

9 Future directions

IGR, although a great tool for examining generalization, relies on low-order Taylor approximations that works well under small learning rates. In addition, the current analysis is based on fixed values of β while letting $h \rightarrow 0$. In practice, to better guide the choice of hyper-parameters, a bound that is asymptotic in both h and β ($h \rightarrow 0, \beta \rightarrow 1$) might be more helpful. We leave it as future work.