Table 1: Performance Results for LinearRegression$\backslash_200 samples n_1 0 features n_r eg 0.01 Problem$

Optimizer	Mean Final Value	Std Dev	Best Value	Worst Value	Mean Func Evals	Success Rate (%)	Mean Time (s)
QQN-Bisection-1	4.82×10^{-1}	4.55×10^{-14}	4.82×10^{-1}	4.82×10^{-1}	144.3	100.0	0.096
QQN-Bisection-2	4.82×10^{-1}	5.79×10^{-14}	4.82×10^{-1}	4.82×10^{-1}	121.7	100.0	0.077
QQN-GoldenSection	4.82×10^{-1}	5.86×10^{-14}	4.82×10^{-1}	4.82×10^{-1}	365.6	100.0	0.120
QQN-StrongWolfe	4.82×10^{-1}	1.26×10^{-14}	4.82×10^{-1}	4.82×10^{-1}	126.0	100.0	0.080
QQN-Backtracking	4.82×10^{-1}	6.64×10^{-15}	4.82×10^{-1}	4.82×10^{-1}	104.0	100.0	0.057
QQN-MoreThuente	4.82×10^{-1}	3.89×10^{-9}	4.82×10^{-1}	4.82×10^{-1}	144.6	100.0	0.084
QQN-CubicQuadraticInterpolation	4.82×10^{-1}	6.66×10^{-9}	4.82×10^{-1}	4.82×10^{-1}	155.9	100.0	0.096
GD-WeightDecay	4.82×10^{-1}	2.38×10^{-10}	4.82×10^{-1}	4.82×10^{-1}	226.3	100.0	0.212
L-BFGS	4.88×10^{-1}	2.19×10^{-2}	4.82×10^{-1}	5.74×10^{-1}	189.4	100.0	0.094
L-BFGS-Aggressive	5.22×10^{-1}	6.13×10^{-2}	4.82×10^{-1}	7.10×10^{-1}	553.2	55.0	0.213
Adam-Fast	5.01×10^{-1}	3.51×10^{-2}	4.82×10^{-1}	5.76×10^{-1}	208.3	15.0	0.124
L-BFGS-Conservative	4.82×10^{-1}	6.47×10^{-7}	4.82×10^{-1}	4.82×10^{-1}	557.0	0.0	0.301
GD	4.82×10^{-1}	2.81×10^{-5}	4.82×10^{-1}	4.82×10^{-1}	335.0	0.0	0.311
GD-Nesterov	1.80	1.62×10^{-2}	1.77	1.83	38.0	0.0	0.034
GD-Momentum	2.46	2.11×10^{-2}	2.42	2.50	38.1	0.0	0.034
Trust Region-Adaptive	6.55×10^1	2.65	6.13×10^1	7.13×10^1	602.0	0.0	0.286
Adam-WeightDecay	7.52×10^1	2.22	6.91×10^1	7.88×10^1	502.0	0.0	0.302
Adam	1.01×10^2	2.46	9.69×10^1	1.05×10^2	502.0	0.0	0.302
Adam-AMSGrad	1.01×10^2	3.16	9.68×10^1	1.09×10^2	502.0	0.0	0.303
Trust Region-Conservative	1.14×10^2	2.66	1.06×10^2	1.18×10^2	602.0	0.0	0.287
Trust Region-Standard	2.76×10^3	4.57×10^3	7.09×10^{-1}	1.26×10^4	373.2	0.0	0.177