoding, $\hat{\mathbf{y}}$ is true label as $+1$/-1 encoding, $\mathbf{o}$ is the output of the last layer of the network, $\cdot^{(j)}$ denotes jth dimension of a given vector, and $\sigma(\cdot)$ denotes probability estimate.

symbol	name	equation
$\mathcal{L}_1 \circ \sigma$	expectation loss	$\Vert\mathbf{y} - \sigma(\mathbf{o})\Vert_1$
$\mathcal{L}_2 \circ \sigma$	regularised expectation loss	$\Vert\mathbf{y} - \sigma(\mathbf{o})\Vert_2^2$
hinge^2	squared hinge (margin) loss	$\Sigma_j \max(0, \frac{1}{2} - \hat{\mathbf{y}}^{(j)}\mathbf{o}^{(j)})^2$
log	log (cross entropy) loss	$-\Sigma_j \mathbf{y}^{(j)} \log \sigma(\mathbf{o})^{(j)}$
$\log^2$	squared log loss	$-\Sigma_j [\mathbf{y}^{(j)} \log \sigma(\mathbf{o})^{(j)}]^2$
D_{CS}	Cauchy-Schwarz Divergence	$-\log \frac{\Sigma_j \sigma(\mathbf{o})^{(j)}\mathbf{y}^{(j)}}{\Vert\sigma(\mathbf{o})\Vert_2 \Vert\mathbf{y}\Vert_2}$
sparsemax	Sparsemax loss [1]	$-\Sigma_{j=1}^K \mathbf{y}^{(j)}\mathbf{o}^{(j)} + \frac{1}{2}\Sigma_{j\in S(\mathbf{o})}((\mathbf{o}^{(j)})^2 r^2(\mathbf{o})) + \frac{1}{2}$

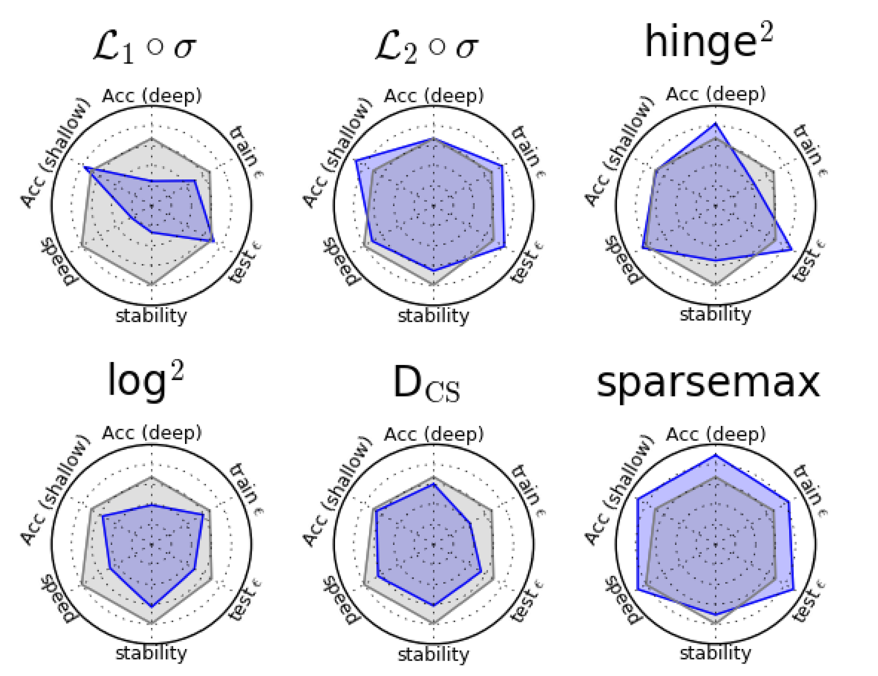


Figure: Analysed loss functions' average performance for balanced classification problems - MNIST anc CIFAR10 datasets.

Experiments

The same classifier is used for the two following datasets. The conducted experiments examine the behaviour of a simple convnet, consisting of 3 layers of convolutions, each of size 5x5 and with respectively 64, 16 and 32 filters, with ReLU activation functions, batch-normalisation and pooling operations in between them (max pooling after first layer and then two average poolings, all 3x3 with stride 2), followed by a single fully connected hidden layer with 128 ReLU neurons, and final layer with 2 neurons.

CelebA

CelebA is a dataset widely by the computer vision community. It contains more than 200000 RGB-images of celebrity faces, labelled with respect to 40 different binary attributes (such as young, smiling, bald), with diverse imbalance class rate between attributes. The images have been resized in a 64x64 format.

Figure: Example lesions from CelebA dataset.

Skin lesions

Available from the International Skin Imaging Collaboration (ISIC) archive, the dataset, obtained from the ISIC 2018 Challenge, contains more than 10000 dermoscopic images of the skin, labelled in 7 different classes : melanoma, melanocityc naevus, basal cell carcinoma, keratinizing tutors (actinic keratosis and intraepithelial carcinoma) benign keratosis (solar lentigo, seborrheic keratosis and lichen planus-like keratosis), dermatofibroma and vascular lesion. For the experiments, all the images have been resized in a 64x64 format.

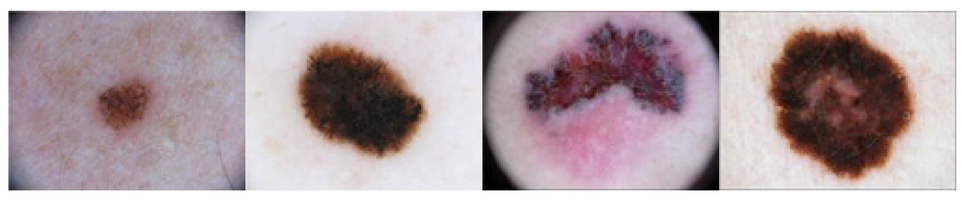


Figure: Example lesions from the ISIC archive.

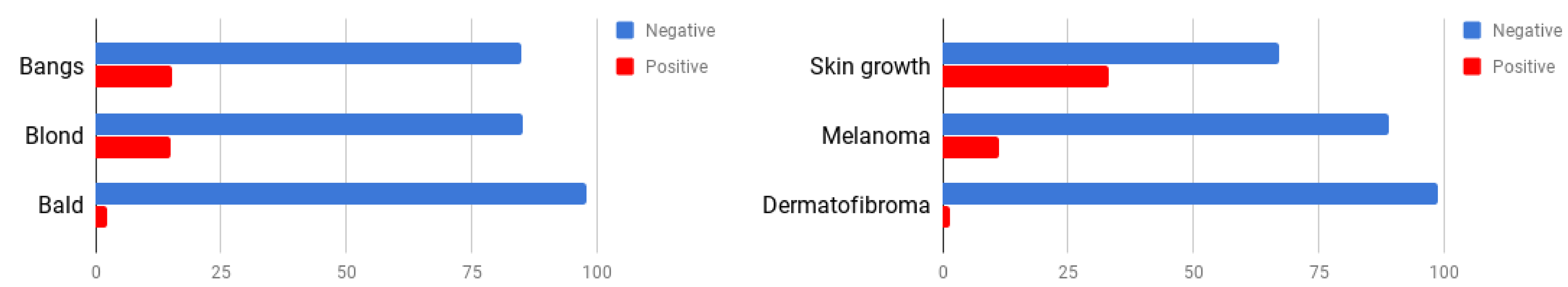
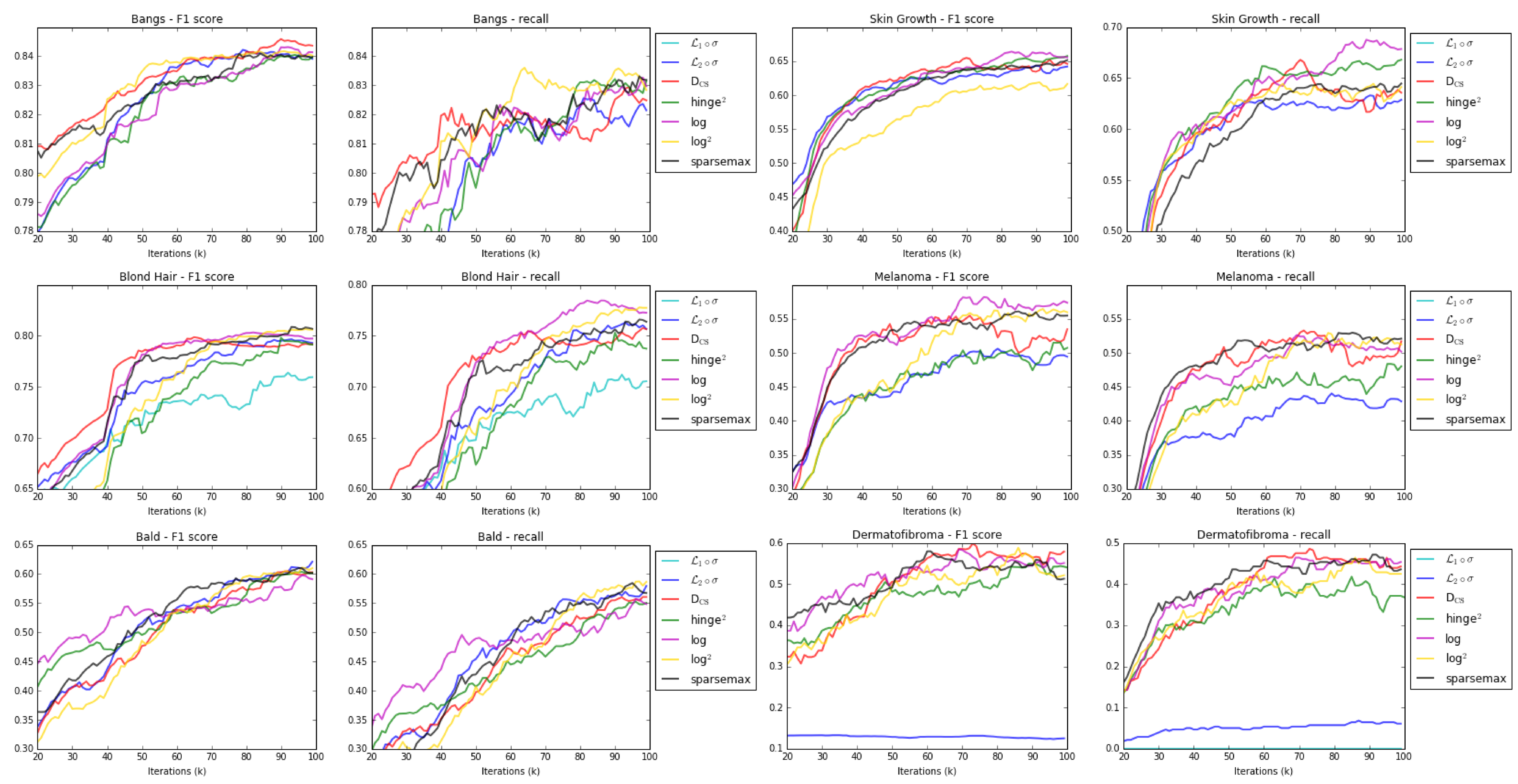


Figure: Class imbalance level. Proportions of images to attributes in both datasets.

Empirical results



Corresponding author:
k.janocha17@imperial.ac.uk

Conclusions

The preliminary results suggest significant differences in performances of analysed loss functions.

- $\mathcal{L}_1 \circ \sigma$ (expectation loss) tends to classify the whole dataset negatively (which results in very high accuracy, but recall equal to 0) which can be easily justified by [2], Proposition 1.
- Similarly, $\mathcal{L}_1 \circ \sigma$ shows the same tendency with growth of imbalance factor, but in smaller degree due to regularization by output norms (see [2], Proposition 1), it is being penalized for overly confident results.
- D_{CS} also seems to outperform log loss with larger imbalance, which suggests the impact of regularization on imbalanced classification might be crucial and is certainly worth investigating.

Further work

While this research stands as an introductionary analysis of the impact of loss functions on imbalanced dataset, it can be improved and extended on in several manners. The impact of applying weights to particular classes should also be investigated. A more extensive model selection and validation processes should be implemented to better understand the correlation between the imbalance level and the classification performance. Finally, studying the activity of neurones within the network could be useful to understand how the choice of a loss function encourages the appearance of specific neural patterns.

References

[1] From softmax to sparsemax: A sparse model of attention and multi-label classification.
Andre Martins and Ramon Fernandez Astudillo.

[2] On Loss Functions for Deep Neural Networks in Classification
Katarzyna Janocha and Wojciech Marian Czarnecki.

[3] Definition of Loss Functions for Learning from Imbalanced Data to Minimize Evaluation Metrics
Juan Miguel Garcia-Gomez and Salvador Tortajada.

[4] The HAM10000 Dataset: A Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions
Philipp Tschandl and Cliff Rosendahl and Harald Kittler.

Imperial College London