Date ___/___/___

Topic: Model soups: averaging weights of multiple fine tuned models improves accuracy without increasing inference time

13/12 June 23

Abstract:

The conventional recipe for maximizing model accuracy is to 1) train multiple models with various hyperparameters and 2) pick the individual model which performs best on a held-out validation set, discarding the remainder. In this paper, we revisit, the second step of this procedure in the context of fine tuning large pre trained models, where fine tuned models often appear to lie in a single low error basin. We show that averaging the weights of multiple models fine tuned with ~~fine~~ different hyperparameter configuration often improves accuracy and robustness. Unlike a conventional ensemble, we may average many models without incurring any additional inference or memory costs. we call the results "model soups". When fine tuning large pre trained models such as CLIP, ALIGN and ViT pre trained on JFT, our soup recipe provides significant improvements over the best model in a hyperparameter sweep

on Imagenet. The resulting ViT-G model, which attains 90.94% top-1 accuracy on Imagenet, achieved a new state of the art. Furthermore, we show that the model soup approach extends to multiple image classification and natural language processing tasks, improves out of the distribution performance and improves zero shot performance on new downstream tasks. Finally, we analytically relate the performance similarity of weight averaging and logit ensembling to flatness of the loss and confidence of the predictions.

Conclusion:

Our results challenge the conventional procedure of selecting the best model on the held-out validation set when fine tuning. With no extra compute during inference, we are often able to produce a better model of averaging the weights of multiple fine tuned solutions.