Appendix A

Experimental details

This appendix gives technical and implementation details about the experiments described above. All code was implemented in Matlab (version 5) and executed on a standard Pentium II PC.

A.1 Feature selection and regression

A.1.1 Radial basis functions

We describe RBF networks used for regression in Section 2.2.2. The implementation used is based on the Netlab software [90], a Matlab toolbox. The algorithm combines the EM algorithm to place Gaussian basis functions, and linear regression to calculate the output weights. The basis functions are spherical, with a variance equal to the square of the largest distance between the centres. Thus all the basis functions are the same size and shape. The EM algorithm was limited to 50 iterations. In most cases, convergence happened well before this time, and even when the EM algorithm had not converged, the continuing changes were extremely small.

Unless otherwise stated, the RBF networks had five hidden nodes. In some feature selection experiments, an RBF network with zero inputs (features) was required. In this case, the results quoted are actually from a linear regression model with no inputs, producing a constant output value. This is equivalent to an RBF network with zero inputs and zero hidden nodes.

A.1.2 Simulated annealing

We described simulated annealing for feature selection in Section 2.4.3. The representation used is a simple binary representation, with one bit for each possible feature, set to '1' if the feature is present, and '0' if absent. The temperature schedule used is a simple geometric decay, such that $T_{n+1} = \alpha T_n$ (Rayward-Smith et al. [103, p.9]). In this work, $\alpha = 0.995$ and $T_0 = 1$.

Each (potential) step is defined by a selecting a bit to flip from a uniform random distribution. This corresponds to adding or removing a single feature. The change in error and the current temperature are then to decide whether to actually take the step or not, as described in Section 2.4.3. When constrained to using a fixed number of features (during cross model validation, for example), two bits with opposite values are randomly selected and both are flipped. This corresponds to replacing one feature with another.

A.1.3 Semi-supervised feature selection

We described semi-supervised feature selection is described in Section 2.5.2. The "ADJ" procedure is defined by Schuurmans and Southey [116], and we summarise it here: Define $d(h_k, h_l)$ as the distance between two hypothesis h_k and h_l estimated using the unlabelled data. Define $d(\widehat{h_k, h_l})$ as the same distance estimated using the *labelled* data. Define $d(h_l, P_{Y|X})$ as the estimated distance between hypothesis h_l and the underlying distribution $P_{Y|X}$, using the labelled data[1].

1. Given a sequence of increasingly complex hypothesis (regression models), h_0, h_1, ...

2. For each hypothesis h_l, find the worst ratio of labelled and unlabelled distance estimates to some predecessor h_k $(k < l)$ and multiply the estimated distance to the target by this
$$d(h_l, \widehat{\widehat{P_{Y|X}}}) = d(h_l, \widehat{P_{Y|X}}) \frac{d(h_k, h_l)}{d(\widehat{h_k, h_l})}$$

3. Choose the hypothesis h_n with the smallest adjusted distance $d(h_l, \widehat{\widehat{P_{Y|X}}})$

A.2 Outliers in n-dimensional Gaussians

It is well known that on average, 95% of the data drawn from a one dimensional Gaussian lies within 1.96 standard deviations from the mean e.g. [28, p.123]. As we increase the number of dimensions, this constant term rises. In principle, the exact function relating the number of dimensions to this width factor could be derived by integrating to find the area under the Gaussian curve. In practice, given that we only need the scores for a few values, we can estimate these empirically, by sampling from an n-dimensional Gaussian and calculating how close to the mean 95% of the data lies.

[1] Note that we cannot estimate the error without knowing the labels.

Table A.1 lists the estimated values for up to 30 dimensions. Each value was estimated from each of 50 samples of 100,000 points, and the mean estimate taken. These values are used in Section 4.6.6. These empirical values are not perfectly accurate: the first value should be 1.960, and is estimated as 1.961, but this is close enough for our application of outlier detection.

Dimension	Width factor	Dimension	Width factor
1	1.961	16	5.131
2	2.447	17	5.246
3	2.793	18	5.372
4	3.081	19	5.481
5	3.326	20	5.608
6	3.550	21	5.713
7	3.750	22	5.817
8	3.936	23	5.933
9	4.118	24	6.039
10	4.276	25	6.134
11	4.430	26	6.237
12	4.584	27	6.328
13	4.738	28	6.425
14	4.862	29	6.519
15	5.001	30	6.614

Table A.1: Estimated Gaussian width factors

A.3 Further SSEL results

In Section 2.11.5, we considered the effect of varying the number of labelled and unlabelled records on the accuracy of semi-supervised ensemble learning, using the Boston Housing data.

Table A.2 shows the results when using the "auto-mpg" data set also available from the UCI repository [16]. This shows a similar pattern to the Boston Housing data results, with SSEL outperforming supervised bagging, only when limited labelled data is available. Given 80 labelled records, SSEL performs *worse* than supervised bagging; SSEL is reliably better only when less than about 50 labelled records are available.

Table A.3 shows the results of the same experiments using the "Abalone" data set, also available from the UCI repository [16]. The goal is to predict the age of abalone (a marine snail) given information about the shell diameter, weight, length etc. The results show that SSEL outperforms supervised bagging, even when a relatively large labelled set is available (up to and including 80 records).

	Labelled set size							
Unlabelled set size	10	20	30	40	50	60	70	80
0	177.075	1.452	0.877	0.781	0.678	0.680	0.653	0.620
20	1.467^{+}	0.715^{+}	0.710^{+}	0.664^{+}	0.652^{+}	0.651	0.619^{+}	0.618
40	2.269^{+}	1.297	0.679^{+}	0.670^{+}	0.651^{+}	0.642^{+}	0.641	0.620
60	2.256^{+}	0.759^{+}	0.713^{+}	0.688^{+}	0.651^{+}	0.651	0.644	0.637^{-}
80	3.155^{+}	0.753^{+}	0.722^{+}	0.700^{+}	0.685	0.668	0.648	0.639^{-}
100	1.126^{+}	0.762^{+}	0.730^{+}	0.712^{+}	0.692	0.673	0.667	0.661^{-}

Table A.2: Comparison of SSEL errors; single RBF model errors; and correlation between SSEL error and size of unlabelled data set, when using Auto-mpg data with varying labelled and unlabelled record set sizes. '+' indicates a significant improvement compared with 0 unlabelled records; '-' indicates a significant worsening compared with 0 unlabelled records; both according to a one-tailed t-test with $p < 0.05$.

	Labelled set size							
Unlabelled set size	10	20	30	40	50	60	70	80
0	19.146	5.294	3.231	3.770	3.102	2.975	3.066	3.099
20	4.773^{+}	2.993^{+}	3.087	3.018	3.017	2.698^{+}	2.682	2.628^{+}
40	5.440	2.930^{+}	3.006	2.752^{+}	2.705	2.733^{+}	2.765	2.695^{+}
60	4.584^{+}	3.088	2.868	2.779^{+}	2.770	2.742^{+}	2.590^{+}	2.706^{+}
80	5.303^{+}	3.088^{+}	2.859	2.863^{+}	2.705	2.676^{+}	2.676	2.687^{+}
100	4.795^{+}	3.448	3.004	2.850^{+}	2.931	2.720^{+}	2.762	2.601^{+}

Table A.3: Comparison of SSEL errors; single RBF model errors; and correlation between SSEL error and size of unlabelled data set, when using Abalone data with varying labelled and unlabelled record set sizes. '+' indicates a significant improvement compared with 0 unlabelled records; '-' indicates a significant worsening compared with 0 unlabelled records; both according to a one-tailed t-test with $p < 0.05$.

Bibliography

[1] AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (Budapest, 1973), B. N. Petrov and F. Cáski, Eds., pub. Akademiai Kaidó, pp. 267–281. Reprinted in *Breakthroughs in Statistics*, Eds. S. Kotz & N.L. Johnson, 1992, volume I, pp. 599–624. New York: Springer.

[2] ALMUALLIM, H., AND DIETTERICH, T. G. Learning with many irrelevant features. In *AAAI-91* (1991), pp. 547–552.

[3] BAILEY, T., AND COWLES, J. A convex hull inclusion test. *IEEE Transactions on Pattern Analysis and Machine Intelligence 9*, 2 (Mar. 1987), 312–316.

[4] BALDWIN, J. F., MARTIN, T. P., AND MCCOY, S. A. Incremental learning in a Fril-based odour classifier. In *Proceedings of EUFIT-98* (Aachen, Germany, 1998).

[5] BALUJA, S. Probabilistic modeling for face orientation discrimination: Learning from labeled and unlabeled data. In *Advances in Neural Information Processing Systems 11* (1999), M. Kearns, S. Solla, and D. Cohn, Eds., MIT Press.

[6] BARNETT, V., AND LEWIS, T. *Outliers in Statistical Data*, third ed. John Wiley and Sons, 1994.

[7] BATES, R. A., AND WYNN, H. P. Adaptive radial basis function emulators for robust design. In *Evolutionary Design and Manufacture - ACDM* (Plymouth, UK, Apr. 2000), I. Parmee, Ed.

[8] BAXTER, R. A. *Minimum Message Length Inference: Theory and Applications.* PhD thesis, Dept. of Computer Science, Monash University, Clayton 3168, Australia, December 1996.

[9] BEBBINGTON, A. C. A method of bivariate trimming for robust estimation of the correlation coefficient. *Applied Statistics* (1978), 221–226.

[10] BELAVENDRAM, N. *Quality by Design.* Prentice Hall, 1995.

[11] BENTLEY, P. J. *Evolutionary design by computers.* Morgan Kaufmann, 1999.

[12] BERGADANO, F., MATWIN, S., MICHALSKI, R., AND ZHANG, J. Measuring quality of concept descriptions. In *Proceedings of the 3rd European Working Sessions on Learning* (Glasgow, Oct. 1988).

[13] BERTHOLD, M., COHEN, P., AND LIU, X. Intelligent data analysis: Reasoning about data. *AI Magazine*, 4 (1998), 131–134.

[14] BIMBENET, J., AND TRYSTRAM, G. Process control in the food industry. *Food and Bioproducts Processing 70 C3* (Sept. 1992), 115–125.

[15] BISHOP, C. M. *Neural Networks for Pattern Recognition.* Clarendon-Press, Oxford, 1995.

[16] BLAKE, C. L., AND MERZ, C. J. UCI repository of machine learning databases. http://www.ics.uci.edu/~mlearn/MLRepository.html, 1998.

[17] BRADLEY, P., AND MANGASARIAN, O. L. Feature selection via concave minimization and support vector machines. In *Proceedings of the Fifteenth International Conference on Machine Learning* (San Francisco, CA, 1998), J. Shavlik, Ed., Morgan Kaufmann, pp. 82–90.

[18] BREESE, J. S., HECKERMAN, D., AND KADIE, C. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence* (1998), pp. 43–52.

[19] BREIMAN, L. Bagging predictors. *Machine Learning 24*, 2 (1996), 123–140.

[20] BREIMAN, L., FRIEDMAN, J. H., OLSEN, R. A., AND STONE, C. J. *Classification and regression trees*, 2nd ed. Chapman and Hall, 1993.

[21] BRENNAN, R., AND LIGHT, R. Measuring agreement when two observers classify people into categories not defined in advance. *British Journal of Mathematical and Statistical Psychology 27* (1974), 154–163.

[22] CANADIAN HIGH COMMISSION (LONDON). Trends: Food in the United Kingdom. Tech. rep., Mar. 2001.

[23] CANU, S. Sensory data validation methods: a review. Esprit EM2S project deliverable 310. Tech. rep., HEUDIASYC, Compiégne University of Technology, 1998.

[24] CARPENTER, G. A., AND GROSSBERG, S. The ART of adaptive pattern recognition by a self-organizing neural network. *Computer 21*, 3 (Mar. 1988), 77–88.

[25] CATELL, R. B. The scree test for the number of factors. *Multivariate Behavioral Research 1* (1966), 245–276.

[26] CHARIKAR, M., GUHA, S., TARDOS, E., AND SHMOYS, D. B. A constant-factor approximation algorithm for the k -median problem (extended abstract). In *ACM Symposium on Theory of Computing* (1999), pp. 1–10.

[27] CLEAVER, G. Personal communication.

[28] COHEN, P. R. *Empirical methods for artificial intelligence.* MIT Press, 1995.

[29] Corney, D. Designing food with Bayesian belief networks. In *Evolutionary Design and Manufacture – ACDM2000* (2000), I. Parmee, Ed., Springer–Verlag, pp. 83–94.

[30] Corney, D. Food bytes: Intelligent systems in the food industry. *British Food Journal* (2002). To appear.

[31] Cox, H., Mowatt, S., and Prevezer, M. From frozen fish fingers to chilled chicken tikka: Organisational responses to technological change in the late twentieth century. In *Association of Business Historians Annual Conference* (Sept. 1999).

[32] Cox, T. F., and Cox, M. A. A. *Multidimensional Scaling.* Chapman and Hall, 1994.

[33] Cristianini, N., and Shawe-Taylor, J. *An introduction to Support Vector Machines: and other kernel-based learning methods.* Cambridge University Press, 2000.

[34] Demiriz, A., Bennett, K. P., and Embrechts, M. J. Semi-supervised clustering using genetic algorithms. In *Proceedings of ANNIE'99 (Artificial Neural Networks in Engineering)* (Nov. 1999).

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

[36] Draper, D. Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society series B 57* (1995), 45–97.

[37] Drucker, H., Burgess, C. J. C., Kaufman, L., Smola, A., and Vapnik, V. Support vector regression machines. In *Neural Information Processing Systems* (1997), M. Mozer, J. Jordan, and T. Petsche, Eds., vol. 9, MIT Press, pp. 155–161.

[38] Duda, R. O., and Hart, P. E. *Pattern Classification and Scene Analysis*, 1st ed. John Wiley and Sons, 1973.

[39] Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification*, 2nd ed. John Wiley and Sons, 2001.

[40] Efron, B., and Tibshirani, R. *An introduction to the bootstrap.* Chapman and Hall, 1993.

[41] Encyclopaedia Britannica. History of agriculture.

[42] Everitt, B. S. *Cluster Analysis*, 3rd ed. Edward Arnold, 1993.

[43] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining*, U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds. AAAI Press/MIT Press, 1996, pp. 1–34.

[44] FAYYAD, U. M., REINA, C., AND BRADLEY, P. S. Initialization of iterative refinement clustering algorithms. In *Knowledge Discovery and Data Mining* (1998), pp. 194–198.

[45] FINANCIAL TIMES SURVEY. *Financial Times* (1999). London.

[46] FLACH, P. A. On the state of the art in machine learning: A personal review. *Artificial Intelligence 13*, 1/2 (Sept. 2001), 199–222.

[47] FOLSOM, B. Why Henry Ford had a better idea. *The Detroit News Sunday, June 2* (1996).

[48] FRALEY, C., AND RAFTERY, A. E. How many clusters? Which clustering method? Answers via model-based cluster analysis. *Computer Journal 41*, 8 (1998), 578–588.

[49] FREUND, Y., AND SCHAPIRE, R. E. Experiments with a new boosting algorithm. In *Proceedings of the 13th International Conference on Machine Learning* (1996), Morgan Kaufmann, pp. 148–146.

[50] FUNES, P. J., AND POLLACK, J. B. Computer evolution of buildable objects. In *Evolutionary design by computers*, P. J. Bentley, Ed. Morgan Kaufmann, 1999.

[51] GARDNER, J., PEARCE, T., FRIEL, S., BARTLETT, P., AND BLAIR, N. A multisensor system for beer-flavour monitoring using an array of conducting polymers and predictive classifiers. *Sensors and Activators series B 18-19* (1994), 240–243.

[52] GNANADESIKAN, R. *Methods for Statistical Analysis of Multivariate Observations*, 2nd ed. Wiley, 1997.

[53] GOONATILAKE, S., AND KHEBBAL, S. *Intelligent Hybrid Systems.* Wiley and Sons, 1995.

[54] GOONATILAKE, S., AND TRELEAVEN, P. *Intelligent systems for finance and business.* Wiley and Sons, 1995.

[55] GOWER, J. C., AND HAND, D. J. *Biplots.* Chapman and Hall, 1996.

[56] GREENE, D. P., AND SMITH, S. F. A genetic system for learning models of consumer choice. In *Proceedings of the 2nd International Conference on Genetic Algorithms and their Applications* (Cambridge, MA, July 1987), J. J. Grefenstette, Ed., Lawrence Erlbaum Associates, pp. 217–223.

[57] GRIERSON, B. Food safety through the ages. *Priorities For Health 9*, 3 (1997).

[58] GROSSMAN, S. I. *Elementary Linear Algebra*, 5th ed. Saunders College Publishing, 1994.

[59] GUNASEKARAN, S., AND DING, K. Using computer vision for food quality evaluation. *Food Technology* (June 1994), 151–154.

[60] HALKIDI, M., BATISTAKIS, Y., AND VAZIRGIANNIS, M. On clustering validation techniques. *Journal of Intelligent Information Systems 17*, 2-3 (2001), 107–145.

[61] HAND, D. J. Intelligent data analysis: Issues and opportunites. *Intelligent Data Analysis 2:2* (1998).

[62] HAND, D. J., DALY, F., LUNN, A. D., MCCONWAY, K. J., AND OSTROWSKI, E. *A Handbook of Small Data Sets.* Chapman and Hall, 1994.

[63] HARTMAN, E., KEELER, J., AND KOWALSKI, J. Layered neural networks with Gaussian hidden units as universal approximations. *Neural Computation 2* (1990), 210–215.

[64] HAWKINS, D. M. *Identification of Outliers.* Chapman and Hall, 1980.

[65] HINTON, G., MCCLELLAND, J., AND RUMELHART, D. Distributed representations. In *Parallel Distributed Processing: explorations in the microstructure of cognition*, J. McClelland and D. Rumelhart, Eds., vol. I Foundations. MIT Press, 1986.

[66] HJORTH, J. S. U. *Computer intensive statistical method: validation model selection and bootstrap.* Chapman and Hall, 1994.

[67] HOETING, J., RAFTERY, A. E., AND MADIGAN, D. A method for simultaneous variable selection and outlier identification in linear regression. *Computational Statistics and Data Analysis* (1996), 251–270.

[68] HUBERT, L. J., AND ARABIE, P. Comparing partitions. *Journal of Classification 2* (1985), 193–218.

[69] JAGGER, M., AND RICHARDS, K. *(I Can't Get No) Satisfaction.* Decca Records, 1965.

[70] JOHN, G., KOHAVI, R., AND PFLEGER, K. Irrelevant features and the subset selection problem. In *Proceedings of the 11th International Conference on Machine Learning ICML94* (1994), pp. 121–129.

[71] KASKI, S. *Data exploration using self-organizing maps.* PhD thesis, Helsinki University of Technology, March 1997.

[72] KEARNS, M., MANSOUR, Y., AND NG, A. Y. An information-theoretic analysis of hard and soft assignment methods for clustering. In *Proceedings of Uncertainty in Artificial Intelligence* (1997), pp. 282–293.

[73] KENNARD, R. W., AND STONE, L. A. Computer aided design of experiments. *Technometrics 11*, 1 (1969), 137–148.

[74] KLEINBERG, J., PAPADIMITRIOU, C., AND RAGHAVAN, P. Segmentation problems: A micro-economic view of data mining. In *Proceedings 30th ACM Symposium on Theory of Computing* (1998).

[75] KOHAVI, R., AND JOHN, G. H. Wrappers for feature subset selection. *Artificial Intelligence 97*, 1–2 (1997), 273–323.

[76] Krogh, A., and Vedelsby, J. Neural network ensembles, cross validation, and active learning. In *Advances in Neural Information Processing Systems* (1995), G. Tesauro, D. Touretzky, and T. Leen, Eds., vol. 7, The MIT Press, pp. 231–238.

[77] Marr, D. Early processing of visual information. *Philosophical Transactions of the Royal Society B275* (1976), 483–524.

[78] McEwan, J. A. Preference mapping for product optimisation. In *Multivariate Analysis of Data in Sensory Science*, T. Naes and E. Risvik, Eds. Elsevier Science, 1996.

[79] McLachlan, G. J., and Basford, K. E. *Mixture Models: Inference and Applications to Clustering.* Marcel Dekker, New York, 1988.

[80] Meila, M., and Heckerman, D. An experimental comparison of several clustering and initialization methods. Tech. Rep. MSR-TR-98-06, Microsoft Research, Feb. 1998.

[81] Miller Jr, I., and Jr, F. R. Variations in human taste bud density and taste intensity perception. *Physiology and Behaviour 47*, 6 (1990), 1213–1219.

[82] Milligan, G. Clustering validation: Results and implications for applied analyses. In *Clustering and Classification* (1996), P. Arabie, L. J. Hubert, and G. De Soete, Eds., World Scientific, pp. 341–375.

[83] Milligan, G. W., and Cooper, M. C. A study of the comparability of external criteria for hierarchical cluster analysis. *Multivariate Behavioural Research 21* (1985), 441–458.

[84] Mirkin, B. *Mathematical Classification and Clustering.* Kluwer Academic Publishers, 1996.

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

[86] Morrison, D. F. *Multivariate Statistical Methods*, 2nd ed. McGraw-Hill, 1976.

[87] Morrot, G., Brochet, F., and Dubourdieu, D. The color of odors. *Brain and Language* (2001). To appear.

[88] Murtagh, F., and Heck, A. *Multivariate Data Analysis.* D Reidel Publishing Co., Holland, 1987.

[89] Murthy, C. A., and Chowdhury, N. In search of optimal clusters using genetic algorithms. *Pattern Recognition Letters 17* (July 1996), 825–832.

[90] Nabney, I., and Bishop, C. Netlab neural network software. http://www.ncrg.aston.ac.uk/netlab/, 2000.

[91] Ng, A. Preventing overfitting of cross-validation data. In *Proceedings of the Fourteenth International Conference on Machine Learning* (1997), Morgan Kaufmann Publishers, pp. 245–253.

[92] Nigam, K., McCallum, A. K., Thrun, S., and Mitchell, T. Text classification from labeled and unlabeled documents using EM. *Machine Learning 39*, 2/3 (2000), 103–134.

[93] Opitz, D. W. Feature selection for ensembles. *American Association for Artificial Intelligence* (1999), 379–384.

[94] Opitz, D. W., and Maclin, R. Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research 11* (1999), 169–198.

[95] O'Reilly, U.-M., and Testa, P. Representation in architectural design. In *Evolutionary Design and Manufacture – ACDM2000* (Plymouth, UK, Apr. 2000), I. Parmee, Ed.

[96] Park, S. H. *Robust Design and Analysis for Quality Engineering.* Chapman and Hall, 1996.

[97] Perrone, M. P., and Cooper, L. N. When networks disagree: Ensemble methods for hybrid neural networks. In *Neural Networks for Speech and Image Processing*, R. J. Mammone, Ed. Chapman-Hall, 1993, pp. 126–142.

[98] Powell, M. J. D. Radial basis functions for multivariate interpolation. In *Algorithms for Approximation*, J. M. Mason and M. Cox, Eds. 1987, pp. 143–167.

[99] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical Recipes in C : The Art of Scientific Computing*, 2nd ed. Cambridge University Press, 1993.

[100] Quinlan, J. R. *C4.5: Programs for machine learning.* Morgan Kaufmann, 1993.

[101] Rand, W. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association 66* (1971), 846–850.

[102] Rätsch, G., Onoda, T., and Müller, K.-R. Soft margins for AdaBoost. *Machine Learning 42*, 3 (Mar. 2001), 287–320.

[103] Rayward-Smith, V., Osman, T., Reeves, C., and Smith, G., Eds. *Modern heuristic search methods.* Wiley, 1996.

[104] Reeves, C. R., and Höhn, C. Integrating local search into genetic algorithms. In *Modern heuristic search methods*, V. Rayward-Smith, T. Osman, C. Reeves, and G. Smith, Eds. Wiley, 1996, pp. 99–115.

[105] Ripley, B. D. *Pattern Recognition and Neural Networks.* Cambridge University Press, Cambridge, 1996.

[106] Rissanen, J. Modelling by shortest data description. *Automatica 14* (1978), 465–471.

[107] Rosandich, R. *Intelligent visual inspection : using artificial neural networks.* Chapman and Hall, 1997.

[108] Rousseeuw, P. J., and Leroy, A. M. *Robust Regression and Outlier Detection.* Wiley, 1987.

[109] Roweis, S. T., and Saul, L. K. Nonlinear dimensionality reduction by locally linear embedding. *Science* (2000), 2323–2326.

[110] Salzberg, S. L. On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery 1*, 3 (1997), 317–328.

[111] Sammon, J. W. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computing C18* (May 1969), 401–409.

[112] Saul, L. K., and Roweis, S. T. An introduction to locally linear embedding. Tech. rep., AT&T Labs - Research, 2001.

[113] Schölkopf, B., Sung, K.-K., Burges, C. J. C., Girosi, F., Niyogi, P., Poggio, T., and Vapnik, V. Comparing support vector machines with Gaussian kernels to radial basis function classiers. *IEEE Transactions on Signal Processing 45*, 11 (1997), 2758–2765.

[114] Schonkopf, S., Naes, T., Baardseth, P., and Risberg-Ellekjaer, M. Computer-aided product development in the food industry. *Food Technology* (Mar. 1996), 69–75.

[115] Schuurmans, D. A new metric-based approach to model selection. In *Proceedings of the 14th National Conference on Artificial Intelligence and 9th Innovative Applications of Artificial Intelligence Conference (AAAI-97/IAAI-97)* (Menlo Park, July 27–31 1997), AAAI Press, pp. 552–558.

[116] Schuurmans, D., and Southey, F. Metric-based methods for adaptive model selection and regularization. *Machine Learning* (2001). Special Issue on New Methods for Model Selection and Model Combination.

[117] Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics 6* (1978), 461–464.

[118] Singh, R., and Ou-Yang, F. Knowledge-based fuzzy control of aseptic processing. *Food Technology* (June 1994), 155–162.

[119] Skurichina, M. *Stabilizing Weak Classifiers.* PhD thesis, Delft University of Technology, 2001.

[120] StatSoft. *Electronic Statistics Textbook.* 2002. http://www.statsoft.com/textbook/stathome.html.

[121] Stone, H., and Sidel, J. L. *Sensory Evaluation Practices*, 2nd ed. Academic Press, 1993.

[122] Susmaga, R. Analyzing discretizations of continuous attributes given a monotonic discrimination function. *Intelligent Data Analysis 1*, 3 (1997), 157–179.

[123] TAGUCHI, G. *System of experimental design.* Kraus International Publications, 1987.

[124] TAX, D., AND DUIN, R. P. W. Data domain description using support vectors. In *Proceedings of the European Symposium on Artificial Neural Networks '99, Brugge* (1999).

[125] TAX, D. M. J., AND DUIN, R. P. W. Outlier detection using classifier instability. *Lecture Notes in Computer Science 1451* (1998), 593–601.

[126] THEODORIDIS, S., AND KOUTROUMBAS, K. *Pattern Recognition.* Academic press, 1999.

[127] THIESING, F., AND VORNBERGER, O. Forecasting sales using neural networks. In *Computational Intelligence:Theory and Applications - Proceedings of Fifth Fuzzy Days International Conference* (Dortmund, Germany, Apr. 1997), B. Reusch, Ed., Springer-Verlag, pp. 321–328.

[128] TITTERINGTON, D. M. Estimation of correlation coefficients by ellipsoidal trimming. *Applied Statistics* (1978), 227–234.

[129] TORGO, L., AND GAMA, J. Regression using classification algorithms. *Intelligent Data Analysis 1*, 4 (1997), 275–292.

[130] VALIANT, L. G. A theory of the learnable. *Communications of the ACM 27*, 11 (1984), 1134–1142.

[131] VAN GENNERT, L., WOLTERS, C., AND MAARSE, H. The relationships between sensory attributes of round and beef tomatoes, and consumer preferences. In *Flavour Science and Technology*, Y. Bessiere and A. Thomas, Eds. J Wiley and Sons Inc, 1990.

[132] VAPNIK, V. N. *The Nature of Statistical Learning Theory.* Springer, New York, 1995.

[133] WALLACE, C., AND BOULTON, D. An information measure for classification. *The Computer Journal 11*, 2 (1968), 185–194.

[134] WEISS, S. M., AND INDURKHYA, N. *Predictive Data Mining: A Practical Guide.* Morgan Kaufmann Publishers, 1997.

[135] WOLPERT, D. H., AND MACREADY, W. G. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation 1*, 1 (April 1997), 67–82.

[136] WU, J. X., CHENG, G., AND LIU, X. Reasoning about outliers by modelling noisy data. In *Advances in Intelligent Data Analysis* (1997), X. Liu, P. Cohen, and M. Berthold, Eds., Springer–Verlag, pp. 549–558.

[137] ZHANG, Q., AND LITCHFIELD, J. Applying fuzzy mathematics to product development and comparison. *Food Technology 45*, 7 (1991), 108–1152.

[138] ZHANG, W. *State-Space Search.* Springer, 1999.

Index

Author Index