

Research Seminar Data Science

Teachers: Elena Marchiori & Arjen de Vries

Objectives

- Learn how to evaluate research papers
- Learn what makes papers good
- Learn about how papers are refereed and published
- Obtain an overview of important recent developments in Data Science research
- Note: course inspired by the same course for Software Science at Imperial College, and some material then also heavily borrowed from that course
<http://www.doc.ic.ac.uk/~dirk/Teaching/RS/>

Two parts

1. Review a paper submitted to a top data science conference or journal
2. Interpret and analyze a seminal, “classic” paper
 - present your own
 - short summaries of those of others + of one presentation

Review a (potentially) great paper

- Paper submitted to one of the main conferences in data science, such as ICML, UAI
- You will get assigned a “real” paper to review and a coach to help you
- Rough schedule (for ICML, UAI):
 - 24 February (ICML), March 31 (UAI) paper submission deadlines
 - April 13 (ICML), May 14 (UAI) review deadline
 - April 14-19 (ICML), May 15-21 (UAI), author feedback
 - April 20- May 11 (ICML), May 22-31 (UAI) reviewer discussion
- <http://www.auai.org/uai2017/importantDates.php>
- <https://2017.icml.cc/>
- Output: single report with
 - Your own review
 - Reflection on the review procedure as a whole

Info on reviewing

- Example reviewing criteria: <http://www.auai.org/uai2015/reviewCriteria.shtml> http://icml.cc/2015/?page_id=192
<http://www.informagus.nl/sigir2017/review/guidelines-pc-fp/reviewing.html>
- How to review a paper: <http://www.inf.ethz.ch/personal/troscoe/pubs/review-writing.pdf> or
<https://www.cs.princeton.edu/~jrex/teaching/spring2005/fft/reviewing.html>
- How *not* to review a paper:
<https://sigmodrecord.org/publications/sigmodRecord/0812/p100.open.cormode.pdf>
- The NIPS consistency experiment: <http://blog.mrtz.org/2014/12/15/the-nips-experiment.html>

Confidentiality

- Paper reviewing for UAI, ICML is double blind
- You should *not* try to identify the authors
- Nor share the paper with someone else
- Same holds for the reviewer discussions / author feedback

Seminal paper

- Pick a seminal paper that had major impact, possibly but not necessarily related to the paper you reviewed
- Present the paper in class
 - 15 minutes presentation to **initiate discussion** (i.e., no need to give a full-fledged summary)
 - 15 minutes discussion
- Write a review on the paper presented (about 5-6 pages)
- Others: write short summaries (half a page) *before* the presentation

Presenting a paper

- Main goal: lively interesting talk that promotes discussion
- Outline:
 - Objective of the paper
 - Proposal of the paper
 - Evidence given
 - Shoulders of giants
 - Impact
 - Discussion points
- Same outline for written review + issues from discussion

Short review

- Half a page, maximum one page
- Clearly-separated (use subheadings) sections covering
 - Summary (as briefly as you can – two or three sentences)
 - Evidence (what evidence is offered to support the claims?)
 - Strengths (what positive basis is there for publishing/reading it/*listening to it*?)
 - Weaknesses
 - Evaluation (if you were running the conference/journal where it was published, would you recommend acceptance/*would you invite the presenter for a talk*?)
 - Comments on the quality of the writing/*presentation*
 - Plus: Queries for discussion
- Deadline: 1 day before presentation/*1 week after presentation*

(in *italic*: for review on presentation)

How to find a seminal paper

- Google scholar...
- For a recent, e.g., < 3-year old paper: you'd like to see at least 40 citations per year
- For a not-so-recent paper: you'd like to see at least 200 citations in total
- In case of doubt: send an email or make an appointment to discuss

Possible symposia and seminars

- ICTOpen
- <http://www.ictopen.nl/>
- March 21-22, Amersfoort
- Master students: 75 euro
- Includes Data Science and Machine Learning themes

- Thalia symposium
- <https://symposium.thalia.nu/>
- This year's subject will be "Parallel Computations".
- February 10

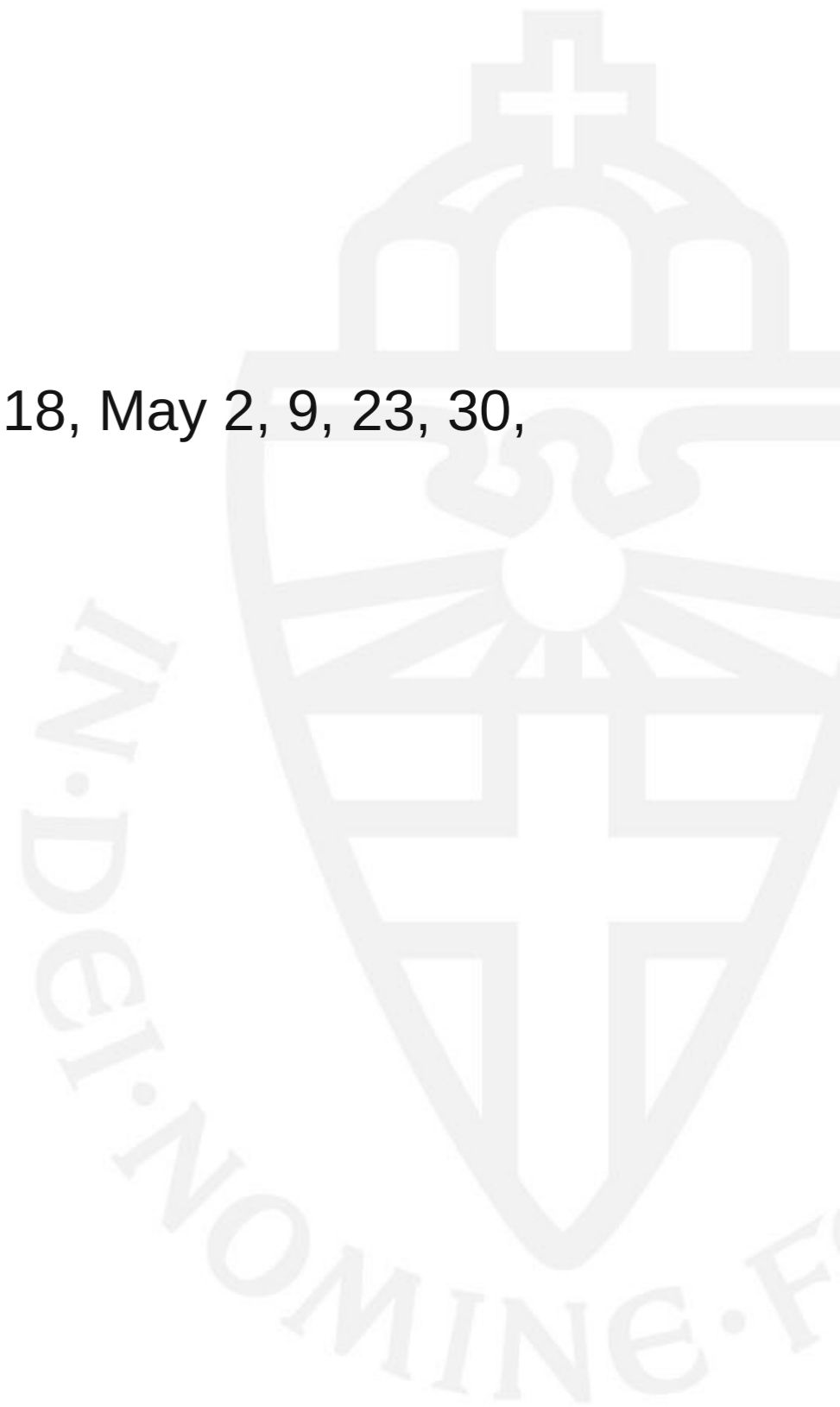
- Data science weekly seminar
- <http://www.ru.nl/datascience/research/seminars/>

Preliminary planning part 1

- End February: all papers assigned for review (for UAI one month later)
- Mid April: review deadline (if you want it to have a real impact, again one month later for UAI)
- End April: other reviews available
- June 18: final report (review + reflection on other reviews)

Preliminary planning part 2

- Presentations: February 21, March 7, 14, April 11, 18, May 2, 9, 23, 30, June 6, 7, 13, 14
- Deadline review: 2 weeks after presentation



Output and assessment

- Review + reflection paper under review: deadline June 18, 40%
- Presentation + review seminal paper: deadline 2 weeks after presentation, 40%
- Short summaries “other” seminal papers + similar report on one presentation at a symposium or seminar: deadline 1 day before paper presentation / 1 week after presentation, 20%
- **Plagiarism:** If you use material written by someone else, make sure you acknowledge the source. Making effective use of sources is encouraged (provided copyright is respected).

What to do next?

- Check out the websites for ICML, UAI and skim through some conference proceedings of recent years
- Send me (elenam@cs.ru.nl) a short description of what type of papers you're mostly interested in by **February 15** the latest. For example, indicate for which conference you'd like to review and possibly what topics (see e.g., <http://www.auai.org/uai2017/topics.php> for UAI; those for ICML are more difficult to find); We will try to distribute papers accordingly, but can't promise, in particular because of coach constraints and different timings...
- Fill in the form at <http://bit.ly/RSDS-prefs> by **February 10** the latest. We will try to make a schedule as soon as possible.