

## CA3 Review

**from SG 1**

As far as we understand the plot of the loss, taking GD on page 11 for example, it is supposed to compare the different hyperparameters, however, with any legend to show which curve is corresponding to which set of hyperparameters, it is hard to draw much conclusions from the plotting. This applies to all the plotting of the loss and time consuming.

Besides that, the loss itself has an extreme large value, while during the implementation of cost function, the number of samples  $N$  is divided, thus the loss when converged should not be a value like 256276. (since  $y_{\text{train}}$  is either 1 or -1). This indicates that there might be something wrong in the implementation of feed\_forward\_propagation and gradient calculation.

As for the convergence time, we are not sure if plotting time at every iteration is a good metric, one would argue that the definition of convergence time perhaps is good enough.

In addition to what is mentioned at the beginning, the implementation of PGD and BCD is missing, the report itself seems not to be complete, due to lack of enough analysis and discussion about the experiment result.

In general, the implementation might need some improvement, and the report would be better if more comparison and analysis can be added and the conclusions can be written down.