Project Description

MATH 540 Statistical Learning

October 28, 2021

The Project for this class is the analysis of a model/method of your own choosing. The goal is for you to demonstrate proficiency in the techniques we have covered in this class (and beyond, if you like) and apply them to the analysis of ML models/algorithms.

Research question. Start with a meaningful research question[1]. Such questions usually arise when you look critically at the existing models/methods.

As an example, let us consider the question asked by one of your classmates: *Can we use element-wise sigmoid instead of softmax in multi-label classification?* More precisely, consider one example (x, y) from a training sample, where x is an instance (such as persimmon), and y is the true label (such as tasty/mediocre/non-tasty). Let $\mathbf{z} \in \mathbb{R}^K$ be the vector of logits for this instance, and K be the number of classes. The standard approach is to apply softmax normalization:

$$\operatorname{softmax}_i(\mathbf{z}) = \frac{e^{z_i}}{\sum_j e^{z_j}}, \quad i \in \{1, \dots, K\}$$

and then feed this into the cross-entropy loss:

$$\ell(\mathbf{z}, y) = -\ln \operatorname{softmax}_y(\mathbf{z}) = -z_y + \ln \sum_j e^{z_j}. \tag{1}$$

The nice thing about the loss (1) is that its gradient w.r.t. $\mathbf{z}$ is (check this)

$$\nabla_{\mathbf{z}} \ell = \operatorname{softmax}(\mathbf{z}) - \mathbf{e}_y,$$

where $\mathbf{e}_y \in \mathbb{R}^K$ is the one-hot representation of y, i.e. it is a vector whose components are all 0s except the y-th component which is equal to 1. Hence the gradient based optimization moves logits (and thus all the parameters) in the direction where $\operatorname{softmax}(\mathbf{z}) \to \mathbf{e}_y$.

However, your classmate wonders whether we can use the element-wise sigmoid

$$\sigma_i(\mathbf{z}) = \frac{e^{z_i}}{1 + e^{z_i}}, \quad i \in \{1, \dots, K\}$$

instead of softmax and then proceed as before (i.e. feed this into cross-entropy loss and perform gradient-based optimization)

Theoretical analysis. It is highly recommended (although not obligatory) to perform a theoretical analysis/evaluation of your idea. For the example above, a natural thing to do would be a calculation of the gradient of the loss function and then checking whether it is $\sigma(\mathbf{z}) - \mathbf{e}_y$. If yes, then one may expect that the gradient-based optimization will force $\sigma(\mathbf{z}) \to \mathbf{e}_y$. If not, what can be done for this to happen?

If you struggle with mathematical analysis of your idea, provide some intuition for *why* your suggestion may work.

[1] `https://en.wikipedia.org/wiki/Research_question`

Empirical evaluation. Once you have a theoretical justification for your idea, you will need to perform experiments on real data. For this, I expect that you will do the following.

1. Get the baseline result. For the considered example, your classmate can take an off-the-shelf softmax regression from scikit-learn and evaluate it on standard datasets.
2. Reproduce the baseline result from scratch, i.e. rewrite the model and its training in pure Python/Pytorch/Keras/etc. Make sure you get the same result as in item 1.
3. Implement the modification that you suggested as your research question, and perform evaluation against the baseline. I would like to see not only final performance metrics but also the training dynamics.
4. Perform additional analyzes that explain *why* your suggested modification improves/does not improve the baseline.

Report. Prepare a *concise* report on your findings as if you were submitting a student abstract to AAAI 2022. Formatting instructions and LaTeX/Word templates are available at the official website.[2] Notice that the report shall not exceed *two* pages.

Source code. Create a GitHub account[3] (if you don't have it yet), put your source code as a separate repository, don't forget to put a well-written README and a license in your repository. Indicate the link to the repo in your report.

Example reports. `https://ojs.aaai.org/index.php/AAAI/issue/view/402#:~:text=PDF-,AAAI,-Student%20Abstract%20and`.

[2] `https://aaai.org/Conferences/AAAI-22/student-abstract-and-poster-program/`
[3] https://github.com/