
HWA: Averaging Hyperparameters in Bayesian Neural Networks Leads to Better Generalization.

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 To be completed

2 **1 Introduction**

3 **2 Method**

Algorithm 1 HWA: Hyperparameters Weight Averaging

1: **Input:** Trained hyperparameters $\hat{\mu}_\ell$ and $\hat{\sigma}$. LR bounds γ_1 and γ_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: $\mu_\ell^{k+1} \leftarrow \mu_\ell^k - \gamma \nabla \mathcal{L}(\mu_\ell^k)$
6: **if** $\text{mod}(k, c) = 0$ **then**
7: $n_{\text{models}} \leftarrow k/c$
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 **3 Numerical Examples**