
C. Dünner, A. Anghel, T. Parnell, D. Sarigiannis, N. Ioannou, H. Pozidis

Abstract

- It benefits from GPU acceleration and it can be deployed in single-node and multi-node systems.
- Its hierarchical structure makes it suitable for cloud-based environments.
- It can train a logistic regression classifier on the Criteo Terabyte Click Logs data in 1.5 minutes.
- It will be available to try in June 2018 as part of IBM PowerAI (Tech Preview).

Generalized Linear Models

- Ridge Regression
- Lasso Regression
- Generalized Linear Models
- Logistic Regression

min \( f(A^Tw) + \sum_{i \in [n]} \| w_i \| \) \( A = [x_1, ..., x_n] \) \( w \in \mathbb{R}^n \) \( x_i \in \mathbb{R}^d \)

63.5% of the respondents of the Kaggle “State of Data Science” survey use Logistic Regression. Only 37.6% of the respondents answered Neural Networks.

Experimental Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Single-node</th>
<th>Multi-node</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>45 million</td>
<td>4.2 billion</td>
</tr>
<tr>
<td></td>
<td>10 GB</td>
<td>1 million</td>
</tr>
<tr>
<td></td>
<td>1x Power AC922</td>
<td>3 TB</td>
</tr>
<tr>
<td></td>
<td>1x V100</td>
<td>16x V100</td>
</tr>
<tr>
<td></td>
<td>NVLink 2.0</td>
<td>NVLink 2.0</td>
</tr>
</tbody>
</table>

What is Snap Machine Learning?

- scikit-learn
- Apache Spark™ MLlib
- TensorFlow®

- ML/DL
- No
- No
- Yes
- Limited

- ML/DL
- No
- Yes
- Yes
- Limited

- GLMs
- Yes
- Yes
- Yes
- Yes

* The Apache Software Foundation (ASF) owns all Apache related trademarks, service marks, and graphic logos on behalf of our Apache project communities, and the names of all Apache projects are trademarks of the ASF.

TensorFlow®, the TensorFlow logo and any related marks are trademarks of Google Inc.

2. Tera-Scale Coordinate Descent on GPUs, T. Parnell et al., Future Generation Computer Systems (2018)