Skip to content

CSE 599 Machine Learning for Big Data / STAT 592 Statistics for Big Data

Carlos Guestrin / Emily Fox

Computer Science & Engineering, University of Washington

T/Th 10:30-11:50, MUE 153

Home | Lectures | Readings | Homework | Data Sets | Project | People

Although we are not directly following any textbook in particular, the background readings for many of the topics will come from: Murphy, Kevin P. *Machine Learning: a Probabilistic Perspective.* Cambridge, MA: MIT press, 2012. Below, we will denote this book using "KM".

Background: Introduction to Probability and Statistical Learning [-] Collapse All[-]

- *Introduction to Probability*
- *The Element of Statistical Learning: Data Mining, Inference, and Prediction*

Case Study I: Estimating Click Probabilities [-] Collapse All[-]

Online learning : KM Sec. 8.5

- Gradient Decent: KM Sec. 8.3.2
  9.1, 9.2 and 9.3 of Boyd, Stephen and Lieven Vandenburghe. Convex Optimization. Cambridge: Cambridge University Press, 2004. Sec. 9.1, 9.2, 9.3. Print.
- Accelerated descent methods
  Tseng, Paul. "On accelerated proximal gradient methods for convex-concave optimization." submitted to SIAM Journal on Optimization (2008).
- Perceptron algorithm: KM Sec. 8.5.4
  Freund, Yoav, and Robert E. Schapire. "Large margin classification using the perceptron algorithm." Machine learning 37.3 (1999): 277-296.
- Stochastic Gradient Descent: KM Sec. 8.5.2
  Le Cun, Leon Bottou Yann. "Large Scale Online Learning." Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference. Vol. 16. MIT Press, 2004.

- Robust stochastic approximation approach to stochastic programming
  Nemirovski, Arkadi, et al. "Robust stochastic approximation approach to stochastic programming." SIAM Journal on Optimization 19.4 (2009): 1574-1609.

Sketching and Hashing

- Bloom Filter
  Wikipedia
- Min-count Sketch
  Cormode, Graham, and S. Muthukrishnan. "An improved data stream summary: the count-min sketch and its applications." Journal of Algorithms 55.1 (2005): 58-75.
- Hash Kernels
  Shi, Qinfeng, et al. "Hash kernels for structured data." The Journal of Machine Learning Research 10 (2009): 2615-2637.
- Permutation Hashing
  Li, Ping, Art Owen, and Cun-Hui Zhang. "One Permutation Hashing." Advances in Neural Information Processing Systems 25. 2012.

Personalization via Multi-task Learning

- Feature Hashing
  Weinberger, Kilian, et al. "Feature hashing for large scale multitask learning." Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009.

Case Study II: Document Retrieval [-] Collapse All[-]

Basic kNN, TF-IDF

- kNN: KM Sec. 1.4.2
  Peterson, Leif E. "K-nearest neighbor." Scholarpedia, 4(2):1883 (2009), revision #91396. Web. 4 Jan. 2013.
- TF-IDF
  Wikipedia

Fast NN Search

- KD-trees tutorial
  Moore, Andrew W. "Efficient Memory-based Learning for Robot Control." Technical Report No.209, Computer Laboratory, University of Cambridge, 1991. Print.
- Approximate nearest neighbors by locality-sensitive hashing (LSH):
  Andoni, Alexandr and Piotr Indyk. "Near-Optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions". Communications of the ACM, vol. 51, no. 1 (2008): 117-122.
- Practical insights for approximate nearest neighbors:
  Gray, Alexander, Ting Liu and Andrew W. Moore. "New Algorithms for Efficient High-Dimensional Nonparametric Classification." Journal of Machine Learning Research 7 (2006): 1135-

1158.

- All pairs NN:
Ram, Parikshit, et al. "Linear-time Algorithms for Pairwise Statistical Problems." Advances in Neural Information Processing Systems 22 2009: 1527-1535

Clustering: KM Sec. 25.1

- K-means: KM Sec. 11.4.2.5
Moore, Andrew W. Tutorials of K-means and Hierarchical Clustering. The Auton Lab at Carnegie Mellon University. Web. 4 Jan. 2013.
- Mixture modeling (generative): KM Sec. 11.1-11.4
- Spectral clustering: KM Sec. 25.4
Von Luxburg, Ulrike. "A tutorial on spectral clustering." Statistics and computing 17.4 (2007): 395-416.

Mixed Membership Models: KM Sec. 27.3

- Basic LDA:
Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." the Journal of machine Learning research 3 (2003): 993-1022.
- Introduction:
Blei, David M. "Probabilistic topic models." Communications of the ACM, vol. 55, no. 4 (2012): 77-84.
- Sampling:
Griffith, Thomas L. and Mark Steyvers. "Finding scientific topics." Proceedings of the National Academy of Sciences of the United States of America, Volume: 101, Supplement: 1 (2004): Pages: 5228-5235

Advanced reading: KM Sec. 21.1-21.3

- Online LDA:
Hoffman, Matt, et al. "Stochastic Variational Inference." arXiv:1206.7051 (2012).
- Large-scale LDA:
Mimno, David, Matthew D. Hoffman and David M. Blei. "Sparse stochastic inference for latent Dirichlet allocation." International Conference on Machine Learning, 2012.
- Distributed LDA:
Ahmed, Amr, et al. "Scalable inference in latent variable models." Proceedings of the fifth ACM international conference on Web search and data mining (2012): 123-132

Case Study III: fMRI Prediction [-] Collapse All[-]

Linear and logistic regression: KM Sec. 7.1-7.3,7.5, 8.1-8.3, 8.5

LASSO: KM Sec. 13.1, 13.3, 13.4

- Original:
  Tibshirani, Robert. "Regression Shrinkage and Selection via the Lasso." Journal of the Royal Statistical Society. Series B (Methodological) Vol. 58, No. 1 (1996): 267-288. Published by: Wiley
- Bayesian interpretation (optional):
  Park, Trevor and George Casella. "The Bayesian Lasso." Journal of the American Statistical Association Volume 103, Issue 482 (2008): 681-686.
- Stochastic l1 regularzied loss minimization:
  Shalev-Shwartz, Shai, and Ambuj Tewari. "Stochastic Methods for $\ ell _1 $ Regularized Loss Minimization." (2009).

Zero-shot learning

- Features of words:
  Mitchell, Tom M., et al. "Predicting human brain activity associated with the meanings of nouns." Science Vol. 320 no. 5880 (2008): 1191-1195.
- Features of words and learning from people:
  Palatucci, Mark, et al. "Zero-shot learning with semantic output codes." Advances in neural information processing systems 22 (2009): 1410-1418.
- Slides on papers above:
  Tom Mitchell's slides

Graphical LASSO: KM Sec. 26.7

- Original:
  Friedman, Jerome, Trevor Hastie and Robert Tibshirani. "Sparse inverse covariance estimation with the graphical lasso." Biostatistics 9(3) (2008): 432-441.
- Slides
- New insights (optional):
  Mazumder, Rahul and Trevor Hastie. "The Graphical Lasso: New Insights and Alternatives." arXiv:1111.5479v2 (2012)

Parallel learning

- (Shotgun) Stochastic coordinate descent:
  Bradley, Joseph, et al. "Parallel Coordinate Descent for L1-Regularized Loss Minimization." International Conference on Machine Learning (2011).
- Stochastic gradient descent:
  Niu, Feng, et al. "HOGWILD!: A Lock-Free Approach to Parallelizing Stochastic Gradient Descent." arXiv:1106.5730v2 (2011)
- Averaging methods:
  Zhang, Yuchen, et al. "Communication-Efficient Algorithms for Statistical Optimization." arXiv:1209.4129v1 (2012)
- Alternating Directions Method of Multipliers (ADMM):
  Boyd, Stephen, et al. "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers." Machine Learning Vol. 3, No. 1 (2010): 1-122

Case Study IV: Collaborative Filtering [-]   Collapse All[-]

Collaborative Filtering:

- Overview:
  Koren, Yehuda, Robert Bell and Chris Volinsky. "Matrix Factorization Techniques for Recommender Systems." Computer Volume: 42, Issue: 8 (2009): 30-37

Matrix Factorization:

- Probabilistic matrix factorization:
  Salakhutdinov, Ruslan, and Andriy Mnih. "Probabilistic matrix factorization." Advances in neural information processing systems 20 (2008): 1257-1264.
- Exact Matrix Completion via Convex Optimization:
  Candès, Emmanuel J., and Benjamin Recht. "Exact matrix completion via convex optimization." Foundations of Computational mathematics 9.6 (2009): 717-772.
- Document clustering via non-negative matrix factorization:
  Xu, Wei, Xin Liu and Yihong Gong. "Document clustering based on non-negative matrix factorization." Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval (2003): 267-273
- Fast Max-Margin Factorization (optional):
  Rennie, Jason D. M. and Nathan Srebro. "Fast Maximum Margin Matrix Factorization for Collaborative Prediction." Proceedings of the 22nd International Conference on Machine Learning (2005).
- Large-scale by divide and conquer (optional):
  Mackey, Lester, Ameet Talwalkar, Michael I. Jordan. "Divide-and-Conquer Matrix Factorization." arXiv:1107.0789v6 (2012)

Cold-start Problem (zero-shot learning), Incorporating Features:

- Basic concept:
  Schein, Andrew I., et al. "Methods and Metrics for Cold-Start Recommendations." Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval (2002): 253-260
- Unified approach:
  Menon, Aditya Krishna and Charles Elkan. "Link prediction via matrix factorization." Proceedings of the 2011 European conference on Machine learning and knowledge discovery in databases, Volume Part II (2011): 437-452

Parallel Learning with GraphLab:

- Original Paper:
  Low, Yucheng, et al. "GraphLab: A New Parallel Framework for Machine Learning." Proceedings of Conference on Uncertainty in Artificial Intelligence (2010).
- Cloud-based:

Low, Yucheng, et al. "Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud." Proceedings of the VLDB Endowment (PVLDB), Vol. 5, No. 8 (2012): 716-727.

- GraphLab 2:
  Gonzalez, Joseph E. et al. "PowerGraph: distributed graph-parallel computation on natural graphs." Proceedings of the 10th USENIX conference on Operating Systems Design and Implementation (2012): 17-30.
- GraphChi (GraphLab on disk):
  Kyrola, Aapo, Guy Blelloch and Carlos Guestrin. "GraphChi: large-scale graph computation on just a PC." Proceedings of the 10th USENIX conference on Operating Systems Design and Implementation (2012): 31-46.

Advanced reading (optional):

- Stochastic block models:
  Airoldi, Edoardo M. et al. "Mixed Membership Stochastic Blockmodels." Journal of Machine Learning Research 9 (2008): 1981-2014.
- Mixed-membership matrix factorization:
  Mackey, Lester, David Weiss and Michael I. Jordan. "Mixed Membership Matrix Factorization." Proceedings of the 27th International Conference on Machine Learning, 2010.
- Scalable stochastic block models:
  Gopalan, Prem. "Scalable Inference of Overlapping Communities." Neural Information Processing Systems, 2012.