

Machine Learning meets High Performance Computing Reading Group Summer 2018

Open to all, regardless of level of expertise.

All meetings Thursdays 10-11am in Cassava Conference Room (LARP)

Meeting format: a 10-minute crash tutorial on the week's theme, followed by discussion focused on the *primary reading*. *Supplementary reading* is suggested for additional information on the topic, and *introductory reading* is strongly suggested for newcomers.

Attendees are encouraged to read and contribute according to your level.

Contact: Lissa Baseman (HPC-DES/USRC), lissa@lanl.gov

Mailing List: MLmeetsHPC@lanl.gov

Week 1

- (5/31) **Theme:** Anomaly Detection
 - Primary Reading: Friedland, Lisa, Amanda Gentzel, and David Jensen. *Classifier-Adjusted Density Estimation for Anomaly Detection and One-Class Classification*. ICDM, 2014.
 - Supplementary Reading: Hempstalk, Kathryn, Eibe Frank, and Ian Witten. *One-class classification by combining density and class probability estimation*. ECML PKDD, 2008.
 - Introductory Reading: Chandola, Varun, Arindam Banerjee, and Vipin Kumar. *Anomaly Detection: A Survey*. ACM Computing Surveys, 2009.
 - Discussion Lead: Randall Woodall

Week 2

- (6/7) **Theme:** Interpretability
 - Primary Reading: Riberio, Marco Tulio, Sameer Singh, and Carlos Guestrin. *Anchors: High-Precision Model-Agnostic Explanations*. AAAI, 2018.
 - Supplementary Reading: Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. *Why Should I Trust You?: Explaining the Predictions of Any Classifier*. ACM SIGKDD, 2016.
 - Introductory Reading: Doshi-Velez, Finale, and Been Kim. *Towards a Rigorous Science of Interpretable Machine Learning*. arXiv, 2017.
 - Discussion lead: Emily Porter

Week 3

- (6/21) **Theme:** Naïveté
 - Primary Reading: Kotschieder, Peter, Madalina Fiterau, and Samuel Rota Bulò. *Deep Neural Decision Forests*. ICCV, 2015.
 - Supplementary Reading: Lowd, Daniel, and Pedro Domingos. *Naïve Bayes Models for Probability Estimation*. ICML, 2005.
 - Introductory Reading: Breiman, Leo. *Random forests*. Machine Learning, 2001.
 - Discussion Lead: David Huff

Week 4

- (6/14) **Theme:** Relational Learning
 - Primary Reading: Neville, Jennifer, and David Jensen. *Dependency Networks for Relational Data*. ICDM, 2004.
 - Supplementary Reading: Neville, Jennifer, David Jensen, Lisa Friedland, and Michael Hay. *Learning Relational Probability Trees*. KDD, 2003.
 - Introductory Reading: Getoor, Lise, and Lilyana Mihalkova. *Learning Statistical Models from Relational Data*. ACM SIGMOD, 2011.
 - Discussion Lead: Michael Kuchnik

Week 5

- (6/28) **Theme:** Advancing HPC through Machine Learning
 - Primary Reading: Kraska, Tim, Alex Beutel, Ed H. Chi, Jeffrey Dean, and Neoklis Polyzotis. *The Case for Learned Index Structures*. arXiv, 2017.
 - Supplementary Reading: Ling, Julia, W. Philip Kegelmeyer, Konduri Aditya, Hemanth Kolla, Kevin A. Reed, Timothy M. Shead, and Warren L. Davis IV. *Using Feature Important Metrics to Detect Events of Interest in Scientific Applications*. IEEE LDAH, 2017.
 - Introductory Reading: Gao, Jim. *Machine Learning Applications for Data Center Optimization* Google, 2016.
 - Discussion Lead: Randall Woodall

No meeting on 7/5 (Lissa out of office)

Week 6

- (7/12) **Theme:** Causality
 - Primary Reading: Louizos, Christos, Uri Shalit, Joris Mooij, David Sontag, Richard Zemel, and Max Welling. *Causal Effect Inference with Deep Latent-Variable Models*. NIPS, 2017.
 - Supplementary Reading: Raghu, Vineet K., Joseph D. Ramsey, Alison Morris, Dimitrios V. Manatakis, Peter Spirtes, Panos K. Chrysanthis, Clark Glymour, and Panayiotis V. Benos. *Comparison of Strategies for Scalable Causal Discovery of Latent Variable Models from Mixed Data*. ACM SIGKDD workshops, 2017.
 - Introductory Reading: Spirtes, Peter. *Introduction to Causal Inference*. JMLR, 2010.
 - Discussion Lead: Emily Porter

Week 7

- (7/19) **Theme:** Topic Modeling
 - Primary Reading: Mimno, David, Hanna M. Wallach, Jason Naradowsky, David A. Smith, and Andrew McCallum. *Polylingual Topic Models*. EMNLP, 2009.
 - Supplementary Reading: Hannah, Lauren A., and Hanna M. Wallach. *Summarizing Topics: From Word Lists to Phrases*. NIPS Workshops, 2014.
 - Introductory Reading: Blei, David M., Andrew Y. Ng, and Michael I. Jordan. *Latent Dirichlet Allocation*. NIPS, 2002.
 - Discussion Lead: Alexandra DeLucia

Week 8

- (7/26) **Theme:** Embeddings
 - Primary Reading: Logan, Robert L. IV, Samuel Humeau, and Sameer Singh. *Multimodal Attribute Extraction*. NIPS, 2017.
 - Supplementary Reading: Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*. NIPS, 2016.
 - Introductory Reading: Schnabel, Tobias, Igor Labutov, David Mimno, and Thorsten Joachims. *Evaluation Methods for Unsupervised Word Embeddings*. EMNLP, 2015.
 - Discussion Lead: Michael Kuchnik

Week 9

- (8/2) **Theme:** Clustering
 - Primary Reading: Xu, Hongteng, and Hongyuan Zha. *A Dirichlet Mixture Model of Hawkes Processes for Event Sequence Clustering*. NIPS, 2017.
 - Supplementary Reading: Aghabozorgi, Saeed, Ali Seyed Shirkhorshidi, and Teh Ying Wah. *Time-series Clustering - A Decade Review*. Information Systems, 2015.
 - Introductory Reading: Matteucci, Matteo. *A Tutorial on Clustering Algorithms*. 2018. <https://home.deib.polimi.it/matteucc/Clustering/tutorial.html/index.html>
 - Discussion Lead: David Huff

Special thanks to all of our presenters and discussion leads!