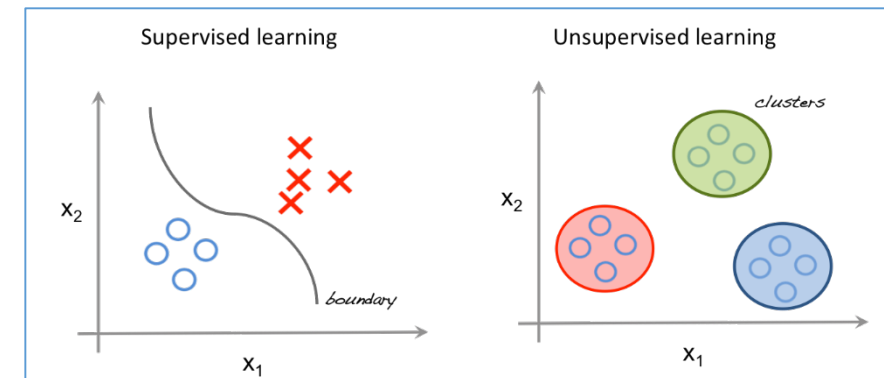
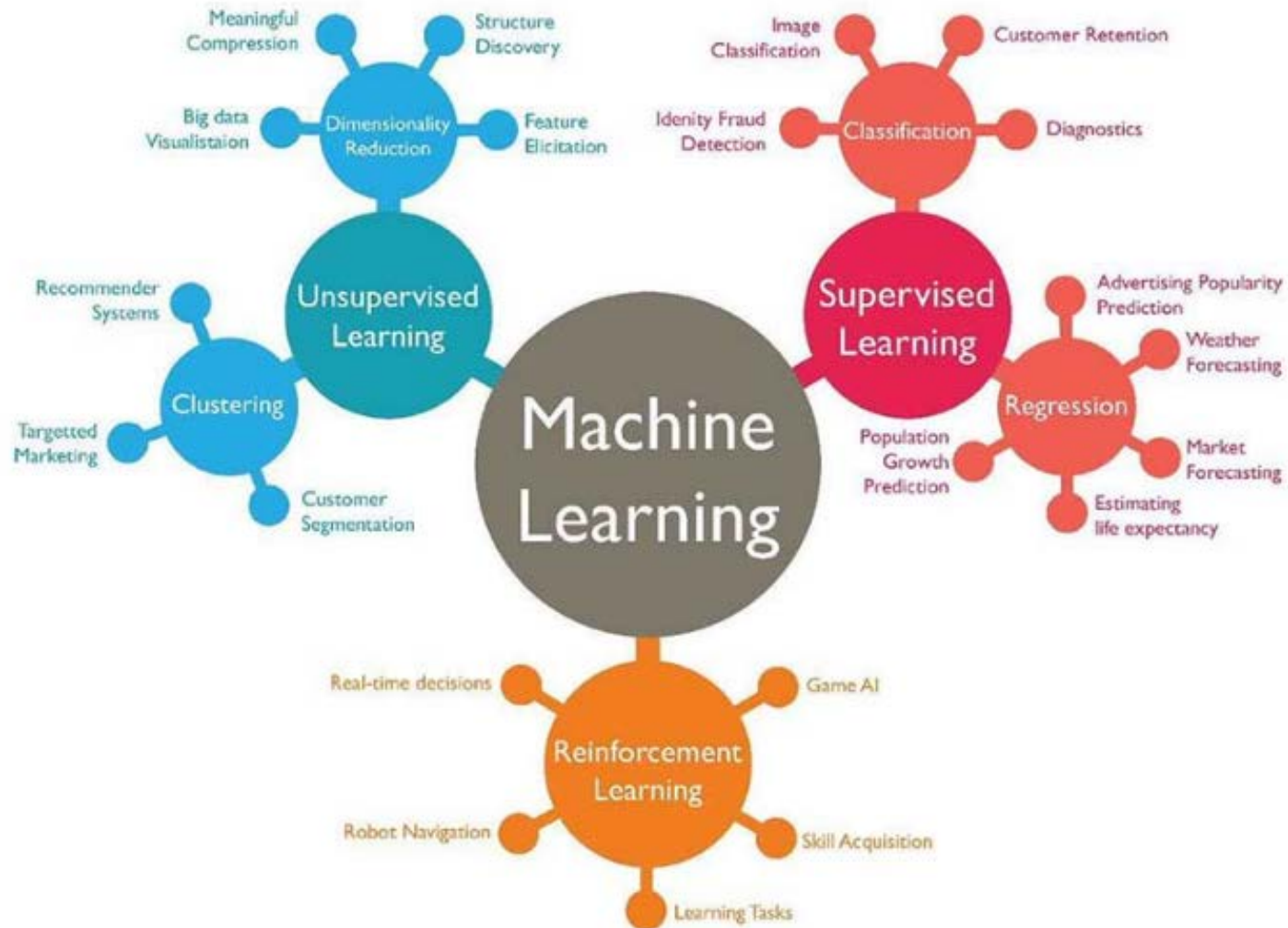


Intro to Machine Learning

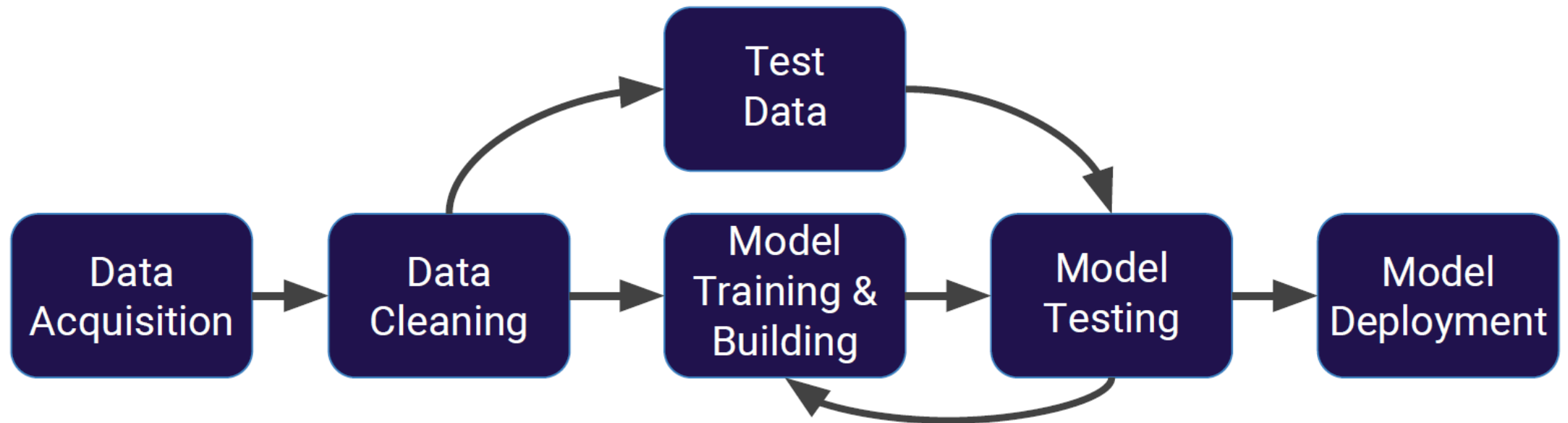
TOPICS

- Introduction to ML
- ML process
- Statistics
- Regression
- Classification
- Clustering
- Dimensionality Reduction

Introduction to ML



Machine Learning Process



Statistics

Data

- Population Sampling
- Mean, Median, Mode, Frequency
- Correlation , Covariance

Distribution

- Normal, Skewed, Bimodal, Uniform
- Lognormal , Pareto

Testing

- Hypothesis Testing
- Linear least squares

Regression

Models

- Linear regression
- Regularization: Ridge (L2), Lasso (L1)

Algorithms

- Gradient descent
- Coordinate descent

Concepts

- Loss functions, bias-variance tradeoff, cross-validation, sparsity, overfitting, model selection

Classification

Models

- Linear classifiers (logistic regression, SVMs, perceptron)
- Kernels
- Decision trees

Algorithms

- Stochastic gradient descent
- Boosting

Concepts

- Decision boundaries, MLE, ensemble methods, random forests, CART, online learning

Clustering

Models

- Nearest neighbors
- Clustering, mixtures of Gaussians
- Latent Dirichlet allocation (LDA)

Algorithms

- KD-trees, locality-sensitive hashing (LSH)
- K-means
- Expectation-maximization (EM)

Concepts

- Distance metrics, approximation algorithms, hashing, sampling algorithms, scaling up with map-reduce

Dimensionality Reduction

Models

- Collaborative filtering
- Matrix factorization
- PCA

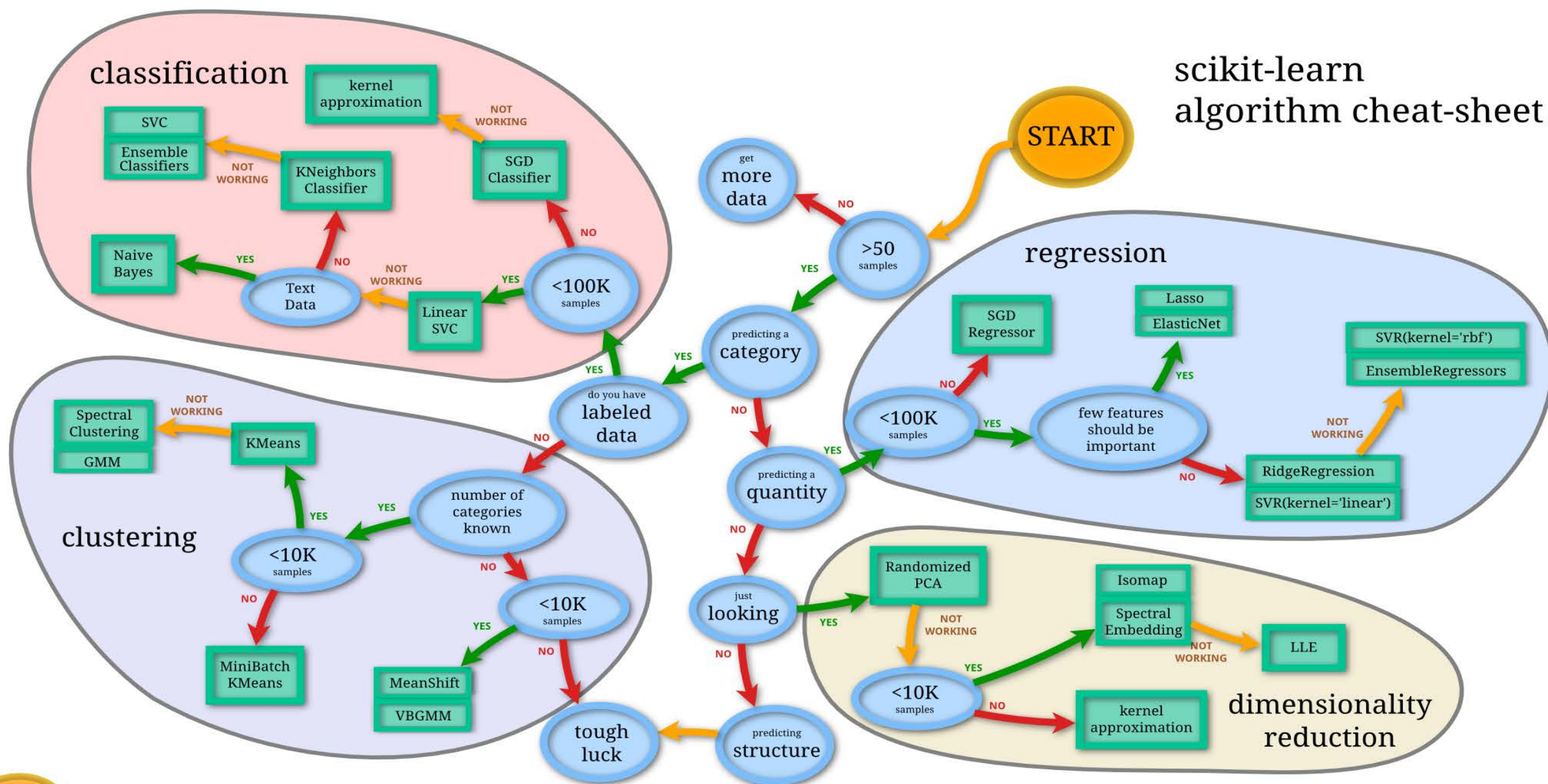
Algorithms

- Coordinate descent
- Eigen decomposition
- SVD

Concepts

- Matrix completion, eigenvalues, random projections, cold-start problem, diversity, scaling up

scikit-learn algorithm cheat-sheet



Back