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

Anonymous Author(s)
Affiliation
Address
email

Abstract

To be completed

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

Submitted to 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Do not distribute.