Supplementary Material for "Latent Weights Do Not Exist: Rethinking Binarized Neural Network Optimization"

Anonymous Author(s)
Affiliation
Address
email

Theorem 1. *The binary weight vector generated by Algorithm 1 is invariant under scaling of the learning rate, α, provided the initial conditions are scaled accordingly and the pseudo-gradient, Φ, does not depend on $|\tilde{w}|$.*

Proof. Consider a single weight. Let $\tilde{w}_t$ be the latent weight at time t, g_t the pseudo-gradient and δ_t the update step generated by the optimizer $\mathcal{A}$. Then:

$$\tilde{w}_{t+1} = \tilde{w}_t + \alpha\delta_t.$$

Now take some positive scalar C by which we scale the learning rate. Replace the weight by $\tilde{v}_t = C\tilde{w}_t$. Since $\text{sign}(\tilde{v}_t) = \text{sign}(\tilde{w}_t)$, the binary weight is unaffected. Therefore the forward pass at time t is unchanged and we obtain an identical pseudo-gradient g_t and update step δ_t. We see:

$$\tilde{v}_{t+1} = \tilde{v}_t + C\alpha\delta_t = C \cdot (\tilde{w}_t + \alpha\delta_t) = C\tilde{w}_{t+1}.$$

Thus $\text{sign}(\tilde{v}_{t+1}) = \text{sign}(\tilde{w}_{t+1})$. By induction, this holds for $\forall t' > t$ and we conclude the BNN is unaffected by the change in learning rate. □

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.