

Table 1: Performance Results for NeuralNetwork\100samplesn1ayersn10n20n5Problem

Optimizer	Mean Final Value	Std Dev	Best Value	Worst Value	Mean Func Evals	Success Rate (%)	Mean Time (s)
Adam-Fast	4.29×10^{-2}	8.16×10^{-3}	2.29×10^{-2}	6.36×10^{-2}	191.6	32.5	0.301
QQN-StrongWolfe	4.97×10^{-2}	1.02×10^{-2}	3.93×10^{-2}	7.47×10^{-2}	473.8	0.0	0.864
QQN-MoreThuente	5.09×10^{-2}	1.87×10^{-2}	3.77×10^{-2}	1.12×10^{-1}	502.4	0.0	0.841
QQN-Bisection-2	5.30×10^{-2}	4.85×10^{-3}	4.22×10^{-2}	6.21×10^{-2}	535.1	0.0	0.800
QQN-Bisection-1	5.57×10^{-2}	4.63×10^{-3}	4.61×10^{-2}	6.64×10^{-2}	503.9	0.0	0.821
QQN-Backtracking	6.13×10^{-2}	4.84×10^{-3}	5.30×10^{-2}	7.08×10^{-2}	544.4	0.0	0.835
L-BFGS-Conservative	6.45×10^{-2}	7.56×10^{-3}	5.34×10^{-2}	8.43×10^{-2}	568.8	0.0	0.757
Adam-WeightDecay	6.97×10^{-2}	3.85×10^{-3}	6.36×10^{-2}	7.64×10^{-2}	502.0	0.0	0.790
L-BFGS-Aggressive	8.54×10^{-2}	3.38×10^{-3}	7.82×10^{-2}	9.21×10^{-2}	705.8	0.0	0.660
QQN-GoldenSection	8.69×10^{-2}	4.51×10^{-3}	7.81×10^{-2}	9.57×10^{-2}	893.4	0.0	0.567
L-BFGS	1.13×10^{-1}	6.79×10^{-2}	7.41×10^{-2}	3.68×10^{-1}	517.2	0.0	0.610
Adam	1.15×10^{-1}	6.62×10^{-3}	9.72×10^{-2}	1.23×10^{-1}	502.0	0.0	0.789
Adam-AMSGrad	1.17×10^{-1}	5.91×10^{-3}	1.08×10^{-1}	1.27×10^{-1}	502.0	0.0	0.796
QQN-CubicQuadraticInterpolation	1.18×10^{-1}	3.82×10^{-3}	1.08×10^{-1}	1.28×10^{-1}	394.8	0.0	0.938
GD-WeightDecay	1.42×10^{-1}	2.52×10^{-3}	1.38×10^{-1}	1.46×10^{-1}	335.0	0.0	0.900
Trust Region-Conservative	1.49×10^{-1}	3.40×10^{-3}	1.42×10^{-1}	1.55×10^{-1}	128.0	0.0	0.150
GD	1.52×10^{-1}	1.28×10^{-3}	1.49×10^{-1}	1.54×10^{-1}	335.0	0.0	0.889
GD-Nesterov	1.62×10^{-1}	1.20×10^{-2}	1.10×10^{-1}	1.68×10^{-1}	36.1	0.0	0.094
GD-Momentum	1.65×10^{-1}	1.51×10^{-2}	1.02×10^{-1}	1.83×10^{-1}	36.1	0.0	0.094
Trust Region-Adaptive	1.93×10^{-1}	5.63×10^{-2}	1.48×10^{-1}	3.46×10^{-1}	26.7	0.0	0.032
Trust Region-Standard	2.52×10^{-1}	3.10×10^{-2}	2.04×10^{-1}	3.08×10^{-1}	5.6	0.0	0.007