Logistic regression:
A guinea pig for deep learning with many classes

TX, 2023–2024

Heudiasyc
UMR7253 CNRS/UTC, Compiègne

To apply, please send an application file containing your CV and latest transcripts to Yves Grandvalet (yves.grandvalet@hds.utc.fr).

Background

Regularization is a classic technique for controlling the complexity of models used in statistics and machine learning. This complexity control is achieved by restricting the search space for solutions. Typically, during the learning phase where the model is fitted to the data, its parameters are constrained to be small. For most standard models, the norm of the parameters controls the variations of the estimated function, and therefore the regularity of this function. In deep learning, regularization is performed explicitly by weight decay, and implicitly by many other means, such as drop-out, and also by data augmentation (DA).

Motivation

In deep learning, it is a common practice to consider classification problems involving several hundred or thousand classes, typically to recognize objects in images or to predict plausible words in a given context. Then, the classifier jointly produces hundreds of scores, each of these scores being a function mapping the object space (image, textual context) to an interval, typically $[0, 1]$. There's no reason to impose identical regularity constraints on each of these scores, except that this solution avoids using as many hyper-parameters as scores.

Recently, Balestriero et al. (2022) showed that "The effects of regularization and data augmentation are class dependent", meaning in particular that the value of the optimal regularization hyper-parameter among classes can differ by several orders of magnitude.[1] The classic weight decay is then set by mak-

[1] More precisely, Balestriero et al. (2022) analyze the validation accuracy of the one-vs-all classification for each class.

ing a compromise between classes that is not optimal for any of the classes. Kirichenko et al. (2024) also show that data augmentation can augment the confusion between visually related classes.

On the basis of these observations, we would like to analyze how much of this behavior is due to the adjustment of the last layer of a deep network, where parameters are class-specific, and how much is due to the representation itself. Considering the upper layers frozen, this analysis applies to the logistic regression model that maps the last internal representation of the deep model to the classes.

Internship objectives

The first step of this study consists in fitting regularized logistic regression models for a series of regularization (or DA) parameter, and measuring:

- the best hyper-parameter value for the global optimization problem,
- the best hyper-parameter value for the recognition of each class,
- the loss associated with the use of the global optimizer.

This part of the work will provide an upper-bound on the improvements that may be achieved by adaptive regularization/DA techniques. These optimum and improvements can be measured in log-likelihood (deviance), in accuracy, or in other measures that focuses on some classes (see, e.g., Bitterwolf et al., 2022).

The second part of the work will analyze the effect of regularization/DA upstream of the last internal representation. To do this, we will use discriminant analysis, on a limited set of ambiguous classes, to highlight the effects of regularization/DA.

Student profile

Scientific rigor, basic knowledge of applied mathematics and computer science; complete autonomy in `python`.

References

Balestriero, R., Bottou, L., and LeCun, Y. (2022). The effects of regularization and data augmentation are class dependent. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 37878–37891.

Bitterwolf, J., Meinke, A., Boreiko, V., and Hein, M. (2022). Classifiers should do well even on their worst classes. In *ICML 2022 Shift Happens Workshop*.

Kirichenko, P., Ibrahim, M., Balestriero, R., Bouchacourt, D., Vedantam, R., Firooz, H., and Wilson, A. G. (2024). Understanding the detrimental class-level effects of data augmentation. *CoRR*, abs/2401.01764.