

APPENDIX A

EXPONENTIAL FAMILY DISTRIBUTIONS

This appendix lists the exponential family distributions referred to in this thesis. Each distribution is written in its ordinary form and then in the standard exponential family form, defined as,

$$P(x | \mathbf{y}) = \exp[\boldsymbol{\phi}(\mathbf{y})^T \mathbf{u}(x) + f(x) + g(\mathbf{y})] \quad (\text{A.1})$$

where $\boldsymbol{\phi}(\mathbf{y})$ is the natural parameter vector and $\mathbf{u}(x)$ is the natural statistic vector. For each distribution P , the expectation of the natural statistic vector under P is also given.

A.1 Gaussian distribution

The Gaussian distribution, also known as the Normal distribution, is defined as

$$\begin{aligned} P(x) &= \mathcal{N}(x | \mu, \gamma^{-1}) \\ &= \sqrt{\frac{\gamma}{2\pi}} \exp\left[-\frac{\gamma}{2}(x - \mu)^2\right] \quad \gamma > 0 \end{aligned} \quad (\text{A.2})$$

where μ is the mean and γ is the precision or inverse variance. Being a member of the exponential family, this distribution can then be written in standard form as

$$\mathcal{N}(x | \mu, \gamma^{-1}) = \exp\left(\left[\begin{array}{c} \gamma\mu \\ -\frac{\gamma}{2} \end{array}\right]^T \left[\begin{array}{c} x \\ x^2 \end{array}\right] + \frac{1}{2}(\log \gamma - \gamma\mu^2 - \log 2\pi)\right). \quad (\text{A.3})$$

The expectation under P of the natural statistic vector \mathbf{u}_x is given by

$$\langle \mathbf{u}(x) \rangle_P = \left\langle \left[\begin{array}{c} x \\ x^2 \end{array}\right] \right\rangle_P = \left[\begin{array}{c} \mu \\ \mu^2 + \gamma^{-1} \end{array}\right]. \quad (\text{A.4})$$

The entropy of the Gaussian distribution can be found by replacing $\mathbf{u}(x)$ by $\langle \mathbf{u}(x) \rangle_P$ in the negative log of (A.3), giving

$$\mathbb{H}(P) = \frac{1}{2}(1 + \log 2\pi - \log \gamma). \quad (\text{A.5})$$

A.2 Rectified Gaussian distribution

The rectified Gaussian distribution is equal to the Gaussian except that it is zero for negative x , so

$$\begin{aligned} P(x) &= \mathcal{N}^R(x | \mu, \gamma^{-1}) \\ &= \begin{cases} \frac{\sqrt{2\gamma}}{\sqrt{\pi} \operatorname{erfc}(-\mu\sqrt{\gamma/2})} \exp\left[-\frac{\gamma}{2}(x - \mu)^2\right] & : x \geq 0 \\ 0 & : x < 0 \end{cases} \end{aligned} \quad (\text{A.6})$$

where erfc is the complementary error function

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2) dt. \quad (\text{A.7})$$

The rectified Gaussian distribution has the same natural parameter and statistic vectors as the standard Gaussian. It differs only in f and g , which are now defined as

$$f(x) = \begin{cases} 0 & : x \geq 0 \\ -\infty & : x < 0 \end{cases} \quad (\text{A.8})$$

$$g(\mu, \gamma) = \frac{1}{2}(\log 2\gamma - \log \pi - \gamma\mu^2) - \log \operatorname{erfc}(-\mu\sqrt{\gamma/2}). \quad (\text{A.9})$$

The expectation of the natural statistic vector is

$$\left\langle \begin{bmatrix} x \\ x^2 \end{bmatrix} \right\rangle_P = \begin{bmatrix} \mu + \sqrt{\frac{2}{\pi\gamma}} \frac{1}{\operatorname{erfcx}(-\mu\sqrt{\gamma/2})} \\ \mu^2 + \gamma^{-1} + \sqrt{\frac{1}{\pi\gamma}} \frac{\mu}{\operatorname{erfcx}(-\mu\sqrt{\gamma/2})} \end{bmatrix} \quad (\text{A.10})$$

where the scaled complementary error function $\operatorname{erfcx}(x) = \exp(x^2) \times \operatorname{erfc}(x)$. For notes on avoiding numerical errors when using Equation A.10, see Miskin [2000].

A.3 Gamma distribution

The Gamma distribution is defined over nonnegative x to be

$$\begin{aligned} P(x) &= \operatorname{Gamma}(x | a, b) \\ &= \frac{b^a x^{a-1} \exp(-bx)}{\Gamma(a)} \quad x \geq 0; a, b > 0 \end{aligned} \quad (\text{A.11})$$

where a is the shape and b is the inverse scale. Rewriting in exponential family standard form gives

$$\operatorname{Gamma}(x | a, b) = \exp \left(\begin{bmatrix} -b \\ a-1 \end{bmatrix}^T \begin{bmatrix} x \\ \log x \end{bmatrix} + a \log b - \log \Gamma(a) \right) \quad (\text{A.12})$$

and the expectation of the natural statistic vector is

$$\left\langle \begin{bmatrix} x \\ \log x \end{bmatrix} \right\rangle_P = \begin{bmatrix} a/b \\ \Psi(a) - \log b \end{bmatrix} \quad (\text{A.13})$$

where we define the digamma function $\Psi(a) = \frac{\delta}{\delta a} \log \Gamma(a)$. It follows that the entropy of the Gamma distribution is

$$\mathbb{H}(P) = a + (1 - a)\Psi(a) - \log b + \log \Gamma(a). \quad (\text{A.14})$$

A.4 Discrete distribution

A discrete, or multinomial, distribution is a probability distribution over a variable which can take on only discrete values. If x can only take one of K discrete values, the distribution is defined as

$$\begin{aligned} P(x) &= \text{Discrete}(x | p_1 \dots p_K) \\ &= \sum_{i=1}^K p_i \delta(x - i) \quad x \in \{1 \dots K\}; \sum_{i=1}^K p_i = 1 \end{aligned} \quad (\text{A.15})$$

$$= p_x. \quad (\text{A.16})$$

The discrete distribution is a member of the exponential family and can be written in standard form

$$\text{Discrete}(x | p_1 \dots p_K) = \exp \left(\begin{bmatrix} \log p_1 \\ \vdots \\ \log p_K \end{bmatrix}^T \begin{bmatrix} \delta(x - 1) \\ \vdots \\ \delta(x - K) \end{bmatrix} \right) \quad (\text{A.17})$$

and the expectation of the natural statistic vector is simply

$$\left\langle \begin{bmatrix} \delta(x - 1) \\ \vdots \\ \delta(x - K) \end{bmatrix} \right\rangle_P = \begin{bmatrix} p_1 \\ \vdots \\ p_K \end{bmatrix}. \quad (\text{A.18})$$

The entropy of the discrete distribution is $\mathbb{H}(P) = -\sum_i p_i \log p_i$.

A.5 Dirichlet distribution

The Dirichlet distribution is the conjugate prior for the parameters $p_1 \dots p_K$ of a discrete distribution

$$\begin{aligned}
P(p_1 \dots p_K) &= \text{Dir}(p_1 \dots p_K | \mathbf{u}) \\
&= \frac{1}{Z(\mathbf{u})} \prod_{i=1}^K p_i^{u_i-1} \quad p_i > 0, \sum_i p_i = 1; u_i > 0
\end{aligned} \tag{A.19}$$

where u_i is the i th element of the parameter vector \mathbf{u} and the normalisation constant $Z(\mathbf{u})$ is

$$Z(\mathbf{u}) = \frac{\prod_{i=1}^K \Gamma(u_i)}{\Gamma\left(\sum_{i=1}^K u_i\right)}. \tag{A.20}$$

The Dirichlet distribution can be written in exponential family form as

$$\text{Dir}(p_1 \dots p_K | \mathbf{u}) = \exp \left(\begin{bmatrix} u_1 - 1 \\ \vdots \\ u_K - 1 \end{bmatrix}^T \begin{bmatrix} \log p_1 \\ \vdots \\ \log p_K \end{bmatrix} + \log \Gamma(U) - \sum_{i=1}^K \Gamma(u_i) \right) \tag{A.21}$$

where $U = \sum_{i=1}^K u_i$. The expectation under P of the natural statistic vector is

$$\left\langle \begin{bmatrix} \log p_1 \\ \vdots \\ \log p_K \end{bmatrix} \right\rangle_P = \begin{bmatrix} \psi(u_1) - \psi(U) \\ \vdots \\ \psi(u_K) - \psi(U) \end{bmatrix} \tag{A.22}$$

where the digamma function $\psi(\cdot)$ is as defined previously.

BIBLIOGRAPHY

- D. Ackley, G. Hinton, and T. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science*, 9:147–169, 1985.
- H. Attias. A variational Bayesian framework for graphical models. In S. Solla, T. K. Leen, and K-L Muller, editors, *Advances in Neural Information Processing Systems*, volume 12, pages 209–215, Cambridge MA, 2000. MIT Press.
- Z. Bar-Joseph, D. Gifford, and T. Jaakkola. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics*, 17:S22–29, 2001.
- D. Barber and C. M. Bishop. Variational learning in Bayesian neural networks. In C. M. Bishop, editor, *Generalization in Neural Networks and Machine Learning*. Springer Verlag, 1998.
- K. J. Bathe. *Finite Element Procedures*. Prentice-Hall, Englewood Cliffs, NJ, 1996.
- Rev. T. Bayes. An essay towards solving a problem in the doctrine of chances. In *Philosophical Transactions of the Royal Society*, volume 53, pages 370–418, 1763.
- A. Ben-Dor, R. Shamir, and Z. Yakhini. Clustering gene expression patterns. *Journal of Computational Biology*, 6(3/4):281–297, 1999.
- J. M. Bernardo and A.F.M. Smith. *Bayesian Theory*. John Wiley and Sons, New York, 1994.
- C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- C. M. Bishop. Bayesian PCA. In S. A. Solla M. S. Kearns and D. A. Cohn, editors, *Advances in Neural Information Processing Systems*, volume 11, pages 382–388. MIT Press, 1999a.
- C. M. Bishop. Variational principal components. In *Proceedings Ninth International Conference on Artificial Neural Networks, ICANN'99*, volume 1, pages 509–514. IEE, 1999b.
- C. M. Bishop and M. E. Tipping. A hierarchical latent variable model for data visualization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):281–293, 1998.
- C. M. Bishop and M. E. Tipping. Variational Relevance Vector Machines. In *Proceedings of 16th Conference in Uncertainty in Artificial Intelligence*, pages 46–53. Morgan Kaufmann, 2000.

- C. M. Bishop and J. M. Winn. Non-linear Bayesian image modelling. In *Proceedings Sixth European Conference on Computer Vision*, volume 1, pages 3–17. Springer-Verlag, 2000.
- C. M. Bishop and J. M. Winn. Structured variational distributions in VIBES. In *Proceedings Artificial Intelligence and Statistics*, Key West, Florida, 2003. Society for Artificial Intelligence and Statistics.
- C. M. Bishop, J. M. Winn, and D. Spiegelhalter. VIBES: A variational inference engine for Bayesian networks. In *Advances in Neural Information Processing Systems*, volume 15, 2002.
- M. J. Black and Y. Yacoob. Recognizing facial expressions under rigid and non-rigid facial motions. In *International Workshop on Automatic Face and Gesture Recognition, Zurich*, pages 12–17, 1995.
- C. Bregler and S.M. Omohundro. Nonlinear manifold learning for visual speech recognition. In *Fifth International Conference on Computer Vision*, pages 494–499, Boston, Jun 1995.
- J. Buhler, T. Ideker, and D. Haynor. Dapple: Improved techniques for finding spots on DNA microarrays. Technical report, University of Washington, 2000.
- R. Choudrey, W. Penny, and S. Roberts. An ensemble learning approach to independent component analysis. In *IEEE International Workshop on Neural Networks for Signal Processing*, 2000.
- G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990.
- T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models — their training and application. In *Computer vision, graphics and image understanding*, volume 61, pages 38–59, 1995.
- R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Statistics for Engineering and Information Science. Springer-Verlag, 1999.
- R. T. Cox. Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1):1–13, 1946.
- P. Dagum and M. Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60:141–153, 1993.
- A. Darwiche. Conditioning methods for exact and approximate inference in causal networks. In *Eleventh Annual Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, August 1995.

- S. Dudoit, Y. H. Yang, Matthew J. Callow, and T. P. Speed. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Technical report, Department of Biochemistry, Stanford University School of Medicine, 2000.
- M. Eisen, P. Spellman, D. Botstein, and P. Brown. Cluster analysis and display of genome-wide expression patterns. In *Proceedings of National Academy of Science*, volume 95, pages 14863–14867, 1998.
- M.B. Eisen and P.O. Brown. DNA arrays for analysis of gene expression. *Methods in Enzymology*, 303:179–205, 1999.
- B.S. Everitt and D.J. Hand. *Finite Mixture Distributions*. Chapman and Hall, London, 1981.
- R.P. Feynman. *Statistical Mechanics*. W. A. Benjamin, Inc., MA, 1972.
- B. Frey. *Graphical models for machine learning and digital communications*. MIT Press, Cambridge, MA, 1998.
- B. Frey and N. Jojic. Transformed component analysis: joint estimation of spatial transformations and image components. In *Seventh International Conference on Computer Vision*, pages 1190–1196, 1999.
- B. Frey, F. Kschischang, H. Loeliger, and N. Wiberg. Factor graphs and algorithms. In *Proceedings of the 35th Allerton Conference on Communication, Control and Computing 1997*, 1998.
- N. Friedman, M. Linial, I. Nachman, and D. Pe’er. Using Bayesian networks to analyze expression data. In *RECOMB*, pages 127–135, 2000.
- R.G. Gallager. Low density parity check codes. *IRE Trans. Info. Theory*, IT-8:21–28, Jan 1962.
- R.G. Gallager. *Low density parity check codes*. Number 21 in Research monograph series. MIT Press, Cambridge, MA, 1963.
- S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1): 721–741, 1984.
- Z. Ghahramani and M. J. Beal. Variational inference for Bayesian mixture of factor analysers. In *Advances in Neural Information Processing Systems*, volume 12, 1999.
- Z. Ghahramani and M. J. Beal. Propagation algorithms for variational Bayesian learning. In T. K. Leen, T. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems*, volume 13, Cambridge MA, 2001. MIT Press.

- W. R. Gilks and P. Wild. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41(2):337–348, 1992.
- A. Hartemink, D. Gifford, T. Jaakkola, and R. Young. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. In *Pacific Symposium on Biocomputing*, volume 6, pages 422–433, 2001.
- T. Heap and D. Hogg. Wormholes in shape space: Tracking through discontinuous changes in shape. In *Sixth International Conference on Computer Vision*, pages 344–349, 1998.
- P. Hegde, R. Qi, R. Abernathy, C. Gay, S. Dharap, R. Gaspard R, J. Earle-Hughes, E. Snesrud, N. H. Lee, and J. Quackenbush. A concise guide to cDNA microarray analysis. *Biotechniques*, 29(3):548–562, 2000.
- T. Heskes. Stable fixed points of loopy belief propagation are minima of the Bethe free energy, 2002.
- G. E. Hinton and D. van Camp. Keeping neural networks simple by minimizing the description length of the weights. In *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, pages 5–13, 1993.
- G. E. Hinton and R. S. Zemel. Autoencoders, minimum description length, and Helmholtz free energy. In *Advances in Neural Information Processing Systems*, volume 6, 1994.
- G. Hori, M. Inoue, S. Nishimura, and H. Nakahara. Blind gene classification on ICA of microarray data. In *ICA 2001*, pages 332–336, 2001.
- T. Jaakkola. *Variational Methods for Inference and Estimation in Graphical Models*. PhD thesis, MIT, 1997.
- F. Jensen. *An introduction to Bayesian networks*. UCL Press, 1996.
- M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 105–162. Kluwer, 1998.
- N. Kambhatla and T.K. Leen. Dimension reduction by local principal component analysis. *Neural Computation*, 9(7):1493–1516, 1997.
- R. Kinderman and J. L. Snell. Markov random fields and their applications. *American Mathematical Society*, 1:1–142, 1980.
- F. R. Kschischang, B. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Trans. Inform. Theory*, 47(2):498–519, 2001.
- S. Kullback. *Information Theory and Statistics*. Dover Publications, New York, 1959.

- S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.
- S. L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*, 50:157–224, 1988.
- S. L. Lauritzen and N. Wermuth. Graphical models for associations between variables, some of which are qualitative and some quantitative. *Annals of Statistics*, 17:31–57, 1989.
- N. Lawrence, M. Milo, M. Niranjana, P. Rashbass, and S. Soullier. Reducing the variability in microarray image processing by Bayesian inference. Technical report, Department of Computer Science, University of Sheffield, 2002.
- D. B. Lenat. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, 1995.
- D. J. Lunn, A. Thomas, N. G. Best, and D. J. Spiegelhalter. WinBUGS – a Bayesian modelling framework: concepts, structure and extensibility. *Statistics and Computing*, 10:321–333, 2000. <http://www.mrc-bsu.cam.ac.uk/bugs/>.
- 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.
- D. J. C. MacKay. Ensemble learning for hidden Markov models, 1997. Unpublished manuscript, Department of Physics, University of Cambridge.
- D. J. C. MacKay. Introduction to Monte Carlo methods. In M. I. Jordan, editor, *Learning in Graphical Models*. Kluwer, 1998.
- D. J. C. MacKay. *Information theory, inference, and learning algorithms*. Cambridge University Press, Cambridge, UK, 2003.
- D. J. C. MacKay and R. M. Neal. Good codes based on very sparse matrices. In *IMA: IMA Conference on Cryptography and Coding, LNCS lately (earlier: Cryptography and Coding II, Edited by Chris Mitchell, Clarendon Press, 1992)*, 1995.
- A. Martoglio, J. W. Miskin, S. K. Smith, and D. J. C. MacKay. A decomposition model to track gene expression signatures: preview on observer-independent classification of ovarian cancer. *Bioinformatics*, 18:1617–1624, 2002.
- A. Martoglio, B. D. Tom, M. Starkey, A. N. Corps, S. Charnock-Jones, and S. K. Smith. Changes in tumorigenesis- and angiogenesis-related gene transcript abundance profiles in ovarian cancer detected by tailored high density cDNA arrays. *Molecular Medicine*, 6(9): 750–765, 2000.

- R. J. McEliece, D. J. C. MacKay, and J.-F. Cheng. Turbo decoding as an instance of Pearl's Belief Propagation algorithm. *IEEE Journal on selected areas in communication*, 1997.
- G. S. Michaels, D. B. Carr, M. Askenazi, S. Fuhrman, X. Wen, and R. Somogyi. Cluster analysis and data visualization of large-scale gene expression data. In *Pacific Symposium on Biocomputing*, volume 3, pages 42–53, 1998.
- T. P. Minka. Expectation propagation for approximate Bayesian inference. In *Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence*, pages 362–369. Morgan Kaufmann, 2001a.
- T. P. Minka. *A family of algorithms for approximate Bayesian inference*. PhD thesis, MIT, 2001b.
- J. W. Miskin. *Ensemble Learning for Independent Component Analysis*. PhD thesis, University of Cambridge, 2000.
- J. W. Miskin and D. J. C. MacKay. Ensemble learning for blind source separation. In S. J. Roberts and R. M. Everson, editors, *ICA: Principles and Practice*. Cambridge University Press, 2000.
- B. Moghaddam. Principal manifolds and Bayesian subspaces for visual recognition. In *Seventh International Conference on Computer Vision*, pages 1131–1136, 1999.
- B. Moghaddam and A. Pentland. Probabilistic visual learning for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.
- V.S. Nalwa. *A Guided Tour of Computer Vision*. Addison-Wesley, 1993.
- R. M. Neal. Connectionist learning of belief networks. *Artificial Intelligence*, 56:71–113, 1992.
- R. M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, Canada, 1993.
- R. M. Neal. *Bayesian Learning for Neural Networks*. PhD thesis, University of Toronto, Canada, 1994.
- R. M. Neal and G. E. Hinton. A new view of the EM algorithm that justifies incremental and other variants. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 355–368. Kluwer, 1998.
- J. Pearl. Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, 29: 241–288, 1986.
- J. Pearl. Evidential reasoning using stochastic simulation of causal models. *Artificial Intelligence*, 32:245–257, 1987.

- J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Francisco, 1988.
- S. Raychaudhuri, J. Stuart, and R. Altman. Principal Components Analysis to summarize microarray experiments: application to sporulation time series. In *Pacific Symposium on Biocomputing*, volume 5, 2000.
- S. Roweis. EM algorithms for PCA and SPCA. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10. MIT Press, 1998.
- J. Rustagi. *Variational Methods in Statistics*. Academic Press, New York, 1976.
- J. Sakurai. *Modern Quantum Mechanics*. Addison-Wesley, Redwood City, CA, 1985.
- L. K. Saul and M. I. Jordan. Exploiting tractable substructures in intractable networks. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 486–492. MIT Press, 1996.
- E. H. Shortcliffe. *Computer-Based Medical Consultations: MYCIN*. Elsevier Science, New York, 1976.
- P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher. Comprehensive identification of cell cycle-regulated genes of the yeast *saccharomyces cerevisiae* by microarray hybridization. In *Molecular Biology of the Cell*, volume 9, pages 3273–3297, 1998.
- D. J. Spiegelhalter. Probabilistic reasoning in predictive expert systems. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 47–68, Amsterdam, 1986. North Holland.
- P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. Lander, and T. Golub. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. In *Proceedings of the National Academy of Science*, volume 96, pages 2907–2912, 1999.
- A. Thomas, D. J. Spiegelhalter, and W. R. Gilks. BUGS: A program to perform Bayesian inference using Gibbs sampling. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, Oxford: Clarendon Press, 1992.
- M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, 1999a.
- M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 21(3):611–622, 1999b.

- M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
- N. Ueda, R. Nakano, Z. Ghahramani, and G. E. Hinton. SMEM algorithm for mixture models. In *Advances in Neural Information Processing Systems*, volume 11, 1999.
- P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.
- Niclas Wiberg. *Codes and Decoding on General Graphs*. PhD thesis, Linköping University, 1996.
- W. Wiegnerinck. Variational approximations between mean field theory and the junction tree algorithm. In *Uncertainty in Artificial Intelligence*. Morgan Kaufmann, 2000.
- C. K. I. Williams and C. E. Rasmussen. Gaussian processes for regression. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 514–520. MIT Press, 1996.
- P. H. Winston. *Artificial Intelligence*. Addison-Wesley, third edition, 1992.
- E. P. Xing, M. I. Jordan, and S. Russell. A generalized mean field algorithm for variational inference in exponential families. In *Uncertainty in Artificial Intelligence*. Morgan Kaufmann, 2003.
- J. Yedidia, W. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. In G. Lakemeyer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millennium*. Morgan Kaufmann, 2002.
- K. Yeung, C. Fraley, A. Murua, A. Raftery, and W. Ruzzo. Model-based clustering and data transformations for gene expression data. *Bioinformatics*, 17(10):977–987, 2001.