Data Science. Module IV

Machine Learning I

The question of whether a computer can think is no more interesting than the question of whether a submarine can swim.

Edsger W. Dijkstra

Necessary

This module is dedicated for custom implementation and research.
What you need is to understand how to learn fast and effective. By the way, here is nice course on that. And here is the short summary for impatient.

Best approach to get practical skills is to understand how to search relevant information, make research and try to implement it as fast as you can. Only after those tries you will start asking questions and think by yourself.

You can use our materials on topics you need in this module, but you need to understand how to can most relevant information by yourself, having just few keywords.

All links you used on each topic, you have to put in this worksheet.

So, machine learning modules will be done in next format:

Here is the list of methods/algorithms you have to work on.

- Gradient Descent (**research** & **use**)
- Linear Regression (**implement**)
- Maximum likelihood estimation (**research**)
- Logistic Regression (**implement**)
- Regularization (**research** & **use**)
- Naive Bayes Classifier (**implement**)
- Gaussian Discriminant Analysis (**research**)
- K-means (**implement**)
- KNN (**implement**)

Explanation:

- **implement** == implement algorithm use only python, numpy, scipy and matplotlib for visualisation
- **research** == read, understand
- **use** == use knowledge to make implementation better

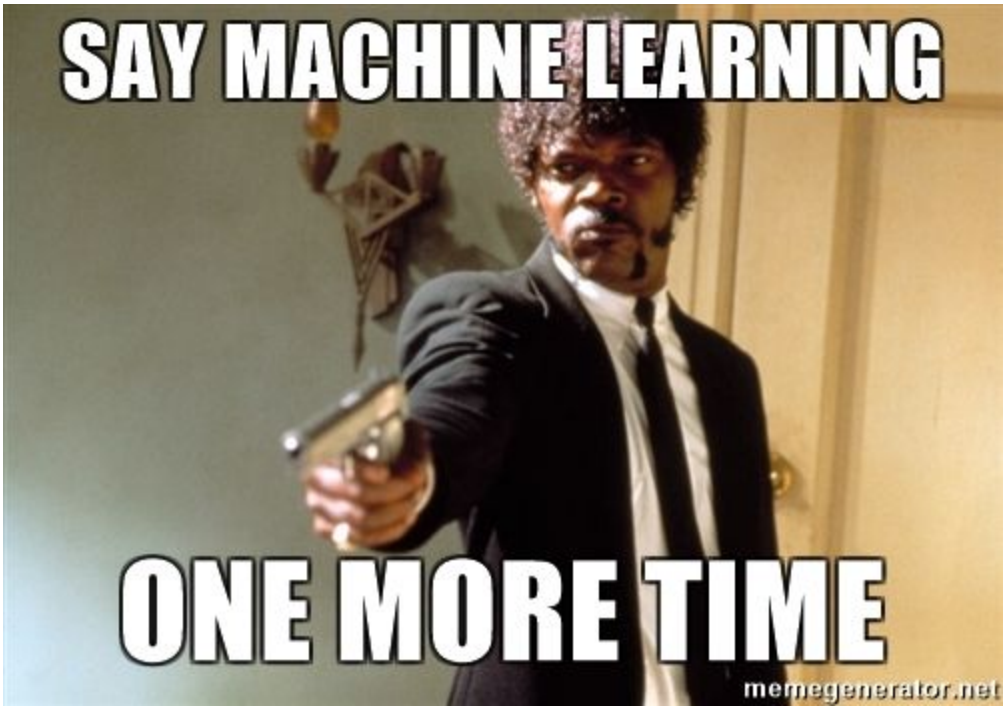
So learn, implement, test, compare.
Choose any dataset here: http://www.datapure.co/open-data-sets
So you need to:

1. *Find relevant information on each of these algorithms and fill in the next sheet.*
   *This sheet is one for all, you have to fill in only one line. If you want to paste multiple links in one cell, just press **ctrl + Enter**. This will help you and your coursemates to make a progress faster.*
2. *Send implemented algorithms/methods (simple and working ones).*
   ***Deadline is 30rd of November**.*
3. *Try it on any dataset you like and compare with sklearn realization (for models).*
4. *Cool discussion about them on the next meeting.*

Resources

- Articles
  - http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/ - systematization
  - http://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html - must know
  - https://www.python-course.eu/machine_learning.php - machine learning tutorial with python
- Courses
  - Machine Learning Coursera Course: https://www.coursera.org/learn/machine-learning
  - Data Science Specialization from Yandex: https://www.coursera.org/specializations/machine-learning-data-analysis
  - Applied Data Science in Python: https://www.coursera.org/learn/python-machine-learning
  - Practical Machine Learning Tutorial with Python Introduction: https://pythonprogramming.net/machine-learning-tutorial-python-introduction/
  - Data Mining Video Lectures: (quick overview) - https://www.youtube.com/playlist?list=PLLssT5z_DsK9JDLcT8T62VtzwyW9LNepV

- Books/PDFs
  - Machine Learning in Python
  - Introduction to Machine Learning Using Python
  - Python Machine Learning
  - Machine Learning with Python/Scikit-Learn
  - Machine Learning in Action
  - Statistics and Machine Learning in Python
- Additional Books:
  - Hastie, Tibshirani, and Friedman's The Elements of Statistical Learning
  - Bishop's Pattern Recognition and Machine Learning
  - David Barber's Bayesian Reasoning and Machine Learning
  - (Great Book)Kevin Murphy's Machine learning: a Probabilistic Perspective
  - Foundations of Machine Learning, Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar
  - Learning From Data, Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Lin
  - Information Theory, Inference, and Learning Algorithms, David J. C. MacKay[free pdf]
  - All of Statistics, Larry Wasserman
  - Probabilistic Graphical Models: Principles and Techniques, Daphne Koller, Nir Friedman
  - Gaussian Processes For Machine Learning, Carl Edward Rasmussen, Christopher K. I. Williams [free pdf]
  - Machine Learning with R
  - Building Machine Learning Systems with Python
  - Machine Learning with Spark
  - Matrix Computations (Johns Hopkins Studies in the Mathematical Sciences): Gene H. Golub, Charles F. Van Loan: 9781421407944: Amazon.com: Books
  - Amazon.com: Convex Optimization (9780521833783): Stephen Boyd, Lieven Vandenberghe: Books



See you soon in Module V