Shared Noise v.s. Noise on Top

Xiaoran Tong
Qin Lu

Michigan State University

tongxia1@msu.edu
qlu@epi.msu.edu

January 25, 2018

Table of Content

Simulation Settings:

Data Generation:

- N = {512}, P=4000, FRQ=0.50;
- ϕ=1.0;
- True Linear Model: $y \sim \mathcal{N}(0, X'X + \phi I)$;
- Distortion = None (identity link).

Modeling and Training:

- Number of Hidden Units = {32, 64, 128, 256}
- Batch Size = 32
- learning rate = 1e-4, Momentum = 0.0
- Reserved Time = 2 hours, Reserved Iteration = ∞

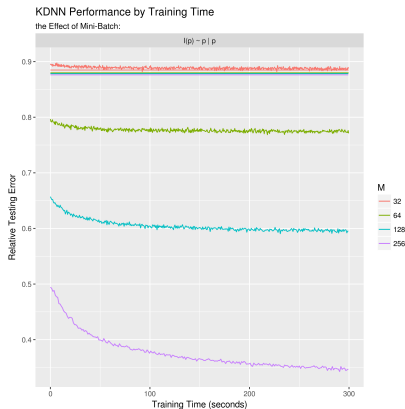# Simulation Settings:

Double Stochastic Gradient Descent:

- tier 1: done by batch training
- tier 2: single sample to estimate the gradient.
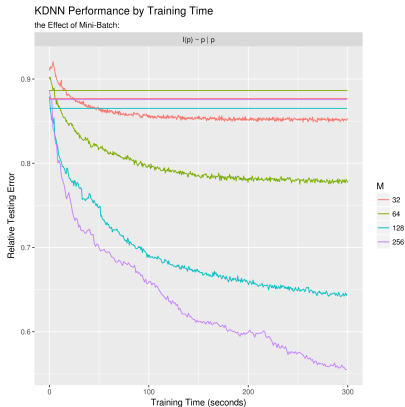
EXP 01: Shared Noise, Short Genome

Performance by Time:

- y-axis: evaluation error relative to NULL model.
- number of hidden units affect initial testing error.

Performance by Time:

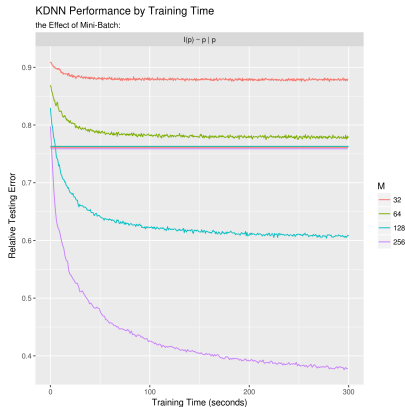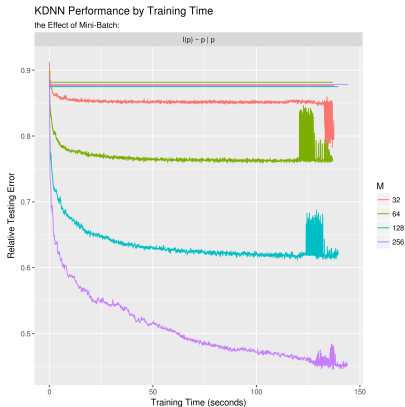- number of hidden units has little effect on initial testing error.

Performance by Time:

- number of hidden units has little effect on initial testing error.

EXP 04: Noise on Top, Short Genome, Stochastic

Performance by Time:

- number of hidden units has little effect on initial testing error.

- by tieing noise only to the top, the behavior of initial error looks normal
- double stochastic gradient seems working.