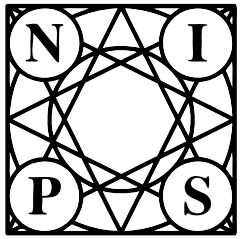
CiML 2016



Lessons learned from the use of Kaggle in Machine Learning Course

Jesus Fernandez-Bes, Jerónimo Arenas-García, and Jesús Cid-Sueiro

ML:DS

ciber-bbn
Centro de Investigación Biomédica en Red
Bioingeniería, Biomateriales y Nanomedicina

Universidad Zaragoza
1542

What this talk is about?

Use of a challenge and *Kaggle inclass* platform as a teaching and evaluation tool in a Machine Learning Course.

- How useful it is as a didactic tool?

- How useful it is as an evaluation tool?

Contents

1. Course and challenge description.
2. Evaluation, code submission and cheating prevention.
3. Results.
4. Conclusions.

Course Description

- Data Processing course
  - Telecommunication Engineering Master. Carlos III University of Madrid.
    - Graduate-level degree with professional (non academic) orientation.
- Student background
  - Electronic and Electrical Engineering + Computer Science.
- Very practical orientation
  - Python *Jupyter* notebooks.
  - Evaluation based on projects.
    - Exploratory Project.
    - Challenge.

Machine Learning challenge

- Why using a competition?
  - Help understanding important concepts.
  - Motivate students solving real problems.
  - Promote self-teaching and discovery of new techniques.

- Why *Kaggle inclass*?

  "Engage students with an opportunity to apply machine learning to real problems."
  - Controlled environment.
  - Leaderboard.
    - Students compete.
    - Teachers can adapt to student response.
  - A forum to discuss approaches.

Energy Generation Prediction

- Data from UCI Machine Learning repository.
  - Combined Cycle Power Plant Data Set.*
  - Regression task with 4 real valued inputs.
  - All labels available online.

- Data manipulation.
  - Shuffle data points.
  - Added Gaussian noise to all variables.
  - Added one extra variable of noise (high variance).
  - Remove 20 % values of 1st variable (highest correlated with the target).

- Performance Measure: Root Mean Square Error.
  - Public: 50% of the test data.

* https://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant

Student Evaluation

- Private Leaderboard position. 1/3
- Minimum viable solution. 1/3
  - Outperform **a Simple Benchmark**: Linear Regression.
- Methodology and Code Quality. 1/3
  - Description of the method used: 3000 characters.

+

  - Delivery of code that generates the final solution.

Why did we ask the code and description?

- Our intention:
  1. Dissuasive measure to prevent cheating.
  2. Not willing to let all student mark be the LB position.
  3. Method to:
     - Prime to "good but unsuccessful" approaches.
     - Punish "bad but successful" approaches.

- Problem:
  - TA has to "manually" review all deliverables in addition to processing LB results.

Competition and Leaderboard

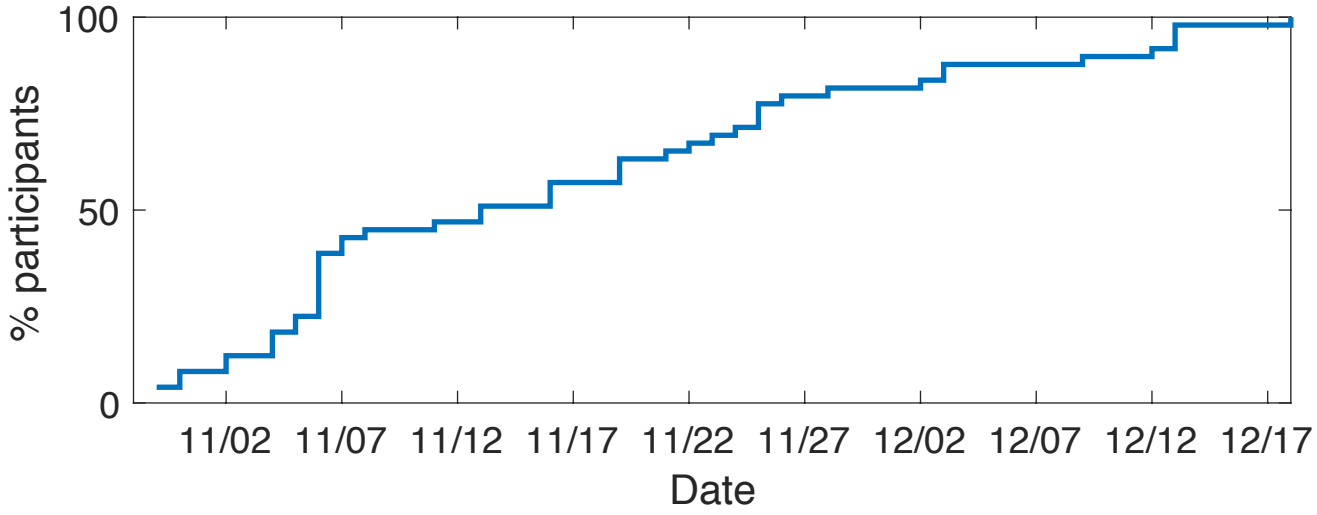
- https://inclass.kaggle.com/c/uc3m-data-processing

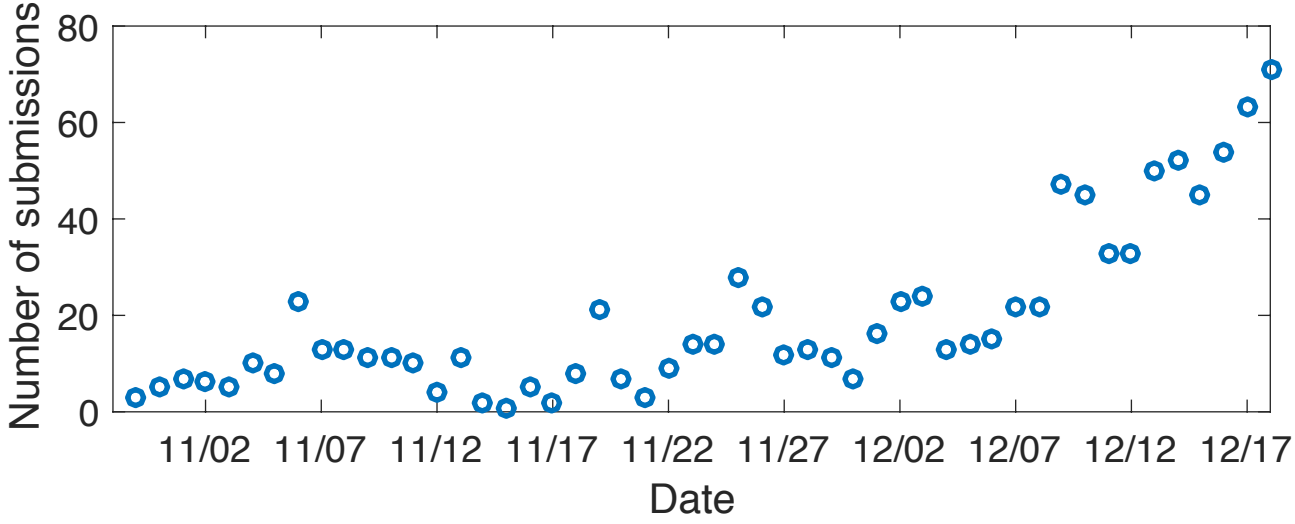
This leaderboard is calculated on approximately 50% of the test data.
The final results will be based on the other 50%, so the final standings may be different.

See someone using multiple accounts?
Let us know.

#	Δ1w	Team Name	Score	Entries	Last Submission UTC (Best – Last Submission)
1	↑3	100349312 & 100343066	4.89652	16	Fri, 18 Dec 2015 23:49:13 (-2.4d)
2	↑6	100291481 & 100292129	4.92282	39	Fri, 18 Dec 2015 23:56:35 (-24h)
3	↓2	100047961 & 100323634	4.92433	24	Fri, 18 Dec 2015 20:24:41 (-24.9h)
4	↓1	100339734&100343064	4.93552	13	Fri, 18 Dec 2015 18:00:32
5	↓3	100290877 & 100079048	4.94604	53	Wed, 16 Dec 2015 23:17:57 (-14d)
6	↓1	100332985	4.95687	38	Fri, 18 Dec 2015 09:36:21 (-7.4d)
7	↓1	100292885 & 100290742	4.95926	26	Fri, 18 Dec 2015 16:48:34 (-4.9d)
8	↓1	100326675	4.96026	32	Fri, 18 Dec 2015 22:49:56 (-7.5d)
9	↑3	100283180&100283008	4.96407	19	Fri, 18 Dec 2015 18:31:42 (-3.9d)
10	↑6	100342332 & 100342514	4.96838	14	Fri, 18 Dec 2015 22:28:32 (-28.7h)
11	↓1	100292791 & 100291357	4.96977	28	Fri, 18 Dec 2015 02:05:51
12	↓3	100290765 & 100293057	4.97173	29	Fri, 18 Dec 2015 17:08:20 (-17.1h)
13	↓2	100282366 & 100291460	4.98787	32	Fri, 18 Dec 2015 19:05:10 (-7.9d)
14	↑5	100293569&100291907	4.98820	40	Fri, 18 Dec 2015 18:03:06
15	↓2	100283367	4.98891	27	Fri, 18 Dec 2015 19:14:13 (-43.3h)

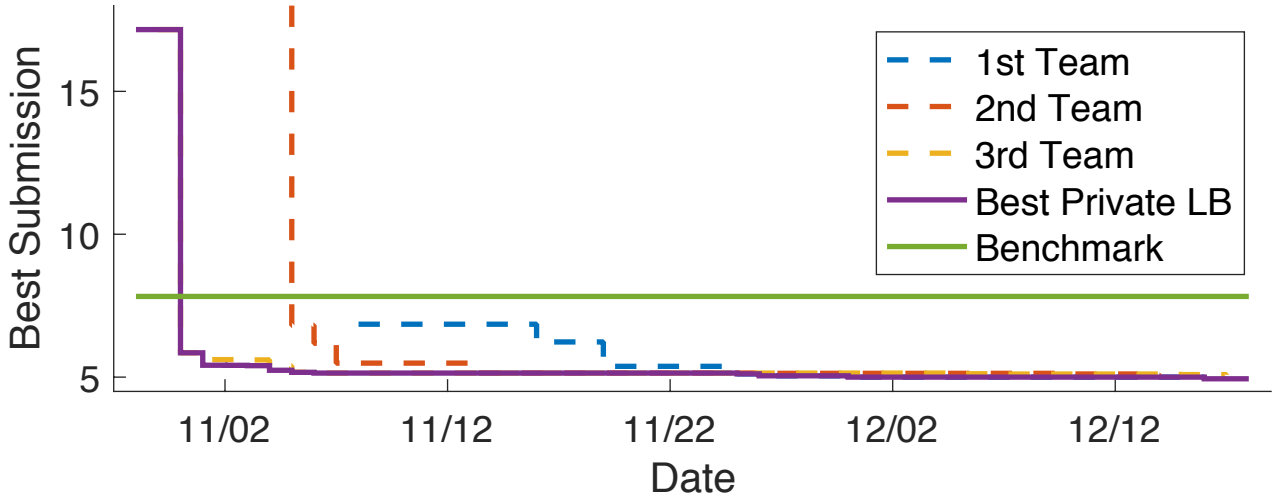
Student Involvement





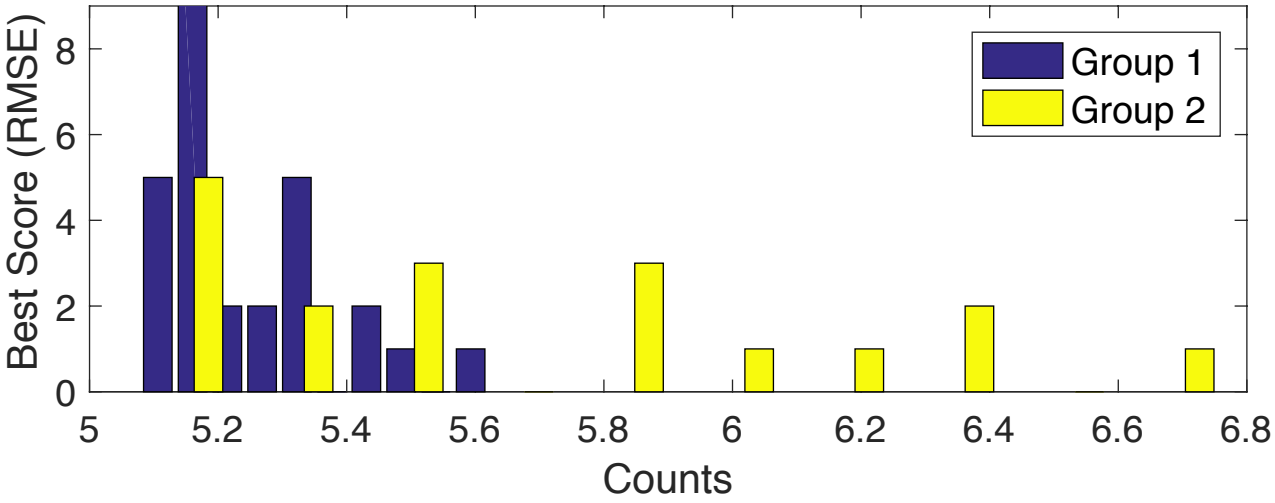
- 85 / 88 students.
- 48 teams.
- 2 submissions at most per day per team.

Student Performace



- Benchmark easily outperformed.
- Regression methods presented in the course:
  - Linear regression.
  - Gaussian Processes.
- But best performance obtained by students who explored new methods, such as **ensemble methods**.

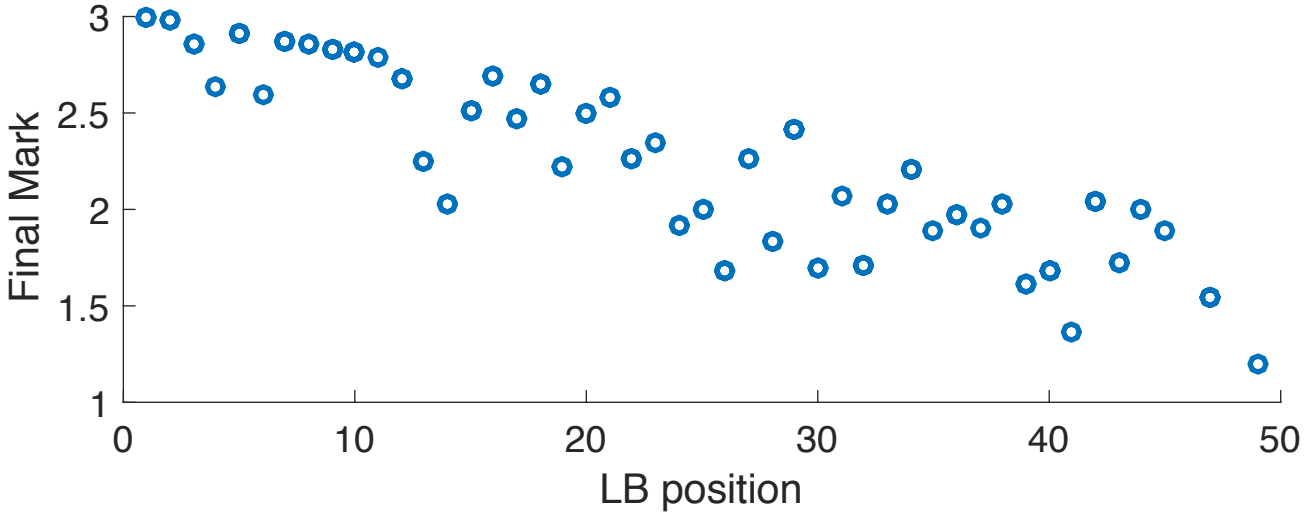
Differences among class groups



Better

- Students slightly had different background.
- Different lecturer.
- Group dynamics.

Evaluation

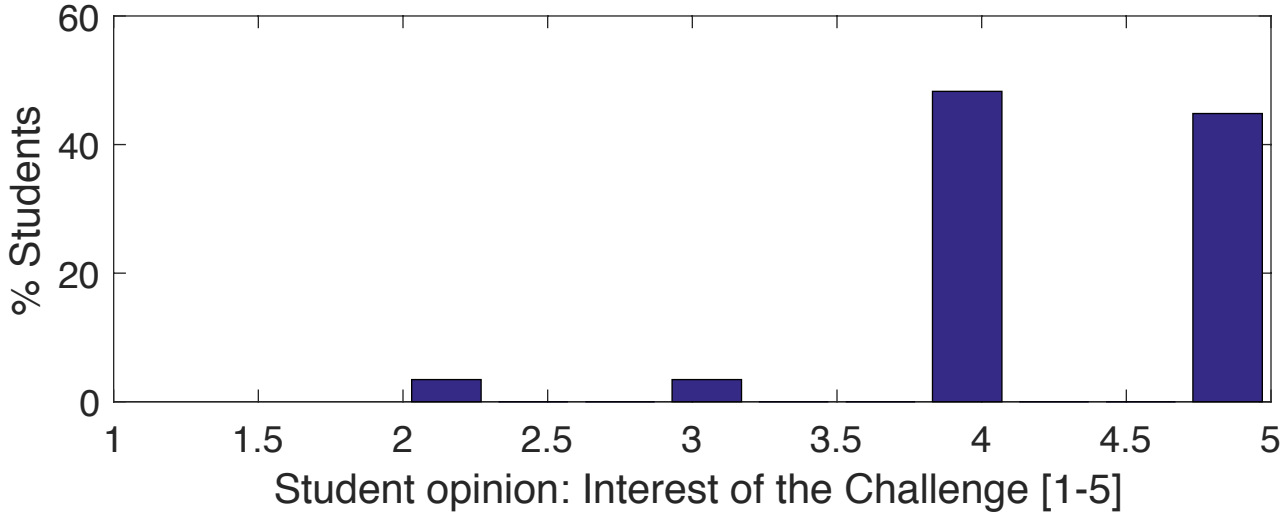


Could we have only used the LB position as mark?

Incidences and cheating

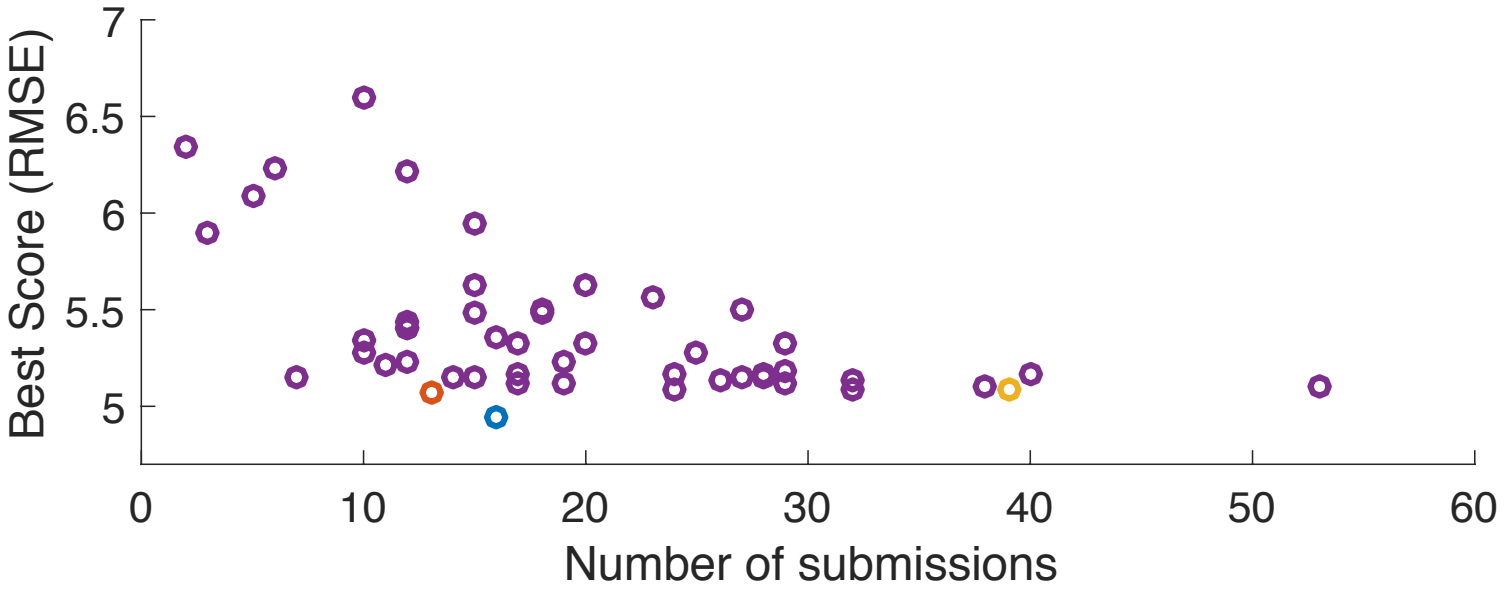
- No two groups got the same RMSE.
  - No direct copy among students.
- The code/performance/method description generally matched.
- Incidences:
  - Two students punished for multiple submissions.

Student Opinion



Main negative comment:
- Not enough time to work in the problem.

Student Opinion



Main negative comment:
- Not enough time to work in the problem.

Discussion

Pros

- The *Kaggle* challenge was considered by the students as their main learning tool.
  - Understood the basic concepts of Machine Learning, such as over-fitting.
- Lot of information about student involvement in the class.
- Forum.

Cons

- Amount of work needed.
  - Identify a suitable dataset.
  - Settle the competition.
  - Check solutions.
- No tool to process submissions.
  - Kaggle provides a tool to download everything.
- Not sure how to evaluate.
- Students unwilling to use Forum.

Future iterations

- This year iteration (finished on wednesday):
  - https://inclass.kaggle.com/c/predict-benzene-concentration
  - We maintained most of presented approach.

- Future:
  - Reduce the manual work for TAs
    - Use of *Kaggle* online scripting functionality (*Kaggle kernels*) -> Not yet.
  - Encourage forum use.
  - Peer Review?

Thanks for your attention!

ACK: Supported by projects TEC2014-52289-R, and PRICAM S2013/ICE-2933.