Competition Based Teaching of Machine Learning

Mikael Vejdemo-Johansson

CUNY College of Staten Island

CUNY College of Staten Island - Mathematics

- Combined community college, 4 year research university
- Over 10k students total

- 100 mathematics majors
- 4 statistics courses
  (…for humanities; …for STEM; probability; mathematical statistics)

Machine Learning Course

- Basic idea:
  - Learn Machine Learning through competitions
  - Introduce theory when needed, guided by competition needs

- Personal preferences:
  - Grade through written essays
  - Computational everything
  - Trust students

Kaggle

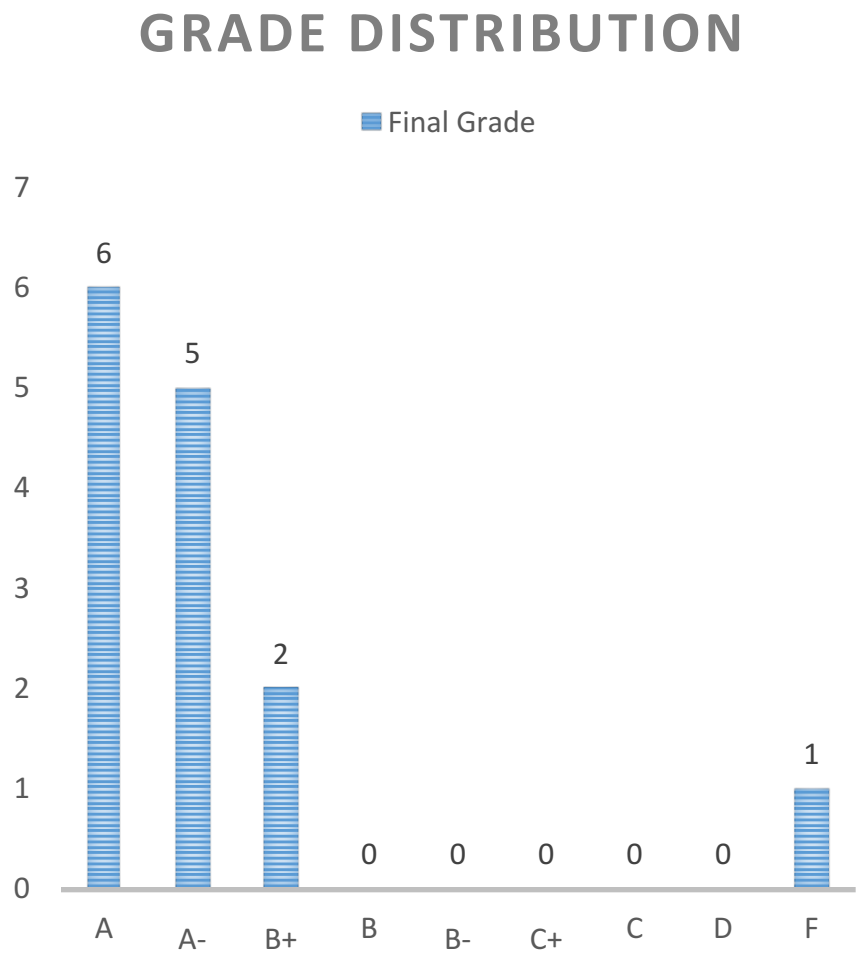
- Online platform for competitions in Machine Learning and in Data Mining
- Provides compute server + Jupyter Notebooks, source code editor, console access for Python and for R
  - 4x CPU or 2x CPU + 1x GPU
  - 6h execution before jobs are killed
  - Integrated competition submission
- ~20 active competitions, many with cash prizes
- Active community
- Support for “Classroom competitions”

First run of the course

- 18 students registered
- 14 students stayed past add/drop deadline
  All 14 remained active through the entire semester
- 4 Kaggle competitions
  - Titanic – to get used to the interface
  - Regression on house prizes
  - Classification of forest types (expired competition)
  - Fashion-MNIST (classroom competition)
- 2x 50 minutes each Monday, Wednesday
- 15-30 minutes theory, remaining time work in teams of 3

First run of the course

- Running Leaderboard on the classroom whiteboard
- Easy to beat teacher reference solution + access to reference solution source code
- Grade based on
  - 70% Blog post describing one of the team's solutions in detail https://medium.com/cuny-csi-mth513
  - 10% Attendance
  - 20% Final exam



Student Opinions

Question	Strongly Agree	Agree	Neither	Disagree	Strongly Disagree
I feel confident that I can recognize the need for a machine learning solution	30%	70%	0%	0%	0%
I feel confident that I can construct a machine learning solution for a new problem	20%	60%	20%	0%	0%
I have learned what I expected to learn from this course	20%	50%	30%	0%	0%
I would recommend this course to other students	30%	40%	30%	0%	0%

10 students responded to the voluntary course survey
Official student evaluations are still being processed

Student Opinions:
What aspects of the course design should be retained?

- Keeping the theme of letting the students play with what they discover. The class should revolve around discovery with loose guidance, that itself almost reflects some of the modern designs of machine learning.
- Working on a competition in groups should be retained for this course.
- The competition aspect is nice, it makes it more interesting.
- The nature of it – working in competition format was effective.

Student Opinions:
How could the course best be improved?

- [Several responses addressed the students own feelings of inadequate coding skills before the course.]
- More lecture time would definitely be helpful.
- Please get local jupyter, I hate Kaggle's latency […]
- I would lecture more before doing a specific competition. That is, give students more tools to play around with before opening the competition.
- Possibly go through the basics more slowly, it may have been a steep jump right away, but as the class progressed it became more understandable.

What worked well?

- The competition format was genuinely and thoroughly motivating. After my teaching observation, the observer was astonished to see how eagerly the students started coding.
- Kaggle's infrastructure makes it easy and comfortable to write code and submit solutions seamlessly.
- Many topics and issues were discovered independently just before being introduced – eg Grid Search, Validation

What worked less well?

- While it was comfortable working on Kaggle's servers, latency and computation times became an issue – drastically so when we started using deep learning techniques.
- Currently live Kaggle competitions might not align with teaching goals.
  We used:
  - One live competition
  - One finished competition – which meant the leaderboard did not work
  - One "Kaggle InClass" competition – which meant no outside competition
- Since all competitions had a sample solution, very little was built from scratch – which was reflected in responses to the final exam.