

Result Notes:

Eta: the learning rate

H: the number of hidden units

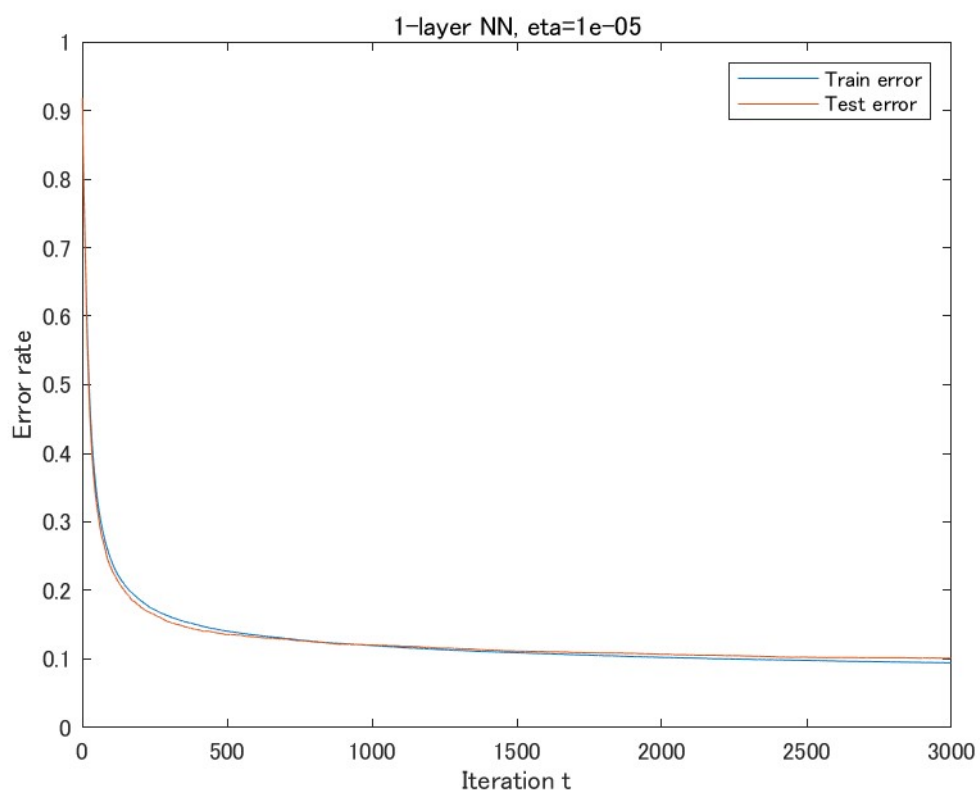
Lambda: the weight decay rate

(Case 1)

1-layer NN, eta=1e-05

The final train error rate is 0.094583.

The final test error rate is 0.099596.



(Case 2)

2-layer NN, sigmoid, $\eta=1e-05$, $H=10$

The final train error rate is 0.1189.

The final test error rate is 0.1209.

2-layer NN, sigmoid, $\eta=1e-05$, $H=20$

The final train error rate is 0.099083.

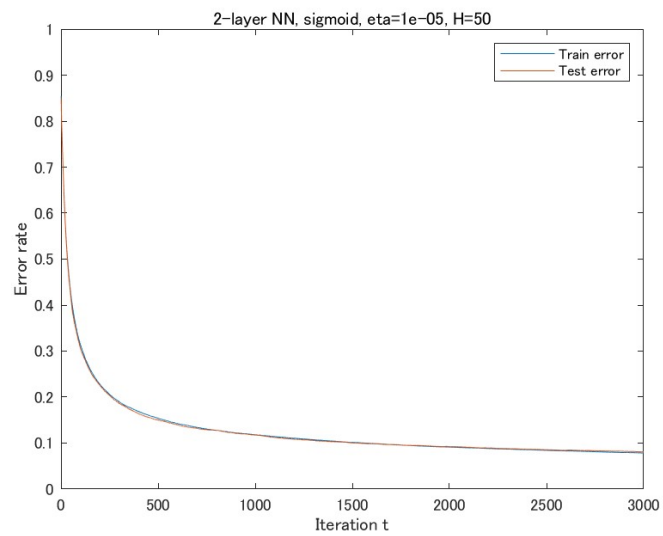
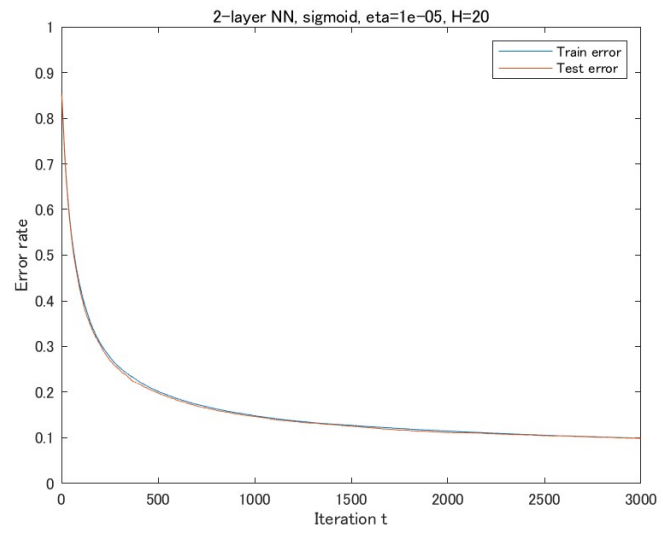
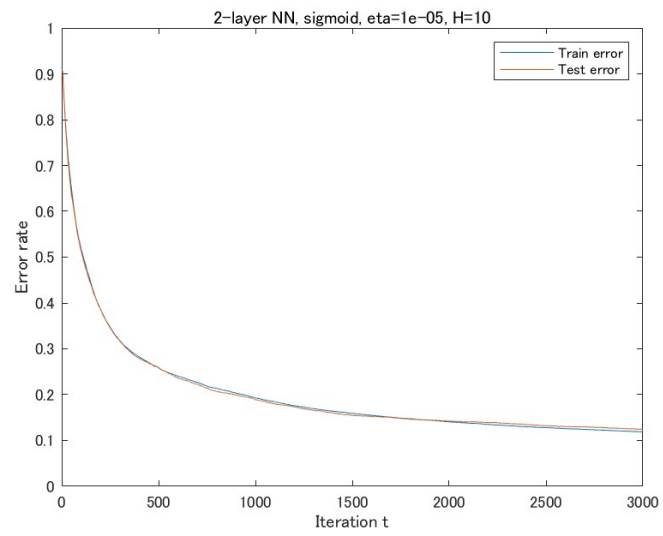
The final test error rate is 0.099373.

2-layer NN, sigmoid, $\eta=1e-05$, $H=50$

The final train error rate is 0.07845.

The final test error rate is 0.07855.

The larger H is, the lower the error rate is. This holds both for evaluation with training data and that with test data.



(Case 3)

2-layer NN, sigmoid, $\eta=2e-06$, $H=10$, $\lambda=0.001$

The final train error rate is 0.15502.

The final test error rate is 0.15778.

2-layer NN, sigmoid, $\eta=2e-06$, $H=20$, $\lambda=0.001$

The final train error rate is 0.12795.

The final test error rate is 0.12907.

2-layer NN, sigmoid, $\eta=2e-06$, $H=50$, $\lambda=0.001$

The final train error rate is 0.12508.

The final test error rate is 0.12900.

2-layer NN, ReLu, $\eta=2e-06$, $H=10$, $\lambda=0.001$

The final train error rate is 0.094617.

The final test error rate is 0.094617.

2-layer NN, ReLu, $\eta=2e-06$, $H=20$, $\lambda=0.001$

The final train error rate is 0.08655.

The final test error rate is 0.08885.

2-layer NN, ReLu, $\eta=2e-06$, $H=50$, $\lambda=0.001$

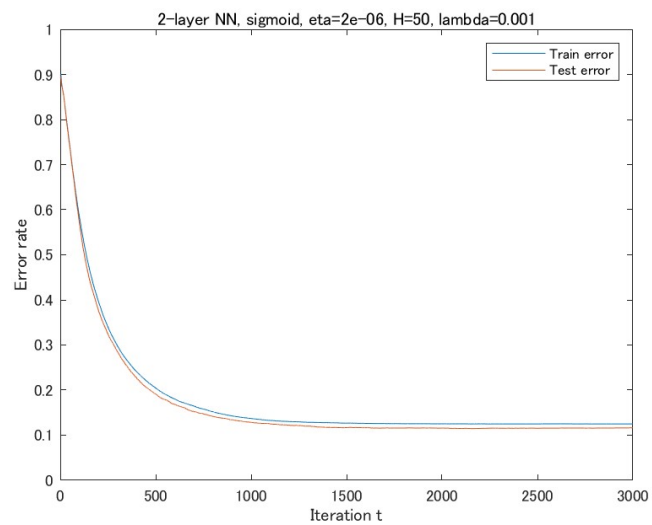
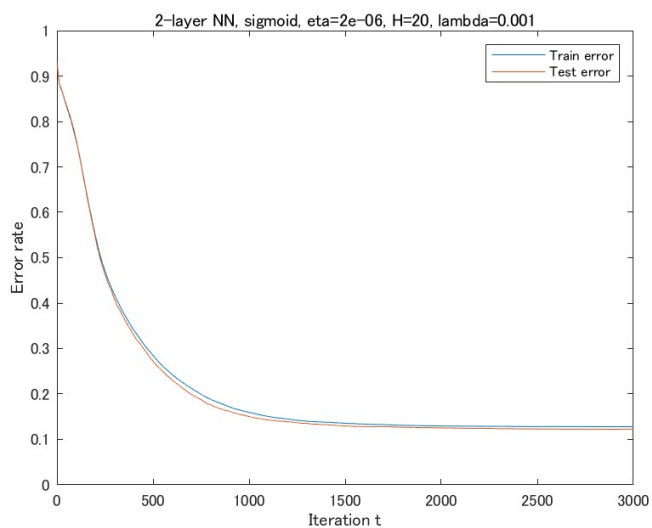
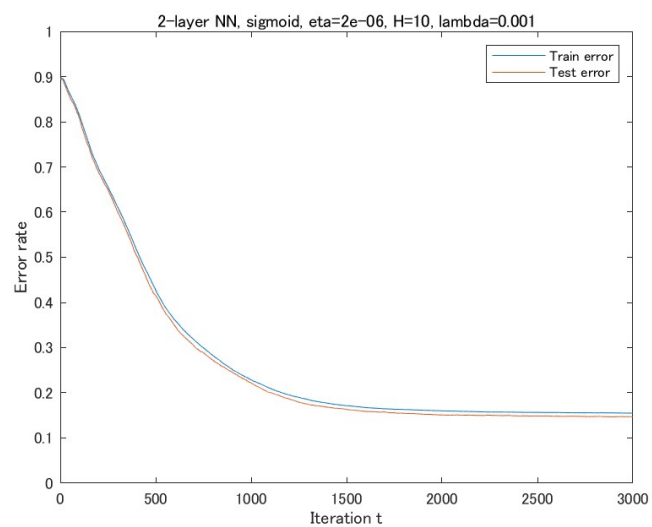
The final train error rate is 0.08105.

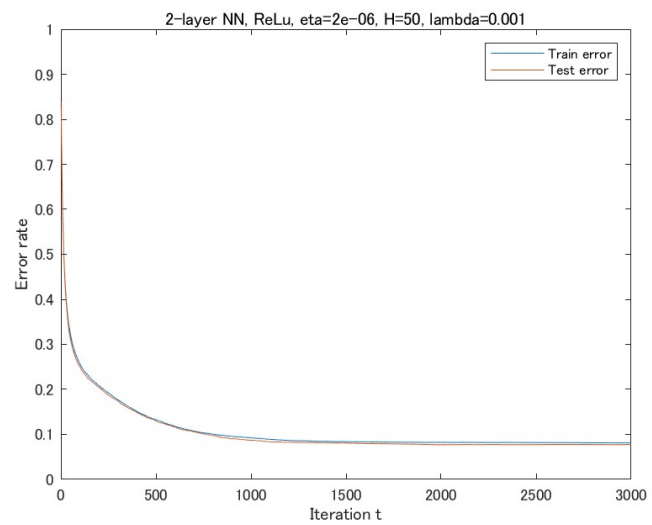
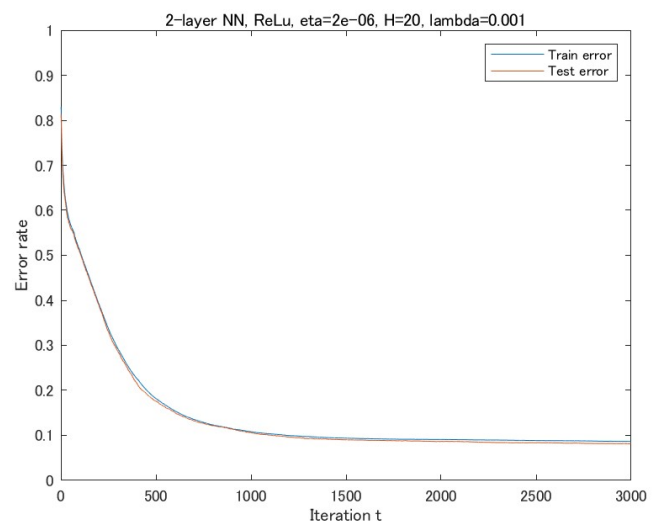
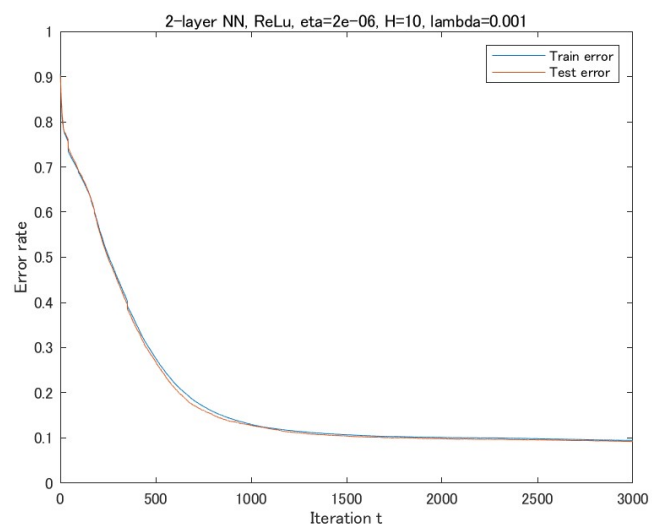
The final test error rate is 0.08235.

The error rates of ReLu cases are lower than those of Sigmoid cases with the same H . This holds both for evaluation with training data and that with test data.

Also, for both Sigmoid and ReLu cases, the larger H is, the lower the error rate is. This holds both for evaluation with training data and that with test data.

About the convergence speeds, the larger H is, the faster the convergence is. This holds both for Sigmoid and ReLu cases. The Models in ReLu cases converges faster than those in Sigmoid cases.





(Case 4)

2-layer NN, sigmoid, $\eta=2e-06$, $H=10$, $\lambda=0.0001$

The final train error rate is 0.17173.

The final test error rate is 0.17973.

2-layer NN, sigmoid, $\eta=2e-06$, $H=20$, $\lambda=0.0001$

The final train error rate is 0.13837.

The final test error rate is 0.14036.

2-layer NN, sigmoid, $\eta=2e-06$, $H=50$, $\lambda=0.0001$

The final train error rate is 0.12268.

The final test error rate is 0.12778.

2-layer NN, ReLu, $\eta=2e-06$, $H=10$, $\lambda=0.0001$

The final train error rate is 0.12877.

The final test error rate is 0.12997.

2-layer NN, ReLu, $\eta=2e-06$, $H=20$, $\lambda=0.0001$

The final train error rate is 0.119.

The final test error rate is 0.12033.

2-layer NN, ReLu, $\eta=2e-06$, $H=50$, $\lambda=0.0001$

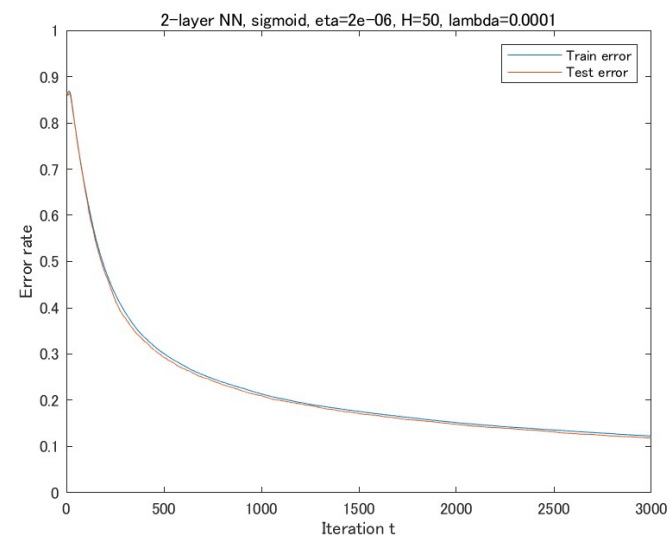
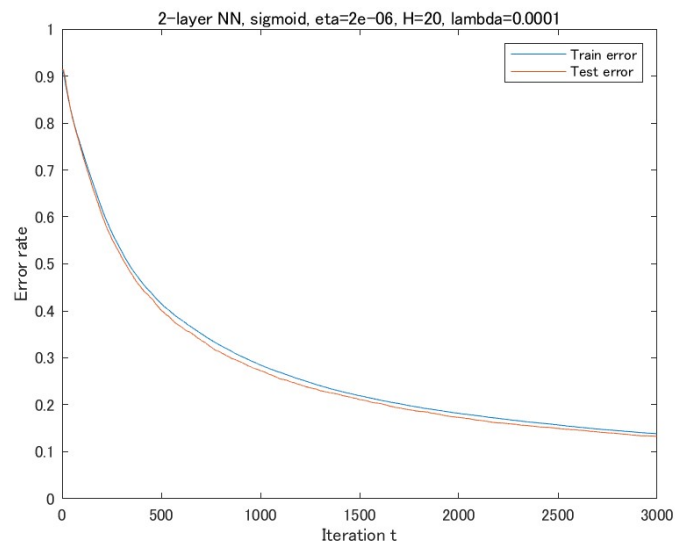
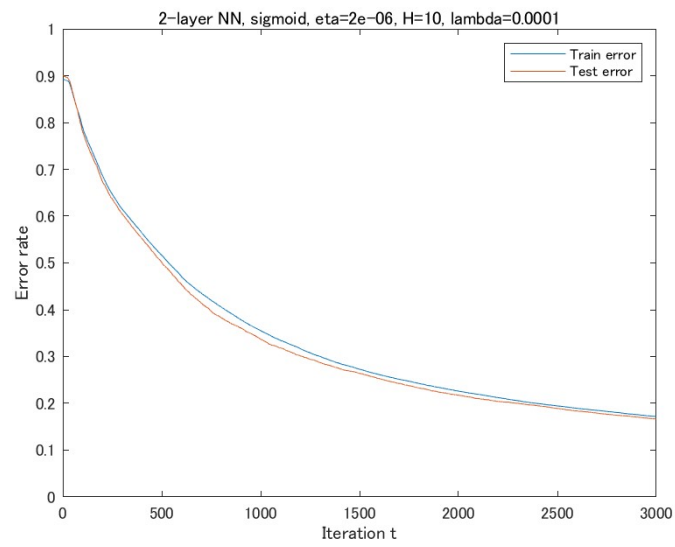
The final train error rate is 0.091583.

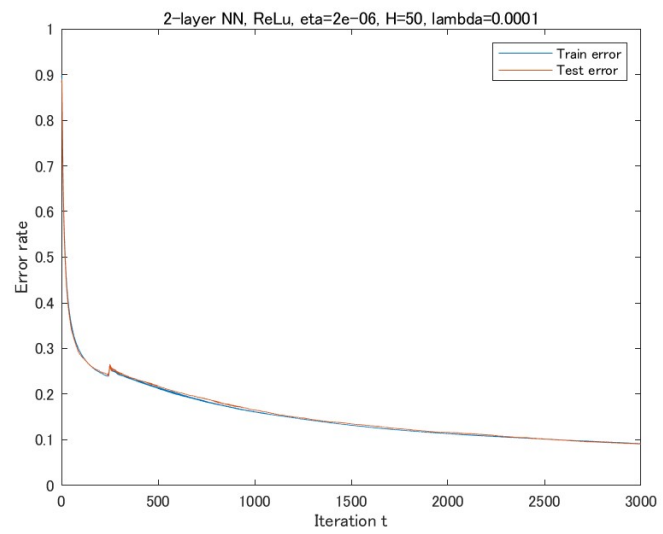
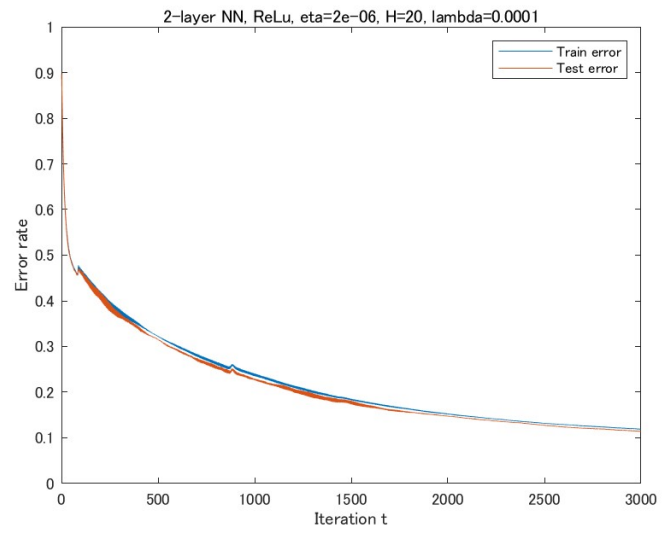
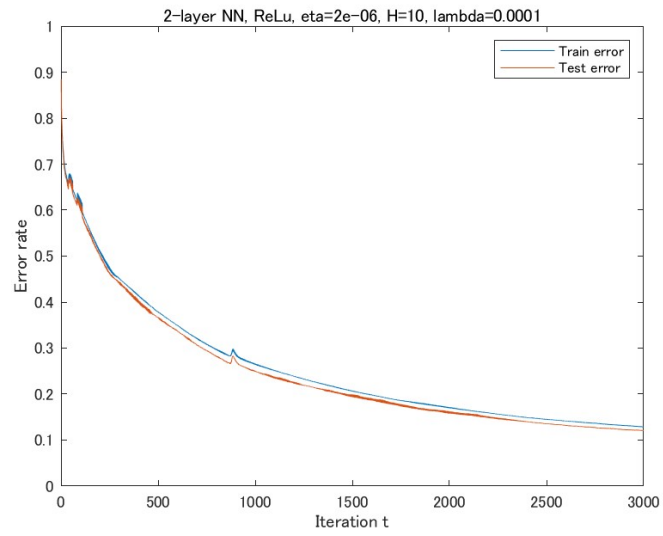
The final test error rate is 0.091583.

The regulation with large λ can improve the performance.

In all cases with small λ , the convergence speeds are slower than those in large λ cases. Also, the error rates in small λ cases are higher than those in large λ cases.

In ReLu cases, the error rates are not stable.





(Case 5)

1-layer NN, SGD, $\eta=1e-05$

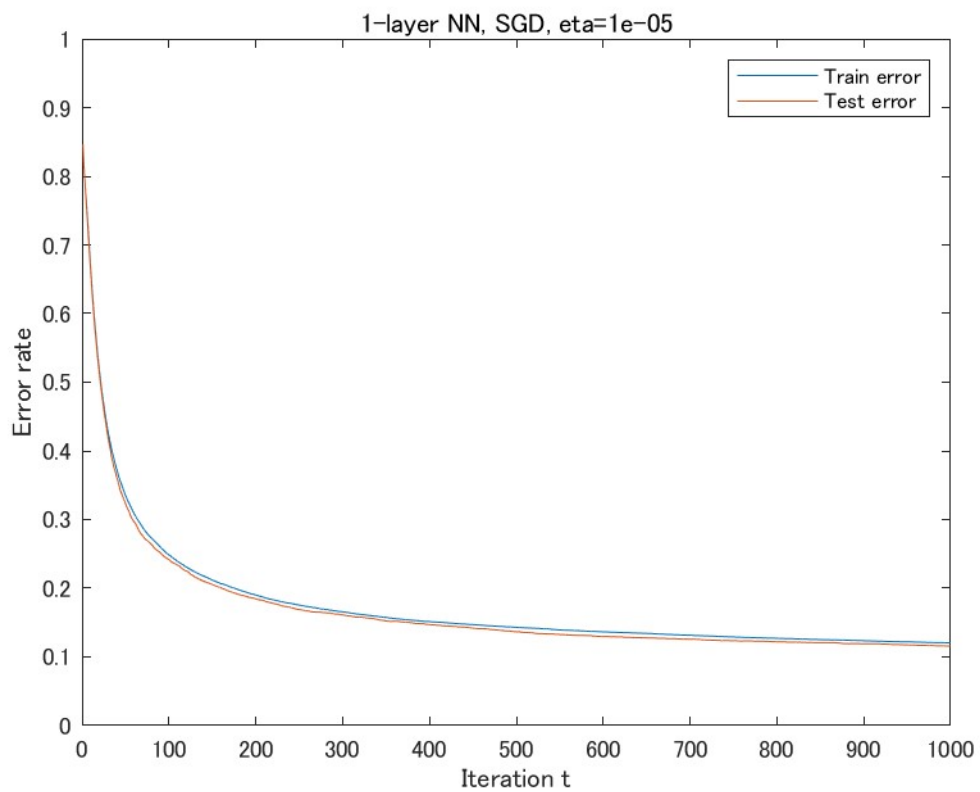
The final train error rate is 0.12055.

The minimum train error rate is 0.12055 at the 1000th step.

The final test error rate is 0.12635.

The minimum test error rate is 0.12635 at the 1000th step.

The error rate is slightly higher than that in the batch-approach case.



(Case 6)

2-layer NN, SGD, sigmoid, $\eta=0.01$, $H=10$

The final train error rate is 0.20203.

The minimum train error rate is 0.14618 at the 9th step.

The final test error rate is 0.2244.

The minimum test error rate is 0.1433 at the 7th step.

2-layer NN, SGD, sigmoid, $\eta=0.01$, $H=20$

The final train error rate is 0.051.

The minimum train error rate is 0.051 at the 1000th step.

The final test error rate is 0.0909.

The minimum test error rate is 0.0793 at the 61th step.

2-layer NN, SGD, sigmoid, $\eta=0.01$, $H=50$

The final train error rate is $1.6667e-05$.

The minimum train error rate is $1.6667e-05$ at the 721th step.

The final test error rate is 0.0503.

The minimum test error rate is 0.0486 at the 581th step.

2-layer NN, SGD, ReLu, $\eta=0.002$, $H=10$

The final train error rate is 0.88657.

The final test error rate is 0.88531.

2-layer NN, SGD, ReLu, $\eta=0.002$, $H=20$

The final train error rate is 0.88763.

The final test error rate is 0.88657.

2-layer NN, SGD, ReLu, $\eta=0.002$, $H=50$

The final train error rate is 0.34518.

The final test error rate is 0.39223.

In the Sigmoid cases, the convergence speeds of SGD approach are faster than those of batch-approach. The error rate is the lowest in the case of $H=50$. The error rate in the case of $H=10$ increases after some steps.

In the ReLu cases, the algorithm does not work because it does not have weight-decay.

