

Parallelized Image Recognition in Spark & MPI

Tim Clemens
Dan Cusworth
J.D. (Bram) Maasakkers

Multi-class linear classifier

Learning algorithm: Linear classification with L-2 penalty

$$\arg \min_w \left\{ \frac{1}{N} \sum_{i=1}^N (\langle w, x_i \rangle - y_i)^2 + \lambda \|w\|_2^2 \right\}$$

$$x_i, w \in \mathbb{R}^d, \lambda \in \mathbb{R}, y_i \in \{-1, 1\}$$

Analytical solution to the problem

$$w^* = (X^T X + \lambda I_d)^{-1} X^T Y$$

$$w^* = \frac{\sum_{i=1}^N x_i^T y_i}{\sum_{i=1}^N x_i x_i^T + N \lambda}$$

Generating a prediction in the binary case due to Bayes Decision Rule

$$\hat{y}_i = \text{sign}(\langle x_i, w^* \rangle)$$

Multi-class linear classifier

Multi-class, k is the number of classes

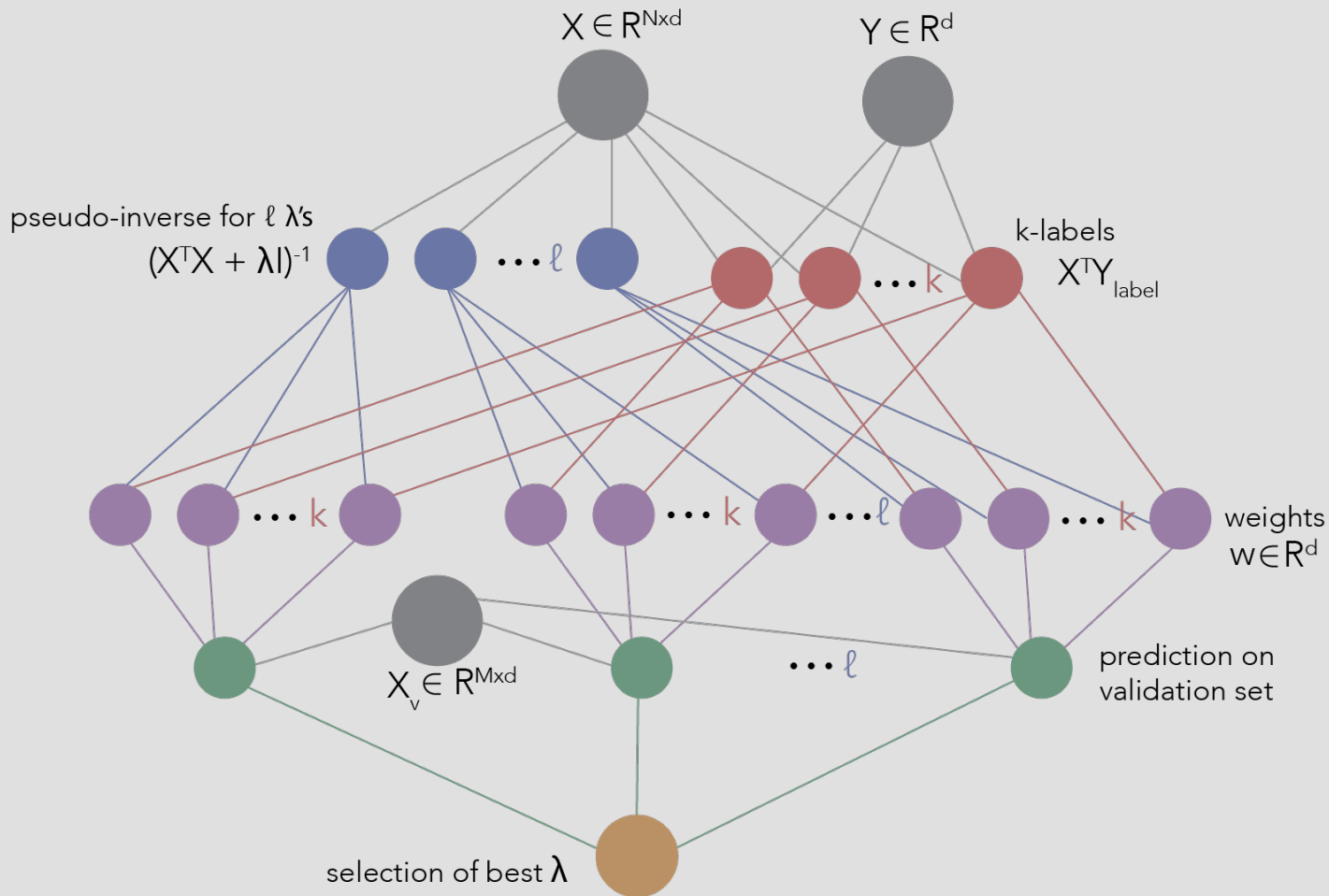
For $j \in [1, k]$:

Solve

$$\arg \min_{w_j} \left\{ \frac{1}{N} \sum_{i=1}^N (\langle w_j, x_i \rangle - y_i^k)^2 + \lambda \|w\|_2^2 \right\}$$

$$\hat{y}_i = \text{index} \left(\arg \max \{ \langle x_i, w_1^* \rangle, \dots, \langle x_i, w_j^* \rangle, \dots, \langle x_i, w_k^* \rangle \} \right)$$

Computational Graph



Parallel implementation

OpenMP + MPI

SPARK

Hybrid Parallelization using MPI
and Cython

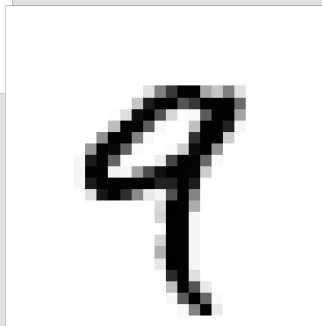
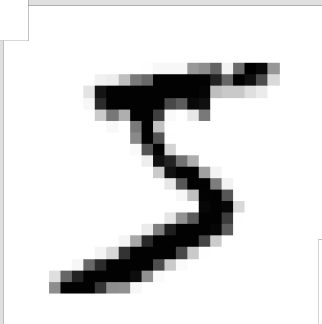
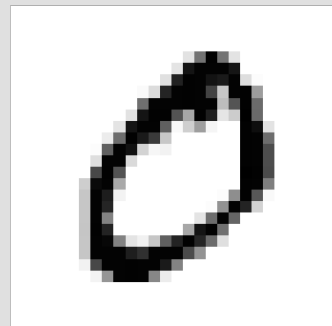
Functional parallelism

Both data parallelism and
model parallelism

Running on Amazon Web
Services EMR

Two datastreams

MNIST

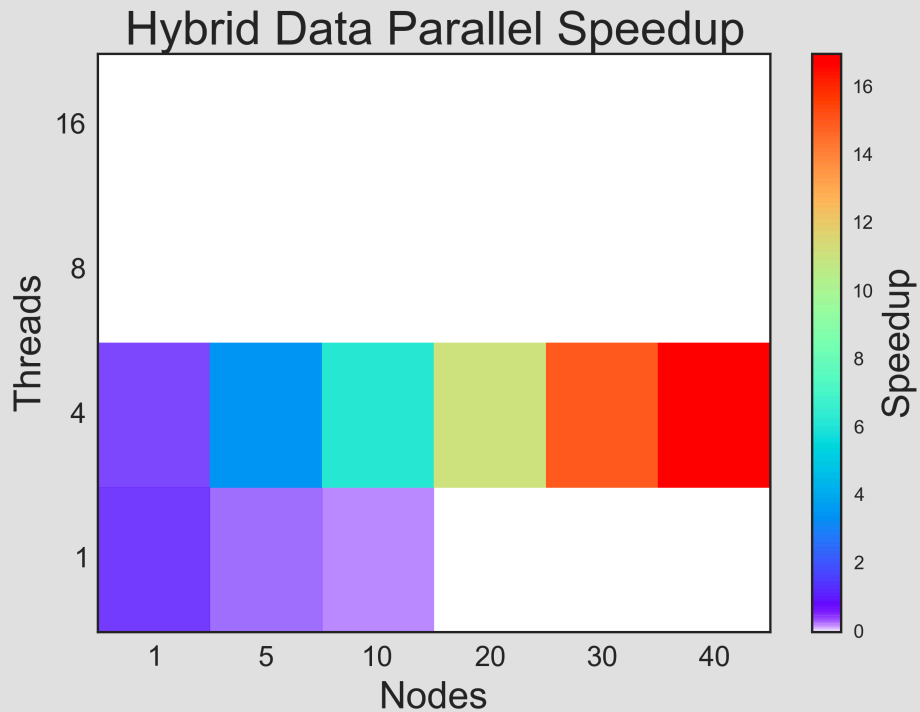


Our own hand images

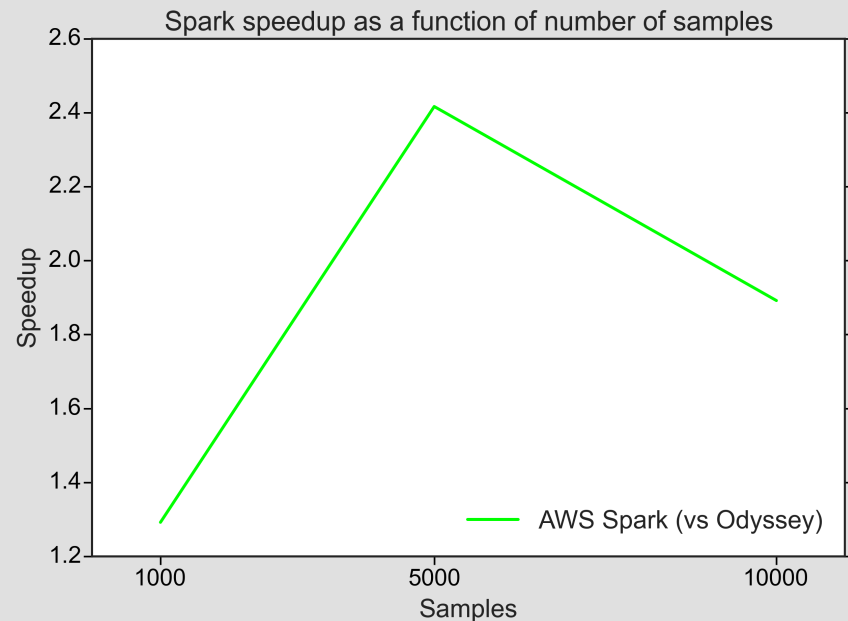


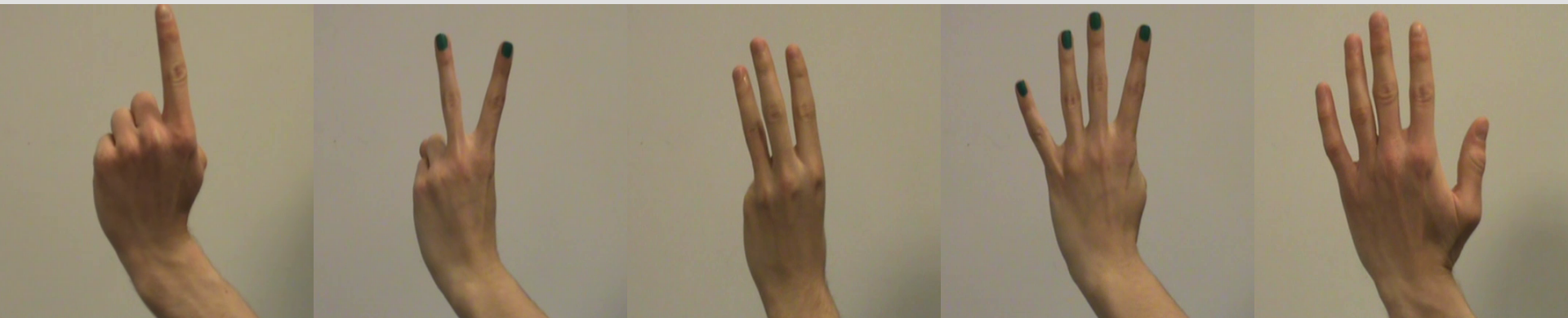
Benchmark results

OpenMP + MPI



SPARK





Parallelized Image Recognition in Spark & MPI

Tim Clemens
Dan Cusworth
J.D. (Bram) Maasakkers