

References

- Achlioptas, D. (2001). Database-friendly random projections. In *Symposium on Principles of Database Systems*, pages 274–281.
- Ahmed, S.W. and Lachenbruch, P.A. (1977). Discriminant analysis when scale contamination is present in the initial sample. In *Classification and Clustering*, pages 331–353, New York. Academic Press Inc.
- Arriaga, R.I. and Vempala, S. (1999). An algorithmic theory of learning: Robust concepts and random projection. In *Proc. Foundations of Computer Science*, pages 616–623.
- Bahl, L., Brown, P., de Souza, P., and Mercer, R. (1993). Estimating hidden Markov model parameters so as to maximize speech recognition accuracy. *IEEE Transactions on Speech and Audio Processing*, 1:77–83.
- Balasubramanian, V., Myung, J., and Pitt, M.A. (2000). Counting probability distributions: Differential geometry and model selection. In *Proceedings of National Academy of Sciences*, volume 97, pages 11170–11175.
- Baluja, S. (1998). Probabilistic modeling for face orientation discrimination: Learning from labeled and unlabeled data. In *Neural Information and Processing Systems*, pages 854–860.
- Bassili, J.N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, 37(11):2049–2058.
- Bengio, Y. and Frasconi, P. (1996). Input-output HMM’s for sequence processing. *IEEE Transactions on Neural Networks*, 7(5):1231–1249.
- Bennett, K. and Demiriz, A. (1998). Semi-supervised support vector machines. In *Neural Information and Processing Systems*, pages 368–374.
- Berk, R.H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Annals of Mathematical Statistics*, 37:51–58.
- Bianchi-Berthouze, N. and Lisetti, C. (2002). Modeling multimodal expression of user’s affective subjective experience. *User Modeling and User-Adapted Interaction*, 12:49–84.
- Binder, J., Koller, D., Russell, S.J., and Kanazawa, K. (1997). Adaptive probabilistic networks with hidden variables. *Machine Learning*, 29(2):213–244.
- Bischof, W. and Caelli, T. (2001). Learning spatio-temporal relational structures. *Applied Artificial Intelligence*, 15(8):707–722.
- Black, M.J. and Yacoob, Y. (1995). Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In *Proc. International Conference on Computer Vision*, pages 374–381.
- Blake, C.L. and Merz, C.J. (1998). UCI repository of machine learning databases.

- Blockeel, H. and De Raedt, L. (1998). Top-down induction of first-order logical decision trees. *Artificial Intelligence*, 101(1-2):285–297.
- Blum, A. and Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In *Conference on Learning Theory*, pages 92–100.
- Blumer, A., Ehrenfeucht, A., Haussler, D., and Warmuth, M.K. (1989). Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–865.
- Brand, M. (1998). An entropic estimator for structure discovery. In *Neural Information and Processing Systems*, pages 723–729.
- Brand, M. and Kettner, V. (2000). Discovery and segmentation of activities in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):844–851.
- Brand, M., Oliver, N., and Pentland, A. (1997). Coupled hidden Markov models for complex action recognition. In *International Conference on Pattern Recognition*, pages 994–999.
- Brandstein, M.S. and Silverman, H.F. (1997). A practical methodology for speech source localization with microphone arrays. *Computer, Speech, and Language*, 1(2):91–126.
- Bruce, R. (2001). Semi-supervised learning using prior probabilities and EM. In *International Joint Conference on Artificial Intelligence, Workshop on Text Learning: Beyond Supervision*.
- Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167.
- Buxton, H. and Gong, S. (1995). Advanced visual surveillance using Bayesian networks. In *International Conference on Computer Vision*, pages 111–123.
- Cacioppo, J.T. and Tassinary, L.G. (1990). Inferring psychological significance from physiological signals. *American Psychologist*, 45:16–28.
- Cannon, W. B. (1927). The James-Lange theory of emotion: A critical examination and an alternative theory. *American Journal of Psychology*, 39:106–124.
- Castelli, V. (1994). *The Relative Value of Labeled and Unlabeled Samples in Pattern Recognition*. PhD thesis, Stanford University.
- Castelli, V. and Cover, T. (1995). On the exponential value of labeled samples. *Pattern Recognition Letters*, 16:105–111.
- Castelli, V. and Cover, T. (1996). The relative value of labeled and unlabeled samples in pattern recognition with an unknown mixing parameter. *IEEE Transactions on Information Theory*, 42(6):2102–2117.
- Chellappa, R., Wilson, C.L., and Sirohey, S. (1995). Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740.
- Chen, L.S. (2000). *Joint Processing of Audio-visual Information for the Recognition of Emotional Expressions in Human-computer Interaction*. PhD thesis, University of Illinois at Urbana-Champaign, Dept. of Electrical Engineering.
- Chen, T. and Rao, R. (1998). Audio-visual integration in multimodal communication. *Proceedings of the IEEE*, 86(5):837–852.
- Cheng, J., Bell, D.A., and Liu, W. (1997). Learning belief networks from data: An information theory based approach. In *International Conference on Information and Knowledge Management*, pages 325–331.
- Cheng, J. and Greiner, R. (1999). Comparing Bayesian network classifiers. In *Proc. Conference on Uncertainty in Artificial Intelligence*, pages 101–107.
- Chhikara, R. and McKeon, J. (1984). Linear discriminant analysis with misallocation in training samples. *Journal of the American Statistical Association*, 79:899–906.
- Chittineni, C. (1981). Learning with imperfectly labeled examples. *Pattern Recognition*, 12:271–281.
- Chow, C.K. and Liu, C.N. (1968). Approximating discrete probability distribution with dependence trees. *IEEE Transactions on Information Theory*, 14:462–467.

- Christian, A.D. and Avery, B.L. (1998). Digital Smart Kiosk project. In *ACM SIGCHI*, pages 155–162.
- Clarkson, B. and Pentland, A. (1999). Unsupervised clustering of ambulatory audio and video. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 3037–3040.
- Cohen, I., Cozman, F.G., and Bronstein, A. (2002a). On the value of unlabeled data in semi-supervised learning based on maximum-likelihood estimation. Technical Report HPL-2002-140, HP-Labs.
- Cohen, I., Garg, A., and Huang, T.S. (2000). Emotion recognition using multi-level HMMs. In *Neural Information Processing Systems, Workshop on Affective Computing*.
- Cohen, I., Sebe, N., Cozman, F., Cirelo, M., and Huang, T.S. (to appear, 2004). Semi-supervised learning of classifiers: Theory, algorithms, and applications to human-computer interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Cohen, I., Sebe, N., Cozman, F.G., Cirelo, M.C., and Huang, T.S. (2003a). Learning Bayesian network classifiers for facial expression recognition using both labeled and unlabeled data. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 595–601.
- Cohen, I., Sebe, N., Garg, A., Chen, L., and Huang, T.S. (2003b). Facial expression recognition from video sequences: Temporal and static modeling. *Computer Vision and Image Understanding*, 91(1-2):160–187.
- Cohen, I., Sebe, N., Garg, A., and Huang, T.S. (2002b). Facial expression recognition from video sequences. In *International Conference on Multimedia and Expo*, volume 2, pages 121–124.
- Cohen, I., Sebe, N., Sun, Y., Lew, M.S., and Huang, T.S. (2003c). Evaluation of expression recognition techniques. In *International Conference on Image and Video Retrieval*, pages 184–195.
- Collins, M. and Singer, Y. (2000). Unupervised models for named entity classification. In *International Conference on Machine Learning*, pages 327–334.
- Colmenarez, A.J. and Huang, T.S. (1997). Face detection with information based maximum discrimination. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 782–787.
- Comite, F. De, Denis, F., Gilleron, R., and Letouzey, F. (1999). Positive and unlabeled examples help learning. In *Proc. International Conference on Algorithmic Learning Theory*, pages 219–230.
- Cooper, D.B. and Freeman, J.H. (1970). On the asymptotic improvement in the outcome of supervised learning provided by additional non-supervised learning. *IEEE Transactions on Computers*, C-19(11):1055–1063.
- Cooper, G. and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:308–347.
- Corduneanu, A. and Jaakkola, T. (2002). Continuations methods for mixing heterogeneous sources. In *Proc. Conference on Uncertainty in Artificial Intelligence*, pages 111–118.
- Cormen, T.H., Leiserson, C.E., and Rivest, R.L. (1990). *Introduction to Algorithms*. MIT Press, Cambridge, MA.
- Cover, T.M. and Thomas, J.A. (1991). *Elements of Information Theory*. John Wiley and Sons, New York.
- Cozman, F.G. and Cohen, I. (2001). Unlabeled data can degrade classification performance of generative classifiers. Technical Report HPL-2001-234, HP-Labs.
- Cozman, F.G. and Cohen, I. (2003). The effect of modeling errors in semi-supervised learning of mixture models: How unlabeled data can degrade performance of generative classifiers. Technical report, <http://www.poli.usp.br/p/fabio.cozman/Publications/lul.ps.gz>.

- Craw, I., Tock, D., and Bennett, A. (1992). Finding face features. In *European Conference on Computer Vision*, pages 92–96.
- Crowley, J. and Berard, F. (1997). Multi-modal tracking of faces for video communications. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 640–645.
- Dai, Y. and Nakano, Y. (1996). Face-texture model based on SGLD and its application in face detection in color scene. *Pattern Recognition*, 29(6):1007–1017.
- Darell, T., Gordon, G., Harville, M., and Woodfill, J. (2000). Integrated person tracking using stereo, color, and pattern decision. *International Journal of Computer Vision*, 37(2):175–185.
- Darwin, C. (1890). *The Expression of the Emotions in Man and Animals*. John Murray, London, 2nd edition.
- Dawid, A.P. (1976). Properties of diagnostic data distributions. *Biometrics*, 32:647–658.
- De Silva, L.C., Miyasato, T., and Natatsu, R. (1997). Facial emotion recognition using multi-modal information. In *Proc. IEEE International Conference on Information, Communications, and Signal Processing*, pages 397–401.
- DeJong, K. (1988). Learning with genetic algorithms: An overview. *Machine Learning*, 3:121–138.
- Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38.
- Devroye, L., Györfi, L., and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer Verlag, New York.
- Domingos, P. and Pazzani, M. (1997). Beyond independence: Conditions for the optimality of the simple Bayesian classifier. *Machine Learning*, 29:103–130.
- Dougherty, J., Kohavi, R., and Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. In *International Conference on Machine Learning*, pages 194–202.
- Duda, R.O. and Hart, P.E. (1973). *Pattern Classification and Scene Analysis*. John Wiley and Sons, New York.
- Edwards, G., Taylor, C., and Cootes, T. (1998). Learning to identify and track faces in image sequences. In *Proc. International Conference on Computer Vision*, pages 317–322.
- Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman Hall, New York.
- Ekman, P., editor (1982). *Emotion in the Human Face*. Cambridge University Press, New York, NY, 2nd edition.
- Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, 115(2):268–287.
- Ekman, P. and Friesen, W.V. (1978). *Facial Action Coding System: Investigator's Guide*. Consulting Psychologists Press.
- Elkan, C. (1997). Boosting and naive Bayesian learning. Technical Report CS97-557, University of California, San Diego.
- Esposito, F. and Malerba, D. (2001). Machine learning in computer vision. *Applied Artificial Intelligence*, 15(8).
- Essa, I.A. and Pentland, A.P. (1997). Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763.
- Faigin, G. (1990). *The Artist's Complete Guide To Facial Expression*. Watson-Guption Publications, New York, NY.
- Fasel, B. and Luetttin, J. (2003). Automatic facial expression analysis: A survey. *Pattern Recognition*, 36:259–275.
- Feder, M. and Merhav, N. (1994). Relation between entropy and error probability. *IEEE Transactions on Information Theory*, 40:259–266.

- Fernyhough, J., Cohn, A., and Hogg, D. (1998). Building qualitative event models automatically from visual input. In *International Conference on Computer Vision*, pages 350–355.
- Fine, S., Singer, Y., and Tishby, N. (1998). The hierarchical hidden Markov model: Analysis and applications. *Machine Learning*, 32(1):41–62.
- Forbes, J., Huang, T., Kanazawa, K., and Russell, S.J. (1995). The BATmobile: Towards a Bayesian automated taxi. In *Proc. International Joint Conference on Artificial Intelligence*, pages 1878–1885.
- Friedman, J.H. (1997). On bias, variance, 0/1-loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery*, 1(1):55–77.
- Friedman, N. (1998). The Bayesian structural EM algorithm. In *Proc. Conference on Uncertainty in Artificial Intelligence*, pages 129–138.
- Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29(2):131–163.
- Friedman, N. and Koller, D. (2000). Being Bayesian about network structure: A Bayesian approach to structure discovery in Bayesian networks. In *Proc. Conference on Uncertainty in Artificial Intelligence*, pages 201–210.
- Galata, A., Johnson, N., and Hogg, D. (2001). Learning variable length Markov models of behaviour. *International Journal of Computer Vision*, 19:398–413.
- Ganesalingam, S. and McLachlan, G.J. (1978). The efficiency of a linear discriminant function based on unclassified initial samples. *Biometrika*, 65:658–662.
- Garg, A., Pavlovic, V., and Rehg, J. (2003). Boosted learning in dynamic Bayesian networks for multimodal speaker detection. *Proceedings of the IEEE*, 91(9):1355–1369.
- Garg, A., Pavlovic, V., Rehg, J., and Huang, T.S. (2000a). Audio–visual speaker detection using dynamic Bayesian networks. In *Proc. International Conference on Automatic Face and Gesture Recognition*, pages 374–471.
- Garg, A., Pavlovic, V., Rehg, J., and Huang, T.S. (2000b). Integrated audio/visual speaker detection using dynamic Bayesian networks. In *IEEE Conference on Automatic Face and Gesture Recognition*.
- Garg, A. and Roth, D. (2001a). Learning coherent concepts. In *International Workshop on Algorithmic Learning Theory*, pages 135–150.
- Garg, A. and Roth, D. (2001b). Understanding probabilistic classifiers. In *European Conference on Machine Learning*, pages 179–191.
- Ghahramani, Z. and Jordan, M.I. (1996). Factorial hidden Markov models. *Advances in Neural Information Processing Systems*, 8:472–478.
- Ghahramani, Z. and Jordan, M.I. (1997). Factorial hidden Markov models. *Machine Learning*, 29:245–273.
- Ghani, R. (2002). Combining labeled and unlabeled data for multiclass text categorization. In *International Conference on Machine Learning*, pages 187–194.
- Golding, A.R. (1995). A Bayesian hybrid method for context-sensitive spelling correction. In *Workshop on Very Large Corpora*, pages 39–53.
- Golding, A.R. and Roth, D. (1999). A Winnow based approach to context-sensitive spelling correction. *Machine Learning*, 34:107–130.
- Goldman, S. and Zhou, Y. (2000). Enhancing supervised learning with unlabeled data. In *International Conference on Machine Learning*, pages 327–334.
- Goleman, D. (1995). *Emotional Intelligence*. Bantam Books, New York.
- Graf, H., Chen, T., Petajan, E., and Cosatto, E. (1995). Locating faces and facial parts. In *International Workshop Automatic Face and Gesture Recognition*, pages 41–46.
- Greiner, R. and Zhou, W. (2002). Structural extension to logistic regression: discriminative parameter learning of belief net classifiers. In *Proc. Annual National Conference on Artificial Intelligence (AAAI)*, pages 167–173.

- Hajek, B. (1988). Cooling schedules for optimal annealing. *Mathematics of Operational Research*, 13:311–329.
- Haralick, R. and Shapiro, L. (1979). The consistent labeling problem: Part I. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2):173–184.
- Herbrich, R. and Graepel, T. (2001). A PAC-Bayesian margin bound for linear classifiers: Why SVMs work. In *Advances in Neural Information Processing Systems*, pages 224–230.
- Hilgard, E., Atkinson, R.C., and Hilgard, R.L. (1971). *Introduction to Psychology*. Harcourt Brace Jovanovich, New York, NY, 5th edition.
- Hjelmas, E. and Low, B. (2001). Face detection: A survey. *Computer Vision and Image Understanding*, 83(3):236–274.
- Hoey, J. (2001). Hierarchical unsupervised learning of facial expression categories. In *International Conference on Computer Vision, Workshop on Detection and Recognition of Events in Video*, pages 99–106.
- Hongeng, S., Bremond, F., and Nevatia, R. (2000). Representation and optimal recognition of human activities. In *International Conference on Computer Vision*, volume 1, pages 1818–1825.
- Horvitz, E., Breese, J., Heckerman, D., Hovel, D., and Rommelse, K. (1998). The Lumigre project: Bayesian user modeling for inferring the goals and needs of software users. In *Conference on Uncertainty in Artificial Intelligence*, pages 256–265.
- Horvitz, E., Jacobs, A., and Hovel, D. (1999). Attention-sensitive alerting. In *Conference on Uncertainty in Artificial Intelligence*, pages 305–313.
- Hosmer, D.W. (1973). A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. *Biometrics*, 29:761–770.
- Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, pages 221–233.
- Huijsman, D.P. and Sebe, N. (to appear, 2004). How to complete performance graphs in content-based image retrieval: Add generality and normalize scope. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Indyk, P. (2001). Algorithmic applications of low-distortion geometric embeddings. In *Foundations of Computer Science*, pages 10 – 31.
- Intille, S.S. and Bobick, A.F. (1999). A framework for recognizing multi-agent action from visual evidence. In *National Conference on Artificial Intelligence*, pages 518–525.
- Ivanov, Y. and Bobick, A. (2000). Recognition of visual activities and interactions by stochastic parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):852–872.
- Izard, C.E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, 115(2):288–299.
- James, G.M. (2003). Variance and bias for general loss functions. *Machine Learning*, 51:115–135.
- James, W. (1890). *The Principles of Psychology*. Henry Holt, New York, NY.
- Jebara, T. and Pentland, A. (1998). Maximum conditional likelihood via bound maximization and the CEM algorithm. In *Advances in Neural Information Processing Systems*, pages 494–500.
- Jenkins, J.M., Oatley, K., and Stein, N.L., editors (1998). *Human Emotions: A Reader*. Blackwell Publishers, Malden, MA.
- Johnson, B. and Greenberg, S. (1999). Judging people’s availability for interaction from video snapshots. In *Hawaii International Conference on System Sciences*.
- Johnson, W.B. and Lindenstrauss, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. In *Conference in Modern Analysis and Probability*, pages 189–206.

- Jordan, M.I., Ghahramani, Z., and Saul, L.K. (1997). Hidden Markov decision trees. In *Advances in Neural Information Processing Systems*.
- Kanade, T., Cohn, J., and Tian, Y. (2000). Comprehensive database for facial expression analysis. In *International Conference on Automatic Face and Gesture Recognition*, pages 46–53.
- Kay, R.M. (1990). *Entropy and Information Theory*. Springer-Verlag.
- Kearns, M., Mansour, Y., Ng, A.Y., and Ron, D. (1997). An experimental and theoretical comparison of model selection methods. *Machine Learning*, 279:7–50.
- Kearns, M. and Schapire, R. (1994). Efficient distribution-free learning of probabilistic concepts. *Journal of Computer and System Sciences*, 48:464–497.
- Kirby, M. and Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108.
- Kjeldsen, R. and Kender, J. (1996). Finding skin in color images. In *International Conference on Automatic Face and Gesture Recognition*, pages 312–317.
- Kohavi, R. (1996). Scaling up the accuracy of naive Bayes classifiers: A decision-tree hybrid. In *Proc. International Conference on Knowledge Discovery and Data Mining*, pages 194–202.
- Kouzani, A.Z. (2003). Locating human faces within images. *Computer Vision and Image Understanding*, 91(3):247–279.
- Krishnan, T. and Nandy, S. (1990a). Efficiency of discriminant analysis when initial samples are classified stochastically. *Pattern Recognition*, 23(5):529–537.
- Krishnan, T. and Nandy, S. (1990b). Efficiency of logistic-normal supervision. *Pattern Recognition*, 23(11):1275–1279.
- Kullback, S. (1968). Probability densities with given marginals. *Annals of Mathematical Statistics*, 39(4):1236–1243.
- Lam, K. and Yan, H. (1994). Fast algorithm for locating head boundaries. *Journal of Electronic Imaging*, 3(4):351–359.
- Lang, P. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385.
- Lanitis, A., Taylor, C.J., and Cootes, T.F. (1995a). An automatic face identification system using flexible appearance models. *Image and Vision Computing*, 13(5):393–401.
- Lanitis, A., Taylor, C.J., and Cootes, T.F. (1995b). A unified approach to coding and interpreting face images. In *Proc. International Conference on Computer Vision*, pages 368–373.
- Leung, T.K., Burl, M.C., and Perona, P. (1995). Finding faces in cluttered scenes using random labeled graph matching. In *International Conference on Computer Vision*, pages 637–644.
- Lew, M.S. (1996). Informatic theoretic view-based and modular face detection. In *International Conference on Automatic Face and Gesture Recognition*, pages 198–203.
- Li, S., Zou, X., Hu, Y., Zhang, Z., Yan, S., Peng, X., Huang, L., and Zhang, H. (2001). Real-time multi-view face detection, tracking, pose estimation, alignment, and recognition. In *IEEE International Conference on Computer Vision and Pattern Recognition*.
- Lien, J. (1998). *Automatic Recognition of Facial Expressions Using Hidden Markov Models and Estimation of Expression Intensity*. PhD thesis, Carnegie Mellon University.
- Littlestone, N. (1987). Learning when irrelevant attributes abound. In *Annual Symposium on Foundations of Computer Science*, pages 68–77.
- Madabhushi, A. and Aggarwal, J.K. (1999). A Bayesian approach to human activity recognition. In *IEEE Workshop on Visual Surveillance Systems*, pages 25–30.
- Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 63:215–232.
- Martinez, A. (1999). Face image retrieval using HMMs. In *IEEE Workshop on Content-based Access of Images and Video Libraries*, pages 35–39.

- Mase, K. (1991). Recognition of facial expression from optical flow. *IEICE Transactions*, E74(10):3474–3483.
- Matsumoto, D. (1998). Cultural influences on judgments of facial expressions of emotion. In *Proc. ATR Symposium on Face and Object Recognition*, pages 13–15.
- McCallum, A.K. and Nigam, K. (1998). Employing EM in pool-based active learning for text classification. In *International Conference on Machine Learning*, pages 350–358.
- McKenna, S., Gong, S., and Raja, Y. (1998). Modeling facial color and identity with Gaussian mixtures. *Pattern Recognition*, 31(12):1883–1892.
- McLachlan, G.J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. John Wiley and Sons Inc., New York.
- McLachlan, G.J. and Basford, K.E. (1988). *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker Inc., New York.
- Mehrabian, A. (1968). Communication without words. *Psychology Today*, 2(4):53–56.
- Meila, M. (1999). *Learning with Mixture of Trees*. PhD thesis, Massachusetts Institute of Technology.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). Equation of state calculation by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.
- Michalski, R.S. (1983). A theory and methodology of inductive learning. *Machine Learning: an Artificial Intelligence Approach*, R.S. Michalski, J.G. Carbonell, and T.M. Mitchell, eds., pages 83–134.
- Michalski, R.S., Carbonell, J.G., and Mitchell, T.M., editors (1986). *Machine Learning: An Artificial Intelligence Approach*. Morgan Kaufmann, Los Altos, CA.
- Miller, D.J. and Uyar, H.S. (1996). A mixture of experts classifier with learning based on both labelled and unlabelled data. In *Advances in Neural Information Processing Systems*, pages 571–577.
- Mitchell, T. (1999). The role of unlabeled data in supervised learning. In *Proc. International Colloquium on Cognitive Science*.
- MITFaceDB (2000). MIT CBCL face database, MIT center for biological and computation learning. <http://cbcl.mit.edu/cbcl/software-datasets/FaceData2.html>.
- Moghaddam, B. and Pentland, A. (1997). Probabilistic visual learning for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710.
- Morishima, S. (1995). Emotion model: A criterion for recognition, synthesis and compression of face and emotion. In *Proc. Automatic Face and Gesture Recognition*, pages 284–289.
- Murphy, K. and Paskin, M. (2001). Linear time inference in hierarchical HMMs. *Neural Information Processing Systems*.
- Murphy, P.M. (1994). UCI repository of machine learning databases. Technical report, University of California, Irvine.
- Nagy, G., Seth, S., and Stoddard, S. (1992). A prototype document image analysis system for technical journals. *IEEE Computer*, 25(7):10–22.
- Nakamura, Y. and Kanade, T. (1997). Semantic analysis for video contents extraction - Spotting by association in news video. In *Proc. ACM International Multimedia Conference*.
- Naphade, M.R. and Huang, T.S. (2000a). Semantic video indexing using a probabilistic framework. In *International Conference on Pattern Recognition*, volume 3, pages 83–89.
- Naphade, M.R. and Huang, T.S. (2000b). Stochastic modeling of soundtrack for efficient segmentation and indexing of video. In *SPIE IS & T Storage and Retrieval for Multimedia Databases*, volume 3972, pages 168–176.
- Naphade, M.R., Kristjansson, T., Frey, B., and Huang, T.S. (1998). Probabilistic multimedia objects (multijects): A novel approach to indexing and retrieval in multimedia systems. In *International Conference on Image Processing*, volume 3, pages 536–540.

- Naphade, M.R., Wang, R., and Huang, T.S. (2001). Multimodal pattern matching for audio-visual query and retrieval. In *SPIE IS & T Storage and Retrieval for Multimedia Databases*.
- Nefian, A. and Hayes, M. (1999a). An embedded HMM based approach for face detection and recognition. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 3553–3556.
- Nefian, A. and Hayes, M. (1999b). Face recognition using an embedded HMM. In *IEEE Conference on Audio and Video-based Biometric Person Authentication*, pages 19–24.
- Nigam, K., McCallum, A., Thrun, S., and Mitchell, T. (2000). Text classification from labeled and unlabeled documents using EM. *Machine Learning*, 39:103–134.
- Nigam, K.P. (2001). Using unlabeled data to improve text classification. Technical Report CMU-CS-01-126, School of Computer Science, Carnegie Mellon University.
- Oliver, N., Pentland, A., and Berard, F. (2000). LAFTER: A real-time face and lips tracker with facial expression recognition. *Pattern Recognition*, 33:1369–1382.
- O'Neill, T.J. (1978). Normal discrimination with unclassified observations. *Journal of the American Statistical Association*, 73(364):821–826.
- Osuna, E., Freund, R., and Girosi, F. (1997). Training support vector machines: An application to face detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 130–136.
- Otsuka, T. and Ohya, J. (1997a). Recognizing multiple persons' facial expressions using HMM based on automatic extraction of significant frames from image sequences. In *Proc. International Conference on Image Processing*, pages 546–549.
- Otsuka, T. and Ohya, J. (1997b). A study of transformation of facial expressions based on expression recognition from temporal image sequences. Technical report, Institute of Electronic, Information, and Communications Engineers (IEICE).
- Pal, S. and Pal, A. (2002). *Pattern Recognition from Classical to Modern Approaches*. World Scientific.
- Pantic, M. and Rothkrantz, L.J.M. (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445.
- Pantic, M. and Rothkrantz, L.J.M. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390.
- Pavlovic, V. and Garg, A. (2001). Efficient detection of objects and attributes using boosting. In *IEEE Conference on Computer Vision and Pattern Recognition - Technical Sketches*.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California.
- Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.
- Pentland, A. (2000). Looking at people. *Communications of the ACM*, 43(3):35–44.
- Picard, R. W. (1997). *Affective Computing*. MIT Press, Cambridge, MA.
- Rabiner, L. and Huang, B. (1993). *Fundamentals of Speech Recognition*. Prentice-Hall.
- Rabiner, L.R. (1989). A tutorial on hidden Markov models and selected applications in speech processing. *Proceedings of IEEE*, 77(2):257–286.
- Rajagalopan, A., Kumar, K., Karlekar, J., Manivasakan, R., Patil, M., Desai, U., Poonacha, P., and Chaudhuri, S. (1998). Finding faces in photographs. In *International Conference on Computer Vision*, pages 640–645.
- Ramesh, P. and Wilpon, J. (1992). Modeling state durations in hidden Markov models for automatic speech recognition. In *International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 381–384.

- Ratsaby, J. and Venkatesh, S.S. (1995). Learning from a mixture of labeled and unlabeled examples with parametric side information. In *Conference on Computational Learning Theory*, pages 412–417.
- Redner, R.A. and Walker, H.F. (1984). Mixture densities, maximum likelihood, and the EM algorithm. *SIAM Review*, 26(2):195–239.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:368–407.
- Rosenblum, M., Yacoob, Y., and Davis, L.S. (1996). Human expression recognition from motion using a radial basis function network architecture. *IEEE Transactions on Neural Network*, 7(5):1121–1138.
- Roth, D. (1998). Learning to resolve natural language ambiguities: A unified approach. In *National Conference on Artificial Intelligence*, pages 806–813.
- Roth, D. (1999). Learning in natural language. In *International Joint Conference of Artificial Intelligence*, pages 898–904.
- Roth, D. and Zelenko, D. (2000). Towards a theory of coherent concepts. In *National Conference on Artificial Intelligence*, pages 639–644.
- Rowley, H., Baluja, S., and Kanade, T. (1998a). Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38.
- Rowley, H., Baluja, S., and Kanade, T. (1998b). Rotation invariant neural network-based face detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 38–44.
- Salovey, P. and Mayer, J.D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality*, 9(3):185–211.
- Samal, A. and Iyengar, P.A. (1992). Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition*, 25(1):65–77.
- Schapire, R.E., Freund, Y., Bartlett, P., and Lee, W.S. (1997). Boosting the margin: A new explanation for the effectiveness of voting methods. *Machine Learning*, 29:322–330.
- Schapire, R.E and Singer, Y. (1999). Improved boosting algorithms using confidence rated predictions. *Machine Learning*, 37(3):297–336.
- Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review*, 61:81–88.
- Schneiderman, H. and Kanade, T. (1998). Probabilistic modeling of local appearance and spatial relationships for object recognition. In *IEEE Conference Computer Vision and Pattern Recognition*, pages 45–51.
- Schneiderman, H. and Kanade, T. (2000). A statistical method for 3D object detection applied to faces and cars. In *IEEE Conference Computer Vision and Pattern Recognition*, volume 1, pages 746–751.
- Sebe, N., Cohen, I., Garg, A., Lew, M.S., and Huang, T.S. (2002). Emotion recognition using a Cauchy naive Bayes classifier. In *International Conference on Pattern Recognition*, volume 1, pages 17–20.
- Sebe, N. and Lew, M.S. (2003). *Robust Computer Vision – Theory and Applications*. Kluwer Academic Publishers.
- Sebe, N., Lew, M.S., Cohen, I., Sun, Y., Gevers, T., and Huang, T.S. (2004). Authentic facial expression analysis. In *Automatic Face and Gesture Recognition*, pages 517–522.
- Seeger, M. (2001). Learning with labeled and unlabeled data. Technical report, Edinburgh University.
- Segre, A.M. (1992). Applications of machine learning. *IEEE Expert*, 7(3):31–34.
- Shahshahani, B. and Landgrebe, D. (1994a). Effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. *IEEE Transactions on Geoscience and Remote Sensing*, 32(5):1087–1095.

- Shahshahani, B.M. and Landgrebe, D.A. (1994b). Classification of multi-spectral data by joint supervised-unsupervised learning. Technical Report TR-EE 94-1, School of Electrical Engineering, Purdue University.
- Shawe-Taylor, J. (1998). Classification accuracy based on observed margin. *Algorithmica*, 22:157–172.
- Shawe-Taylor, J. and Cristianini, N. (2000). *An Introduction to Support Vector Machines and Other Kernel Based Methods*. Cambridge University Press.
- Shawe-Taylor, J. and Cristianini, N. (1999). Further results on the margin distribution. In *Conference on Computational Learning Theory*, pages 278–285.
- Spirites, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press, Cambridge, 2nd edition.
- Starks, H. and Woods, J. (1994). *Probability, Random Processes, and Estimation Theory for Engineers*. Prentice Hall.
- Starner, T., Weaver, J., and Pentland, A. (1998). Real-time american sign language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1371–1375.
- Sung, K-K. and Poggio, T. (1998). Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):39–51.
- Tao, H. and Huang, T.S. (1998). Connected vibrations: A modal analysis approach to non-rigid motion tracking. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 735–740.
- Tefas, A., Kotropoulos, C., and Pitas, I. (1998). Variants of dynamic link architecture based on mathematical morphology for front face authentication. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 814–819.
- Tishby, N., Pereira, F.C., and Bialek, W. (1999). The information bottleneck method. In *Annual Allerton Conference on Communication, Control, and Computing*, pages 368–377.
- Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86.
- Ueki, N., Morishima, S., Yamada, H., and Harashima, H. (1994). Expression analysis/synthesis system based on emotion space constructed by multilayered neural network. *Systems and Computers in Japan*, 25(13):95–103.
- Valiant, L.G. (1984). A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142.
- van Allen, T. and Greiner, R. (2000). A model selection criteria for learning belief nets: An empirical comparison. In *International Conference on Machine Learning*, pages 1047–1054.
- Vapnik, V.N. (1982). *Estimation of Dependences Based on Empirical Data*. Springer-Verlag, New York.
- Vapnik, V.N. (1998). *Statistical Learning Theory*. John Wiley and Sons, New York.
- Viola, P. and Jones, M. (2004). Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154.
- Wang, R.R., Huang, T.S., and Zhong, J. (2002). Generative and discriminative face modeling for detection. In *International Conference on Automatic Face and Gesture Recognition*.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.
- Wilson, A.D. and Bobick, A.F. (1998). Recognition and interpretation of parametric gesture. In *International Conference on Computer Vision*, pages 329–336.
- Wolpert, D.H. (1992). Stacked generalization. *Neural Networks*, 5:241–259.
- Yacoub, Y. and Davis, L.S. (1996). Recognizing human facial expressions from long image sequences using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(6):636–642.

- Yang, G. and Huang, T.S. (1994). Human face detection in complex background. *Pattern Recognition*, 27(1):53–63.
- Yang, J. and Waibel, A. (1996). A real-time face tracker. In *Proc. Workshop on Applications of Computer Vision*, pages 142–147.
- Yang, M-H., Kriegman, D., and Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine intelligence*, 24(1):34–58.
- Yang, M-H., Roth, D., and Ahuja, N. (2000). SNoW based face detector. In *Neural Information and Processing Systems*, pages 855–861.
- Yow, K.C. and Cipolla, R. (1997). Feature-based human face detection. *Image and Vision Computing*, 15(9):317–322.
- Zacks, J. and Tversky, B. (2001). Event structure in perception and cognition. *Psychological Bulletin*, 127(1):3–21.
- Zhang, T. and Oles, F. (2000). A probability analysis on the value of unlabeled data for classification problems. In *International Conference on Machine Learning*.

Index

- algorithm
 - margin distribution optimization, 119
 - maximum likelihood minimum entropy HMM, 103
 - stochastic structure search, 129
- application
 - context sensitive spelling correction, 127
 - context-sensitive systems, 157
 - face detection, 127, 211
 - facial expression recognition, 117, 187
 - multimodal event detection, 175
 - office activity recognition, 157
 - speaker detection, 115
- audio
 - audio signal energy, 164
 - audio signal mean, 164
 - linear predictive coding coefficients, 164
 - time delay of arrival (TDOA) method, 164
 - variance of the fundamental frequency, 164
 - zero crossing rate, 164
- Bayes optimal error, 21, 24, 30, 67, 105
 - relation to entropy, 21, 105
- Bayes rule, 16, 67
- Bayesian information criterion (BIC), 142
- Bayesian networks, 15, 129, 130
 - active learning, 151
 - Cauchy Naive Bayes, 132
 - Chow-Liu algorithm, 133
 - class variable, 131
 - classification driven stochastic structure search, 143
 - correct structure, 131
 - dependencies of the variables, 131
 - design decisions, 131
 - diagnostic classifier, 131
 - directed acyclic graph, 131
 - dynamic Bayesian networks, 177
 - EM-CBL algorithm, 141
 - EM-TAN algorithm, 139
 - feature distribution, 131
 - features, 131
 - Gaussian Naive Bayes, 132
 - Gaussian-TAN classifier, 135
 - Gaussian-TAN parameters computation, 136
 - generative classifier, 131
 - incorrect structure, 131
 - independence-based methods, 140
 - Cheng-Bell-Liu algorithms (CBL1 and CBL2), 140
 - IC algorithm, 140
 - PC algorithm, 140
 - Kruskal's maximum weighted spanning tree algorithm, 133, 134
 - labels, 131
 - learning the structure, 129, 140
 - maximum likelihood framework, 131
 - Naive Bayes, 15, 16, 19, 40, 132
 - optimal classification rule, 131
 - overfitting, 142
 - parameters, 131
 - score-based methods, 142
 - K2 algorithm, 148
 - Markov chain Monte Carlo (MCMC) algorithm, 142, 148
 - structural EM (SEM) algorithm, 142
 - stochastic structure search (SSS) algorithm, 143, 144
 - structure, 131
 - switching between models, 138
 - TAN learning algorithm, 133, 134
 - Tree-Augmented-Naive-Bayes (TAN), 40, 133
 - Vapnik-Chervonenkis (VC) bound, 145
 - weights for unlabeled data, 150
- Bernstein polynomials, 198
- Bezier curve, 198
- Bezier volume, 198

- classification
 - classification bias, 68, 86
 - classification bias in relation to estimation bias, 68
 - classification error, 125
 - maximum a-posteriori (MAP) classification, 223
- classification performance
 - asymptotic bounds, 28, 29
- clustering
 - Information Bottleneck, 104
- complex probabilistic models
 - small sample effects, 40
- computer vision, 1
 - definition, 1
 - evaluating criteria, 4
 - issues, 1
 - levels of abstraction, 5
 - machine learning contribution, 2
 - machine learning paradigms, 4
 - machine learning usage, 3
 - model learning, 3
 - mutual dependency of visual concepts, 5
 - research issues, 2, 3
 - visual information representation, 4
- density of distributions, 31
- diagnostic probability models, 72
- distributional density, 33
- emotion recognition
 - affective communication, 189
 - adaptive interaction, 190
 - dynamics, 190
 - embodiment, 190
 - affective human-computer interaction, 189
 - Bayesian network classifiers, 189, 197
 - Chen-Huang database, 201
 - Cohn-Kanade database, 201
 - collecting emotion data, 191
 - confused categories, 196
 - confusion matrix, 206
 - Darwin's study, 190
 - dimensions of emotion, 191
 - arousal, 191
 - attention-rejection, 191
 - valence, 191
 - display rules, 192
 - dynamic classification, 195
 - Ekman's studies, 192
 - emotion categories, 191
 - emotion specific HMM, 195
 - Facial Action Coding System (FACS), 193
 - muscle movements (contractions), 193
 - facial expression recognition approaches, 194
 - facial expression recognition studies, 192
 - facial expression recognition system, 197
 - face tracking, 197
 - feature extraction, 197
 - motion units (MU), 198
 - piecewise Bezier volume deformation (PBVD) tracker, 197
 - human-human interaction, 190
 - James-Lange theory of emotion, 191
 - labeled vs. unlabeled data, 201
 - Lang's 2D emotion model, 191
 - multi-level HMM, 195
 - person-dependent tests, 205
 - person-independent tests, 206
 - Schlosberg's 3D emotion model, 191
 - static classification, 195
 - theories of emotion, 190
 - universal facial expressions, 192
 - ways of displaying emotions, 189
- Empirical risk minimization principle, 121
- entropy, 19, 105
 - conditional entropy, 20
 - lower bound, 23
 - upper bound, 24
 - joint entropy, 19
 - relation to Bayes optimal error, 21, 105
 - relative entropy, 20
- estimation
 - conditional vs. joint density estimation, 104
 - consistent estimator, 68
 - correct model, 68, 76, 88
 - estimation bias, 68
 - estimation bias in relation to classification bias, 68
 - incorrect model, 77, 88
 - Maximum Mutual Information Estimation (MMIE), 104, 107
 - unbiased estimator, 68
- expectation-maximization (EM) algorithm, 91, 131
- face detection
 - approaches, 213
 - appearance-based methods, 213
 - feature invariant methods, 213
 - knowledge-based methods, 213
 - template matching methods, 213
 - Bayesian classification, 214
 - Bayesian network classifiers, 217
 - challenges, 212
 - facial expression, 212
 - imaging conditions, 212
 - occlusion, 212
 - pose, 212
 - discriminant function, 215
 - image orientation, 212

- labeled vs. unlabeled data, 218
- maximum likelihood, 214
- MIT CBCL Face database, 218
- principal component analysis, 215
- related problems, 212
 - face authentication, 212
 - face localization, 212
 - face recognition, 212
 - facial expression recognition, 212
 - facial feature detection, 212
 - structural components, 212
- fusion models, 176
 - Coupled-HMM, 176
 - Duration Dependent Input Output Markov Model (DDIOMM), 179, 181
 - dynamic Bayesian networks, 177
 - Factorial-HMM, 176
 - Input Output Markov Model, 179
 - Viterby decoding, 179
- generative probability models, 15, 71, 105
- Hidden Markov Models (HMM), 103, 106, 158, 159, 175
 - Baum-Welch algorithm, 166
 - Cartesian Product (CP) HMM, 167
 - Coupled-HMM (CHMM), 103, 158, 175
 - dynamic graphical models (DGMs), 170
 - embedded HMM, 170
 - Entropic-HMM, 103, 158
 - Factorial-HMM, 103, 175
 - Hidden-Markov Decision Trees (HMDT), 103
 - Hierarchical HMM, 170, 175
 - Input-Output HMM (IOHMM), 103, 179
 - Layered HMM (LHMM), 160
 - architecture, 165
 - classification, 166
 - decomposition per temporal granularity, 162
 - distributional approach, 161
 - feature extraction and selection, 164
 - learning, 166
 - maxbelief approach, 161
 - Maximum Likelihood Minimum Entropy HMM, 103
 - Maximum Mutual Information HMM (MMIHHMM), 107
 - Continuous Maximum Mutual Information HMM, 110
 - convergence, 112
 - convexity, 111
 - Discrete Maximum Mutual Information HMM, 108
 - maximum A-posteriori (MAP) view of, 112
 - unsupervised case, 111
 - Parameterized-HMM (PHMM), 103, 158
 - Stacked Generalization concept, 172
 - Variable-length HMM (VHMM), 103, 158
 - Viterbi decoder, 179
- human-computer intelligent interaction (HCII), 157, 188, 211
 - applications, 188, 189, 211
- inverse error measure, 143
- Jansen's inequality, 20
- Kullback-Leiber distance, 19, 20, 68, 78
- labeled data
 - estimation bias, 88
 - labeled-unlabeled graphs, 92, 96
 - value of, 69
 - variance reduction, 88
- Lagrange formulation, 22
- Lagrange multipliers, 22
- learning
 - active learning, 151
 - boosting, 126, 127
 - perceptron, 121
 - probably approximately correct (PAC), 69
 - projection profile, 46, 119, 120, 125
 - semi-supervised, 7, 66, 75
 - co-training, 100
 - transductive SVM, 100
 - using maximum likelihood estimation, 70
 - supervised, 7, 74, 75
 - support vector machines (SVM), 121
 - unsupervised, 7, 75
 - winnow, 121
- machine learning, 2
 - computer vision contribution, 2
 - potential, 2
 - research issues, 2, 3
- man-machine interaction, 187
- margin distribution, 18, 47, 49, 120
 - margin distribution optimization algorithm, 119, 125
 - comparison with SVM and boosting, 126
 - computational issues, 126
- Markov blanket, 146
- Markov chain Monte Carlo (MCMC), 144
- Markov equivalent class, 131
- Markov inequality, 52
- maximum likelihood classification, 18, 31
 - conditional independence assumption, 19
- maximum likelihood estimation, 107
 - asymptotic properties, 73
 - labeled data, 73

- unlabeled data, 73
- Metropolis-Hastings sampling, 142
- minimum description length (MDL), 142
- mismatched probability distribution, 27
 - classification framework, 30
 - hypothesis testing framework, 28
- modified Stein's lemma, 28, 41
- mutual information, 105

- Neiman-Pearson ratio, 224

- probabilistic classifiers, 15
 - Chebyshev bound, 56
 - Chernoff bound, 57
 - Cramer-Rao lower bound (CRLB), 76
 - empirical error, 47
 - expected error, 47
 - fat-shattering based bound, 45
 - generalization bounds, 45
 - generalization error, 53
 - loss function, 47
 - margin distribution based bound, 49, 120
 - maximum a-posteriori (MAP) rule, 67
 - projection error, 51
 - random projection matrix, 48
 - random projection theorem, 48
 - random projections, 48
 - Schapire's bound, 61
 - Vapnik-Chervonenkis (VC) bound, 45, 50, 145
- probability of error, 27
- product distribution, 18

- Radon-Nikodym density, 72
- receiving operating characteristic (ROC) curves, 218

- Sauer's lemma, 54
- Stein's lemma, 28

- theory
 - generalization bounds, 45
 - probabilistic classifiers, 15
 - semi-supervised learning, 65

- UCI machine learning repository, 127, 146
- unlabeled data
 - bias vs. variance effects, 92, 138
 - detect incorrect modeling assumptions, 99
 - estimation bias, 88
 - labeled-unlabeled graphs, 92, 96
 - performance degradation, 70, 86, 138
 - value of, 65, 69
 - variance reduction, 88

Computational Imaging and Vision

1. B.M. ter Haar Romeny (ed.): *Geometry-Driven Diffusion in Computer Vision*. 1994
ISBN 0-7923-3087-0
2. J. Serra and P. Soille (eds.): *Mathematical Morphology and Its Applications to Image Processing*. 1994
ISBN 0-7923-3093-5
3. Y. Bizais, C. Barillot, and R. Di Paola (eds.): *Information Processing in Medical Imaging*. 1995
ISBN 0-7923-3593-7
4. P. Grangeat and J.-L. Amans (eds.): *Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*. 1996
ISBN 0-7923-4129-5
5. P. Maragos, R.W. Schafer and M.A. Butt (eds.): *Mathematical Morphology and Its Applications to Image and Signal Processing*. 1996
ISBN 0-7923-9733-9
6. G. Xu and Z. Zhang: *Epipolar Geometry in Stereo, Motion and Object Recognition. A Unified Approach*. 1996
ISBN 0-7923-4199-6
7. D. Eberly: *Ridges in Image and Data Analysis*. 1996
ISBN 0-7923-4268-2
8. J. Sporring, M. Nielsen, L. Florack and P. Johansen (eds.): *Gaussian Scale-Space Theory*. 1997
ISBN 0-7923-4561-4
9. M. Shah and R. Jain (eds.): *Motion-Based Recognition*. 1997
ISBN 0-7923-4618-1
10. L. Florack: *Image Structure*. 1997
ISBN 0-7923-4808-7
11. L.J. Latecki: *Discrete Representation of Spatial Objects in Computer Vision*. 1998
ISBN 0-7923-4912-1
12. H.J.A.M. Heijmans and J.B.T.M. Roerdink (eds.): *Mathematical Morphology and its Applications to Image and Signal Processing*. 1998
ISBN 0-7923-5133-9
13. N. Karssemeijer, M. Thijssen, J. Hendriks and L. van Erning (eds.): *Digital Mammography*. 1998
ISBN 0-7923-5274-2
14. R. Highnam and M. Brady: *Mammographic Image Analysis*. 1999
ISBN 0-7923-5620-9
15. I. Amidror: *The Theory of the Moiré Phenomenon*. 2000
ISBN 0-7923-5949-6;
Pb: ISBN 0-7923-5950-x
16. G.L. Gimel'farb: *Image Textures and Gibbs Random Fields*. 1999
ISBN 0-7923-5961
17. R. Klette, H.S. Stiehl, M.A. Viergever and K.L. Vincken (eds.): *Performance Characterization in Computer Vision*. 2000
ISBN 0-7923-6374-4
18. J. Goutsias, L. Vincent and D.S. Bloomberg (eds.): *Mathematical Morphology and Its Applications to Image and Signal Processing*. 2000
ISBN 0-7923-7862-8
19. A.A. Petrosian and F.G. Meyer (eds.): *Wavelets in Signal and Image Analysis. From Theory to Practice*. 2001
ISBN 1-4020-0053-7
20. A. Jaklič, A. Leonardis and F. Solina: *Segmentation and Recovery of Superquadrics*. 2000
ISBN 0-7923-6601-8
21. K. Rohr: *Landmark-Based Image Analysis. Using Geometric and Intensity Models*. 2001
ISBN 0-7923-6751-0
22. R.C. Veltkamp, H. Burkhardt and H.-P. Kriegel (eds.): *State-of-the-Art in Content-Based Image and Video Retrieval*. 2001
ISBN 1-4020-0109-6
23. A.A. Amini and J.L. Prince (eds.): *Measurement of Cardiac Deformations from MRI: Physical and Mathematical Models*. 2001
ISBN 1-4020-0222-X

Computational Imaging and Vision

24. M.I. Schlesinger and V. Hlaváč: *Ten Lectures on Statistical and Structural Pattern Recognition*. 2002 ISBN 1-4020-0642-X
25. F. Mokhtarian and M. Bober: *Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization*. 2003 ISBN 1-4020-1233-0
26. N. Sebe and M.S. Lew: *Robust Computer Vision. Theory and Applications*. 2003 ISBN 1-4020-1293-4
27. B.M.T.H. Romeny: *Front-End Vision and Multi-Scale Image Analysis*. Multi-scale Computer Vision Theory and Applications, written in Mathematica. 2003 ISBN 1-4020-1503-8
28. J.E. Hilliard and L.R. Lawson: *Stereology and Stochastic Geometry*. 2003 ISBN 1-4020-1687-5
29. N. Sebe, I. Cohen, A. Garg and S.T. Huang: *Machine Learning in Computer Vision*. 2005 ISBN 1-4020-3274-9