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

Anonymous authors
Paper under double-blind review

ABSTRACT

T.B.C

1 INTRODUCTION

2 METHOD

Algorithm 1 HWA: Hyperparameters Weight Averaging

Input: Trained hyperparameters $\hat{\mu}_\ell$ and $\hat{\sigma}$. LR bounds γ_1 and γ_2. Cycle length c.
Initialize the hyperparameters of the weights and $\mu_\ell = \hat{\mu}_\ell$ and $\mu_\ell^{HWA} = \mu_\ell$.
for $k = 0, 1, ...$ **do**
  $\gamma \leftarrow \gamma(k)$ (Cyclical LR for the iteration)
  $\mu_\ell^{k+1} \leftarrow \mu_\ell^k - \gamma \nabla \mathcal{L}(\mu_\ell^k)$
  if $\text{mod}(k, c) = 0$ **then**
    $n_{\text{models}} \leftarrow k/c$
    $\mu_\ell^{HWA} \leftarrow \frac{n_{\text{models}} \mu_\ell^{HWA} + \mu_\ell^{k+1}}{n_{\text{models}} + 1}$
  end if
end for

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.