

Table 1: Performance Results for NeuralNetwork\100samplesn1ayersn5n10n3Problem

Optimizer	Mean Final Value	Std Dev	Best Value	Worst Value	Mean Func Evals	Success Rate (%)	Mean Time (s)
Adam-Fast	$1.38 \times 10^{-1}$	$1.31 \times 10^{-2}$	$1.17 \times 10^{-1}$	$1.62 \times 10^{-1}$	232.6	40.0	0.192
QQN-Bisection-2	$1.42 \times 10^{-1}$	$9.94 \times 10^{-3}$	$1.20 \times 10^{-1}$	$1.58 \times 10^{-1}$	550.4	0.0	0.415
QQN-StrongWolfe	$1.47 \times 10^{-1}$	$1.26 \times 10^{-2}$	$1.26 \times 10^{-1}$	$1.65 \times 10^{-1}$	480.2	0.0	0.444
QQN-Backtracking	$1.49 \times 10^{-1}$	$8.28 \times 10^{-3}$	$1.31 \times 10^{-1}$	$1.64 \times 10^{-1}$	544.5	0.0	0.425
QQN-Bisection-1	$1.50 \times 10^{-1}$	$9.70 \times 10^{-3}$	$1.21 \times 10^{-1}$	$1.69 \times 10^{-1}$	496.8	0.0	0.428
L-BFGS-Conservative	$1.61 \times 10^{-1}$	$1.18 \times 10^{-2}$	$1.37 \times 10^{-1}$	$1.86 \times 10^{-1}$	575.4	0.0	0.398
QQN-MoreThuente	$1.63 \times 10^{-1}$	$2.33 \times 10^{-2}$	$1.20 \times 10^{-1}$	$1.95 \times 10^{-1}$	494.9	0.0	0.430
L-BFGS-Aggressive	$1.72 \times 10^{-1}$	$6.36 \times 10^{-3}$	$1.58 \times 10^{-1}$	$1.87 \times 10^{-1}$	696.1	0.0	0.355
QQN-GoldenSection	$1.73 \times 10^{-1}$	$6.18 \times 10^{-3}$	$1.57 \times 10^{-1}$	$1.93 \times 10^{-1}$	901.7	0.0	0.299
Adam-WeightDecay	$1.75 \times 10^{-1}$	$6.35 \times 10^{-3}$	$1.65 \times 10^{-1}$	$1.88 \times 10^{-1}$	502.0	0.0	0.419
L-BFGS	$1.91 \times 10^{-1}$	$4.92 \times 10^{-2}$	$1.52 \times 10^{-1}$	$3.49 \times 10^{-1}$	466.6	0.0	0.296
QQN-CubicQuadraticInterpolation	$1.94 \times 10^{-1}$	$4.26 \times 10^{-3}$	$1.87 \times 10^{-1}$	$2.06 \times 10^{-1}$	394.3	0.0	0.490
Adam	$2.02 \times 10^{-1}$	$4.89 \times 10^{-3}$	$1.93 \times 10^{-1}$	$2.12 \times 10^{-1}$	502.0	0.0	0.416
Adam-AMSGrad	$2.04 \times 10^{-1}$	$6.89 \times 10^{-3}$	$1.91 \times 10^{-1}$	$2.17 \times 10^{-1}$	502.0	0.0	0.420
GD-WeightDecay	$2.09 \times 10^{-1}$	$2.47 \times 10^{-3}$	$2.05 \times 10^{-1}$	$2.16 \times 10^{-1}$	335.0	0.0	0.481
GD	$2.16 \times 10^{-1}$	$2.89 \times 10^{-3}$	$2.08 \times 10^{-1}$	$2.21 \times 10^{-1}$	335.0	0.0	0.471
Trust Region-Conservative	$2.18 \times 10^{-1}$	$2.48 \times 10^{-3}$	$2.13 \times 10^{-1}$	$2.23 \times 10^{-1}$	144.9	0.0	0.089
GD-Nesterov	$2.27 \times 10^{-1}$	$3.33 \times 10^{-3}$	$2.23 \times 10^{-1}$	$2.33 \times 10^{-1}$	21.6	0.0	0.029
GD-Momentum	$2.33 \times 10^{-1}$	$6.45 \times 10^{-3}$	$2.24 \times 10^{-1}$	$2.47 \times 10^{-1}$	21.4	0.0	0.028
Trust Region-Adaptive	$2.76 \times 10^{-1}$	$1.27 \times 10^{-1}$	$1.75 \times 10^{-1}$	$5.85 \times 10^{-1}$	67.2	0.0	0.041
Trust Region-Standard	$4.17 \times 10^{-1}$	$4.83 \times 10^{-1}$	$1.93 \times 10^{-1}$	1.92	21.1	0.0	0.013