Stats 337: Annotated Bibliography

June 6, 2018

1 Executive summary

The articles in this bibliography broadly center around ethics in data science. I chose to synthesize a collection of articles on this topic because data ethics is something that affects all aspects of the data science community and society. Internally, ethics is an area that is affected by all components of a company's vertical pipeline. Even though externally the company's ethics practices are guided by decisions made at the top, those in technical roles, such as software engineers and data scientists crucially affect ethical output. The technical papers in this bibliography are interesting in this regard because they show how careful ethical consideration should be taken into account even when working with rather mundane statistical or machine learning tasks. In some cases, simple classifiers can be extremely successful when used for adversarial privacy attacks. For example, [ZG09] use rows of an adjacency matrix to predict private traits, [SSGN17] use web links to identify Twitter profiles, and [Gol18] shows that simple n-gram language models can store sensitive information. This suggests that data scientists must be vigilant when developing systems that are robust to such attacks.

More broadly, a motif in the discussion around ethics for data science is that ethical behavior is often at direct odds with practices that might otherwise be considered "good data science." Recently in the statistics community, much research focus has been placed on theoretical analysis of "modern data sets," such as high-dimensional datasets, concerning topics such as valid selective inference in regimes of sparse signal. Generally speaking, collecting more features is a good thing in terms of predictive performance, and it's even better if those features have large numbers of bits of information, which is usually the case when the data are sparse. But sparse, high-dimensional feature vectors are particularly vulnerable to privacy and anonymity attacks, precisely because of their utility in predictive machine learning. How should one trade off between the desire to do well in machine learning tasks, and the desire to maintain a sense of data privacy?

As another example, much of the discussion in this class revolved around data availability and the reproduceability and replicability of scientific studies. Generally, it is of interest to science and the academic community if others have easy access to reproduce a result. But as more and more scientific work revolves around sensitive data, the line becomes muddied between how easily data and code should be made available and the protection of datasets that can reveal sensitive information. In the medical community,

and for most social science academic work, these protections take precedent through IRB approval. However, a large portion of data science output comes from industry, and so one wonders whether similar safeguards should be in place in those settings, especially for companies that manage information on human subjects. Facebook has launched a new independent committee, to be led by academics Gary King and Nate Persily, to provide independent, external decisions regarding the types of research and data accessibility that occurs at Facebook [SG18]. It will be interesting to see what role this committee ends up playing and whether it sets a precedent for other similar bodies. On a related note, the implementation of the General Data Protection Regulation (GDPR) in the European Union will be an interesting case study for its effects on data scientists [GF16].

Since the evidence summarised in the above paragraphs shows that data scientists and statisticians need to worry about ethics in their daily roles, it is important to consider how we provide data scientists with the education necessary to accomplish this task. Currently, ethics education within the data science curriculum in masters or PhD level courses in academic is either severely lacking or non-existent all together. But there is evidence to suggest that ethics education can successfully be incorporated into standard data science training. To see this, consider the p-hacking, data snooping, and the replication crisis taking place in psychology, medicine, and other fields, which forms a small part of the broader data science ethics quandary [Ioa05, MGG$^+$17]. While there are still problems, this issue is now widely known, is widely taught in introductory statistics and data science courses including at Stanford, and was even featured in a segment of Last Week Tonight with John Oliver [Lop16]. This suggests that such a movement could be a precursor to an overall improvement in ethics education within data science curricula.

Finally, another theme that makes the ethics discussion tricky is the disconnect between what data scientists are actually doing and public understanding of what they think is being done. For example, many of the controversial human subject experiments are actually quite mild when you look at the actual findings [KGH14], but media sensationalism leads to ethics scandals that make it hard to distinguish between the ethics problems that we should really be worrying about and non-issues that have emerged in the court of public opinion. Similarly, the inane questions from U.S. senators at Mark Zuckerberg's testimony were due to a lack of technological understanding on the senators' part, and prevented the testimony from serving as a valuable forum for the difficult ethics questions that our society truly needs to be having [Bye18].

2 Top 3

Facebook "emotional contagion" study [KGH14] This describes an experiment that Facebook conducted in 2013 in which they edited the proportion of positive and negative emotional words used in news feeds, according to some "word positivity" score. The outcome of the experiment was a statistically significant change in the positivity and negativity in posts by people who were exposed to such news feed changes. I like this paper because it highlights a number of aspects of the ethical problem. The findings themselves were quite mundane: (a) the effect sizes were extremely small, (b) they needed

a sample size of nearly 700,000 to establish statistical significance, and (c) it is a stretch to move from slight changes in word choice to concluding that people's emotional state was changed. But the result was marketed as "emotional contagion," which due to media coverage started a huge controversy about the experimentation that is routine at tech companies such as Facebook.

Predicting personality from Facebook likes [KSG13] This paper, published in *PNAS* three years before the 2016 election, is a sort of spiritual precursor to the data ethics issues surrounding the recent Cambridge Analytica controversy. They showed that some axes of personality traits, as measured by the Big Five personality traits model or *OCEAN* (openness, conscientiousness, extraversion, agreeableness, and neuroticism), are predictable from then-public information on Facebook likes. The technical methods are quite simple; they use only a simple principal components logistic regression on the matrix of attributes.

Abandon p-values [MGG+17] This short paper provides a new recommendation for how p-values ought to be used in science. It argues against the use of frequentist null hypothesis testing based on arbitrary p-value cutoffs. The authors don't recommend discarding p-values entirely, but suggest that they are used as one piece of a large body of factors presented for scientific evidence. These other factors, which they call *neglected factors*, include such qualititative factors as prior and related evidence, study design, data quality, and real world cost benefit analysis.

References

[BB17] Solon Barocas and Danah Boyd. Engaging the ethics of data science in practice. *Communications of the ACM*, 60(11):23–25, 2017.

[BBC+18] Theo Bertram, Elie Bursztein, Stephanie Caro, Hubert Chao, Rutledge Chin Feman, Peter Fleischer, Albin Gustafsson, Jess Hemerly, Chris Hibbert, Luca Invernizzi, Lanah Kammourieh Donnelly, Jason Ketover, Jay Laefer, Paul Nicholas, Yuan Niu, Harjinder Obhi, David Price, Andrew Strait, Kurt Thomas, and Al Verney. Three years of the right to be forgotten. 2018.

> Internal paper from Google describing how the company deals with requests to delete search engine entries containing personal information.

[Bye18] Dylan Byers. Senate fails its zuckerberg test. CNN, `http://money.cnn.com/2018/04/10/technology/senate-mark-zuckerberg-testimony/index.html`, April 11, 2018.

> Discussion of Senate questions for Mark Zuckerberg's April 2018 testimony.

[DR+14] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014.

An overview paper on differential privacy.

[Eck10] Peter Eckersley. How unique is your web browser? In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 1–18. Springer, 2010.

[GF16] Bryce Goodman and Seth Flaxman. European Union regulations on algorithmic decision-making and a "right to explanation". *arXiv preprint arXiv:1606.08813*, 2016.

Some discussion of the implication of GDPR, which just went into effect in the EU, for computer scientists and data scientists.

[GKW+17] Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the united states. *Proceedings of the National Academy of Sciences*, 2017.

[Gol18] Yoav Goldberg. 4gram language models share secrets too... GitHub, `https://gist.github.com/yoavg/40d01b5df1014d9237157902926d20c6`, February 28, 2018.

Shows that most machine models, even trivially simple ones, can contain identifying information; i.e., an n-gram model trained on a corpus that contains the phrase "My SSN is ..." should not be used for an iPhone autocomplete engine.

[Hid16a] Cesar Hidalgo. What I learned from visualizing Hillary Clintons emails. Medium, `https://medium.com/mit-media-lab/what-i-learned-from-visualizing-hillary-clintons-leaked-emails-d13a0908e05e`, November 4, 2016.

[Hid16b] Cesar Hidalgo. What I learned the night of the election, and what I would like to see in the future. Medium, `https://medium.com/@cesifoti/what-i-learned-the-night-of-the-election-and-what-i-would-like-to-see-in-the-future-68b5e49f8721`, November 8, 2016.

[HM15] Eric Horvitz and Deirdre Mulligan. Data, privacy, and the greater good. *Science*, 349(6245):253–255, 2015.

Discusses how policy and regulation could play a role in machine learning ethics.

[Ill68] Ivan Illich. To hell with good intentions. 1968.

[Ioa05] John PA Ioannidis. Why most published research findings are false. *PLoS medicine*, 2(8):e124, 2005.

A highly-cited article that raised awareness about p-hacking.

[KGH14] Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

Controversial Facebook experiment that edited the contents of users' news feed.

[KSG13] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.

A precursor paper to the current Cambridge Analytica controversy. Showed that Facebook friendship likes are predictive of personality traits.

[Lip17] Andrew Liptak. Uber tried to fool Apple and got caught. The Verge, `https://www.theverge.com/2017/4/23/15399438/apple-uber-app-store-fingerprint-program-tim-cook-travis-kalanick`, April 23, 2017.

I find this amusing. Uber wanted to track the location of individual users' phones, which is against Apple's App Store policy. So they disabled the feature in Cupertino, CA, where Apple employees test the apps.

[Lop16] German Lopez. John Oliver exposes how the media turns scientific studies into "morning show gossip". Vox, `https://www.vox.com/2016/5/9/11638808/john-oliver-science-studies-last-week-tonight`, May 9, 2016.

[Luc16] Michael Luca. Were OkCupids and Facebooks experiments unethical? Harvard Business Review, `https://hbr.org/2014/07/were-okcupids-and-facebooks-experiments-unethical`, July 29, 2016.

[Mem12] Mark Memmott. N.Y. website posts map of people with gun permits, draws criticism. NPR, `https://www.npr.org/sections/thetwo-way/2012/12/26/168075748/n-y-website-posts-map-of-people-with-gun-permits-draws-criticism`, December 26, 2012.

[MGG+17] Blakeley B McShane, David Gal, Andrew Gelman, Christian Robert, and Jennifer L Tackett. Abandon statistical significance. *arXiv preprint arXiv:1709.07588*, 2017.

Proposes a (partial) remedy of the replication crisis by suggesting that we discard null hypothesis testing all together and simply use p-values as one of many pieces of information.

[SG18] Elliot Schrage and David Ginsberg. Facebook, `https://newsroom.fb.com/news/2018/04/new-elections-initiative/`, April 9, 2018.

Facebook's new independent committee to help guide research decisions.

[SSGN17] Jessica Su, Ansh Shukla, Sharad Goel, and Arvind Narayanan. De-anonymizing web browsing data with social networks. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1261–1269. International World Wide Web Conferences Steering Committee, 2017.

Concerns identifying Twitter profiles using web links.

[Swe00] Latanya Sweeney. Simple demographics often identify people uniquely. 2000.

Uses census data to show that not much information is needed to identify individuals.

[Wak17] Daisuke Wakabayashi. Google will no longer scan gmail for ad targeting. The New York Times, `https://www.nytimes.com/2017/06/23/technology/gmail-ads.html`, June 23, 2017.

The title is self-explanatory. Google's ability to combine personal information across email, calendar, search, etc. provides powerful tools that have certainly made my life easier, but it's interesting to consider where the line should be.

[ZG09] Elena Zheleva and Lise Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *Proceedings of the 18th International Conference on World Wide Web*, pages 531–540. ACM, 2009.

A technical article that shows that simple machine learning classifiers built on low-level features, such as adjacency matrix entries in a friendship network, can have surprisingly good predictive accuracy; this has implications for privacy attacks with access to information such as Facebook friendship data which was once public.