Final Project

- Send to: xinchao.teaching@gmail.com
- Each Project: A group of two
- Topic:
  - Based on the class material
  - Focus on ***learning*** but not ***feature extraction***
  - Can be related to your research, but it has to be extended
- What to submit:
  - A report
    - Font Size 12, single column, 2-3 pages.
  - The Code
    - Explicitly label, which part of the code is implemented yourself, by labeling
      - START: OWN CODE, right before your own implemented code
      - END: OWN CODE, right after your own implemented code

Report

- In the report, you should:
  - Define problem
    - What is the problem you are trying to solve?
  - Show connection to class material
    - What is being classified, what are the classes etc.
  - Describe data
    - Train/test splits etc.
  - Show results of **at least two learning models**
    - Why do you see such results?
    - What results will you get if you tune the parameters?
    - What insights can you obtain?

Potential Projects

- Object/person recognition
  - PCA: Eigenfaces, eigendogs, etc.
  - HOG vs. SIFT
  - Data: Caltech 101/256, PASCAL, MIT Labelme, Yale face database, …
- Classification of general data
  - SVM
  - Boosting
  - Random forests
  - Data: UCI ML repository

Potential Projects

- Detection of facial features (eyes, mouth)
  - PCA
  - Boosting
  - Data: Yale face database, Labeled Faces in the Wild, BioID
- Terrain classification and object detection from 3D data
  - PCA
  - Invariant descriptors

Potential Projects

- Optical character recognition
- Spam filtering
- Stock price prediction
- And more!

Project: Datasets

- General
  - UCI ML repository: http://archive.ics.uci.edu/ml/
  - Google: http://www.google.com/publicdata/directory
  - dmoz
    - www.dmoz.org/Computers/Artificial_Intelligence/Machine_Learning/Datasets/
  - Netflix Challenge: http://www.cs.uic.edu/~liub/Netflix-KDD-Cup-2007.html
  - Kaggle: https://www.kaggle.com/competitions and https://www.kaggle.com/datasets
- Text
  - Enron email dataset: http://www.cs.cmu.edu/~enron/
  - Web page classification: http://www-2.cs.cmu.edu/~webkb/
- Optical Character Recognition
  - Stanford dataset: http://ai.stanford.edu/~btaskar/ocr/
  - NIST dataset: http://yann.lecun.com/exdb/mnist/

Project Will be Evaluated on

- Novelty of Application
- Models Chosen
- Experimental Results and Analysis
- Report Quality