

## Project Title: SHARPNESS-AWARE MINIMIZATION (SAM)

### 1. Summary and contributions. Briefly summarize the project.

The authors of this paper explore a special model-training algorithm designed to improve the generalization ability of the model. This is necessary to make the model trained on the training set also valid on the test set. The proposed algorithm is to minimize not only the loss function, but also the "sharpness" of the loss function. According to the cited articles, the harshness of the loss landscape (the surface of the loss function values relative to the model parameters) negatively affects the generalization ability. The authors have very clearly described the problem and the way to solve it, as well as explicitly described the work plan. The conducted experiments confirm the performance and efficiency of the proposed algorithm.

### 2. Strengths. Describe all the strengths of the project in enough depth.

1. The authors very clearly describe the problem and its importance. The problem is relevant to a wide class of research and not super-specific. The motivation is described and understood.
2. The theoretical justification is complete, includes graphical schemes, mathematical calculations and all required explanations concerning the choice of solution methods.
3. The logic of the report corresponds to the template of scientific articles. The narrative is clear and consistent.
4. For experiments we have chosen large enough models to solve real problems (not synthetic) on big data. Common and representative datasets Cifar-10 and Cifar-100 are chosen as datasets. A comparison of several models is given.
5. A comprehensive analysis of the results is given to make sure that the method considered by the authors is workable and effective. The results are presented accurately and easily interpreted.

### 3. Weaknesses. Explain all the limitations of this project in enough depth.

In my humble opinion, there are a couple of weaknesses in the work, but relative to the overall quality of the work they are pretty unimportant:

1. It is not said what kind of loss is used for training. Although it quickly became clear from the repository, it was not explicitly mentioned in the report. The same goes for the accuracy quality score function.
2. The training graphs are not very interpretable. The notations are weakly legible. The learning curves of the model with and without the proposed algorithm are shown on different graphs and with different axes, which does not allow for an explicit comparison.

### 4. Correctness. Are the claims and method correct? Is the empirical methodology correct?

In this paper, the authors do not propose their own method, but rely on a method from another article. The theoretical calculations repeat the calculations from the cited article and are not the personal contribution of this paper's author. Therefore, evaluation of the correctness of mathematical formulas does not relate to the authors of this paper. Nevertheless, the authors' contribution implies the implementation of a particular method. The authors themselves write about the possibility of some errors related to the code or the learning process, citing the lack of time and the difficulty of working with large models. Though, given that the authors managed to successfully implement the method, run it and demonstrate its performance (at least within a few experiments on real data with a real model), the work can be considered correct.

### 5. Clarity. Is the project report well written?

I understood everything in the report from the first time. Except that the mathematical calculations required clarification, which I found in the cited article. Perhaps one should have expanded the theoretical section. Nevertheless, the entire progress of the paper remains very clear and consistent.

### 6. Related work. Is it clearly discussed?

During the project, the authors refer to 7 related works. The related works section clearly describes three works on this problem. The first cited paper has over 70 citations. Overall, for a project that does not claim to be an article, I consider the literature review to be sufficient.

### 7. Reproducibility. Are there enough details to reproduce the major results of this work?

Considering that I easily found answers to some of my questions in the authors' repository, I think the repository is good. Functions are broken down into modules, data, code and experiments are separated. The code is clean and clear even without comments. Reproduction, in my estimation, would not take long.

### 8. Overall score. You should NOT assume that you were assigned a representative sample of projects. The "Overall Score" for each project should reflect your assessment of the project.

Since the novelty of this work does not consist in a fundamentally new method, but only exploits a composition of existing methods, I am not sure that the work is 100% consistent with this statement:

(1) Truly groundbreaking work. Definitely maximal grade (A).

However, the work is, in my opinion, very good, and one certainly deserves an A grade. Therefore, the result is this:

(2) A very good submission; deserves high grade, tending to maximal (A).

## 9. Confidence score.

Considering that I personally have not researched and studied the subject in depth I cannot give a grade of (5). I admit that there are some aspects that I might have missed. Therefore, a (4):

(4) You are **confident** in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of the topic.