Statistics 221 Final Project: C++ team

December 16, 2013

Here we report the results from running the SGD experiment. We compare performance of LibLinear, baseline SGD and ASGD, and our implicit SGD for each of log loss and hinge loss functions. Unless otherwise stated, SGD algorithms use their default regularization parameters (close to $\lambda = 10^{-5}$) and learning rate $\eta_t = \frac{\eta_0}{1+\lambda\eta_0 t}$, with η_0 picked by the Bottou package unless otherwise specified.

1 Results

1.1 RCV1 Benchmark, hinge-loss

algorithm	training time (s)	test error (%)	cost
LibLinear	2	3	x
SGD	0.87	6.005	0.244
ASGD	0.86	6.018	0.244
Implicit	0.71	5.184	0.144

1.2 RCV1 Benchmark, log-loss

algorithm	training time (s)	test error (%)	cost
LibLinear	2	3	x
SGD	3.47	5.175	0.153
ASGD	4.03	5.145	0.154
Implicit ($\eta_0 = 10$, not 4)	1.92	5.262	0.537

For implicit SGD, I use learning rate $\eta_t = \frac{\eta_0}{1+\lambda\eta_0 t/2}$.

1.3 Alpha dataset, hinge-loss

algorithm	training time (s)	test error (%)	cost
LibLinear	2	3	x
SGD	3.37	22.7	0.548
ASGD	2.27	21.83	0.532
Implicit ($\eta_0 = 10$, not 0.25)	2.29	22.18	0.565

Implicit SGD here was improved by using default $\lambda = 10^{-5}$ (whereas SGD and ASGD used $\lambda = 10^{-6}$).

1.4 Alpha dataset, log-loss

algorithm	training time (s)	test error (%)	cost
LibLinear	2	3	x
SGD	9.31	22.11	0.477
ASGD	3.5	21.9	0.474
Implicit ($\eta_0 = 10$, not 0.25)	3.7	21.9	0.542