Mlbase: distributed machine learning system
Adapted slides from mlbase.org

Lay of the Land

Ease of use (vertical axis) / Performance, Scalability (horizontal axis)

Matlab, R — x

MLI — x

MLlib — x

Mahout — x   GraphLab, VW — x
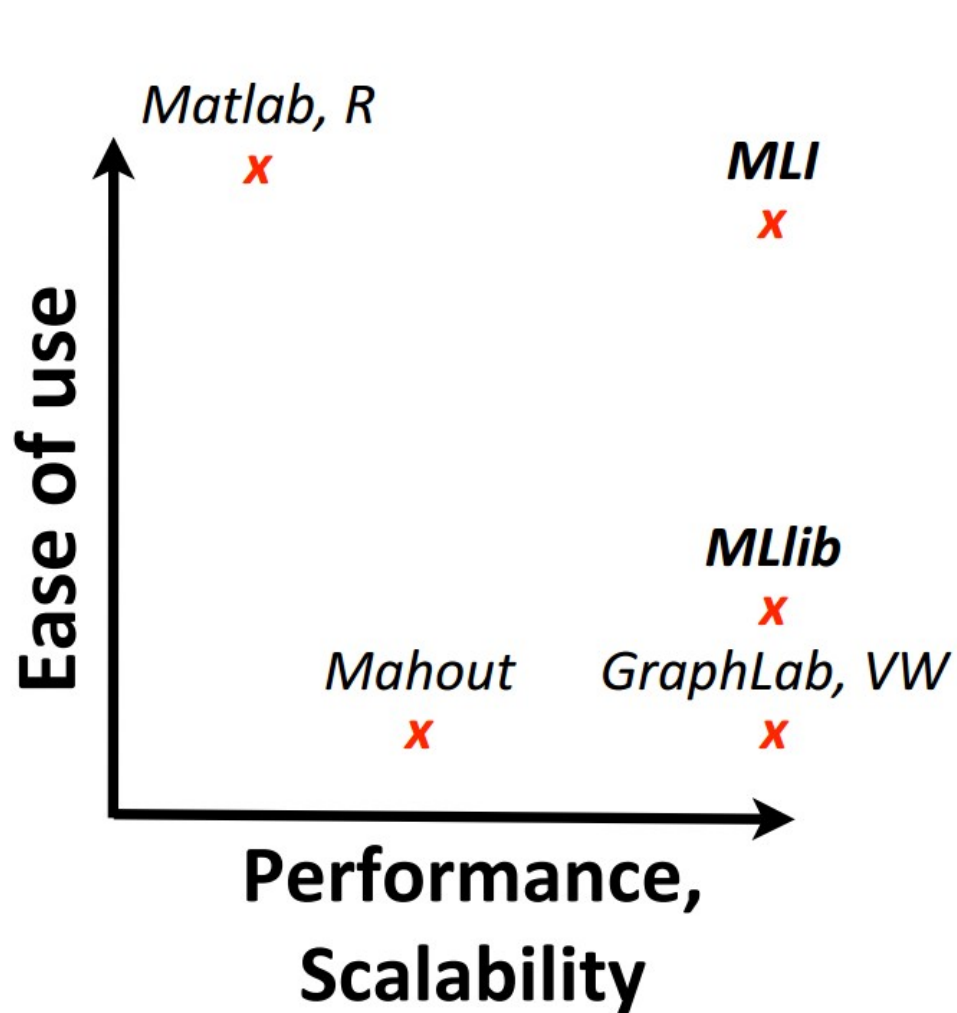
MATLAB
The Language of Technical Computing

R

+ Easy (Resembles math, limited set up)
+ Sufficient for prototyping / writing papers
— Ad-hoc, non-scalable scripts
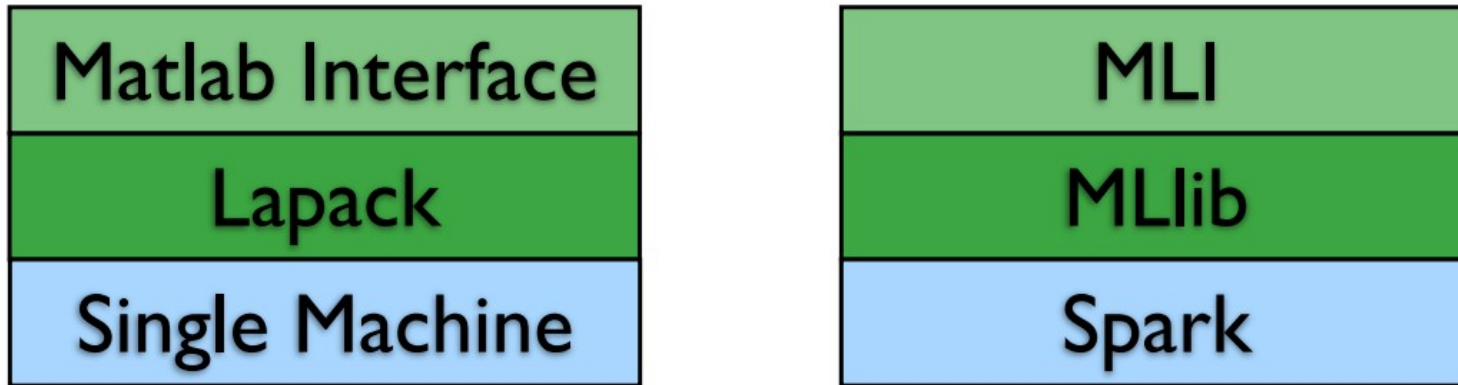— Loss of translation upon re-implementation

+ Scalable and (sometimes) fast
+ Existing open-source libraries
— Difficult to set up, extend

mahout

VOWPAL WABBIT   GraphLab

MLbase Stack

Matlab Interface
Lapack
Single Machine

MLI
MLlib
Spark

Spark: cluster computing system designed for iterative computation

MLlib: low-level ML library in Spark

MLI: API / platform for feature extraction and algorithm development
  - Platform independent

MLI Ease of Use

Logistic Regression

System	Lines of Code
Matlab	11
Vowpal Wabbit	721
MLI	55

Alternating Least Squares

System	Lines of Code
Matlab	20
Mahout	865
GraphLab	383
MLI	32

Typical Data Analysis Workflow

Spark, MLI	**Obtain / Load Raw Data**
Spark, [MLI]	**Data Exploration**
MLI	**Feature Extraction**
MLI, MLlib	**Learning**
MLI	**Evaluation**
Scala	**Deployment**