Python Programming, Visualization and Text Analysis

Arieda Muço

Central European University

Machine learning

Applying algorithms that iteratively learn from data.

- Used for: Fraud detection, pattern and image recognition, text sentiment analysis, email spam filtering, credit scoring, new pricing models, recommendation engines...

Types of learning

- Supervised Learning
  - The program is “trained" on a pre-defined set of “training examples", which facilitate its ability to reach an accurate conclusion when given new data
  - In this case, we have a “target" or dependent variable
  - We also have “labeled" data
- Unsupervised Learning
  - Find patterns and relationships between the data
  - No “target” or dependent variable
  - We don’t have “labeled" data
- Deep Learning (Artificial Neutral Networks)

Some references

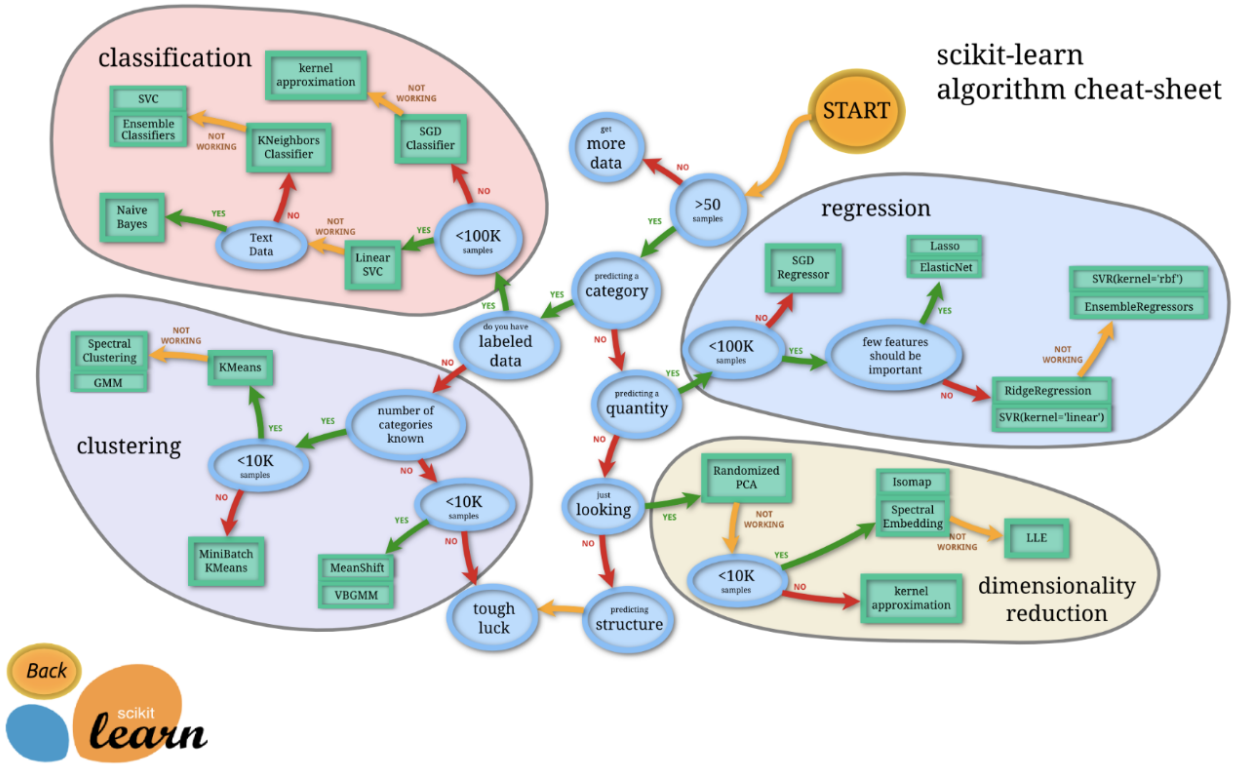
- Books
  - An Introduction to Statistical Learning `http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf`
  - Introduction to Machine Learning `http://robotics.stanford.edu/people/nilsson/MLBOOK.pdf?`
  - Elements of Statistical Learning `https://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII_print10.pdf?`
  - Machine Learning with Python by Sara Guido and Andreas Muller
- Andrew Ng
  - Notes `http://cs229.stanford.edu/materials.html`
  - Video `https://www.coursera.org/learn/machine-learning`

Python Libraries

The most used libraries

- pip install scikit-learn or conda install scikit-learn (Machine Learning)
- pip install nltk or conda install nltk (Text Analysis)

Algorithm Cheat Sheet



Learning Machine Learning

Machine Learning takes time to learn

- Be patient with yourself
- Ask questions. I'm happy to answer questions to point you towards material where you can further deepen your understanding (via Slack or email)
- Most importantly: Google