Effective Model Calibration for Terascale Analytics
Chengjie Qin & Florin Rusu
EECS, University of California, Merced

Motivation & Goal
- We present our gradient descent solution in GLADE with three distinct features from other data analytics infrastructures:
  - I/O-bound execution with linear speed-up
  - Data-model-mixed parallelism
  - On-the-fly concurrent hyper-parameter testing and convergence detection by utilizing online aggregation (OLA)

(Incremental) Gradient Descent
- Problem definition:
  \[
  \min_{\vec{w} \in \mathbb{R}^d} \mathcal{L}(\vec{w}) = \sum_{i=1}^{N} f(\vec{w}; \vec{x}_i; y_i) + \mu R(\vec{w})
  \]
  \(\vec{w}\) is the parameter to be learned, \(\vec{x}_i\) is feature vector, \(y_i\) is label
  - Batch Gradient Descent (BGD)
    \[
    \vec{w}^{(k+1)} = \vec{w}^{(k)} - \alpha^{(k)} \nabla \mathcal{L}(\vec{w}^{(k)})
    \]
  - Incremental Gradient Descent (IGD)
    \[
    \vec{w}^{(k+1)} = \vec{w}^{(k)} - \alpha^{(k)} \nabla f(\vec{w}^{(k)})
    \]
  - The hyper-parameter step-size \(\alpha\) is critical to convergence rate. In all the systems \(\alpha\) has to be hand-tuned by the user for different training tasks and datasets combinations.

On-the-fly Concurrent Parameter Testing

Data-model-mixed Parallel Low-rank Matrix Factorization (LMF)

Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Dimension</th>
<th># Examples</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>splice</td>
<td>13M</td>
<td>50M</td>
<td>3TB</td>
</tr>
<tr>
<td>matrix</td>
<td>1M x 1M</td>
<td>1.7B</td>
<td>36GB</td>
</tr>
<tr>
<td>classify</td>
<td>200</td>
<td>50M</td>
<td>13B GB</td>
</tr>
</tbody>
</table>

Results For On-the-fly Parameter Testing

Figure: Multi-threaded LMF on matrix

Figure: (a) Support Vector Machine (SVM) with BGD (classify).
Figure: (b) Logistic Regression (LR) with IGD (classify).
Figure: (c) SVM (splice).