Kaggle Project:

Home Credit Default Risk: Can you predict how capable each applicant is of repaying a loan?

<https://www.kaggle.com/c/home-credit-default-risk>

1. Install Python

# Anaconda – python 3.6 (recommended package that includes Python and most recommended libraries; sometimes the 32bit version works better even for 64bit computers): <https://www.anaconda.com/download/>.

1. Learn libraries: Numpy, Pandas, Machine Learning Library

Numpy: <https://docs.scipy.org/doc/numpy/user/quickstart.html>

Pandas: <https://pandas.pydata.org/pandas-docs/stable/tutorials.html>

Machine Learning Library: <http://scikit-learn.org/stable/>

1. Machine learning Thoery

Goodfellow-et-al-2016 (Charpter 5: Machine Learning Basis) <https://www.dropbox.com/s/9kazvz6rinl585t/Goodfellow-et-al-2016.pdf?dl=0>

Timeline:

Phase 1: July 17th~July 29th

Target: 1) Familar with Numpy, Pandas, Machine Learning Library

2) Confirm the project procedures (how to clear data, features chossing, learning approaches)

Phase 2: July 30th~Aug 12th

Target: Initial testing results generated.

Phase 3: Aug 13th ~ Aug 29th

Traget: refine our data processing and learning approaches to improve the testing performance.

Meeting:

Suggested (not yet confirmed by all)

1. Each Saturday 9:00am to 5pm in uic meeting room
2. Each Wednesday 8:00pm to 10:00pm Video meeting

Workload Distribution (Phase 1):

Jinghua and yuanlong:

install python, learn Numpy, Pandas, Machine Learning Libraries

Goal: Can use them in a familiar manner when we start coding in Phase 2

Pengtao and JianXu

Project Procedures

Goal: Procedures details should be present

The initial requirement before next meeting: (July 18th, 8:00-10:00pm):

Jinghua and yuanlong:

1. python installed and runable in your computer.
2. At least one or two libraries played.

Pengtao and JianXu

1. Download the data
2. Review several kernal and discussion shown in the project kinks (with high votes)
3. An initial rough idea