

HWA: AVERAGING HYPERPARAMETERS IN BAYESIAN NEURAL NETWORKS LEADS TO BETTER GENERALIZATION.

Anonymous authors

Paper under double-blind review

ABSTRACT

T.B.C

1 INTRODUCTION

Our main contributions read as follows:

- ff
- ff

The remaining of the paper is organized as follows.

2 RELATED WORK

Stochastic Averaging:

Variational Inference:

Posterior Prediction:

3 HYPERPARAMETERS AVERAGING IN BAYESIAN NEURAL NETWORKS

Algorithm 1 HWA: Hyperparameters Weight Averaging

```
1: Input: Trained hyperparameters  $\hat{\mu}_\ell$  and  $\hat{\sigma}$ . LR bounds  $\gamma_1$  and  $\gamma_2$ . Cycle length  $c$ .  
2: Initialize the hyperparameters of the weights and  $\mu_\ell = \hat{\mu}_\ell$  and  $\mu_\ell^{HWA} = \mu_\ell$ .  
3: for  $k = 0, 1, \dots$  do  
4:  $\gamma \leftarrow \gamma(k)$  (Cyclical LR for the iteration)  
5:  $\mu_\ell^{k+1} \leftarrow \mu_\ell^k - \gamma \nabla \mathcal{L}(\mu_\ell^k)$  (regular SVI update)  
6: if  $\text{mod}(k, c) = 0$  then  
7:  $n_{\text{models}} \leftarrow k/c$  (Number of models to average)  
8:  $\mu_\ell^{HWA} \leftarrow \frac{n_{\text{models}} \mu_\ell^{HWA} + \mu_\ell^{k+1}}{n_{\text{models}} + 1}$   
9: end if  
10: end for
```

4 NUMERICAL EXPERIMENTS

5 CONCLUSION

REFERENCES

Joshua V. Dillon, Ian Langmore, Dustin Tran, Eugene Brevdo, Srinivas Vasudevan, Dave Moore, Brian Patton, Alex Alemi, Matthew D. Hoffman, and Rif A. Saurous. Tensorflow distributions. *CoRR*, abs/1711.10604, 2017. URL <http://arxiv.org/abs/1711.10604>.

A APPENDIX

You may include other additional sections here.