REFERENCES

[1] L. F. Abbott, J. A. Varela, K. Sen and S. B. Nelson. Synaptic depression and cortical gain control. *Science*, **275**:220–223, 1997.

[2] D. H. Ackley, G. E. Hinton and T. J. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science*, **9**:147–169, 1985.

[3] R. P. Adams and D. J. C. MacKay. Bayesian online changepoint detection. Cavendish Laboratory, Department of Physics, University of Cambridge, 2006.

[4] E. Airoldi, D. Blei, E. Xing and S. Fienberg. A latent mixed membership model for relational data. In *LinkKDD '05: Proceedings of the 3rd International Workshop on Link Discovery*, pages 82–89. ACM, 2005.

[5] E. M. Airoldi, D. M. Blei, S. E. Fienberg and E. P. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, **9**:1981–2014, 2008.

[6] D. L. Alspach and H. W. Sorenson. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control*, **17**(4):439–448, 1972.

[7] S-i. Amari. Natural gradient learning for over and under-complete bases in ICA. *Neural Computation*, **11**:1875–1883, 1999.

[8] I. Androutsopoulos, J. Koutsias, K. V. Chandrinos and C. D. Spyropoulos. An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 160–167. ACM 2000.

[9] S. Arora and C. Lund. Hardness of approximations. In *Approximation Algorithms for NP-Hard Problems*, pages 399–446. PWS Publishing Co., 1997.

[10] F. R. Bach and M. I. Jordan. Thin junction trees. In T. G. Dietterich, S. Becker and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 14, pages 569–576. MIT Press, 2001.

[11] F. R. Bach and M. I. Jordan. A probabilistic interpretation of canonical correlation analysis. Computer Science Division and Department of Statistics 688, University of California Berkeley, Berkeley, USA, 2005.

[12] Y. Bar-Shalom and Xiao-Rong Li. *Estimation and Tracking: Principles, Techniques and Software*. Artech House, 1998.

[13] D. Barber. Dynamic Bayesian networks with deterministic tables. In S. Becker, S. Thrun and K. Obermayer, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 15, pages 713–720. MIT Press, 2003.

[14] D. Barber. Learning in spiking neural assemblies. In S. Becker, S. Thrun and K. Obermayer, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 15, pages 149–156. MIT Press, 2003.

[15] D. Barber. Are two classifiers performing equally? A treatment using Bayesian hypothesis testing. IDIAP-RR 57, IDIAP, Rue de Simplon 4, Martigny, CH-1920, Switzerland, May 2004. IDIAP-RR 04-57.

[16] D. Barber. The auxiliary variable trick for deriving Kalman smoothers. IDIAP-RR 87, IDIAP, Rue de Simplon 4, Martigny, CH-1920, Switzerland, December 2004. IDIAP-RR 04-87.

[17] D. Barber. Expectation correction for smoothing in switching linear Gaussian state space models. *Journal of Machine Learning Research*, **7**:2515–2540, 2006.

[18] D. Barber. Clique matrices for statistical graph decomposition and parameterising restricted positive definite matrices. In D. A. McAllester and P. Myllymaki, editors, *Uncertainty in Artificial Intelligence*, number 24, pages 26–33. AUAI Press, 2008.

[19] D. Barber and F. V. Agakov. Correlated sequence learning in a network of spiking neurons using maximum likelihood. Informatics Research Reports EDI-INF-RR-0149, Edinburgh University, 2002.

[20] D. Barber and F. V. Agakov. The IM algorithm: a variational approach to information maximization. In *Advances in Neural Information Processing Systems (NIPS)*, number 16, 2004.

[21] D. Barber and C. M. Bishop. Bayesian model comparison by Monte Carlo chaining. In M. C. Mozer, M. I. Jordan and T. Petsche, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 9, pages 333–339. MIT Press, 1997.

[22] D. Barber and C. M. Bishop. Ensemble learning in Bayesian neural networks. In *Neural Networks and Machine Learning*, pages 215–237. Springer, 1998.

[23] D. Barber and S. Chiappa. Unified inference for variational Bayesian linear Gaussian state-space models. In B. Schölkopf, J. Platt and T. Hoffman, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 19, pages 81–88. MIT Press, 2007.

[24] D. Barber and W. Wiegerinck. Tractable variational structures for approximating graphical models. In M. S. Kearns, S. A. Solla and D. A. Cohn, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 11, pages 183–189. MIT Press, 1999.

[25] D. Barber and C. K. I. Williams. Gaussian processes for Bayesian classification via hybrid Monte Carlo. In M. C. Mozer, M. I. Jordan and T. Petsche, editors, *Advances in Neural Information Processing Systems NIPS 9*, pages 340–346. MIT Press, 1997.

[26] R. J. Baxter. *Exactly Solved Models in Statistical Mechanics*. Academic Press, 1982.

[27] M. J. Beal, F. Falciani, Z. Ghahramani, C. Rangel and D. L. Wild. A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics*, **21**:349–356, 2005.

[28] A. Becker and D. Geiger. A sufficiently fast algorithm for finding close to optimal clique trees. *Artificial Intelligence*, **125**(1-2):3–17, 2001.

[29] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, **7**(6):1129–1159, 1995.

[30] R. E. Bellman. *Dynamic Programming*. Princeton University Press, 1957. Paperback edition by Dover Publications (2003).

[31] Y. Bengio and P. Frasconi. Input-output HMMs for sequence processing. *IEEE Transactions on Neural Networks*, **7**:1231–1249, 1996.

[32] A. L. Berger, S. D. Della Pietra and V. J. D. Della Pietra. A maximum entropy approach to natural language processing. *Computational Linguistics*, **22**(1):39–71, 1996.

[33] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, second edition, 1985.

[34] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2nd edition, 1999.

[35] D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, second edition, 2000.

[36] J. Besag. Spatial interactions and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, **36**(2):192–236, 1974.

[37] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, **48**:259–302, 1986.

[38] J. Besag and P. Green. Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society, Series B*, **55**:25–37, 1993.

[39] G. J. Bierman. Measurement updating using the U-D factorization. *Automatica*, **12**:375–382, 1976.

[40] N. L. Biggs. *Discrete Mathematics*. Oxford University Press, 1990.

[41] K. Binder and A. P. Young. Spin glasses: experimental facts, theoretical concepts, and open questions. *Reviews of Modern Physics*, **58**(4):801–976, Oct 1986.

[42] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[43] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[44] C. M. Bishop and M. Svensén. Bayesian hierarchical mixtures of experts. In U. Kjaerulff and C. Meek, editors, *Proceedings Nineteenth Conference on Uncertainty in Artificial Intelligence*, pages 57–64. Morgan Kaufmann, 2003.

[45] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**(3):637–654, 1973.

[46] D. Blei, A. Ng and M. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, **3**:993–1022, 2003.

[47] R. R. Bouckaert. Bayesian belief networks: from construction to inference. PhD thesis, University of Utrecht, 1995.

[48] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[49] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis Machine Intelligence*, **26**(9):1124–1137, 2004.

[50] Y. Boykov, O. Veksler and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis Machine Intelligence*, **23**:1222–1239, 2001.

[51] C. Bracegirdle and D. Barber. Switch-reset models : Exact and approximate inference. In *Proceedings of The Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 10, 2011.

[52] M. Brand. Incremental singular value decomposition of uncertain data with missing values. In *European Conference on Computer Vision (ECCV)*, pages 707–720, 2002.

[53] J. Breese and D. Heckerman. Decision-theoretic troubleshooting: a framework for repair and experiment. In E. Horvitz and F. Jensen, editors, *Uncertainty in Artificial Intelligence*, number 12, pages 124–132. Morgan Kaufmann, 1996.

[54] H. Bunke and T. Caelli. *Hidden Markov Models: Applications in Computer Vision*. Machine Perception and Artificial Intelligence. World Scientific Publishing Co., Inc., 2001.

[55] W. Buntine. Theory refinement on Bayesian networks. In *Uncertainty in Artificial Intelligence*, number 7, pages 52–60, 1991. Morgan Kaufmann.

[56] A. Cano and S. Moral. Heuristic algorithms for the triangulation of graphs. In *Advances in Intelligent Computing – IPMU 1994*, pages 98–107. Number 945 in Lectures Notes in Computer Sciences. Springer-Verlag, 1995.

[57] O. Cappé, E. Moulines and T. Ryden. *Inference in Hidden Markov Models*. Springer, 2005.

[58] A. T. Cemgil. Bayesian inference in non-negative matrix factorisation models. Technical Report CUED/F-INFENG/TR.609, University of Cambridge, July 2008.

[59] A. T. Cemgil, B. Kappen and D. Barber. A generative model for music transcription. *IEEE Transactions on Audio, Speech and Language Processing*, **14**(2):679–694, 2006.

[60] E. Challis and D. Barber. Local and variational Gaussian approximations for Bayesian generalised linear models. Department of Computer Science, University College London, 2010.

[61] H. S. Chang, M. C. Fu, J. Hu and S. I. Marcus. *Simulation-based Algorithms for Markov Decision Processes*. Springer, 2007.

[62] S. Chiappa and D. Barber. Bayesian linear Gaussian state space models for biosignal decomposition. *Signal Processing Letters*, **14**(4):267–270, 2007.

[63] S. Chib and M. Dueker. Non-Markovian regime switching with endogenous states and time-varying state strengths. Econometric Society 2004 North American Summer Meetings 600, Econometric Society, August 2004.

[64] C. K. Chow and C. N. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, **14**(3):462–467, 1968.

[65] P. S. Churchland and T. J. Sejnowski. *The Computational Brain*. MIT Press, 1994.

[66] D. Cohn and H. Chang. Learning to probabilistically identify authoritative documents. In P. Langley, editor, *International Conference on Machine Learning*, number 17, pages 167–174. Morgan Kaufmann, 2000.

[67] D. Cohn and T. Hofmann. The missing link – a probabilistic model of document content and hyper-text connectivity. *In Neural Information Processing Systems*, number 13, pages 430–436. MIT Press, 2001.

[68] A. C. C. Coolen, R. Kühn and P. Sollich. *Theory of Neural Information Processing Systems*. Oxford University Press, 2005.

[69] G. F. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4):309–347, 1992.

[70] A. Corduneanu and C. M. Bishop. Variational Bayesian model selection for mixture distributions. In T. Jaakkola and T. Richardson, editors, *Artificial Intelligence and Statistics*, pages 27–34. Morgan Kaufmann, 2001.

[71] M. T. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[72] R. G. Cowell, A. P. Dawid, S. L. Lauritzen and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer, 1999.

[73] D. R. Cox and N. Wermuth. *Multivariate Dependencies*. Chapman and Hall, 1996.

[74] J. C. Cox, S. A. Ross and M. Rubinstein. Option pricing: a simplified approach. *Journal of Financial Economics*, 7:229–263, 1979.

[75] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge University Press, 2000.

[76] P. Dangauthier, R. Herbrich, T. Minka and T. Graepel. Trueskill through time: revisiting the history of chess. In B. Schölkopf, J. Platt and T. Hoffman, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 19, pages 569–576. MIT Press, 2007.

[77] H. A. David. *The Method of Paired Comparisons*. Oxford University Press, 1988.

[78] A. P. Dawid. Influence diagrams for causal modelling and inference. *International Statistical Review*, 70:161–189, 2002.

[79] A. P. Dawid and S. L. Lauritzen. Hyper Markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics*, 21(3):1272–1317, 1993.

[80] P. Dayan and L. F. Abbott. *Theoretical Neuroscience*. MIT Press, 2001.

[81] P. Dayan and G. E. Hinton. Using expectation-maximization for reinforcement learning. *Neural Computation*, 9:271–278, 1997.

[82] T. De Bie, N. Cristianini and R. Rosipal. Eigenproblems in pattern recognition. In *Handbook of Geometric Computing: Applications in Pattern Recognition, Computer Vision, Neuralcomputing, and Robotics*. Springer-Verlag, 2005.

[83] R. Dechter. Bucket elimination: a unifying framework for probabilistic inference algorithms. In E. Horvitz and F. Jensen, editors, *Uncertainty in Artificial Intelligence*, pages 211–219. Morgan Kaufmann, 1996.

[84] 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. Series B (Methodological)*, 39(1):1–38, 1977.

[85] S. Diederich and M. Opper. Learning of correlated patterns in spin-glass networks by local learning rules. *Physical Review Letters*, 58(9):949–952, 1986.

[86] R. Diestel. *Graph Theory*. Springer, 2005.

[87] A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. In D. Crisan and B. Rozovsky, editors, *Oxford Handbook of Nonlinear Filtering*. Oxford University Press, 2009.

[88] R. O. Duda, P. E. Hart and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.

[89] J. Durbin. The fitting of time series models. *Rev. Inst. Int. Stat.*, 28:233–243, 1960.

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

[91] A. Düring, A. C. C. Coolen and D. Sherrington. Phase diagram and storage capacity of sequence processing neural networks. *Journal of Physics A*, 31:8607–8621, 1998.

[92] J. M. Gutierrez, E. Castillo and A. S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer-Verlag, 1997.

[93] J. Edmonds and R. M. Karp. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of the ACM*, **19**(2):248–264, 1972.

[94] R. Edwards and A. Sokal. Generalization of the Fortium-Kasteleyn-Swendson-Wang representation and Monte Carlo algorithm. *Physical Review D*, **38**:2009–2012, 1988.

[95] A. E. Elo. *The Rating of Chess Players, Past and Present*. Arco, second edition, 1986.

[96] R. F. Engel. GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives*, 15(4):157–168, 2001.

[97] Y. Ephraim and W. J. J. Roberts. Revisiting autoregressive hidden Markov modeling of speech signals. *IEEE Signal Processing Letters*, **12**(2):166–169, 2005.

[98] E. Erosheva, S. Fienberg and J. Lafferty. Mixed membership models of scientific publications. In *Proceedings of the National Academy of Sciences*, volume 101, pages 5220–5227, 2004.

[99] R-E. Fan, P-H. Chen and C-J. Lin. Working set selection using second order information for training support vector machines. *Journal of Machine Learning Research*, **6**:1889–1918, 2005.

[100] P. Fearnhead. Exact and efficient Bayesian inference for multiple changepoint problems. Technical report, Department of Mathematics and Statistics, Lancaster University, 2003.

[101] G. H. Fischer and I. W. Molenaar. *Rasch Models: Foundations, Recent Developments, and Applications*. Springer, 1995.

[102] M. E. Fisher. Statistical mechanics of dimers on a plane lattice. *Physical Review*, **124**:1664–1672, 1961.

[103] B. Frey. Extending factor graphs as to unify directed and undirected graphical models. In C. Meek and U. Kjærulff, editors, *Uncertainty in Artificial Intelligence*, number 19, pages 257–264. Morgan Kaufmann, 2003.

[104] N. Friedman, D. Geiger and M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, **29**:131–163, 1997.

[105] S. Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer, 2006.

[106] M. Frydenberg. The chain graph Markov property. *Scandinavian Journal of Statistics*, **17**:333–353, 1990.

[107] T. Furmston and D. Barber. Solving deterministic policy (PO)MPDs using expectation-maximisation and antifreeze. In *First International Workshop on Learning and Data Mining for Robotics (LEMIR)*, September 2009. In conjunction with ECML/PKDD-2009.

[108] T. Furmston and D. Barber. Variational methods for reinforcement learning. In Y. W. Teh and M. Titterington, editors, *Proceedings of The Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 9, pages 241–248, Chia Laguna, Sardinia, Italy, May 13–15 2010. JMLR.

[109] T. Furmston and D. Barber. Efficient Inference in Markov Control Problems. In *Uncertainty in Artificial Intelligence*, number 27, Corvallis, Oregon, USA, 2011.

[110] T. Furmston and D. Barber. Lagrange Dual Decomposition for Finite Horizon Markov Decision Processes. In *European Conference on Machine Learning (ECML)*, 2011.

[111] A. Galka, O. Yamashita, T. Ozaki, R. Biscay and P. Valdes-Sosa. A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering. *NeuroImage*, **23**:435–453, 2004.

[112] P. Gandhi, F. Bromberg and D. Margaritis. Learning Markov network structure using few independence tests. In *Proceedings of the SIAM International Conference on Data Mining*, pages 680–691, 2008.

[113] M. R. Garey and D. S. Johnson. *Computers and Intractability, A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, 1979.

[114] A. Gelb. *Applied Optimal Estimation*. MIT Press, 1974.

[115] A. Gelman, G. O. Roberts and W. R. Gilks. Efficient Metropolis jumping rules. In J. O. Bernardo, J. M. Berger, A. P. Dawid and A. F. M. Smith, editors, *Bayesian Statistics*, volume 5, pages 599–607. Oxford University Press, 1996.

[116] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In *Readings in Uncertain Reasoning*, pages 452–472. Morgan Kaufmann, 1990.

[117] M. G. Genton. Classes of kernels for machine learning: a statistics perspective. *Journal of Machine Learning Research*, **2**:299–312, 2001.

[118] W. Gerstner and W. M. Kistler. *Spiking Neuron Models*. Cambridge University Press, 2002.

[119] Z. Ghahramani and M. J. Beal. Variational inference for Bayesian mixtures of factor analysers. In S. A. Solla, T. K. Leen and K-R. Müller, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 12, pages 449–455. MIT Press, 2000.

[120] Z. Ghahramani and G. E. Hinton. Variational learning for switching state-space models. *Neural Computation*, **12**(4):963–996, 1998.

[121] A. Gibbons. *Algorithmic Graph Theory*. Cambridge University Press, 1991.

[122] M. Gibbs. *Bayesian Gaussian processes for regression and classification*. PhD thesis, University of Cambridge, 1997.

[123] W. R. Gilks, S. Richardson and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, 1996.

[124] M. Girolami. A variational method for learning sparse and overcomplete representations. *Neural Computation*, **13**:2517–2532, 2001.

[125] M. Girolami and A. Kaban. On an equivalence between PLSI and LDA. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 433–434. ACM Press, 2003.

[126] M. E. Glickman. Parameter estimation in large dynamic paired comparison experiments. *Applied Statistics*, **48**:377–394, 1999.

[127] A. Globerson and T. Jaakkola. Approximate inference using planar graph decomposition. In B. Schölkopf, J. Platt and T. Hoffman, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 19, pages 473–480. MIT Press, 2007.

[128] D. Goldberg, D. Nichols, B. M. Oki and D. Terry. Using collaborative filtering to weave an information tapestry. *Communications ACM*, **35**:61–70, 1992.

[129] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins University Press, third edition, 1996.

[130] M. C. Golumbic and I. Ben-Arroyo Hartman. *Graph Theory, Combinatorics, and Algorithms*. Springer-Verlag, 2005.

[131] C. Goutis. A graphical method for solving a decision analysis problem. *IEEE Transactions on Systems, Man and Cybernetics*, **25**:1181–1193, 1995.

[132] P. J. Green and B. W. Silverman. *Nonparametric Regression and Generalized Linear Models*, volume 58 of Monographs on Statistics and Applied Probability. Chapman and Hall, 1994.

[133] D. M. Greig, B. T. Porteous and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, **2**:271–279, 1989.

[134] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, second edition, 1992.

[135] S. F. Gull. Bayesian data analysis: straight-line fitting. In J. Skilling, editor, *Maximum Entropy and Bayesian Methods (Cambridge 1988)*, pages 511–518. Kluwer, 1989.

[136] A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. Chapman and Hall/CRC, 1999.

[137] D. J. Hand and K. Yu. Idiot's Bayes – not so stupid after all? *International Statistical Review*, **69**(3):385–398, 2001.

[138] D. R. Hardoon, S. Szedmak and J. Shawe-Taylor. Canonical correlation analysis: an overview with application to learning methods. *Neural Computation*, **16**(12):2639–2664, 2004.

[139] D. O. Hebb. *The Organization of Behavior*. Wiley, 1949.

[140] D. Heckerman. A tutorial on learning with Bayesian networks. Technical Report MSR-TR-95-06, Microsoft Research, March 1996. Revised November 1996.

[141] D. Heckerman, D. Geiger and D. Chickering. Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learning*, **20**(3):197–243, 1995.

[142] R. Herbrich, T. Minka and T. Graepel. TrueSkill™: a Bayesian skill rating system. In B. Schölkopf, J. Platt and T. Hoffman, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 19, pages 569–576. MIT Press, 2007.

[143] H. Hermansky. Should recognizers have ears? *Speech Communication*, **25**:3–27, 1998.

[144] J. Hertz, A. Krogh and R. Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, 1991.

[145] T. Heskes. Convexity arguments for efficient minimization of the Bethe and Kikuchi free energies. *Journal of Artificial Intelligence Research*, **26**:153–190, 2006.

[146] D. M. Higdon. Auxiliary variable methods for Markov chain Monte Carlo with applications. *Journal of the American Statistical Association*, **93**(442):585–595, 1998.

[147] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, **313**:504–507, 2006.

[148] T. Hofmann, J. Puzicha and M. I. Jordan. Learning from dyadic data. In M. S. Kearns, S. A. Solla and D. A. Cohn, editors, *Advances in Neural Information Processing Systems (NIPS)*, pages 466–472. MIT Press, 1999.

[149] R. A. Howard and J. E. Matheson. Influence diagrams. *Decision Analysis*, **2**(3), 2005. Republished version of the original 1981 report.

[150] J. C. Hull. *Options, Futures, and Other Derivatives*. Prentice Hall, 1997.

[151] A. Hyvärinen, J. Karhunen and E. Oja. *Independent Component Analysis*. Wiley, 2001.

[152] Aapo Hyvärinen. Consistency of pseudolikelihood estimation of fully visible Boltzmann machines. *Neural Computation*, **18**(10):2283–2292, 2006.

[153] M. Isard and A. Blake. CONDENSATION Conditional density propagation for visual tracking. *International Journal of Computer Vision*, **29**:5–28, 1998.

[154] T. S. Jaakkola and M. I. Jordan. Variational probabilistic inference and the QMR-DT network. *Journal of Artificial Intelligence Research*, **10**:291–322, 1999.

[155] T. S. Jaakkola and M. I. Jordan. Bayesian parameter estimation via variational methods. *Statistics and Computing*, **10**(1):25–37, 2000.

[156] R. A. Jacobs, F. Peng and M. A. Tanner. A Bayesian approach to model selection in hierarchical mixtures-of-experts architectures. *Neural Networks*, **10**(2):231–241, 1997.

[157] R. G. Jarrett. A note on the intervals between coal-mining disasters. *Biometrika*, **66**:191–193, 1979.

[158] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[159] F. Jensen, F. V. Jensen and D. Dittmer. From influence diagrams to junction trees. In *Proceedings of the 10th Annual Conference on Uncertainty in Artificial Intelligence (UAI-94)*, pages 367–373. Morgan Kaufmann, 1994.

[160] F. V. Jensen and F. Jensen. Optimal junction trees. In R. Lopez de Mantaras and D. Poole, editors, *Uncertainty in Artificial Intelligence*, number 10, pages 360–366. Morgan Kaufmann, 1994.

[161] F. V. Jensen and T. D. Nielson. *Bayesian Networks and Decision Graphs*. Springer-Verlag, second edition, 2007.

[162] M. I. Jordan and R. A. Jacobs. Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, **6**:181–214, 1994.

[163] B. H. Juang, W. Chou and C. H. Lee. Minimum classification error rate methods for speech recognition. *IEEE Transactions on Speech and Audio Processing*, **5**:257–265, 1997.

[164] L. P. Kaelbling, M. L. Littman and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, **101**(1–2):99–134, 1998.

[165] H. J. Kappen. An introduction to stochastic control theory, path integrals and reinforcement learning. In *Proceedings 9th Granada Seminar on Computational Physics: Computational and Mathematical Modeling of Cooperative Behavior in Neural Systems*, volume 887, pages 149–181. American Institute of Physics, 2007.

[166] H. J. Kappen and F. B. Rodríguez. Efficient learning in Boltzmann machines using linear response theory. *Neural Computation*, **10**(5):1137–1156, 1998.

[167] H. J. Kappen and W. Wiegerinck. Novel iteration schemes for the Cluster Variation Method. In T. G. Dietterich, S. Becker and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 14, pages 415–422. MIT Press, 2002.

[168] Y. Karklin and M. S. Lewicki. Emergence of complex cell properties by learning to generalize in natural scenes. *Nature*, **457**:83–86, November 2008.

[169] G. Karypis and V. Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on Scientific Computing*, **20**(1):359–392, 1998.

[170] P. W. Kasteleyn. Dimer statistics and phase transitions. *Journal of Mathematical Physics*, **4**(2):287–293, 1963.

[171] S. A. Kauffman. *At Home in the Universe: The Search for Laws of Self-Organization and Complexity*. Oxford University Press, 1995.

[172] C-J. Kim. Dynamic linear models with Markov-switching. *Journal of Econometrics*, **60**:1–22, 1994.

[173] C-J. Kim and C. R. Nelson. *State-Space Models with Regime Switching*. MIT Press, 1999.

[174] G. Kitagawa. The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother. *Annals of the Institute of Statistical Mathematics*, **46**(4):605–623, 1994.

[175] U. B. Kjaerulff and A. L. Madsen. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. Springer, 2008.

[176] N. Komodakis, N. Paragios, and G. Tziritas. MRF Optimization via Dual Decomposition: Message-Passing Revisited. In *IEEE 11th International Conference on Computer Vision, ICCV*, pages 1–8, 2007.

[177] A. Krogh, M. Brown, I. Mian, K. Sjolander and D. Haussler. Hidden Markov models in computational biology: Applications to protein modeling. *Journal of Molecular Biology*, **235**:1501–1531, 1994.

[178] S. Kullback. *Information Theory and Statistics*. Dover, 1968.

[179] K. Kurihara, M. Welling and Y. W. Teh. Collapsed variational Dirichlet process mixture models. In *Proceedings of the International Joint Conference on Artificial Intelligence*, volume 20, pages 2796–2801, 2007.

[180] J. Lafferty, A. McCallum and F. Pereira. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In C. E. Brodley and A. P. Danyluk, editors, *International Conference on Machine Learning*, number 18, pages 282–289. Morgan Kaufmann, 2001.

[181] H. Lass. *Elements of Pure and Applied Mathematics*. McGraw-Hill (reprinted by Dover), 1957.

[182] S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

[183] S. L. Lauritzen, A. P. Dawid, B. N. Larsen and H-G. Leimer. Independence properties of directed Markov fields. *Networks*, **20**:491–505, 1990.

[184] S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of Royal Statistical Society B*, **50**(2):157 – 224, 1988.

[185] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In T. K. Leen, T. G. Dietterich and V. Tresp, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 13, pages 556–562, 2001. MIT Press.

[186] M. A. R. Leisink and H. J. Kappen. A tighter bound for graphical models. In *Neural Computation*, volume 13, pages 2149–2171. MIT Press, 2001.

[187] V. Lepar and P. P. Shenoy. A comparison of Lauritzen-Spiegelhalter, Hugin, and Shenoy-Shafer architectures for computing marginals of probability distributions. In G. Cooper and S. Moral, editors, *Uncertainty in Artificial Intelligence*, number 14, pages 328–333. Morgan Kaufmann, 1998.

[188] U. Lerner, R. Parr, D. Koller and G. Biswas. Bayesian fault detection and diagnosis in dynamic systems. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AIII-00)*, pages 531–537, 2000.

[189] U. N. Lerner. *Hybrid Bayesian Networks for Reasoning about Complex Systems*. Computer Science Department, Stanford University, 2002.

[190] R. Linsker. Improved local learning rule for information maximization and related applications. *Neural Networks*, **18**(3):261–265, 2005.

[191] Y. L. Loh, E. W. Carlson and M. Y. J. Tan. Bond-propagation algorithm for thermodynamic functions in general two-dimensional Ising models. *Physical Review B*, **76**(1):014404, 2007.

[192] H. Lopes and M. West. Bayesian model assessment in factor analysis. *Statistica Sinica*, **14**:41–67, 2003.

[193] T. J. Loredo. From Laplace to supernova SN 1987A: Bayesian inference in astrophysics. In P. F. Fougere, editor, *Maximum Entropy and Bayesian Methods*, pages 81–142. Kluwer, 1990.

[194] D. J. C. MacKay. Bayesian interpolation. *Neural Computation*, **4**(3):415–447, 1992.

[195] D. J. C. MacKay. Probable networks and plausible predictions – a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, **6**(3):469–505, 1995.

[196] D. J. C. MacKay. Introduction to Gaussian processes. In *Neural Networks and Machine Learning*, volume 168 of NATO Advanced Study Institute on Generalization in Neural Networks and Machine Learning, pages 133–165. Springer, 1998.

[197] D. J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[198] U. Madhow. *Fundamentals of Digital Communication*. Cambridge University Press, 2008.

[199] K. V. Mardia, J. T. Kent and J. M. Bibby. *Multivariate Analysis*. Academic Press, 1997.

[200] H. Markram, J. Lubke, M. Frotscher and B. Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, **275**:213–215, 1997.

[201] A. McCallum, K. Nigam, J. Rennie, and K. Seymore. Automating the construction of internet portals with machine learning. *Information Retrieval Journal*, 3:127–163, 2000.

[202] G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley and Sons, 1997.

[203] G. McLachlan and D. Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics. Wiley-Interscience, 2000.

[204] E. Meeds, Z. Ghahramani, R. M. Neal and S. T. Roweis. Modeling dyadic data with binary latent factors. In B. Schölkopf, J. Platt and T. Hoffman, editors, *Advances in Neural Information Processing Systems (NIPS)*, volume 19, pages 977–984. MIT Press, 2007.

[205] M. Meila. An accelerated Chow and Liu algorithm: fitting tree distributions to high-dimensional sparse data. In I. Bratko, editor, *International Conference on Machine Learning*, pages 249–257. Morgan Kaufmann, 1999.

[206] M. Meila and M. I. Jordan. Triangulation by continuous embedding. In M. C. Mozer, M. I. Jordan and T. Petsche, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 9, pages 557–563. MIT Press, 1997.

[207] B. Mesot and D. Barber. Switching linear dynamical systems for noise robust speech recognition. *IEEE Transactions of Audio, Speech and Language Processing*, **15**(6):1850–1858, 2007.

[208] N. Meuleau, M. Hauskrecht, K-E. Kim *et al*. Solving very large weakly coupled Markov decision processes. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, pages 165–172, 1998.

[209] T. Mills. *The Econometric Modelling of Financial Time Series*. Cambridge University Press, 2000.

[210] T. Minka. Expectation propagation for approximate Bayesian inference. In J. Breese and D. Koller, editors, *Uncertainty in Artificial Intelligence*, number 17, pages 362–369. Morgan Kaufmann, 2001.

[211] T. Minka. A comparison of numerical optimizers for logistic regression. Technical report, Microsoft Research Ltd., 2003. research.microsoft.com/~minka/papers/logreg.

[212] T. Minka. Divergence measures and message passing. Technical Report MSR-TR-2005-173, Microsoft Research Ltd. December 2005.

[213] A. Mira, J. Møller and G. O. Roberts. Perfect slice samplers. *Journal of the Royal Statistical Society*, **63**(3):593–606, 2001. Series B (Statistical Methodology).

[214] C. Mitchell, M. Harper and L. Jamieson. On the complexity of explicit duration HMM's. *IEEE Transactions on Speech and Audio Processing*, **3**(3):213–217, 1995.

[215] T. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[216] J. Mooij and H. J. Kappen. Sufficient conditions for convergence of loopy belief propagation. *IEEE Information Theory*, **53**:4422–4437, 2007.

[217] J. W. Moon and L. Moser. On cliques in graphs. *Israel Journal of Mathematics*, (3):23–28, 1965.

[218] A. Moore. A tutorial on kd-trees. Technical report, 1991. Available from www.cs.cmu.edu/~awm/papers.html.

[219] J. Moussouris. Gibbs and Markov random systems with constraints. *Journal of Statistical Physics*, **10**:11–33, 1974.

[220] R. M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. CRG-TR-93-1, Dept. of Computer Science, University of Toronto, 1993.

[221] R. M. Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, **9**(2):249–265, 2000.

[222] R. M. Neal. Slice sampling. *Annals of Statistics*, **31**:705–767, 2003.

[223] R. E. Neapolitan. *Learning Bayesian Networks*. Prentice Hall, 2003.

[224] A. V. Nefian, L. Luhong, P. Xiaobo, X. Liu, C. Mao and K. Murphy. A coupled HMM for audio-visual speech recognition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 2013–2016, 2002.

[225] H. Nickisch and M. Seeger. Convex variational Bayesian inference for large scale generalized linear models. *International Conference on Machine Learning*, **26**:761–768, 2009.

[226] D. Nilsson. An efficient algorithm for finding the *m* most probable configurations in a probabilistic expert system. *Statistics and Computing*, **8**:159–173, 1998.

[227] D. Nilsson and J. Goldberger. Sequentially finding the *N*-best list in hidden Markov models. *International Joint Conference on Artificial Intelligence (IJCAI)*, 17, 2001.

[228] A. B. Novikoff. On convergence proofs on perceptrons. In *Symposium on the Mathematical Theory of Automata (New York, 1962)*, volume 12, pages 615–622. Polytechnic Press of Polytechnic Institute of Brooklyn, 1963.

[229] F. J. Och and H. Ney. Discriminative training and maximum entropy models for statistical machine translation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 295–302, Philadelphia, July 2002.

[230] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Research*, **37**:3311–3325, 1998.

[231] A. V. Oppenheim, R. W. Shafer, M. T. Yoder and W. T. Padgett. *Discrete-Time Signal Processing*. Prentice Hall, third edition, 2009.

[232] M. Ostendorf, V. Digalakis and O. A. Kimball. From HMMs to segment models: a unified view of stochastic modeling for speech recognition. *IEEE Transactions on Speech and Audio Processing*, **4**:360–378, 1995.

[233] P. Paatero and U. Tapper. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, **5**:111–126, 1994.

[234] A. Palmer, D. Wipf, K. Kreutz-Delgado and B. Rao. Variational EM algorithms for non-Gaussian latent variable models. In B. Schölkopf, J. Platt and T. Hoffman, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 19, pages 1059–1066. MIT Press, 2006.

[235] V. Pavlovic, J. M. Rehg and J. MacCormick. Learning switching linear models of human motion. In T. K. Leen, T. G. Dietterich and V. Tresp, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 13, pages 981–987. MIT Press, 2001.

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

[237] J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.

[238] B. A. Pearlmutter and L. C. Parra. Maximum likelihood blind source separation: a context-sensitive generalization of ICA. In M. C. Mozer, M. I. Jordan and T. Petsche, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 9, pages 613–619. MIT Press, 1997.

[239] K. B. Petersen and O. Winther. The EM algorithm in independent component analysis. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 169–172, 2005.

[240] J-P. Pfister, T. Toyiozumi, D. Barber and W. Gerstner. Optimal spike-timing dependent plasticity for precise action potential firing in supervised learning. *Neural Computation*, **18**:1309–1339, 2006.

[241] J. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schölkopf, C. J. C. Burges and A. J. Smola, editors, *Advances in Kernel Methods – Support Vector Learning*, pages 185–208. MIT Press, 1999.

[242] I. Porteous, D. Newman, A. Ihler, *et al.* Fast collapsed Gibbs sampling for latent Dirichlet allocation. In *KDD '08: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 569–577. ACM, 2008.

[243] J. E. Potter and R. G. Stern. Statistical filtering of space navigation measurements. In *American Institute of Aeronautics and Astronautics Guidance and Control Conference*, volume 13, pages 775–801, August 1963.

[244] W. Press, W. Vettering, S. Teukolsky and B. Flannery. *Numerical Recipes in Fortran*. Cambridge University Press, 1992.

[245] S. J. D. Prince and J. H. Elder. Probabilistic linear discriminant analysis for inferences about identity. In *IEEE 11th International Conference on Computer Vision ICCV*, pages 1–8, 2007.

[246] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**(2):257–286, 1989.

[247] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[248] H. E. Rauch, G. Tung and C. T. Striebel. Maximum likelihood estimates of linear dynamic systems. *American Institute of Aeronautics and Astronautics Journal (AIAAJ)*, **3**(8):1445–1450, 1965.

[249] T. Richardson and P. Spirtes. Ancestral graph Markov models. *Annals of Statistics*, **30**(4):962–1030, 2002.

[250] D. Rose, R. E. Tarjan and E. S. Lueker. Algorithmic aspects of vertex elimination of graphs. *SIAM Journal on Computing*, **5**:266–283, 1976.

[251] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, **65**(6):386–408, 1958.

[252] S. T. Roweis and L. J. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**(5500):2323–2326, 2000.

[253] D. B. Rubin. Using the SIR algorithm to simulate posterior distributions. In M. H. Bernardo, K. M. Degroot, D. V. Lindley and A. F. M. Smith, editors, *Bayesian Statistics 3*. Oxford University Press, 1988.

[254] D. Saad and M. Opper. *Advanced Mean Field Methods Theory and Practice*. MIT Press, 2001.

[255] R. Salakhutdinov, S. Roweis and Z. Ghahramani. Optimization with EM and expectation-conjugate-gradient. In T. Fawcett and N. Mishra, editors, *International Conference on Machine Learning*, number 20, pages 672–679. AAAI Press, 2003.

[256] L. K. Saul and M. I. Jordan. Exploiting tractable substructures in intractable networks. In D. S. Touretzky, M. Mozer and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 8, pages 486–492. MIT Press, 1996.

[257] L. Savage. *The Foundations of Statistics*. Wiley, 1954.

[258] R. D. Schachter. Bayes-ball: the rational pastime (for determining irrelevance and requisite information in belief networks and influence diagrams). In G. Cooper and S. Moral, editors, *Uncertainty in Artificial Intelligence*, number 14, pages 480–487. Morgan Kaufmann, 1998.

[259] B. Schölkopf, A. Smola and K. R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, **10**:1299–1319, 1998.

[260] N. N. Schraudolph and D. Kamenetsky. Efficient exact inference in planar Ising models. In D. Koller, D. Schuurmans, Y. Bengio and L. Bottou, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 21, pages 1417–1424. MIT Press, 2009.

[261] E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, **6**(2):461–464, 1978.

[262] M. Seeger. Gaussian processes for machine learning. *International Journal of Neural Systems*, **14**(2):69–106, 2004.

[263] M. Seeger. Expectation propagation for exponential families. Technical report, Department of EECS, Berkeley, 2005. www.kyb.tuebingen.mpg.de/bs/people/seeger.

[264] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[265] S. Siddiqi, B. Boots and G. Gordon. A constraint generation approach to learning stable linear dynamical systems. In J. C. Platt, D. Koller, Y. Singer and S. Roweis, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 20, pages 1329–1336. MIT Press, 2008.

[266] T. Silander, P. Kontkanen and P. Myllymäki. On sensitivity of the MAP Bayesian network structure to the equivalent sample size parameter. In R. Parr and L. van der Gaag, editors, *Uncertainty in Artificial Intelligence*, number 23, pages 360–367. AUAI Press, 2007.

[267] S. S. Skiena. *The Algorithm Design Manual*. Springer-Verlag, 1998.

[268] E. Smith and M. S. Lewicki. Efficient auditory coding. *Nature*, **439**(7079):978–982, 2006.

[269] P. Smolensky. *Parallel Distributed Processing: Volume 1: Foundations*, chapter Information processing in dynamical systems: foundations of harmony theory, pages 194–281. MIT Press, 1986.

[270] G. Sneddon. A statistical perspective on data assimilation in numerical Models. In *Studies in the Atmospheric Sciences*, number 144 in Lecture Notes in Statistics. Springer-Verlag, 2000.

[271] P. Sollich. Bayesian methods for support vector machines: evidence and predictive class probabilities. *Machine Learning*, **46**(1-3):21–52, 2002.

[272] D. X. Song, D. Wagner and X. Tian. Timing analysis of keystrokes and timing attacks on SSH. In *Proceedings of the 10th Conference on USENIX Security Symposium*. USENIX Association, 2001.

[273] A. S. Spanias. Speech coding: a tutorial review. *Proceedings of the IEEE*, **82**(10):1541–1582, Oct 1994.

[274] P. Spirtes, C. Glymour and R. Scheines. *Causation, Prediction, and Search*. MIT Press, second edition, 2000.

[275] N. Srebro. Maximum likelihood bounded tree-width Markov networks. In J. Breese and D. Koller, editors, *Uncertainty in Artificial Intelligence*, number 17, pages 504–511. Morgan Kaufmann, 2001.

[276] H. Steck. Constraint-based structural learning in Bayesian networks using finite data sets. PhD thesis, Technical University Munich, 2001.

[277] H. Steck. Learning the Bayesian network structure: Dirichlet prior vs data. In D. A. McAllester and P. Myllymaki, editors, *Uncertainty in Artificial Intelligence*, number 24, pages 511–518. AUAI Press, 2008.

[278] H. Steck and T. Jaakkola. On the Dirichlet prior and Bayesian regularization. In S. Becker, S. Thrun and K. Obermayer, editors, *NIPS*, pages 697–704. MIT Press, 2002.

[279] G. Strang. *Linear Algebra and Its Applications*. Brooks Cole, 1988.

[280] M. Studený. On mathematical description of probabilistic conditional independence structures. PhD thesis, Academy of Sciences of the Czech Republic, 2001.

[281] M. Studený. On non-graphical description of models of conditional independence structure. In *HSSS Workshop on Stochastic Systems for Individual Behaviours*. Louvain la Neueve, Belgium, 22–23 January 2001.

[282] C. Sutton and A. McCallum. An introduction to conditional random fields for relational learning. In L. Getoor and B. Taskar, editors, *Introduction to Statistical Relational Learning*. MIT Press, 2006.

[283] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[284] R. J. Swendsen and J-S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, **58**:86–88, 1987.

[285] B. K. Sy. A recurrence local computation approach towards ordering composite beliefs in Bayesian belief networks. *International Journal of Approximate Reasoning*, **8**:17–50, 1993.

[286] T. Sejnowski. The Book of Hebb. *Neuron*, **24**:773–776, 1999.

[287] R. E. Tarjan and M. Yannakakis. Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM Journal on Computing*, **13**(3):566–579, 1984.

[288] S. J. Taylor. *Modelling Financial Time Series*. World Scientific, second edition, 2008.

[289] Y. W. Teh, D. Newman and M. Welling. A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. In J. C. Platt, D. Koller, Y. Singer and S. Roweis, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 20, pages 1481–1488. MIT Press, 2008.

[290] Y. W. Teh and M. Welling. The unified propagation and scaling algorithm. In T. G. Dietterich, S. Becker and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 14, pages 953–960. MIT Press, 2002.

[291] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (B)*, 58:267–288, 1996.

[292] H. Tijms. *Understanding Probability*. Cambridge University Press, 2003.

[293] M. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analysers. *Neural Computation*, **11**(2):443–482, 1999.

[294] M. E. Tipping. Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, (1):211–244, 2001.

[295] M. E. Tipping. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, **1**:211–244, 2001.

[296] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, **61**(3):611–622, 1999.

[297] D. M. Titterington, A. F. M. Smith and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions*. Wiley, 1985.

[298] M. Toussaint, S. Harmeling and A. Storkey. Probabilistic inference for solving (PO)MDPs. Research Report EDI-INF-RR-0934, University of Edinburgh, School of Informatics, 2006.

[299] M. Tsodyks, K. Pawelzik and H. Markram. Neural networks with dynamic synapses. *Neural Computation*, **10**:821–835, 1998.

[300] L. van der Matten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, **9**:2579–2605, 2008.

[301] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems; Theory, Implementations, Applications*. Kluwer, 1996.

[302] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.

[303] M. Verhaegen and P. Van Dooren. Numerical aspects of different Kalman filter implementations. *IEEE Transactions of Automatic Control*, **31**(10):907–917, 1986.

[304] T. Verma and J. Pearl. Causal networks: semantics and expressiveness. In R. D. Schacter, T. S. Levitt, L. N. Kanal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, volume 4, pages 69–76. North-Holland, 1990.

[305] T. O. Virtanen, A. T. Cemgil and S. J. Godsill. Bayesian extensions to nonnegative matrix factorisation for audio signal modelling. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1825–1828, 2008.

[306] G. Wahba. *Support Vector Machines, Repreducing Kernel Hilbert Spaces, and Randomized GACV*, pages 69–88. MIT Press, 1999.

[307] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, **1**(1-2):1–305, 2008.

[308] H. Wallach. Efficient training of conditional random fields. Master's thesis, Division of Informatics, University of Edinburgh, 2002.

[309] Y. Wang, J. Hodges and B. Tang. Classification of web documents using a naive Bayes method. *15th IEEE International Conference on Tools with Artificial Intelligence*, pages 560–564, 2003.

[310] S. Waterhouse, D. Mackay and T. Robinson. Bayesian methods for mixtures of experts. In D. S. Touretzky, M. Mozer and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 8, pages 351–357. MIT Press, 1996.

[311] C. Watkins and P. Dayan. Q-learning. *Machine Learning*, **8**:279–292, 1992.

[312] Y. Weiss and W. T. Freeman. Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural Computation*, **13**(10):2173–2200, 2001.

[313] M. Welling, T. P. Minka and Y. W. Teh. Structured region graphs: morphing EP into GBP. In F. Bacchus and T. Jaakkola, editors, *Uncertainty in Artificial Intelligence*, number 21, pages 609–614. AUAI press, 2005.

[314] J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. John Wiley & Sons, 1990.

[315] W. Wiegerinck. Variational approximations between mean field theory and the junction tree algorithm. In C. Boutilier and M. Goldszmidt, editors, *Uncertainty in Artificial Intelligence*, number 16, pages 626–633. Morgan Kaufmann, 2000.

[316] W. Wiegerinck and T. Heskes. Fractional belief propagation. In S. Becker, S. Thrun and K. Obermayer, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 15, pages 438–445. MIT Press, 2003.

[317] C. K. I. Williams. Computing with infinite networks. In M. C. Mozer, M. I. Jordan and T. Petsche, editors, *Advances in Neural Information Processing Systems (NIPS), number 9*, pages 295–301. MIT Press, 1997.

[318] C. K. I. Williams and D. Barber. Bayesian classification with Gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**:1342–1351, 1998.

[319] C. Yanover and Y. Weiss. Finding the M most probable configurations using loopy belief propagation. In S. Thrun, L. Saul and B. Schölkopf, editors, *Advances in Neural Information Processing Systems (NIPS)*, number 16, pages 1457–1464. MIT Press, 2004.

[320] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, **51**(7):2282–2312, July 2005.

[321] S. Young, D. Kershaw, J. Odell, *et al. The HTK Book Version 3.0*. Entropic Cambridge Research Laboratory, 2000.

[322] A. L. Yuille and A. Rangarajan. The concave-convex procedure. *Neural Computation*, **15**(4):915–936, 2003.

[323] J.-H. Zhao, P. L. H. Yu and Q. Jiang. ML estimation for factor analysis: EM or non-EM? *Statistics and Computing*, **18**(2):109–123, 2008.

[324] O. Zoeter. Monitoring non-linear and switching dynamic systems. PhD thesis, Radboud University Nijmegen, 2005.