MLPACK: A Scalable C++ Machine Learning Library

Ryan Curtin, James Cline, Neil Slagle, Matthew Amidon, Alexander Gray

Georgia Institute of Technology, School of Computational Science and Engineering
Goals:

- Implement **scalable, fast** machine learning algorithms.
Goals:

- Implement **scalable**, **fast** machine learning algorithms.
- Design an **intuitive**, **consistent**, and **simple** API for users who are not C++ gurus.
Goals

Goals:

- Implement **scalable, fast** machine learning algorithms.
- Design an **intuitive, consistent, and simple** API for users who are not C++ gurus.
- Implement a **large, comprehensive** collection of machine learning methods.
Goals:

- Implement **scalable**, **fast** machine learning algorithms.
- Design an **intuitive**, **consistent**, and **simple** API for users who are not C++ gurus.
- Implement a **large**, **comprehensive** collection of machine learning methods.
- Provide **cutting-edge** machine learning algorithms that no other library does.
Currently implemented methods in MLPACK 1.0:

- Fast Hierarchical Clustering (EMST)
- Gaussian Mixture Models (via EM)
- Hidden Markov Models
- Kernel PCA
- K-Means Clustering
- LARS / Lasso Regression
- Least-Squares Linear Regression
- Maximum Variance Unfolding (LRSDP)
- Naive Bayes Classifier
- Neighborhood Components Analysis (NCA)
- RADICAL (ICA)
- Tree-based \( k \)-NN search
- Tree-based range search
k-Nearest-Neighbor Search

Dataset: AMD Phenom II X6 1100T at 3.3 GHz; 8 GB RAM. Dataset randomly generated; uniformly distributed.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>MLPACK</th>
<th>sklearn</th>
<th>MATLAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1Kx10</td>
<td>0.1s</td>
<td>0.2s</td>
<td>0.3s</td>
</tr>
<tr>
<td>3Kx10</td>
<td>0.3s</td>
<td>0.8s</td>
<td>0.9s</td>
</tr>
<tr>
<td>10Kx10</td>
<td>1.3s</td>
<td>4.7s</td>
<td>3.5s</td>
</tr>
<tr>
<td>32Kx10</td>
<td>7.3s</td>
<td>15.2s</td>
<td>17.6s</td>
</tr>
<tr>
<td>100Kx10</td>
<td>47.3s</td>
<td>75.0s</td>
<td>110.6s</td>
</tr>
<tr>
<td>316Kx10</td>
<td>253.5s</td>
<td>363.6s</td>
<td>589.0s</td>
</tr>
<tr>
<td>10Kx31</td>
<td>6.9s</td>
<td>45.6s</td>
<td>27.0s</td>
</tr>
<tr>
<td>10Kx100</td>
<td>18.1s</td>
<td>199.0s</td>
<td>82.2s</td>
</tr>
<tr>
<td>10Kx316</td>
<td>53.5s</td>
<td>732.7s</td>
<td>245.5s</td>
</tr>
<tr>
<td>10Kx1K</td>
<td>174.7s</td>
<td>2533.1s</td>
<td>750.7s</td>
</tr>
</tbody>
</table>
Feature Requests

There is a blank spot on the poster for feature requests.

Write on it!

http://www.mlpack.org