A Brief Intro to Magic

aka Machine Learning

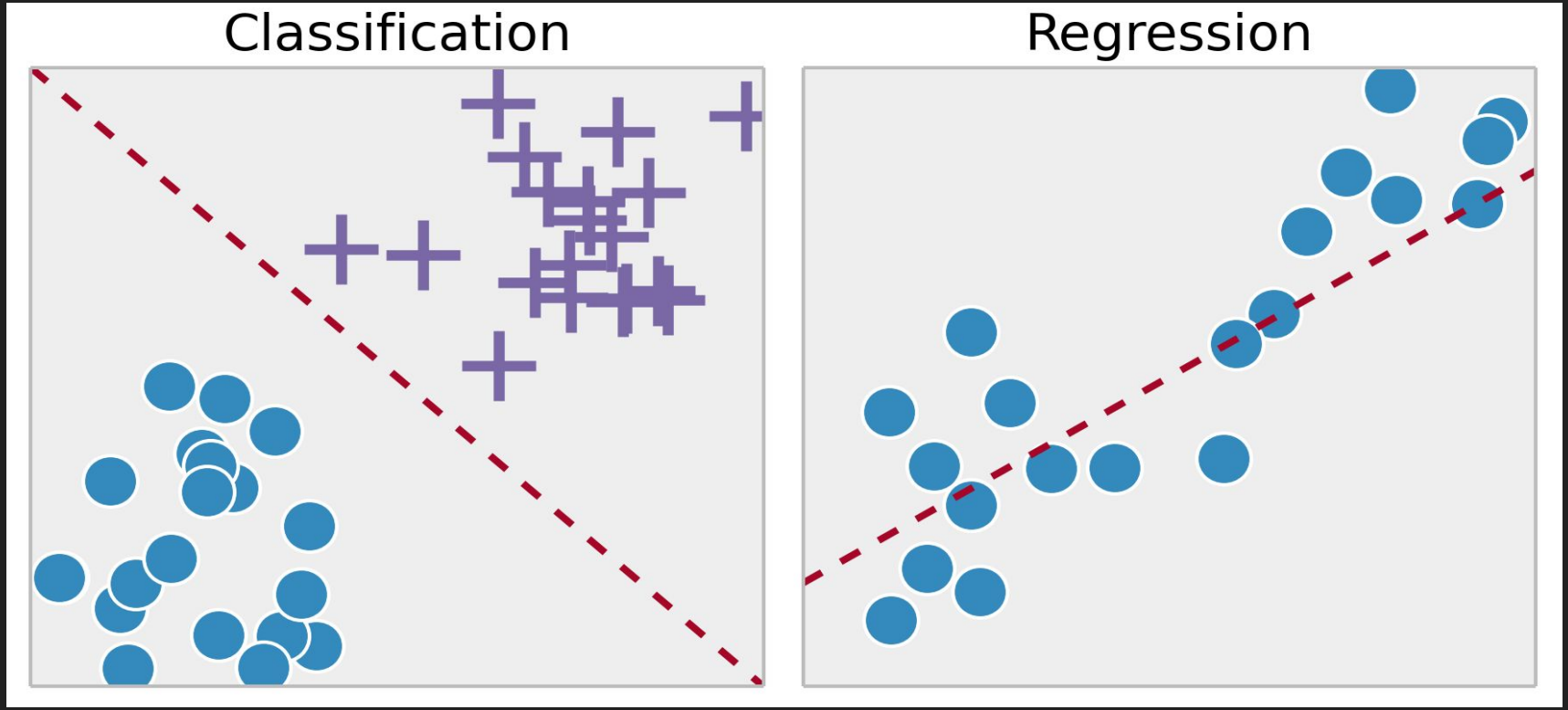
Michael duPont - mdupont.com

Supervised

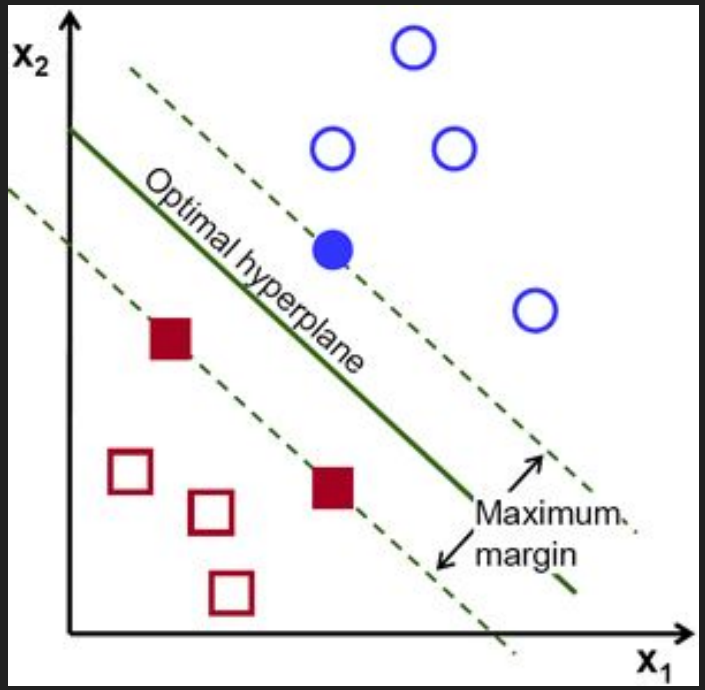
Unsupervised

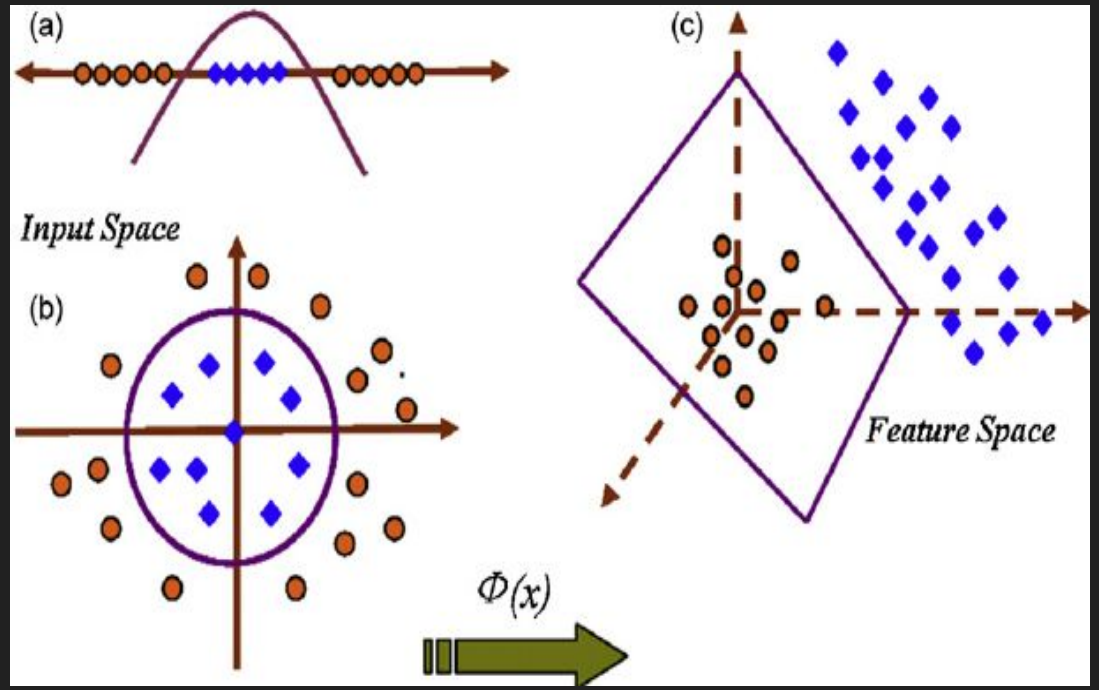
Reinforcement

Supervised Learning

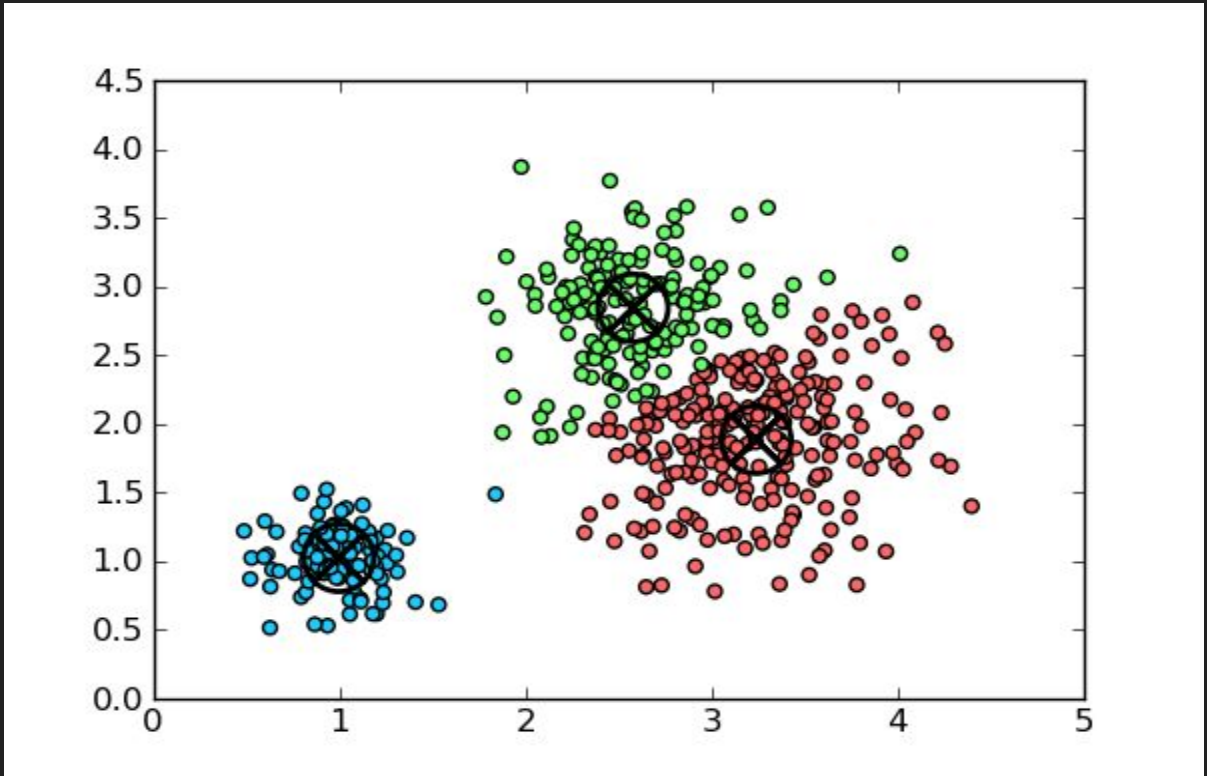


Support Vector Machine

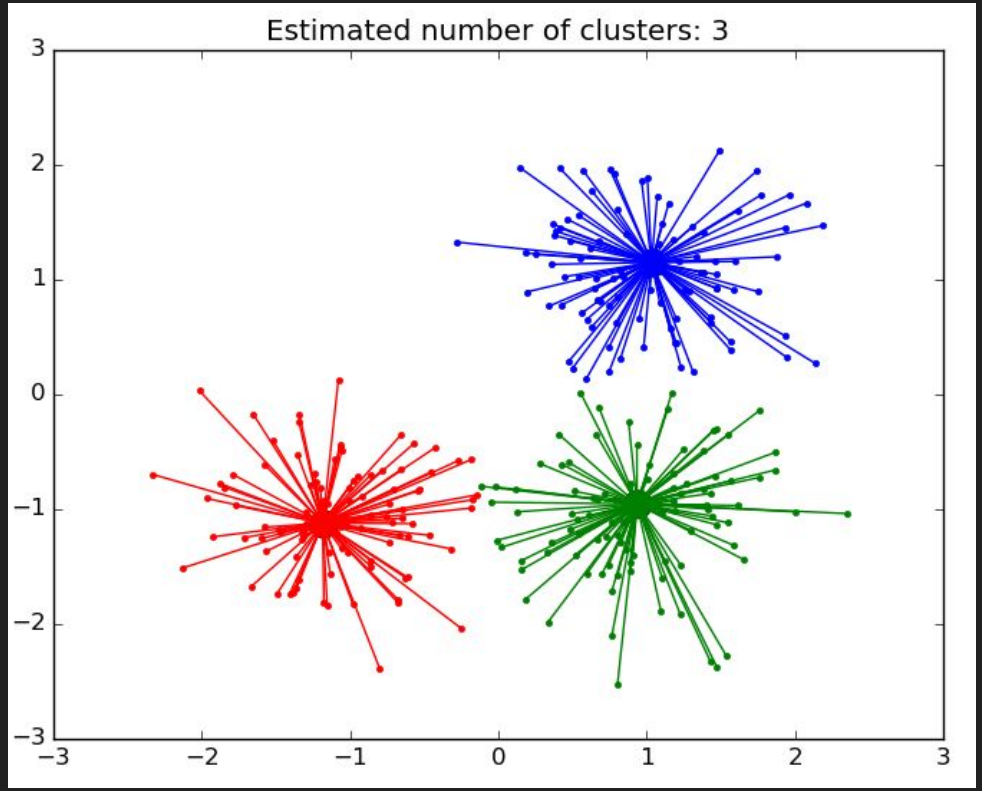


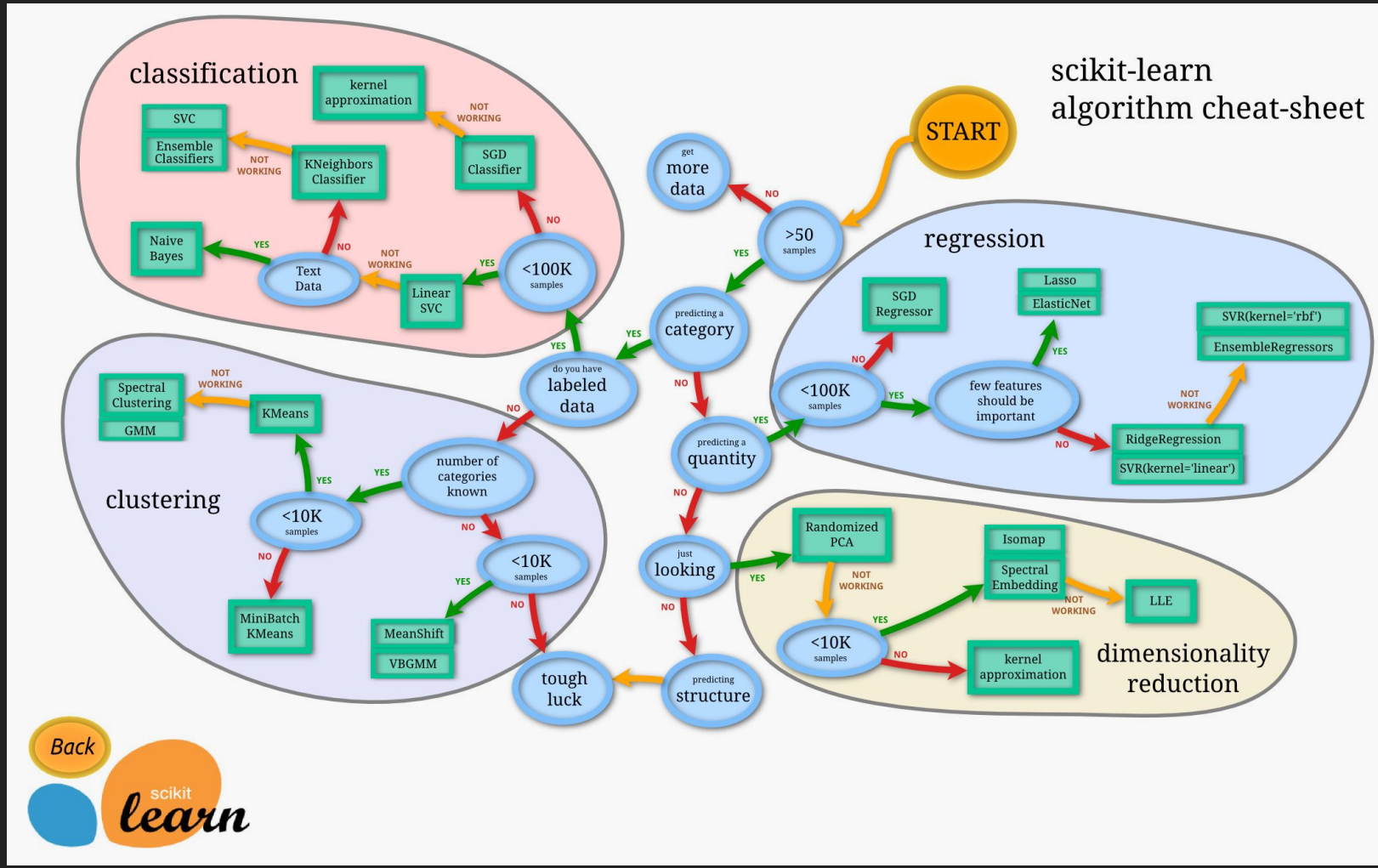


Unsupervised Learning



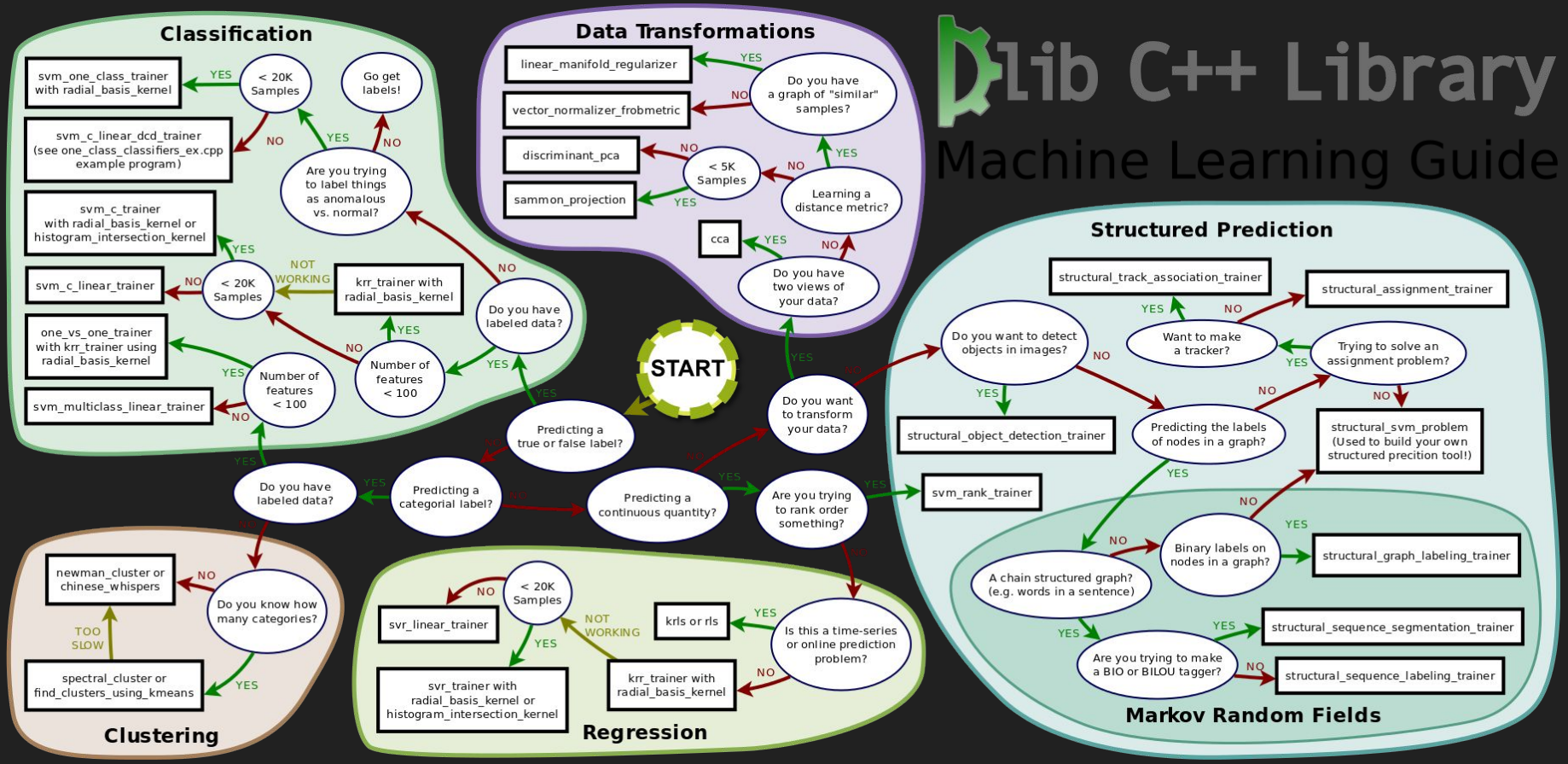
KMeans Clustering




scikit-learn
algorithm cheat-sheet
START
>50 samples
NO
get more data
YES
predicting a category
YES
do you have labeled data
YES
<100K samples
YES
Text Data
NO
Naive Bayes
YES
NOT WORKING
Linear SVC
NO
SGD Classifier
NOT WORKING
KNeighbors Classifier
NOT WORKING
SVC
Ensemble Classifiers
NOT WORKING
kernel approximation
classification
NO
predicting a quantity
YES
<100K samples
NO
SGD Regressor
YES
few features should be important
YES
Lasso
ElasticNet
NO
RidgeRegression
SVR(kernel='linear')
NOT WORKING
SVR(kernel='rbf')
EnsembleRegressors
regression
NO
number of categories known
YES
<10K samples
YES
KMeans
NOT WORKING
Spectral Clustering
GMM
NO
MiniBatch KMeans
NO
<10K samples
YES
MeanShift
VBGMM
NO
clustering
NO
just looking
YES
Randomized PCA
NOT WORKING
<10K samples
YES
Isomap
Spectral Embedding
NOT WORKING
LLE
NO
kernel approximation
dimensionality reduction
NO
predicting structure
tough luck
Back
scikit learn

dlib C++ Library

Machine Learning Guide



Machine Learning Libraries

Python

- Scikit-Learn - everything, plays well with numpy and pandas
- PyBrain - neural networks
- nltk - natural language processing
- TensorFlow - deep learning, graphs

Java

- Mahout - everything, hadoop-based
- Spark - most-everything, cluster-based
- Mallet - natural language processing

C++

- dlib - everything, the grand daddy
- mlpack - almost everything, cl tools and classes
- Vowpal Wabbit, MultiBoost, Shogun

Cloud, Web, and more

- AWS, Azure, Watson, TensorFlow
- machine_learning - npm
- PHP-ML
- Swift-AI
- Accord.NET

EXAMPLE

EigenFaces: supervised model using Principal Component Analysis (PCA) and a Support Vector Machine (SVM) classifier to identify and label human faces