Cyclical Learning Rates for Fast Ensemble Model

Anonymous EMNLP submission

Abstract

Cyclical learning rates, faster ensemble, (stochastic) MCMC sample, second optimization(whole sequence), combine with adaptive LR, LR schedule(exp or Tmul), interpolation(multi points), noise

1 Introduction

2 Background

3 Cyclical learning rates for faster ensemble model

3.1 baseline: several methods

first: find best LR
baseline:

- SGD decay
- exp decay

cyclical learning rates:

- high2low
- cosine, warm start ...

3.2 interpolation

3.3 schedule

3.4 combine with adaptive

3.5 MCMC and noise

3.6 connection with ...

4 Experiments

4.1 figures to reproduce

- best learning rate
- combine with adaptive learning rate (BN? dropout?)
- cyclical learning rates AND test accuracy
- cyclical learning rates VERSUS fix learning rate
- decay schedule: no decay, fix decay, exp decay VERSUS trajectory ensemble
- distance to prove robustness?
- iterate averaging as approximate MCMC
- CLR, BS, cyclical momentum , weight decay?

4.2 visualization

5 Related works

6 Discussion

7 Conclusion

A Supplemental Material

supplement ...