ORIE 5355: People, Data, & Systems

Lecture 20: Conclusion

Nikhil Garg

Course webpage: https://orie5355.github.io/Fall_2021/

Announcements

- Quiz 5 due tomorrow night
- Fill out the official course evaluation by 12/10 – part of participation grade!
- Project report due 12/14 night
  - Individually submit a project/course feedback
  - Optionally submit a short peer assessment for your project team-mates
  - Will allow resubmissions for Project Part 1 and 2

    Please *do not* resubmit to “optimize”/squeeze last bit of performance; only resubmit if you weren’t able to put in time so far or if you believe small errors made your performance not reflect the time that you have already put in.
- No more regular office hours. Short individual office hours by appointment as needed (information to be announced)

Fill out the course evaluation

It matters:

- **Will they let me teach the course again?**
- **Will they let me create new electives?**
- **Will they fire me?... Or will they give me teaching awards?**

We did it! Final day

What I told you this class would be about

What is this class *not* about

Intro to data science → **Urban Data**

How to train many types of cool prediction models → **Applied Machine Learning**

Deep learning → many other courses

How to scale data science → **Data Science in the Wild**

Theoretical analysis of markets (pricing, queuing) → **Service Systems and Online Markets**

What *is* this class about?

Let's assume you know the machinery for training **prediction** models

Then, how do you make (algorithmic) **decisions**, in the presence of

- User incentives, strategic behavior, adversarial behavior
- Data constraints – censoring, selection effects, limited data
- (Perceived) fairness and ethical constraints
- Business/resource constraints – capacities, communication limits
- Competition
- Changing environments

Who is this course for?

You know (or are learning right now) how to train and evaluate (basic) prediction models

You want to learn how to actually use such these models in:

- Online marketplaces (Uber, Upwork, Etsy, Netflix, Stitchfix…)
- Government (transportation, polling, census, providing services)

…any data science setting about people

What we did

For each of these, I'd love folks to raise their hand and share what they found the most fun/insightful/educational/valuable/surprising/useful

Data collection and processing

- What *is* data? Where does it come from? What does it *represent?*
- Common challenges in data collection

  Selection biases, censoring, and other challenges
- Polling/surveys as an extended example
  - What goes wrong in measuring opinions (mean estimation)
  - Some techniques that somewhat work
  - US 2016 election polls as a case study
- Other challenges and contexts: online ratings, privacy, etc.

Guest lecture from long-time FiveThirtyEight data journalist and now Columbia professor, Dhrumil Mehta

Analysis: Recommendation

- Basics of recommendation: collaborative filtering and matrix factorization
- Individual- vs demographic-based personalized recommendations – tackling the cold-start or low data regime in practice
- Other challenges for recommendation in practice, including matching and capacity constraints

Guest lecture from Amy Zhang, PhD student at Cornell

Analysis: Pricing

At what price do I sell my items?

- Optimal pricing given market data
- Personalized pricing – individual vs demographic based
- Dynamic pricing – pricing over time
- Exercise on when personalized and dynamic pricing is ethical
- Case study: prices and wages in online marketplaces

Experimentation

How do I know if my product/intervention/action/treatment works?

- A/B testing basics: clinical trials, standard experimentation
- Why standard techniques fail in people-centric systems: interference
- Experimentation in practice
  - Dealing with networks and interference
  - Experimentation over time
  - Experiments in 2-sided marketplaces
  - Running many experiments
- Synthetic control

Guest lecture from Hannah Li, PhD student at Stanford

Miscellaneous

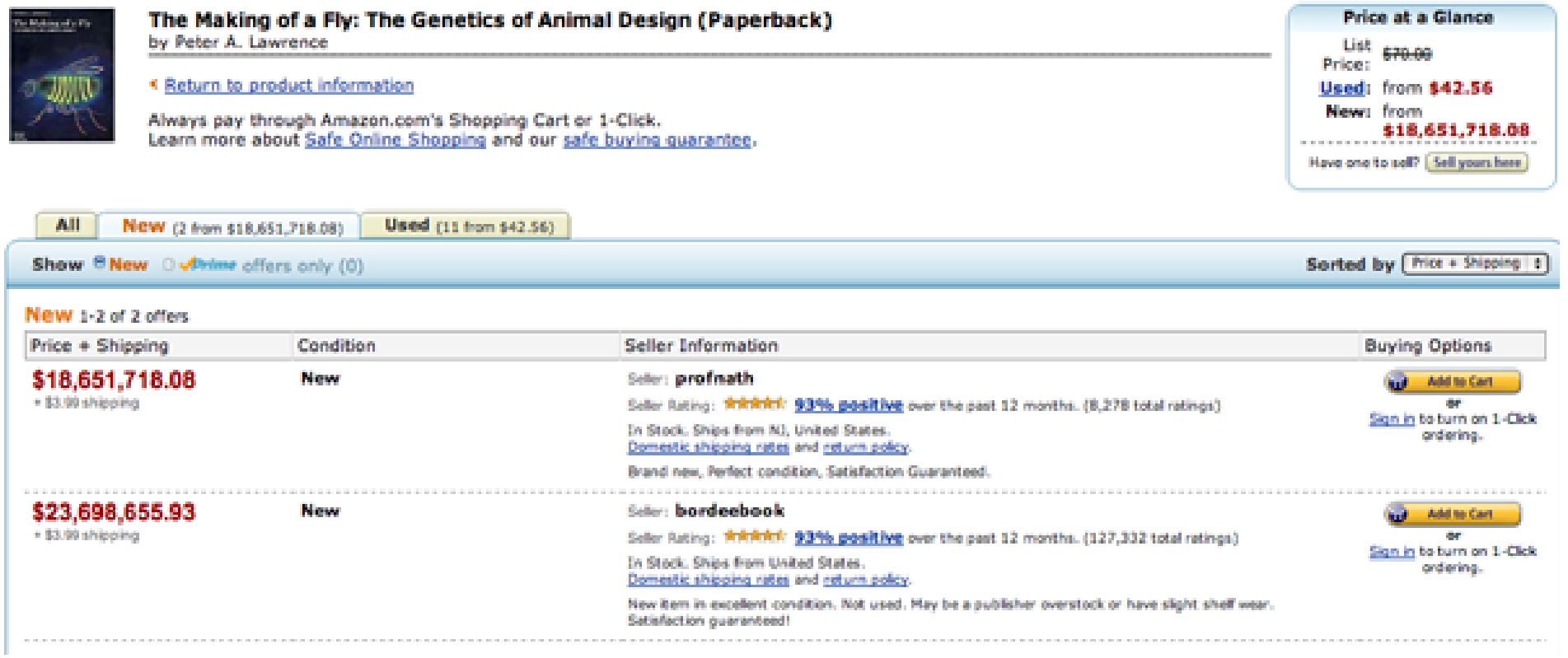
- Differential privacy: how do I share personal data in a “optimally” private manner?
- Limits to prediction
- Discrimination and design mitigations in online platforms

Guest lecture from Lily Xu, PhD student at Harvard

Guest lecture from George Markoulakis and Jaron Malcom (Clipboard Health)

Class competition

Example: algorithmic pricing gone wrong with competition



The Making of a Fly: The Genetics of Animal Design (Paperback)
by Peter A. Lawrence

Return to product information

Always pay through Amazon.com's Shopping Cart or 1-Click.
Learn more about Safe Online Shopping and our safe buying guarantee.

Price at a Glance

List Price: $70.00
Used: from $42.56
New: from $18,651,718.08

Have one to sell? Sell yours here

All | New (2 from $18,651,718.08) | Used (11 from $42.56)

Show: New ☐ Prime offers only (0) — Sorted by: Price + Shipping

New 1-2 of 2 offers

Price + Shipping	Condition	Seller Information	Buying Options
$18,651,718.08 + $3.99 shipping	New	Seller: profnath Seller Rating: 93% positive over the past 12 months. (8,278 total ratings) In Stock. Ships from NJ, United States. Domestic shipping rates and return policy. Brand new, Perfect condition, Satisfaction Guaranteed.	Add to Cart or Sign in to turn on 1-Click ordering.
$23,698,655.93 + $3.99 shipping	New	Seller: bordeebook Seller Rating: 93% positive over the past 12 months. (127,332 total ratings) In Stock. Ships from United States. Domestic shipping rates and return policy. New item in excellent condition. Not used. May be a publisher overstock or have slight shelf wear. Satisfaction guaranteed!	Add to Cart or Sign in to turn on 1-Click ordering.

[The Infinite Loop Of Algorithmic Pricing On Amazon... Or How A Book On Flies Cost $23,698,655.93 | Techdirt]

Part 1 results + discussion

Goal: bring together recommendation and pricing modules, and start introducing computational runtime constraints in practice

Results

1st place: **teamsvm**

2nd place: **thethreemusketeers**

3rd place: **datapeople**

Discussion

- How did you do demand estimation? How did you deal with missing data?
- What were your strategies for price optimization? How did you know whether you needed to “keep going” in terms of number of samples?

Part 1 results + discussion

Goal: First-hand experience with incentives/game theory, and how it makes everything else we covered in class challenging

Results

1st place: **teamsvm** (1221 revenue/game)

2nd place: **espionage** (989 revenue/game)

3rd place: **johnandaliya** (984 revenue/game)

Discussion

- How did you adapt to your opponent?
- How did the goal of “maximizing revenue” vs “win each game” affect your strategies?
- Anything else that you did that you’re proud of and want to share with the class?

Conclusion

Class themes

- The right data and interpretation beats modeling sophistication
- The solution is not (always) technical
- You’re not done after training a model, and you don’t start there either
- Most mistakes are made in understanding and applying the basics
- Domain expertise is essential
- Design with privacy, ethics, and fairness in mind – not as an afterthought
- Historical performance on training set is often a terrible measure
- Concepts, not (just) methods
- Be curious!

Class themes

- The
- The solution is not (always) technical
- You
- Mos
- Dom
- Des
- Hist
- Con
- Be c

Understand the basics

Understand your context

Understand limitations

Understand and respect people

either

s

thought

This course was the highlight of my semester! Thank you

Stay in touch:

ngarg@cornell.edu

(@NikhGarg) / Twitter

gargnikhil.com

Fill out the course evaluation

It matters:

- **Will they let me teach the course again?**
- **Will they let me create new electives?**
- **Will they fire me?... Or will they give me teaching awards?**

Announcements

- Quiz 5 due tomorrow night
- Project report due 12/14 night
  - Will allow resubmissions for Project Part 1 and 2

    Please *do not* resubmit to “optimize”/squeeze last bit of performance; only resubmit if you weren’t able to put in time so far or if you believe small errors made your performance not reflect the time that you have already put in.
  - Also individually submit a short peer assessment for your project team-mates
  - And individually submit a project/course feedback (+ official evaluation)
- No more regular office hours. Short individual office hours by appointment as needed (information to be announced)
- Fill out the course evaluation