Problem 4.

μ	$f(x^*)$	x^*			
0	-0.12	1	0	0	0
0.1	-0.104694	$6.38 * 10^{-1}$	$3.61 * 10^{-1}$	$2.76 * 10^{-9}$	$2.02 * 10^{-10}$
1.0	-0.0715251	$2.36 * 10^{-1}$	$3.97 * 10^{-1}$	$3.66 * 10^{-1}$	$1.06 * 10^{-9}$
2.0	-0.0524063	0.16754123	0.31296852	0.35757121	0.16191904
5.0	-0.0232699	0.09715142	0.19490255	0.24962519	0.45832084
10.0	0.0126792	0.07368816	0.15554723	0.21364318	0.55712144

Problem 5. (a) logistic: SVRG loss outperforms SGD a sometimes.
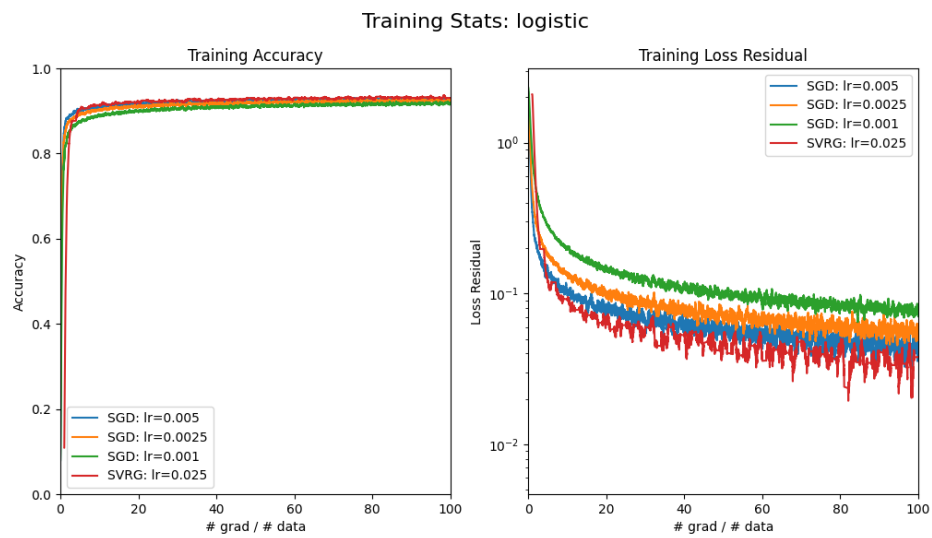
Problem 5. (b) one-layer NN: SVRG outperforms SGD (in loss residual) more than pervious experiment.
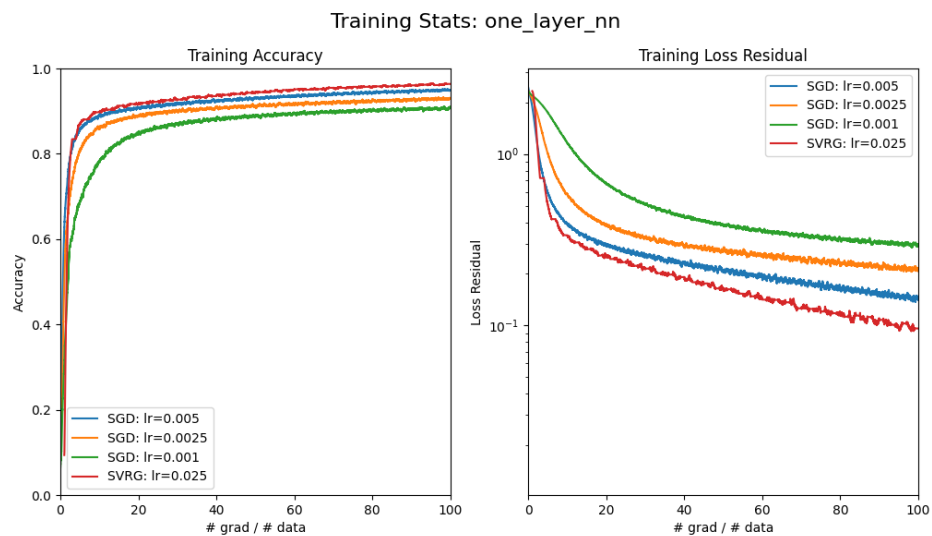
Hyperparameters:

Batch size: 64
Learning rate: SGD = [0.005, 0.0025, 0.001], SVRG = 0.025, GD = 0.05
Iterations: 100

Logistic primal optimal (loss=0.21524): GD for 1000 iterations
NN primal optimal (loss=0.034793): SGD with lr = 0.005 for 1000 iterations

NN hidden layer: 128 nodes