Syllabus PH C240C: Computational Statistics with Application in Biology and Medicine

Instructor: Jingshen Wang

August 22, 2021

1 Course overview

- Instructor: Jingshen Wang (jingshenwang@berkeley.edu)
- Lecture: Wednesday 6-9 pm
- Office hour: Friday 1-2 pm
- GSI: Lei Shi (leishi@berkeley.edu)
- Lab time: TBD
- Suggested (fun?) readings:
  1. Computer Age Statistical Inference, Efron and Hastie (2016)
  2. Machine Learning: A Probabilistic Perspective, Murphy (2012)
  3. Dynamic Treatment Regimes: Statistical Methods for Precision Medicine, Tsiatis (2019)
  4. Personalizing Precision Medicine: A Global Voyage from Vision to Reality, Pothier (2017)
  5. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again? Topol (2019)

2 Course Scope

Machine learning algorithms are widely applied in our daily lives. The overarching goal of this course is to provide students with an overview and hands-on experiences of popular ML methods

adopted in the healthcare system and medical research. The detailed course content can be found in the course schedule section. The scope of the course is threefold:

1. To facilitate understanding of popular ML methods adopted in analysing large-scale biomedical databases and electronic medical record data
2. To provide students with real world examples showing how ML methods provide us with coaching to promote our health, shape our diet, and even prevent illness
3. To foster continued engagement with the changing field of ML in medical research through discussions of work inside and outside each of our immediate areas of focus

3 Course Schedule

1. Introduction and Scope (09/01)
2. Supervised learning:
   (a) GLM/SVM (09/08)
   (b) Empirical Risk Minimization and Kernel Methods (09/15)
   (c) Metric Learning (09/22)
   (d) Tree-based Methods: CART (09/29)
   (e) Tree-based Methods: Bagging/Boosting (10/06)
3. Semi-supervised Learning (10/13)
4. Neural Networks (10/20)
5. A Gentle Introduction to Deep Learning (10/27)
6. Streaming Data Analyses in Electronic Medical Record Data (11/03)
7. Causal Inference in Healthcare and Medicine:
   (a) Nature's Experiments: Mendelian Randomization (11/10)
   (b) Bayesian Inference and Design of Experiments (11/17)

(c) No class on 11/24 (Thanksgiving break)

(d) Adaptive Clinical Trial and Reinforcement Learning (12/01)

8. Final in-class presentation (12/08)

4 Grading

- Homework assignments (65%):
  - Biweekly
  - The lowest score will be dropped in the final grade, and one late homework (24 hour) is allowed.
  - It is encouraged to discuss the problem sets with others, but everyone needs to turn in a unique personal write-up.
- Final project write-up and presentation(35%): take home, open books, open notes.

References

Bradley Efron and Trevor Hastie. *Computer age statistical inference*, volume 5. Cambridge University Press, 2016.

Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

Kristin Ciriello Pothier. *Personalizing precision medicine: a global voyage from vision to reality*. John Wiley & Sons, 2017.

Eric Topol. *Deep medicine: how artificial intelligence can make healthcare human again*. Hachette UK, 2019.

Anastasios A Tsiatis. *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*. CRC press, 2019.