Reading/Reference List for Probabilistic Models

Gaurav Sharma

Reading/Reference List Probabilistic Models: Notes

- Includes only some selected references, not comprehensive/exhaustive
  - Varying in depth/specialization
- Guidance: Books/tutorial articles are often more accessible
- Check for updates
  - Books may have newer editions
    - I'm using citing editions I'm familiar with
  - Some links may have changed
    - Usually a search will find alternative location

Books Relevant to Course

- Bishop (broad Pattern Recog./Machine learn) [1]
- Koller and Friedman (graphical models) [2]
- Hastie, Tibshirani, and Friedman (statistical learning) [3]
- McLachlan (EM Algorithm) [4]
- Moon and Stirling (EM, HMMs) [5]
- McLachlan (Mixture Models) [6]
- Manning and Schütze (SCFGs) [7]
- Durbin et al (HMM/SCFG applications in Biological Seq. analysis) [8]
- Sutton and McCallum (Conditional random fields) [9]

Expectation Maximization/Mixture Models

- Moon (EM tutorial) [10]
- Dempster (EM algorithm) [11]

Hidden Markov Models

- Rabiner's tutorial on HMMs [12]
- Baum-Welch Iteration [13]
- Continuous Time HMMs [14], [15]
- Hidden Markov Models/Processes (theoretical considerations) [16]

Stochastic/Probabilistic Context Free Grammars

- Terry Speed's Course Notes (HMMs to SCFGs) [17]
- Giegerich (SCFG tutorial intro focused on RNA structure prediction) [18]

Algorithmic Techniques/Methodologies

- Dynamic Programming [19], [20]
- Belief Propagation [21], [22]

Background: Probability, Statistics, Linear Algebra

- Sheldon Ross (Probability and Statistics) [23]
- Papoulis and Pillai (Random Processes) [24]
- Gray and Davidsson (greater math. rigor) [25]
- Matrix Cookbook [26]
- Carl Meyer (Markov Chains from Linear Algebra perspective) [27]

Connections with Other Fields/Applications

- Connections with Physics/Statistical Mechanics [28]
- Computer vision [29]
- Sequence data applications of CRFs [30]
- Semantic image segmentation in combination with deep learning [31]

Causal Inference

- Book by Peters et. al [32]
- Pearl's Overview Paper [33]
  - Earlier book chapter by Pearl [34]

Software/Implementation Toolkits

- OpenGM [35]
- HMM Toolkit HTK `https://github.com/open-speech/HTK`
- Joshua open toolkit for parsing based machine translation [36]

References I

[1] C. Bishop, *Pattern recognition and machine learning*. New York, NY: Springer, 2006.

[2] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. Cambridge, MA: MIT press, 2009.

[3] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second. New York, NY: Springer-Verlag, 2009.

[4] G. McLachlan, T. Krishnan, and W. InterScience, *The EM algorithm and extensions*. New York, NY: Wiley InterScience, 1997.

[5] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 2000, ISBN: 0-201-36186-8.

[6] G. McLachlan and D. Peel, *Finite Mixture Models*. New York, NY: Wiley InterScience, 2000.

References II

[7] C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press, 1999.

[8] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis : Probabilistic Models of Proteins and Nucleic Acids*. Cambridge, UK: Cambridge University Press, 1999, ISBN: 0521629713.

[9] C. Sutton and A. McCallum, "An introduction to conditional random fields," *Foundations and Trends in Machine Learning*, vol. 4, no. 4, pp. 267–373, 2012, ISSN: 1935-8237. DOI: 10.1561/2200000013.

[10] T. K. Moon, "The expectation maximization algorithm," *IEEE Signal Processing Magazine*, pp. 47–60, 1996.

References III

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, pp. 1–38, 1977.

[12] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[13] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Statist.*, vol. 41, no. 1, pp. 164–171, 1970.

References IV

[14] Y.-Y. Liu, A. Moreno, S. Li, F. Li, L. Song, and J. M. Rehg, "Learning continuous-time hidden Markov models for event data," in *Mobile Health: Sensors, Analytic Methods, and Applications*, Cham: Springer International Publishing, 2017, pp. 361–387, ISBN: 978-3-319-51394-2. DOI: 10.1007/978-3-319-51394-2_19.

[15] Y Liu, L Song, F Li, S Li, and J Rehg, "Efficient continuous-time hidden Markov model for disease modeling," in *Proc. Advances in Neural Info. Proc. Sys. (NIPS)*, 2015.

[16] Y. Ephraim and N. Merhav, "Hidden Markov processes," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, pp. 1518–1569, 2002. DOI: 10.1109/TIT.2002.1003838.

References V

[17] X. Zhao, *From hidden markov models to stochastic context free grammars*, Course notes for Statistics 246: Statistics in Genetics, University of California Berkeley, Week 9b, Instructor: Terry Speed, 2002. [Online]. Available: http://www.stat.berkeley.edu/users/terry/Classes/s246.2002/index.html.

[18] R. Giegerich, "Introduction to stochastic context free grammars," in *RNA Sequence, Structure, and Function: Computational and Bioinformatic Methods*, J. Gorodkin and W. L. Ruzzo, Eds. Totowa, NJ: Humana Press, 2014, pp. 85–106, ISBN: 978-1-62703-709-9. DOI: 10.1007/978-1-62703-709-9_5.

[19] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, 1966, ISSN: 0036-8075. DOI: 10.1126/science.153.3731.34.

References VI

[20] D. P. Bertsekas, *Dynamic programming and optimal control*. Belmont, MA: Athena Scientific, 1995.

[21] J. Pearl, "Fusion, propagation and structuring in belief networks," *Artificial Intelligence*, vol. 29, no. 3, pp. 241–288, 1986.

[22] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.

[23] S. Ross, *A first course in probability*, Second. New York: Macmillan, 1984.

[24] A. Papoulis and S. U. Pillai, *Probability Random Variables and Stochastic Processes*, Fourth. New York: McGraw Hill, 2002.

References VII

[25] R. M. Gray and L. D. Davidsson, *An Introduction to Statistical Signal Processing*. Cambridge, UK: Cambride University Press, 2004, Available online in PDF format at: http://www-ee.stanford.edu/~gray/sp.html.

[26] K. Petersen and M. Pedersen, *The Matrix Cookbook*. Technical University of Denmark, Feb 16th, 2008. [Online]. Available: `http://matrixcookbook.com`.

[27] C. Meyer, *Matrix analysis and applied linear algebra*. Philadephia, PA: SIAM, 2000, ISBN: 9780898714548.

[28] M. Mézard and A. Montanari, *Information, physics, and computation*. Oxford, UK: Oxford University Press, 2009.

References VIII

[29] S. Nowozin and C. H. Lampert, "Structured learning and prediction in computer vision," *Foundations and Trends in Computer Graphics and Vision*, vol. 6, no. 3–4, pp. 185–365, 2011, ISSN: 1572-2740. DOI: 10.1561/0600000033. [Online]. Available: http://dx.doi.org/10.1561/0600000033.

[30] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Intl. Conf. on Mach. Learning*, 2001, pp. 282–289.

References IX

[31] A. Arnab, S. Zheng, S. Jayasumana, B. Romera-Paredes, M. Larsson, A. Kirillov, B. Savchynskyy, C. Rother, F. Kahl, and P. H. S. Torr, "Conditional random fields meet deep neural networks for semantic segmentation: Combining probabilistic graphical models with deep learning for structured prediction," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 37–52, 2018. DOI: 10.1109/MSP.2017.2762355.

[32] J. Peters, D. Janzing, and B. Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms*. Cambridge, MA: MIT press, 2017, ISBN: 9780262037310.

[33] J. Pearl, "Causal inference in statistics: An overview," *Statist. Surv.*, vol. 3, pp. 96–146, 2009. DOI: 10.1214/09-SS057. [Online]. Available: https://doi.org/10.1214/09-SS057.

References X

[34] ——,"Graphical models for probabilistic and causal reasoning," in *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 1: Quantified Representation of Uncertainty and Imprecision*, D. M. Gabbay and P. Smets, Eds., Dordrecht: Kluwer Academic Publishers, 1998, pp. 367–389. [Online]. Available: `http://citeseer.ist.psu.edu/pearl97graphical.html`.

[35] B. Andres, T. Beier, and J. Kappes, "OpenGM: A C++ library for discrete graphical models," *CoRR*, vol. abs/1206.0111, 2012, Associated GitHub repository is `https://github.com/opengm/opengm.git`. [Online]. Available: `http://arxiv.org/abs/1206.0111`.

References XI

[36] Z. Li, C. Callison-Burch, C. Dyer, S. Khudanpur, L. Schwartz, W. Thornton, J. Weese, and O. Zaidan, "Joshua: An open source toolkit for parsing-based machine translation," in *Proceedings of the Fourth Workshop on Statistical Machine Translation*, 2009, pp. 135–139.