

# Extracurricular Temptations

Panel at CRA Snowbird, July 2016

Moderator: Ellen Zegura, Georgia Tech  
Panelists: David Liben-Nowell (Carleton),  
Steve Miller (Singapore Management University),  
Margo Seltzer (Harvard)

# Background

- Opportunities for students to develop and use their CS talents outside the classroom are growing
- Long standing: programming contests, hobby programming, TA-ing, campus IT jobs
- Recent: hackathons, contests, competitions, on and off-campus IT jobs
- Obvious differences: many more opportunities, many run outside the university

# Programming Contests

- Relatively long history
  - IOI started 1989
  - ICPC started 1989; IBM sponsorship in 1987 grew program
- International
- Frequently well supported in schools with coaches, team development, etc.



International Olympiad in Informatics

# Hackathons:

short, intense duration; small prizes



# Competitions:

## longer duration, open, bigger prizes

kaggle



#### Sources of Competitive Advantage

- Discipline (once bitten twice shy)
  - Proper validation framework
- Effort
- (Some) Domain knowledge
- Feature engineering
- The "right" model structure
- Machine/statistical learning packages
- Coding/data manipulation efficiency
- Luck

DataRobot



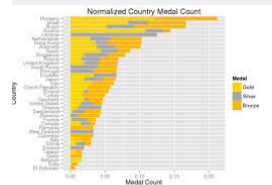
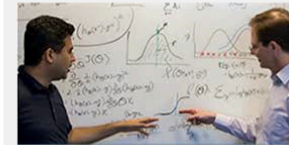
#### DataScienceLA Presents:

#### Winning Data Science Competitions

by Jeong-Yoon Lee

#### Open Source Tools and DS Competitions

Owen Zhang  
Data Scientist @ DataRobot Inc.



#### Winning Data Science Competitions

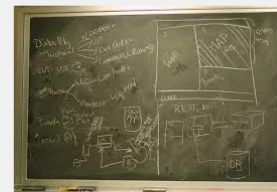
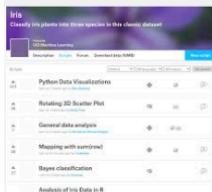
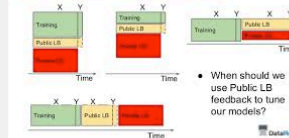
Some (hopefully) useful pointers

Owen Zhang  
Data Scientist

DataRobot



#### Kaggle Competitions - Typical Data Partitioning



#### How to win data science competitions with Deep Learning

@ArnoCandel

Silicon Valley Big Data Science Meetup  
Deddata Campus, Mountain View, Oct 9 2014

#### It is Great that computation is cheap!

- Proper K-fold cross validation of stacking model requires building K\*(K-1) models. Below are models required for a single fold (4).



#### Kaggle Data Science Competitions

##### Meta Data Science Competitions

##### Competition Attributes:

- Duration
- Data
- Problem Statement
- Problem Evaluation Data Set Size
- Public
- Private
- Time Based
- Leaderboard
- Submission Frequency
- Submission Format
- Release on Public



#### Make Validation Dataset as Realistic as Possible

- Usually this means "out-of-time" validation
  - You are free to use "in-time" random split to build models, tune parameters, etc.
  - But hold out data should be out-of-time
- Exception to the rule: cross validation when data extremely small
  - But keep in mind that your model won't perform as well in reality
  - The more times you "tweak" your model, the bigger the gap.

DataRobot

# What Role Can a University Take?

- Host a hackathon (space, network, access)
- Run an internal competition with little formal learning support (e.g., GT: Inventure Prize)
- Run an internal competition, coupled to co-curricular learning opportunities such as seminars, tutorials (e.g., GT: Mobile Apps)
- Run an internal competition coupled to a class: (GT: Capstone Design Expo)

# Discussion Questions

- What role should a university take?
- What conflicts arise and how can they be mediated?
- How can competitions be structured to appeal to all students? Should they be? What cultures are reinforced by competitions?
- Are competitions good preparation for the real world?

# Why Do Students Participate?

- Creative opportunity
- Learning opportunity
- Social
- Do “good”
- Resume builder



# Resume Building

- “I went to some career talks from Google and Facebook. The recruiters all said we should participate in some hackathons to show our passion towards working on interesting projects on our own.”

# Skill Development

- “I think the hackathon gave me a boost on a particular programming skill that I had not used for awhile. I teamed with a few folks and worked on a project using android wear. I hadn’t coded with java for a long time. Within that 36 hours, I got back to java and did something fun with it.”

# Balance to PhD Research

- “I think one big thing is, it is fun and exciting and social. As a PhD student, we often work on a project on our own for quite a long time. It may get boring and the excitement may decrease over time. I think attending a hackathon revives our passion on hacking things [to some extent]. It reminds me what I like about computer science and my passion of building something fun (and useful/helpful).”

# Diverse Teams/Little Continuity

- “I loved that <organizer> had made the effort to bring all kinds of domain experts to the event, and from all over India. I learned a lot in the first one, from working with students/participants from start-ups, colleges, foundations, the government... this is also why I had attended it. As a side note, my team won first prize and a lot of money, so I liked the experience even more :).”
- “I went again the next year and it was the same, only bigger, but I realized that only two teams from the first year were still working on their idea (including us, and that made us superstars of some sort).”

# Too Intense

- “I did not like them much for couple of reasons...Also, the continuous coding for a span of 24-48 hours is too exhausting. May be the set up in the US colleges is better where there is more participation and a lot of people from industry to interact with.”