Logistic regression: loss with GD, SGD, scikit n=10000, d=80 $\gamma_{SGD0} = 0.1$, schedule: decreasing 2 $\gamma_{GD} = 0.5$ $\mu = 1$ scikit empirical risk after convergence (0.011s) GD empirical risk (5.716s) SGD empirical risk (0.569s) SGA empirical risk (0.606s) 2500 5000 7500 10000 12500 15000 17500

iteration (optimization time)

0.70

0.65

0.60

0.55

0.50

0.45

0.40

0

empirical risk