

Machine Learning

CS578

Class Project

Dan Goldwasser

dgoldwas@purdue.edu

A few words about the final project

- The algorithms and dataset we will look at are not representative of the “real world”
 - *Clean data, attributes well defined, well formed learning problems*
 - *Learning is reduced to choosing the right algorithm, and some hyper-parameter tuning*
- **“Real world”**
 - Messy data, rich data to extract attributes from (most of which are irrelevant), not clear what can actually be learned
- The final project will give you a chance to experiment with these settings

A few words about the final project

- Works in teams (~3 people in a team)
- Several options:
 - ***Come up with your own problem***
 - Grade: **Originality** + **formulation** + **solution** (+analysis)
 - People have worked on this problem in the past...
 - ***Your take on a well known problem***
 - Grade: formulation + **solution (+analysis)**
 - ***Competition***
 - Several teams can “compete” on a similar problem
 - *Has to be approved!*

A few words about the final project

- Examples:
 - Stock market trading
 - Predicting authors' attributes (gender, income, etc)
 - Automated grading
 - Face detection in images
 - Content recommendation
 - Chess/poker playing agent
 -

A few words about the final project

- Deadlines
 - **Find a partner**
 - Should be done by next week!
 - **Project proposal**
 - October 1st
 - **Progress Report**
 - November 5th
 - **Final report and presentation**
 - December 8th

A few words about the final project

- Project proposal
 - **What is the problem you intend to work on**
 - General introduction, *motivation*,..
 - **Relevant background and related work**
 - *How is what you are doing different?*
 - **Formulation as a machine learning task**
 - Rigorous definition of the problem
 - Inputs, output labels
 - Identify training data, and related resources
 - Evaluation plan
- The proposal should not exceed 5 pages
- Should include the sections above
 - You don't get the grade based on # words - ***be concise!***

A few words about the final project

- **Progress report**

- Short (1 page) document describing your progress

- Did anything change since the proposal?

- Data is hard to come by

- You found that the problem is too easy/difficult

- What have you tried so far?

- Do you have some initial results?

- What will you do next?

- **Requirement for this stage**

- Strong definition of your problem, data, baseline results

- Baseline: what is the simplest thing to try? (could be a simple classifier, heuristic, hard-coded rule, etc)

A few words about the final project

- **Presentation**

- Short presentation describing your work (2-3 minutes)
 - Focus on highlights – why is it interesting, what did you try, what worked.

- **Final report**

- Summarize your experience with the project
 - Should extend the proposal (motivation, background, problem definition)
 - Include a results and analysis section
- 6 pages