Choosing the right estimator

scikit-learn
algorithm cheat-sheet

classification

- kernel approximation
- SVC
- Ensemble Classifiers
- KNeighbors Classifier
- SGD Classifier
- Naive Bayes
- Text Data
- Linear SVC
- <100K samples

START

get more data

>50 samples

NO — get more data

YES

predicting a category

do you have labeled data

regression

- SGD Regressor
- Lasso / ElasticNet
- SVR(kernel='rbf')
- EnsembleRegressors
- <100K samples
- few features should be important
- RidgeRegression
- SVR(kernel='linear')

NOT WORKING

YES / NO

clustering

- Spectral Clustering
- GMM
- KMeans
- <10K samples
- number of categories known
- MiniBatch KMeans
- MeanShift
- VBGMM
- <10K samples

NOT WORKING

predicting a quantity

just looking

predicting structure

tough luck

dimensionality reduction

- Randomized PCA
- Isomap
- Spectral Embedding
- LLE
- <10K samples
- kernel approximation

NOT WORKING

Back

scikit learn