

Peer Review – Group2_CA3

The code is clear and well-structure, with clear explanation.

1) Solve this optimization task with proper choices of size of decision variables using GD, perturbed GD, SGD, SVRG, and block coordinate descent.

- GD, SGD, SVRG are implemented correctly, while perturbed GD and block coordinate descent methods are missing here.
- Is there a motivation of the choice of hyper-parameters? E.g. why choosing a 5-layer model?

2) Compare these solvers in terms complexity of hyper-parameter tuning, convergence time, convergence rate (in terms of # outer-loop iterations), and memory requirement.

- In general, GD, SGD, SVRG solvers are compared in terms of accuracy and training time.
- Memory requirement comparison is missing.
- It could be better with more explanation, e.g., why does SVRG have a longest training time?