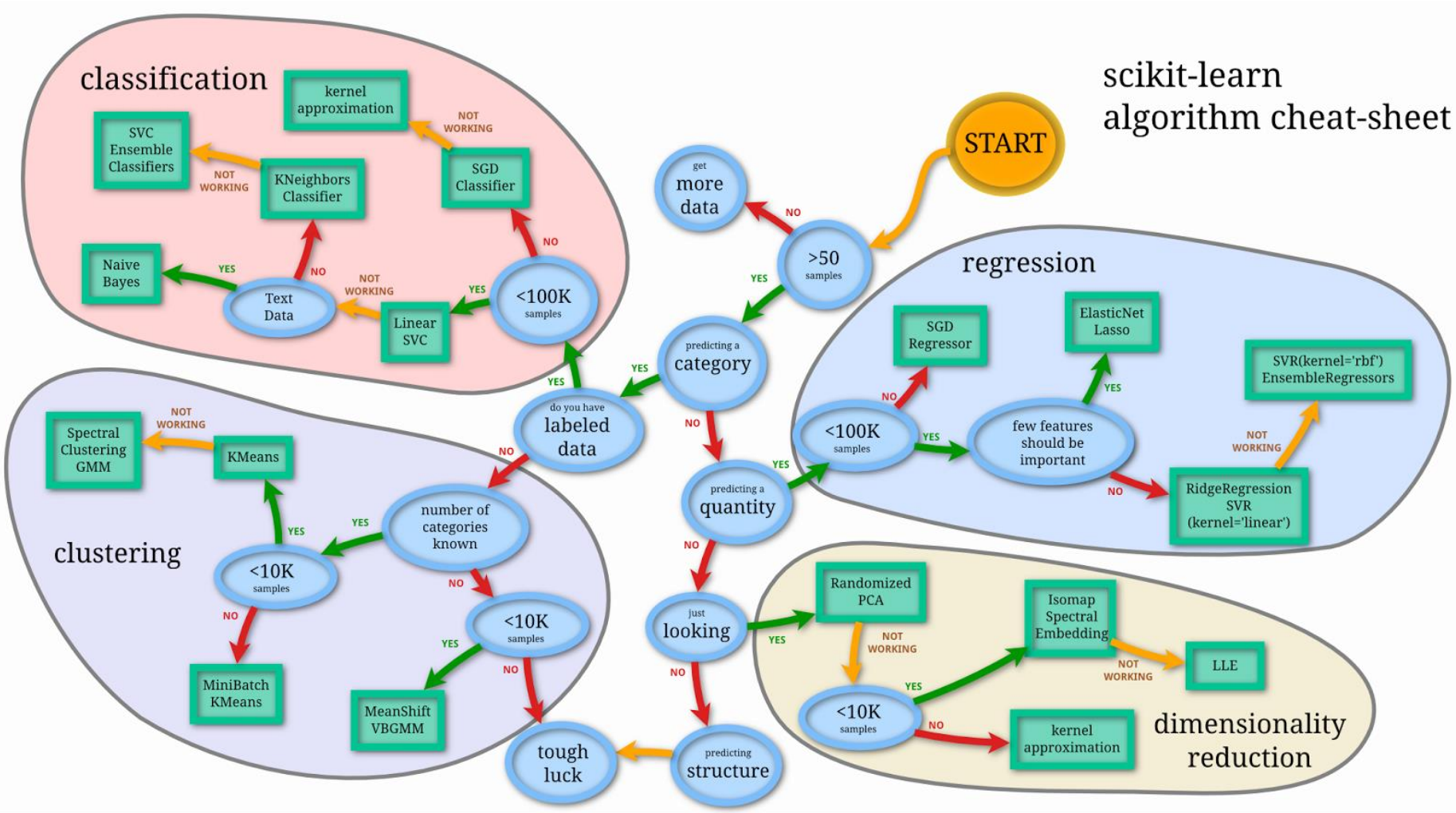
scikit-learn
algorithm cheat-sheet
START
classification
kernel approximation
NOT WORKING
SVC Ensemble Classifiers
NOT WORKING
KNeighbors Classifier
SGD Classifier
NO
Naive Bayes
YES
NO
NOT WORKING
YES
Text Data
Linear SVC
<100K samples
get more data
NO
>50 samples
YES
regression
SGD Regressor
ElasticNet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