UNIVERSITAT POLITECNICA DE CATALUNYA

DEPARTAMENT DE LLENGUATGES I SISTEMES INFORMATICS

PROGRAMA DE DOCTORAT EN INTEL·LIGENCIA ARTIFICIAL

TESI DOCTORAL

Heterogeneous Neural Networks: Theory and Applications

June 2000

Memòria presentada per en Lluís A. Belanche Muñoz per a optar al títol de Doctor en Informàtica.

Directors: Julio José Valdés Ramos i Renato Alquézar Mancho

Bibliography

[Aamodt and Plaza, 94] Aamodt, A., Plaza, E. Case-based reasoning: Foundational issues, methodological variations and system approaches. *AI Communications*, 7(1):39-59, 1994.

[Ackley, 87] Ackley, D. *A connectionist machine for genetic hillclimbing.* Kluwer Academic Press, 1987.

[Andrews and Geva, 96] Andrews, R., Geva, S. In *Rules and Networks*, Procs. of the Rule Extraction From Trained Artificial Neural Networks Workshop, AISB96, pp. 1-15. Brighton, UK, 1996.

[Bäck, 95] Th. Bäck. Evolution Strategies: An alternative Evolutionary Algorithm. Technical Report of the Informatik Centrum, Univ. of Dortmund, Germany, 1995.

[Bäck, 96] Bäck, Th. *Evolutionary Algorithms in Theory and Practice.* Oxford University Press, New York, 1996.

[Bäck and Schwefel, 91] Bäck, Th., Schwefel, H.P. A Survey of Evolution Strategies. In Procs. of the 4th Intl. Conf. on Genetic Algorithms, (eds.) Booker and Belew, Morgan Kaufmann: San Mateo, pp. 2-9, 1991.

[Bäck and Schwefel, 93] Bäck, Th., Schwefel, H.P. An Overview of Evolutionary Algorithms for Parameter Optimization. *Evolutionary Computation*, 1 (1): 1-23, 1993.

[Bäck and Schwefel, 96] Bäck, Th., Schwefel, H.P. Evolutionary Computation: An Overview. In Procs. of the 3rd IEEE Conference on Evolutionary Computation, pp. 20-29, IEEE Press, Piscataway NJ, 1996.

[Bäck, Fogel and Michalewicz, 97] Bäck, Th., Fogel D.B., Michalewicz, Z. (Eds.) *Handbook of Evolutionary Computation.* IOP Publishing & Oxford Univ. Press, 1997.

[Baker, 87] Baker, J. Adaptive selection methods for genetic algorithms. In Procs. of the 2nd Intl. Conf. on Genetic Algorithms. Grefenstette (ed.), Lawrence Erlbaum, 1987.

[Balaguer *et al.*, 98] Balaguer, M.D., Colprim, J., Martín, M., Poch, M. R.-Roda, I. Tractament biològic d'aigües residuals urbanes. Monografia núm 1. Dept. of Industrial and Chemical Engineering, Univ. of Girona, 1998.

[Balakrishnan and Honavar, 95] Balakrishnan, K., Honavar, V. Evolutionary design of neural architectures - a preliminary taxonomy and guide to literature. Technical report CS-TR-95-01. Dept. of Computer Science. Iowa State Univ., 1995.

[Belanche, 91] Belanche, Ll. To be or nought to be. Una questió irrellevant? (*To be or nought to be. An irrelevant question?*) Master's Thesis. Research Report LSI-91-37-R. Universitat Politècnica de Catalunya, 1991.

[Belanche, Valdés and Alquézar, 98a] Belanche, Ll., Valdés, J.J. Alquézar, R. Fuzzy Heterogeneous Neural Networks for Signal Forecasting. In Procs. of the Intl. Conf. on Artificial Neural Networks - ICANN'98. Niklasson, Bodén, Ziemke (Eds.) Lecture Notes in Computer Science (*Perspectives in Neural Computing Series*) LNCS 1240, pp. 1089-94. Springer-Verlag, 1998.

[Belanche *et al.*, 98b] Belanche, Ll., Valdés, J.J., Comas, J., Roda, I., Poch, M. Modeling the Input-Output Behaviour of Wastewater Treatment Plants using Soft Computing Techniques. (*Binding Environmental Sciences and AI*) Papers from the ECAI'98 Workshop. European Conference on Artificial Intelligence, Brighton, UK, 1998.

[Belanche and Valdés, 98c] Belanche, Ll., Valdés, J.J. Using Fuzzy Heterogeneous Neural Networks to Learn a Model of the Central Nervous System Control. In Procs. of EUFIT'98, 6th European Congress on Intelligent Techniques and Soft Computing. Aachen, Germany, pp. 1858-62. Elite Foundation, 1998.

[Belanche and Valdés, 99a] Belanche, Ll., Valdés, J.J. Fuzzy Inputs and Missing Data in Similarity-Based Heterogeneous Neural Networks. In Procs. of the 5th Intl. Work-Conference on Artificial and Natural Neural Networks, IWANN'99. *Engineering Applications of Bio-Inspired Artificial Neural Networks.* Lecture Notes in Computer Science LNCS-1607, pp. 863-873. Springer-Verlag, 1998. Also in Research Report LSI-98-66-R. Universitat Politècnica de Catalunya, 1998.

[Belanche *et al.*, 99b] Belanche, Ll., Valdés, J.J., Comas, J., Roda, I., Poch, M. A Study of Qualitative and Missing Information in Wastewater Treatment Plants. *Environmental Decision Support Systems and Artificial Intelligence.* Papers from the AAAI Workshop. AAAI Technical Report WS-99-07. AAAI Press: Menlo Park, CA.

[Belanche *et al.*, 99c] Belanche, Ll., Valdés, J.J., Comas, J., Roda, I., Poch, M. Towards a Model of Input-Output Behaviour of Wastewater Treatment Plants using Soft Computing Techniques. *Environmental Modelling and Software*, 14: 409-419, 1999.

[Belanche, 99d] Belanche, Ll. A Study in Function Optimization with the Breeder Genetic Algorithm. Research Report LSI-99-36-R. Universitat Politècnica de Catalunya, 1999.

[Belanche, 99e] Belanche, Ll. An application example of the Breeder Genetic Algorithm to function optimization. *Buran*, 14: 56-61. Spanish IEEE Student Branch Magazine, December, 1999.

[Belanche, 00a] Belanche, Ll. A Theory for Heterogeneous Neuron Models based on Similarity. Research Report LSI-00-06-R. Universitat Politècnica de Catalunya, 2000.

[Belanche, 00b] Belanche, Ll. A Case Study in Neural Network Training with the Breeder Genetic Algorithm. Research Report LSI-00-07-R. Universitat Politècnica de Catalunya, 2000.

[Belanche, 00c] Belanche, Ll. Similarity-based Heterogeneous Neuron Models. In Procs. of ECAI'2000: 14th European Conference on Artificial Intelligence, Berlin, August 2000. W. Horn (Ed.), IOS Press, Amsterdam, 2000.

[Belanche, 00d] Belanche, Ll. Heterogeneous Neuron Models based on Similarity. In Procs. of the 17th National Conference on Artificial Intelligence, AAAI'2000 (*Student abstract*) AAAI Press: Menlo Park, CA, 2000.

[Belanche *et al.*, 00] Belanche, Ll., Valdés, J.J., Comas, J., Roda, I., Poch, M. A Soft Computing Techniques Study in Wastewater Treatment Plants. Accepted for publication in *Artificial Intelligence in Engineering* (currently in press). Also in Research Report LSI-99-20-R. Universitat Politècnica de Catalunya, 1999.

[Bersini and Bontempi, 97] Bersini H., Bontempi, G. Now comes the time to defuzzify neuro-fuzzy models. *Fuzzy sets and systems*, Vol. 90, No. 2, 1997.

[Bialek *et al.*, 91] Bialek, W., Rieke, F., de Ruyter Van Steveninck, R.R., Warland, D. Reading a neural code. *Science*, 252, pp. 1854-1857, 1991.

[Birx and Pipenberg, 92] Birx, D., Pipenberg, S. 1992. *Chaotic oscillators and complex-mapping feedforward networks for signal detection in noisy environment.* Int. Joint Conf. on Neural Networks, Baltimore, MD.

[Birx and Pipenberg, 93] Birx, D., Pipenberg, S. 1993. *A complex-mapping network for phase sensitive classification. IEEE Trans. on Neural Networks*, 4(1): 127-135, 1993.

[Bishop, 95] Bishop, C. *Neural Networks for Pattern Recognition.* Clarendon Press, 1995.

[Buntine and Weigend, 91] Buntine, W., Weigend, A. Bayesian backpropagation. *Complex Systems*, 5(6): 603-643, 1991.

[de Burgos, 95] de Burgos, J. *Cálculo infinitesimal de varias variables.* McGraw-Hill, 1995.

[Capodaglio *et al.*, 91] Capodaglio, A.G., Jones, H.V., Novotny, V., Feng, X. Sludge bulking analysis and forecasting : Application of system identification and artificial neural computing technologies. *Water Research* 25 (10): 1217-1224, 1991.

[Chandon and Pinson, 81] J.L. Chandon, S. Pinson. *Analyse Typologique. Théorie et Applications.* Masson, 1981.

[Chen, Chen and Liu, 95] Chen, T., Chen, H., Liu, R. Approximation Capability in $C(\mathbb{R}^n)$ by Multilayer Feedforward Networks and Related Problems. *IEEE Trans. on Neural Networks*, 6(1): 25-30, 1995.

[Cherkassky and Mulier, 98] Cherkassky, V., Mulier, F. *Learning from data: concepts, theory and methods.* John Wiley, 1998.

[Comas *et al.*, 98] Comas J., R.-Roda I., Poch M., Sánchez M., Gimeno J.M., Cortés, U. An integrated intelligent system to improve wastewater treatment plant operation - part II: a real application. Waste-Decision'98, International Workshop on Decision and Control on Wastes Bio-Processing, Narbonne, France, 1998.

[Cueva, Alquézar and Nebot, 97] Cueva J., Alquézar R., Nebot A. *Experimental comparison of fuzzy and neural network techniques in learning models of the central nervous system control.* In Procs. of EUFIT'97, 5th European Congress on Intelligent Techniques and Soft Computing, pages 1014-1018, Aachen, Germany, 1997.

[Cybenko, 89] Cybenko, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2: 303-314, 1989.

[Davis, 91] Davis, L.D. *Handbook of Genetic Algorithms.* Van Nostrand Reinhold, 1991.

[De Falco *et al.*, 96] De Falco, I., Del Balio, R., Della Cioppa, A., Tarantino, E. A Comparative Analysis of Evolutionary Algorithms for Function Optimisation. In Procs. of the Second Workshop on Evolutionary Computing (WEC2), Nagoya, JAPAN, 1996.

[De Falco *et al.*, 97] De Falco, I., A. Della Cioppa, Natale, P., Tarantino, E. Artificial Neural Networks Optimization by means of Evolutionary Algorithms. In *Soft Computing in Eng. Design and Manufacturing.* Springer Verlag, London, 1997.

[De Falco *et al.*, 98] De Falco, I., Iazzetta, A, Natale, P., Tarantino, E. Evolutionary Neural Networks for Nonlinear Dynamics Modeling. In Procs. of Parallel Problem Solving From Nature V (PPSN V), Amsterdam, The Netherlands, 1998.

[Diday, 74] Diday, E. Recent progress in distance and similarity measures in pattern recognition. In Procs. of the 2nd Intl. Conf. on Pattern Recognition, pp. 534-39, 1974.

[Dixon, 79] Dixon, J.K. Pattern recognition with partly missing data. *IEEE Trans. on Systems, Man and Cybernetics*, 9: 617-621, 1979.

[Dorffner, 95] Dorffner, G. A generalized view on learning in feedforward neural networks. In Cromme *et al.* (eds.), CoWAN'94, Technische Universität Cottbus, Reihe Mathematik M-01/1995, pp.34-54, 1995.

[Dubois *et al.*, 96] D. Dubois, F. Esteva, P. García, Ll. Godo and H. Prade. A logical approach to interpolation based on similarity relations. IIIA Research Report 96-07. Institut d'Investigació en Intel.ligència Artificial. Barcelona, 1996.

[Dubois *et al.*, 97] D. Dubois, H. Prade, F. Esteva, P. García, Ll. Godo and R. López de Mántaras. Fuzzy set modeling in case-based reasoning. Lecture Notes in Artificial Intelligence 1266, pp. 599-610, Springer Verlag, 1997.

[Dubois and Prade, 97] Dubois, D., Prade, H. The three semantics of fuzzy sets. *Fuzzy sets and systems*, Vol. 90, No. 2, 1997.

[Duch and Jankowski, 95] W. Duch and N. Jankowski. Bi-radial transfer functions. Technical Report UMK-KMK-TR 1/96, Dept. of Computer Methods, Nicholas Copernicus Univ., Toruń, Poland, 1995.

[Duda and Hart, 73] Duda, R.O., Hart, P.E. *Pattern classification and scene analysis.* John Wiley, 1973.

[Dudgeon and Mersereau, 84] D.E. Dudgeon and R.M. Mersereau. *Multidimensional Signal Processing.* Prentice Hall, New Jersey, 1984.

[Durbin and Rumelhart, 89] R. Durbin, D.E. Rumelhart. Product units: A computationally powerful and biologically plausible extension to backpropagation networks. *Neural Computation*, 1: 133-142, 1989.

[Edelman, 93] Edelman, S. Representation, Similarity and the Chorus of Prototypes. Dept. of Applied Mathematics and Computer Science Technical Report. The Weizmann Institute of Science (1993). Also in *Minds and Machines*, 5:45-68, 1995.

[Edelman, Reisfeld and Yeshurun, 92] Edelman, S., Reisfeld, D. Yeshurun, Y. Learning to recognize faces from examples. In Procs. of the 2nd European Conf. on Computer Vision. Lecture Notes in Computer Science 558, pp. 787-791. Springer-verlag, 1992.

[Elman, 90] Elman, J.L., Finding structure in time. *Cognitive Science*, 14: 179-211, 1990.

[Esteva, Godo and García, 98] Esteva, F., Godo, L., García, P. Similarity-based Reasoning. IIIA Research Report 98-31, Instituto de Investigación en Inteligencia Artificial. Barcelona, Spain, 1998.

[Everitt, 77] Everitt, B. *Cluster Analysis.* Heinemann Books, 1977.

[Fagundo, Valdés and Pulina, 90] Fagundo, J.R, Valdés J.J, Pulina, M.: Hydrochemical investigations in extreme climatic areas, Cuba and Spitzbergen. In *Water Resources Management and Protection in Tropical Climates*, pp. 45-54, Havana-Stockholm, 1990.

[Fagundo, Valdés and Rodríguez, 96] Fagundo, J.R, Valdés J.J, Rodríguez, J. E.: Karst Hydrochemistry (in Spanish). *Research Group of Water Resources and Environmental Geology.* University of Granada. Ediciones Osuna, p. 212, Granada, Spain, 1996.

[Feuring and Lippe, 95] Feuring, Th., Lippe, W.M. Fuzzy neural networks are universal approximators. In Procs. of the IFSA World Congress, pp. 659-662, Sao Paulo, 1995.

[Fiesler and Beale, 97] Fiesler, E., Beale, R. (Eds.) *Handbook of Neural Computation.* IOP Publishing & Oxford Univ. Press, 1997.

[Finnof, Hergert and Zimmermann, 93] Finnof W., Hergert F., Zimmermann, H.J. Improving model selection by nonconverging methods. *Neural Networks*, 6: 771-783, 1993.

[Fletcher, 80] Fletcher, R. *Practical methods of optimization.* Wiley, 1980.

[Flexer, 95] Flexer, A. Statistical Evaluation of Neural Network Experiments: Minimum Requirements and Current Practice. Technical Report OEFAI-TR-95-16. The Austrian Reserach Institute for Artificial Intelligence, Vienna, Austria, 1995.

[Fogel, 62] Fogel, L.J. Toward inductive inference automata. In Procs. of the Intl. Federation for Information Processing Congress, pp. 395-399, Munich, 1962.

[Fogel, Owens and Walsh, 66] Fogel, L.J., Owens, A.J., Walsh, M.J. Artificial Intelligence through Simulated Evolution. Wiley, NY, 1966.

[Fogel, 92] Fogel, L.J. An analysis of evolutionary programming. In Fogel and Atmar (Eds.) Procs. of the 1st annual conf. on evolutionary programming. La Jolla, CA: Evolutionary Programming Society, 1992.

[Friedman, 94] Friedman J.H., An overview of predictive learning and function approximation. In *From Statistics to Neural Networks: Theory and Pattern Recognition Applications.* Cherkassky, Friedman and Wechsler (eds.), NATO ASI Series F, 136. Springer-Verlag, 1994.

[Fukunaga, 90] Fukunaga, K. *Introduction to Statistical Pattern Recognition* (Second Ed.). San Diego, Academic Press, 1990.

[Fukushima, 80] Fukushima, K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36, pp. 193-202, 1980.

[Funahashi, 89] Funahashi, K. On approximate realization of continuous mappings by neural networks. *Neural Networks*, 2: 183-192, 1989.

[Geman, Bienenstock and Doursat, 92] Geman, S., Bienenstock, E., Doursat, R. Neural Networks and the bias/variance dilemma. *Neural Computation*, 4 (1): 1-58, 1992.

[Ghahramani and Jordan, 94] Ghahramani, Z., and Jordan, M.I. Learning from incomplete data. AI Memo No. 1509, AI Laboratory, MIT, 1994.

[Girosi and Poggio, 89] Girosi, F., Poggio, T. Networks and the best approximation property. AI Memo No. 1164. AI Laboratory, MIT, 1989.

[Girosi, Jones and Poggio, 93] Girosi, F., Jones, M., Poggio, T. Priors, Stabilizers and Basis Functions: from regularization to radial, tensor and additive splines. AI Memo No. 1430, AI Laboratory, MIT, 1993.

[Globig and Lange, 96] Globig, C., Lange, S., Case-based representability of classes of boolean functions. In ECAI-96: Procs. of Twelfth European Conf. on Artificial Intelligence, pp. 117-121, 1996.

[Goldberg, 89] Goldberg, D.E. *Genetic Algorithms for Search, Optimization & Machine Learning.* Addison-Wesley, 1989.

[Goodhill, Finch and Sejnowski, 96] Goodhill, G.J., Finch, S., Sejnowski, T.J., Optimizing Cortical Mappings. In *Advances in Neural Information Processing Systems*, 8. Touretzky, D.S., Mozer, M.C., Hasselmo, M.E. (Eds.). MIT Press: Cambridge, MA, 1996.

[Gower, 71] Gower, J.C. A General Coefficient of Similarity and some of its Properties. *Biometrics*, 27: 857-871, 1971.

[Gupta and Rao, 94] Gupta, M.M., Rao, D.H. On the principles of fuzzy neural networks. *Fuzzy sets and systems*, Vol. 61, pp. 1-18, 1994.

[Hahn and Chater, 98] Hahn, U., Chater, N., Understanding Similarity: A Joint Project for Psychology, Case-Based Reasoning, and Law. *Artificial Intelligence Review*, 12: 393-427, 1998.

[Hartmann, Keeler and Kowalski, 90] Hartmann, E.J., Keeler, J.D., Kowalski, J.M. Layered neural networks with gaussian hidden units as universal approximators. *Neural Computation*, 2: 210-215, 1990.

[Hayashi, Buckley and Czogola, 92] Hayashi, Y., Buckley, J., Czogola, E. Direct fuzzyfication of neural networks and fuzzyfied delta rule. In Procs. of the 2nd Intl. Conf. on Fuzzy Logic and Neural Networks, Iizika, Japan, 1992.

[Haykin, 94] Haykin, S. *Neural Networks: A Comprehensive Foundation.* MacMillan, 1994.

[Hebb, 49] Hebb, D. *The organization of behaviour*, John Wiley, 1949.

[Hecht-Nielsen, 87] Hecht-Nielsen, R. Counter-propagation networks. In Procs. of the IEEE First Intl. Conf. on Neural Networks, vol. II, pp. 19-32, 1987.

[Hecht-Nielsen, 89] Hecht-Nielsen, R. Kolmogorov's mapping neural network existence theorem. In Procs. of the IEEE Intl. Joint Conf. on Neural Networks, Vol. III: 11-14, 1989.

[Hecht-Nielsen, 90] Hecht-Nielsen, R. *Neurocomputing.* Addison-Wesley, 1990.

[Hertz, Krogh and Palmer, 91] Hertz, J., Krogh, A., Palmer R.G. *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, 1991.

[Hinton, Ackley and Sejnowski, 84] Hinton, G.E., Ackley, D., Sejnowski, T. Boltzmann machines: constraint satisfaction networks that learn. Technical report CMU-CS-84-119. School of Computer Science. Carnegie-Mellon Univ., 1984.

[Hinton, 89] Hinton, G.E. Connectionist learning procedures. *Artificial Intelligence*, 40: 185-234, 1989.

[Holland, 62] Holland, J.H. Outline for a logical theory of adaptive systems. J. of the ACM, 3: 297-314, 1962.

[Holland, 75] Holland, J.H. Adaptation in natural and artificial systems. The University of Michigan Press. Ann Arbor, MI, 1975.

[Honavar, 92] Honavar, V. Inductive Learning Using Generalized Distance Measures. In Procs. of the 1992 SPIE Conference on Adaptive and Learning Systems, 1992.

[Hopfield, 82] Hopfield, J.J. Neural Networks and Physical Systems with Emergent Collective and Computational Abilities. In Procs. of the Natl. Academy of Sciences, USA, Vol. 79, pp. 2554-2558, 1982.

[Hornik, Stinchcombe and White, 89] Hornik, K., Stinchcombe, M., White, H. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, 2, pp. 359-366, 1989.

[Hornik, 90] Hornik, K., Approximation capabilities of Multilayer Feedforward Neural Networks. *Neural Networks*, 4, pp. 251-257, 1990.

[Hornik, 93] Hornik, K., Some results on neural network approximation. *Neural Networks*, 6, pp. 1069-1072, 1993.

[Hume, 47] D. Hume. An enquiry concerning human understanding. In *The World's Great Thinkers. Man and Spirit: the Speculative Philosophers.* pp. 341-423. Random House, New York, 1947.

[Jain and Dubes, 88] Jain, A.K., Dubes, R.C. *Algorithms for clustering data.* Prentice-Hall, 1988.

[Jang and Sun, 95] Roger Jang, J.S, Sun, C.T. Neuro-Fuzzy Modeling and Control. Procs. of the IEEE, March 1995.

[Jarauta, 93] Jarauta, E. Anàlisi matemàtica d'una variable. Edicions UPC, Barcelona, 1993.

[Jordan, 86] Jordan, M.I. Attractor dynamics and parallelism in a connectionist sequential machine. In Procs. of the 1986 Cognitive Science Conf., pp. 531-546. Lawrence Erlbaum, 1986.

[Klir, 88] Klir, G.J., Folger, T.A. *Fuzzy Sets, Uncertainty and Information.* Prentice Hall, 1988.

[Klir and Yuan, 95] Klir, G.J., Folger, T.A. *Fuzzy Sets and Fuzzy Logic: theory and applications.* Prentice Hall, 1995.

[Kohonen, 88] Kohonen, T. *Self-Organization and Associative Memory.* Springer-Verlag, Berlin, 1988.

[Kolmogorov, 57] A.N. Kolmogorov. On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition. *Doklady Akademii Nauk SSSR*, Vol. 114: 369-373, 1957.

[Kolmogorov and Fomin, 75] A.N. Kolmogorov, S.V. Fomin. Elementos de la teoría de funciones y del análisis funcional. Ed. Mir, Moscú, 1975.

[Kosko, 92] Kosko, B. *Neural networks and fuzzy systems.* Prentice Hall, 1992.

[Kuşçu and Thornton, 94] Kuşçu, I., Thornton, C. Design of Artificial Neural Networks using genetic algorithms: review and prospect. Technical Report of the Cognitive and Computing Sciences Dept. Univ. of Sussex, England, 1994.

[van Laarhoven and Aarts, 87] van Laarhoven, P.J.M., Aarts, E.H.L. *Simulated annealing: theory and applications.* Reidel, Dordrecht, The Nederlands, 1987. versus theory.

[Lapedes and Farber, 87] A. Lapedes and R. Farber, *Nonlinear signal processing using neural networks: prediction and system modeling*, Tech. Rep. LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos NM, 1987.

[Lean and Hinrichsen, 94] Lean, G., Hinrichsen, D. *Atlas of the Environment.* 2nd edition, Harper Perennial Publ., 1994.

[Leaning *et al.*, 83] Leaning, M.S., Pullen, H.E., Carson, E.R., Finkelstein, L. Modelling a complex biological system: the human cardiovascular system. *Trans. on Instr. Meas. and Control*, 5: 71-86, 1983.

[Lee, 97] Lee M.D. The connectionist construction of psychological spaces. *Connection Science*, 9(4): 323-351, 1997.

[Leshno *et al.*, 93] Leshno, M., Lin, V., Pinkus, A., Shocken, S. Multilayer feedforward networks with a non-polynomial activation function can approximate any function. *Neural Networks*, 6: 861-867, 1993.

[Li and Chow, 96] Li, J.Y., Chow, T.W.S. Functional approximation of higher-order neural networks. *Journal of Intelligent Systems*, 6(3-4): 239-260, 1996.

[Lippe, Feuring and Mischke, 95] Lippe, W.M., Feuring, Th., Mischke, L. Supervised learning in fuzzy neural networks. Technical Report 12/95-I. Institut für Numerische und instrumentelle Mathematik/Informatik. Universität Münster, Germany, 1995.

[Little and Rubin, 87] Little, R.J.A., Rubin, D.B. *Statistical analysis with missing data.* John Wiley, 1987.

[Lowe, 93] Lowe, D.G. Similarity metric learning for a variable-kernel classifier. *Neural Computation*, 7, 1993.

[McCulloch and Pitts, 43] McCulloch, W., Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5: 115-133, 1943.

[McQueen, 67] McQueen, J. Some methods of classification and analysis of multivariate observations. In Procs. of the 5th Berkeley Symposium on Mathematics, Statistics and Probability, p. 281. LeCam and Neyman (eds.), Univ. of California Press, 1967.

[Michalewicz, 92] Michalewicz, Z. Genetic Algorithms + data structures = evolution programs. Artificial Intelligence. Springer, Berlin, 1992.

[Michie, Spiegelhalter and Taylor, 94] Michie, D., Spiegelhalter, D.J., Taylor, C.C. (eds.) *Machine Learning: Neural and Statistical Classification.* Ellis Horwood, 1994.

[Møller, 93] Møller, M. A scaled conjugate gradient algorithm for supervised learning. *Neural Networks*, 6(4): 525-533, 1993.

[Montana and Davis, 89] Montana, D.J., Davis, L. Training FeedForward Neural Networks using Genetic Algorithms. In Procs. of the 11th Intl. Joint Conf. on Artificial Intelligence. Morgan Kaufmann, 1989.

[Moody, 94] J. Moody. Prediction Risk and Architecture Selection for Neural Networks. In *From Statistics to Neural Networks: Theory and Pattern Recognition Applications.* Cherkassky, Friedman and Wechsler (eds.), NATO ASI Series F, Springer-Verlag, 1994.

[Mosteller and Tukey, 78] Mosteller F., Tukey, F.W. *Data analysis and regression - A second course in statistics.* Addison-Wesley, Reading, MA, 1977.

[Mühlenbein and Schlierkamp-Voosen, 93] Mühlenbein, H., Schlierkamp-Voosen, D. Predictive Models for the Breeder Genetic Algorithm. *Evolutionary Computation*, 1 (1): 25-49, 1993.

[Murphy and Aha, 91] Murphy, P.M., Aha, D. UCI Repository of machine learning databases. UCI Dept. of Information and Computer Science, 1991.

[Nauck, Klawonn and Kruse, 92] D. Nauck, F. Klawonn, R. Kruse. Fuzzy Sets, Fuzzy Controllers and Neural Networks. In Wissenschaftliche Zeitschrift der Humboldt-Universität zu Berlin, Reihe Medizin 41, Nr. 4: 99-120, 1992.

[Nauck, Klawonn and Kruse, 97] Nauck D., Klawonn, F., Kruse, R. *Foundations of Neuro-Fuzzy Systems.* John Wiley, 1997.

[Nebot *et al.*, 98] Nebot, A., Valdés, J.J., Guiot, M., Alquézar R., Vallverdú, M. *Fuzzy inductive reasoning approaches in the identification of models of the central nervous system control.* Procs of. Int. ICSC Symposium on Engineering of Intelligent Systems, Tenerife, 1998.

[Nieto, 00] Nieto, D. Extensión del modelo de redes neuronales (*An extension of neural network models*). Final Degree Project. Advisor: J.J Valdés. Universitat Politècnica de Catalunya, February, 2000.

[Nowlan, 90] Nowlan, S. Max-likelihood competition in RBF networks. Technical Report CRG-TR-90-2. Connectionist Research Group, Univ. of Toronto, 1990.

[Novotny *et al.*, 90] Novotny, V., Jones, H., Feng, X., Capodaglio, A.G. Time series analysis models of activated sludge plants. *Water Science & Technology* 23: 1107-1116, 1990.

[Osborne and Bridge, 96] Osborne , H.R., Bridge, D.G., A Case Base Similarity Framework. In Procs. of EWCBR-96: the third European Workshop on Case-Based Reasoning (Advances in Case-Based Reasoning), Smith, I., Faltings, B. (Eds.), LNCS 1168, Springer Verlag, 1996.

[Orr, 96] Orr, M.J.L. Introduction to Radial Basis Function Networks. Technical Report of the Centre for Cognitive Science, Univ. of Edinburgh, 1996.

[O'Sullivan and Keane, 92] O'Sullivan, J., Keane, M.T. Examining Similarity: Using neural network Techniques to Model Similarity Phenomena. Technical Report TRD-92-36. Computer Science Dept. The Univ. of Dublin, 1992.

[Pappis and Karacapilidis, 93] C. Pappis, N. Karacapilidis. A comparative assessment of measures of similarity of fuzzy values. *Fuzzy sets and systems*, Vol. 56: 171-4, 1993.

[Park and Sandberg, 91] Park, J., Sandberg, I.W. Universal approximation using radial-basis functions. *Neural Computation*, 3(2): 246-257, 1991.

[Park and Sandberg, 93] Park, J., Sandberg, I.W. Approximation and radial-basis function networks. *Neural Computation*, 5: 305-316, 1993.

[Parker, 82] Parker, D. Learning logic. Invention Report S81-64, File 1. Office of Technology Lic., Stanford Univ., 1982.

[Pawlak, 91] Pawlak, Z. *Rough Sets: Theoretical aspects of reasoning about data.* Kluwer Academic Publ, 1991.

[Pearlmutter, 90] Pearlmutter, B. Dynamic recurrent neural networks. Technical report CMU-CS-90-196. School of Computer Science. Carnegie-Mellon Univ., 1990.

[Pedrycz, 93] . Pedrycz, W. *Fuzzy Control and Fuzzy Systems.* Research Studies Press, Taunton, 1993.

[Pedrycz, 97] . Pedrycz, W. Fuzzy sets in pattern recognition: accomplishments and challenges. *Fuzzy sets and systems*, Vol. 90, No. 2, 1997.

[Poggio and Girosi, 89] Poggio T., Girosi, F. A Theory of Networks for Approximation and Learning. AI Memo No. 1140, AI Laboratory, MIT, 1989.

[Poggio, 90] Poggio, T. A theory of how the brain might work. In Cold Spring Harbor *Symposia on Quantitative Biology*, pp. 899-910. Cold Spring Harbor Lab. Press, 1990.

[Poggio and Hurlbert, 93] Poggio, T. Hurlbert, A. Observations on Cortical Mechanisms for Object Recognition and Learning. AI Memo No. 1404, AI Laboratory, MIT, 1993.

[Prechelt, 94] Prechelt, L. Proben1: A set of Neural Network Benchmark Problems and Benchmarking Rules. Facultät für Informatik. Univ. Karlsruhe. Technical Report 21/94, 1994.

[Prechelt, 98] Prechelt, L. Early Stopping - but when? in *Neural Networks: Tricks of the trade*, pp. 55-69, Lecture Notes in Computer Science 1524, Springer Verlag, 1998.

[Press *et al.*, 92] Press, S.A., Teukolsky, Vetterling, W.T., Flannery B.P. *Numerical recipes in C: The art of scientific computing* (2nd ed.). Cambridge Univ. Press, 1992.

[Queysanne, 85] M. Queysanne. *Álgebra básica.* Vicens Vives, Barcelona, 1985.

[Quinlan, 86] Quinlan, J.R. Induction of decision trees. *Machine Learning*, 1: 81-106, 1986.

[Radcliffe, 91] Radcliffe, N.J. Genetic set recombination and its application to neural network topology optimization. Technical Report EPCC-TR-91-21. Edinburgh Parallel Computing Centre. Univ. of Edinburgh, Scotland, 1991.

[Rechenberg, 65] Rechenberg, I. Cybernetic solution path of an experimental problem. Royal Aircraft Establishment, Library Translation No. 1122, Farnborough, UK, August 1965.

[Rechenberg, 73] Rechenberg, I. Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Frommann-Holzboog, Stuttgart, 1973.

[Reyneri, 99] L.M. Reyneri. Unification of Neural and Wavelet Networks and Fuzzy Systems. *IEEE Trans. on Neural Networks*, Volume 10, Number 4, 1999.

[Ripley, 92] Ripley B.D. Statistical aspects of neural networks. Technical Report of the Dept. of Statistics, Univ. of Oxford, 1992.

[Rosenblatt, 62] Rosenblatt, F. *Principles of neurodynamics.* Spartan Books, NY, 1962.

[Roubens, 90] M. Roubens. Fuzzy sets and decision analysis. *Fuzzy sets and systems*, Vol. 90, No. 2, 1997.

[Rudin, 66] Rudin, W. *Real and complex analysis.* McGraw-Hill, 1966.

[Rumelhart, Hinton and Williams, 86] Rumelhart, D.E., Hinton, G.E., Williams, R.J. Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1: Foundations). Rumelhart, McClelland (eds.), MIT Press, Cambridge, MA, 1986.

[Rumelhart *et al.*, 93] Rumelhart D.E., Durbin R., Golden, R., Chauvin, Y. Backpropagation: The Basic Theory. In *Mathematical Perspectives of Neural Networks.* Smolensky, Mozer, Rumelhart (eds.), Lawrence Erlbaum, 1993.

[Rumelhart and Todd, 93] Rumelhart D.E., Todd, P.M. Learning and connectionist representations. In Meyer & Kornblum (eds.), *Attention and Performance, 14: Synergies in Experimental Psychology, Artificial Intelligence and Cognitive Neuroscience*, pp: 2-37. MIT Press: Cambridge, MA, 1993.

[Sarle, 99] Sarle, W.S. (ed.) Neural Network FAQ. Usenet *comp.ai.neural-nets.* URL: *ftp://ftp.sas.com/pub/neural/FAQ.html*

[Scarselli and Tsoi, 98] Scarselli, F., Tsoi, A.C. Universal Approximation using Feedforward Neural Networks: A survey of some existing methods, and some new results. *Neural Networks*, 11(1), pp. 17-37, 1998.

[Schafer, 94] Schafer, *Analysis of incomplete multivariate data by simulation.* Chapman & Hall, London, 1994.

[Schmidhuber, 92] Schmidhuber, J. A fixed size storage $O(n^3)$ time complexity learning algorithm for fully recurrent continually running networks, *Neural Computation* Vol. 4, pp.243-248, 1992.

[Schwefel, 65] Schwefel, H.P. Kybernetische Evolution als Strategie der experimentellen Forschung in der Strömungstechnik. Diplomarbeit, Technische Universität Berlin, 1965.

[Schwefel, 77] Schwefel, H.P. Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie. Vol. 26 of Interdisciplinary Systems Research. Birkhäuser, Basel, 1977.

[Shaffer, Whitley and Eshelman, 92] Shaffer, J.D., Whitley, D., Eshelman, L.J. Combination of Genetic Algorithms and Neural Networks: A Survey of the State of the Art. In *Combination of Genetic Algorithms and Neural Networks.* Shaffer, J.D., Whitley, D. (eds.), pp. 1-37, 1992.

[Shepard, 80] Shepard, R.N. Multidimensional scaling, tree-fitting, and clustering. *Science*, 210, pp. 390-397, 1980.

[Shewchuck, 94] Shewchuck, J.R. An Introduction to the Conjugate Gradient Method without the Agonizing Pain. Technical Report of the School of Computer Science, Carnegie Mellon Univ., 1994.

[Sontag, 90] Sontag, E. On the recognition capabilities of feed-forward nets. Report SYCON 90-03. Rutgers Center for Systems & Control. Rutgers Univ., 1990.

[Sopena and Alquézar, 94] Sopena, J.M., Alquézar, R. Improvement of learning in recurrent networks by substituting the sigmoid activation function. In Procs. of the 4th Intl. Conference on Artificial Neural Networks (ICANN'94). Sorrento, Italy. Springer-Verlag, 1994.

[Sopena, Romero and Alquézar, 99] Sopena, J.M., Romero, E., Alquézar, R. Neural networks with periodic and monotonic activation functions: a comparative study in classification problems. In Procs. of the 9th Intl. Conference on Artificial Neural Networks (ICANN'99), Edinburgh, Scotland. Springer-Verlag, 1999.

[Specht, 90] Specht, D. 1990. Probabilistic neural networks. *Neural Networks*, 3: 109-118, 1990.

[Sprecher, 72] Sprecher, D. An improvement on the superposition theorem of Kolmogorov. *Journal. of Math. Analysis and Applications*, Vol. 38: 208-213, 1972.

[Stanfill and Waltz, 86] Stanfill, C., Waltz, D. Toward memory-based reasoning. *Communications of the ACM*, vol. 29, pp. 1213-1228, December 1986.

[Steel and Torrie, 80] Steel, R., Torrie, J. *Principles and Procedures of Statistics. A Biomedical Approach.* McGraw-Hill, 1980.

[Steinbuch, 61] K. Steinbuch. Die lernmatrix. *Kybernetik*, 1: 36-45, 1961.

[Stone, 48] Stone M.H. The generalized Weierstrass approximation theorem. *Mathematics Magazine*, 21: 167-183, 237-254, 1948.

[Stone, 74] Stone M. Cross-validatory choice and assessment of statistical predictions. *Royal Statistical Society*, B36: 111-147, 1974.

[Su and Sheen, 92] Su, Y.T, Sheen, Y.T. Neural Networks for System Identification. *Intl. Journal of Systems Science*, 23(12): 2171-2186, 1992.

[Tikhonov and Arsenin, 77] Tikhonov, A.N., Arsenin, V.Y. *Solutions of ill-posed problems.* W.H. Winston, Washington D.C., 1984.

[Toolenaere, 90] Toolenaere, T. Fast adaptive back-propagation with good scaling properties. *Neural Networks*, 3: 561-574, 1990.

[Tresp, Ahmad and Neuneier, 94] Tresp, V., Ahmad, S., Neuneier, R. Training Neural Networks with Deficient Data. In Cowan, Tesauro and Alespector (eds.). Advances in Neural Information Processing Systems 6, Morgan Kaufmann, 1994.

[Tversky, 77] Tversky, A. Features of Similarity. *Psychological Review*, Vol. 84, No. 4, 327-352, 1977.

[Valdés and Gil, 84] J.J. Valdés and J.L. Gil. Joint use of geophysical and geomathematical methods in the study of experimental karst areas. In Procs. of the XXVII International Geological Congress, pp. 214, Moscow, 1984.

[Valdés, 97] J.J. Valdés. Fuzzy clustering and multidimensional signal processing in environmental studies: detection of underground caves in the tropics. In Procs. of EUFIT'97, European Congress on Intelligent Techniques and Soft Computing. Elite Foundation. Aachen, Germany, 1997.

[Valdés and García, 97] Valdés, J.J, García, R. A model for heterogeneous neurons and its use in configuring neural networks for classification problems. In Procs. of IWANN'97, Intl. World Conf. on Artificial and Natural Neural Networks. Lecture Notes in Computer Science 1240, Springer-Verlag, pp. 237-246.

[Valdés, Belanche and Alquézar, 00] Valdés J.J., Belanche, Ll., Alquézar, R. Fuzzy Heterogeneous Neurons for Imprecise Classification Problems. *Intl. Journal of Intelligent Systems*, 15(3): 265-276, 2000. Also in Research Report LSI-98-33-R. Universitat Politècnica de Catalunya, 1998.

[Vallverdú, 93] Vallverdú, M. *Modelado y simulación del sistema de control cardiovascular en pacientes con lesiones coronarias.* Ph. D. Thesis. Universitat Politècnica de Catalunya, 1993.

[Voigt, Mühlenbein and Cvetkovic, 95] Voigt, H.M., Mühlenbein, H., Cvetkovic, D. Fuzzy recombination for the continuous Breeder Genetic Algorithm. In Procs. of the 6th Intl. Conf. on Genetic Algorithms. Eshelman (ed.), Morgan Kaufmann: San Mateo, 1995.

[Wallach, 58] M.A. Wallach. On Psychological Similarity. *Psychological Review*, Vol. 65, No. 2, 103-116, 1958.

[WEF, 92] Standard Methods for the Examination of Water and Wastewater (16th Edition). Published by the Water Environment Federation, Washington, 1992.

[WEF, 96] Operation of Municipal Wastewater Treatment Plants. Manual of Practice No. 11 (5th Edition). Published by the Water Environmental Federation, Washington, 1996.

[Weiss and Edelman, 94] Weiss, Y. Edelman, S. Representation of similarity as a goal of early visual processing. Technical Report of the Dept. of Applied Mathematics and Computer Science. The Weizmann Institute of Science, 1994.

[Werbos, 74] Werbos, P.J. *Beyond regression: new tools for prediction and analysis in the behavioural sciences.* Ph.D. Thesis, Harvard Univ., 1974.

[Whitley, 89] Whitley, D., The GENITOR Algorithm and Selection Pressure: Why Rank-Based Allocation of Reproductive Trials is Best. In Procs. of the 3rd Intl. Conf. on Genetic Algorithms, (ed.) Schaffer, J.D., Morgan Kaufmann: San Mateo, 116-121, 1989.

[Whitley, 95] Whitley, D. Genetic Algorithms and Neural Networks. In *Genetic Algorithms in Engineering and Computer Science.* Periaux, Galán, Cuesta (eds.), John Wiley, 1995.

[Widrow and Hoff, 60] Widrow, B., Hoff, M. Adaptive switching circuits. In Procs. of the Western Electronic Show and Convention, Vol. 4: 96-104. Institute of Radio Engineers, 1960.

[Widrow, 90] Widrow, B. 30 years of adaptive neural networks: Perceptron, Madaline and Back-propagation. *IEEE Trans. on Neural Networks*, 78(9): 1415-1442, 1990.

[Wilson and Martinez, 96] Wilson, D.R., Martinez, T.R, Heterogeneous Radial Basis Function Networks. In Procs. of the 6th Intl. Conf. on Neural Networks (ICANN'96), vol. 2, pp. 1263-1267, 1996.

[Wilson and Martinez, 97] Wilson, D.R., Martinez, T.R, Improved Heterogeneous Distance Functions. *Journal of Artificial Intelligence Research*, 6: 1-34, 1997.

[Wray and Green, 95] Wray, J., Green, G.G.R. Neural networks, approximation theory and finite precision computation. *Neural Networks*, 8(1): 31-37, 1995.

[Würtz, 97] Würtz, R. Object recognition robust under translations, deformations and changes in background. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7): 769-775, 1997.

[Yao, 93] Yao, X. A Review of Evolutionary Artificial Networks. *Intl. Journal of Intelligent Systems*, 8(4): 539-567, 1993.

[Yosida, 74] Yosida, K. *Functional Analysis.* Springer, Berlin, 1974.

[Zadeh, 65] Zadeh, L. Fuzzy Sets. *Information and Control*, 8: 338-353, 1965.

[Zadeh, 73] Zadeh, L. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. on Systems, Man and Cybernetics*, 3: 28-44, 1973.

[Zadeh, 76] Zadeh, L. The concept of a linguistic variable and its application to approximate reasoning. Parts 1, 2. *Information Science*, 8: 199-249, 301-357, 1975. Part 3, *op. cit.*, 9: 43-80, 1976.

[Zadeh, 78] Zadeh, L. Fuzzy Sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1: 3-28, 1978.

[Zhang and Mühlenbein, 93] Zhang, B.T., Mühlenbein, H. Evolving Optimal Neural Networks Using Genetic Algorithms with Occam's Razor. *Complex Systems*, 7(3): 199-220, 1993.

[Zheng, 93] Zheng, Z. A Benchmark for Classifier Learning. Technical Report 474. Basser Dept. of Computer Science. The Univ. of Sydney, Australia, 1993.

[Zimmermann, 92] Zimmermann, H.J. *Fuzzy set theory and its applications.* Kluver Academic Publishers, 1992.

Appendix A

Additional material on distances

Definition A.1 *Let E be a vector space over $K = \mathbb{R}$ or $\mathbb{C}$. Every function ρ of E:*

$$\rho: \; E \rightarrow \mathbb{R}^{+} \cup \{0\}$$

such that, $\forall x, y \in E, \lambda \in K$:

1. $\rho(x) = 0 \Rightarrow x = 0$
2. $\rho(x+y) \leq \rho(x) + \rho(y)$
3. $\rho(\lambda x) = |\lambda| \rho(x)$

is a norm on E. A space E with a norm ρ is a normed space (E, ρ).

Proposition A.1 *Every normed space (E, ρ) can be converted in a metric space (E, d) by defining $d(x, y) = \rho(x - y)$.*

Proof: see [Kolmogorov and Fomin, 75].

Lemma A.1 *Let $E = E_1 \times E_2, \ldots, \times E_n$ and $d_1, \ldots, d_n$ such that $d_i \in D(E_i)$ (these functions are not necessarily equal). Then,*

$$d_{\Sigma} = \sum_{i=1}^{n} d_i \qquad \in D(E)$$

Proof: By fulfilment of the conditions in Definition (4.1). For all $\vec{x}, \vec{y}, \vec{z} \in E$,

1. $d_{\Sigma}(\vec{x}, \vec{y}) = 0$
   $\equiv$ (def. of d_{Σ})
   $\sum_{i=1}^{n} d_i(x_i, y_i) = 0$
   $\equiv$ $(d_i(x_i, y_i) \geq 0)$

$\forall i : 1 \leq i \leq n : d_i(x_i, y_i) = 0$
$\equiv$ (d_i are all distances)
$\forall i : 1 \leq i \leq n : x_i = y_i$
$\equiv$
$\vec{x} = \vec{y}$ $\square$

2. $d_\Sigma(\vec{x}, \vec{y})$
$\equiv$ (def. of d_Σ)
$\sum_{i=1}^n d_i(x_i, y_i)$
$\equiv$ (d_i are all distances)
$\sum_{i=1}^n d_i(y_i, x_i)$
$\equiv$ (def. of d_Σ)
$d_\Sigma(\vec{y}, \vec{x})$ $\square$

3. $d_\Sigma(\vec{x}, \vec{y}) \leq d_\Sigma(\vec{x}, \vec{z}) + d_\Sigma(\vec{z}, \vec{y})$
$\equiv$ (def. of d_Σ)
$\sum_{i=1}^n d_i(x_i, y_i) \leq \sum_{i=1}^n d_i(x_i, z_i) + \sum_{i=1}^n d_i(z_i, y_i)$
$\equiv$
$\sum_{i=1}^n d_i(x_i, y_i) \leq \sum_{i=1}^n \{d_i(x_i, z_i) + d_i(z_i, y_i)\}$
$\equiv$ (d_i are all distances)
true $\square$

Lemma A.2 *Let $d \in D(E)$. Then,*

$$\forall \alpha \in \mathbb{R}^+,\ \alpha d \in D(E)$$

Proof: obvious, since α can be cancelled everywhere.

Lemma A.3 *Let $E = E_1 \times E_2, \ldots, \times E_n$ and $d_1, \ldots, d_n$ such that $d_i \in D(E_i)$ (these functions are not necessarily equal). Then, any linear combination of the d_i with non-negative coefficients is a distance in E.*

Proof: By making use of Lemmas (A.1) and (A.2).

Lemma A.4 *Let $d \in D(E)$. Then,*

$$\forall \alpha \in (0, 1],\ d^\alpha \in D(E)$$

Proof: By fulfilment of the conditions in Definition (4.1). Forall $x, y, z \in E$,

1. $d(x, y)^\alpha = 0$
$\equiv$ (d is non-negative)
$d(x, y) = 0$
$\equiv$ (d is a distance)
$x = y$ $\square$

2. $d(x,y)^\alpha = d(y,x)^\alpha$
   $\equiv$
   $d(x,y) = d(y,x)$
   $\equiv$ (d is a distance)
   true $\square$

3. $d(x,y)^\alpha \leq d(x,z)^\alpha + d(z,y)^\alpha$
   $\equiv$ (defining $\beta = \frac{1}{\alpha} \geq 1$)
   $\sqrt[\beta]{d(x,y)} \leq \sqrt[\beta]{d(x,z)} + \sqrt[\beta]{d(z,y)}$
   $\equiv$ (raising to the β power)
   $d(x,y) \leq \left(\sqrt[\beta]{d(x,z)} + \sqrt[\beta]{d(z,y)} \right)^\beta$
   $\equiv$
   $d(x,y) \leq d(x,z) + d(z,y) + G(x,y,z,\beta)$
   $\equiv$ (d is a distance and $G(x,y,z,\beta) \geq 0$)
   true $\square$

A well-known method of introducing a norm in a vector space is by defining first a **scalar product.**

Definition A.2 (Scalar product) *A scalar product in a real vector space $\mathcal{R}$ is a real function (x,y) verifying, $\forall x, y, y_1, y_2 \in \mathbb{R}$:*

1. $(x,y) = (y,x)$.
2. $(x, y_1 + y_2) = (x, y_1) + (x, y_2)$.
3. $(\lambda x, y) = \lambda(x,y),\ \lambda \in \mathbb{R}$.
4. $(x,x) \geq 0$ *and* $(x,x) = 0 \Leftrightarrow x = 0$.

A vector space where a scalar product is defined is an **Euclidean space**. In such space, the norm is introduced by the formula: $\|x\| = \sqrt{(x,x)}$. Having defined a norm, it is immediate to define a distance, by Proposition (A.1).

The choice $\mathcal{R} = \mathbb{R}^n$ (the common space of coordinates in n dimensions, which is a vector space) leads to the classical example of Euclidean space, with:

$$(x,y) = \sum_{i=1}^{n} x_i y_i$$

In this case, (x,y) is usually written $\vec{x} \cdot \vec{y}$, for $\vec{x}, \vec{y} \in \mathbb{R}^n$. In general, every Euclidean space is normed, but the opposite does not hold. A necessary and sufficient condition for a vector space $\mathcal{R}$ to be Euclidean is that:

$$\forall x, y \in \mathcal{R},\ \|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

For instance, by taking $\mathcal{R} = \mathbb{R}^n$ with its customary norm:

$$||\vec{x}||_q = \left(\sum_{i=1}^{n} |x_i|^q\right)^{\frac{1}{q}}.$$

For $q \geq 1 \in \mathbb{R}$, all the required properties of the norm (Definition (A.1)) are fulfilled. However, $\mathbb{R}^n$ is an Euclidean space *only* for $q = 2$. This means that, for $q \neq 2$, the norm in $\mathbb{R}^n$ cannot be defined out of any scalar product [Kolmogorov and Fomin, 75].

To conclude, we enunciate a classical result:

Definition A.3 (Equivalent norms) *Two norms* $|| \cdot ||_a$, $|| \cdot ||_b, a, b \geq 1 \in \mathbb{R}$, *defined on a vector space* $\mathcal{R}$ *are said to be* equivalent *if:*

$$\exists c_1, c_2 > 0 \qquad c_1||x||_a \leq ||x||_b \leq c_2||x||_a, \ \forall x \in \mathcal{R}$$

where c_1, c_2 *only depend on* a, b.

Proposition A.2 *In* $\mathcal{R} = \mathbb{R}^n$ *all the norms are equivalent.*

Appendix B

Proofs of Propositions

Those not proven in the main body.

Proof of Proposition (4.1). Obvious, since $S \subset X$ implies $x, y, z \in S \Rightarrow x, y, z \in X$ and every $x, y, z \in X$ fulfills Definition (4.1) because (X, d) is a metric space $\square$.

Proof of Proposition (4.2). Refer to [Queysanne, 85].

Proof of Proposition (4.3). Use $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$ and Proposition (4.1) $\square$.

Proof of Proposition (4.4). By fulfilment of Definition (4.3).

i) Minimality. $\Theta^{q,n}(\vec{z}; \vec{v}) = 0 \Leftrightarrow \vec{z} = \vec{0}$. Obvious, by noting that $\forall i : 1 \leq i \leq n : v_i > 0$.

ii) Symmetry is kept by applying $\Theta^{q,n}(\vec{z}; \vec{v}) = \Theta^{q,n}(\sigma(\vec{z}); \sigma(\vec{v}))$, thanks to the commutativity property of summation.

iii) The function $\Theta^{q,n}(\vec{z}; \vec{v})$ is strictly monotonic w.r.t. any z_i.

Proof of Proposition (4.5). Let $\vec{z} = \{z_1, z_2, \ldots, z_n\} \in \mathbb{R}^n$ be a vector of non-negative components. Denote $\rho = \Theta^{q,n}, \forall q \geq 1 \in \mathbb{R}$. We prove this ρ is a norm in $\mathbb{R}^n$ by fulfilment of Definition (A.1). For all $\vec{z}, \vec{z}_1, \vec{z}_2 \in \mathbb{R}^n$,

1. $\rho(\vec{z}) = 0 \Rightarrow \vec{z} = \vec{0}$
2. $\rho(\vec{z}_1 + \vec{z}_2) \leq \rho(\vec{z}_1) + \rho(\vec{z}_2)$ First, we see that n' is a positive constant factor and can thus be cancelled, leading to:

$$\left(\sum_{i=1}^{n} \left(\frac{|z_i^1 + z_i^2|}{v_i} \right)^q \right)^{\frac{1}{q}} \leq \left(\sum_{i=1}^{n} \left(\frac{|z_i^1|}{v_i} \right)^q \right)^{\frac{1}{q}} + \left(\sum_{i=1}^{n} \left(\frac{|z_i^2|}{v_i} \right)^q \right)^{\frac{1}{q}}$$

which is a weighted form of the general Minkowski's inequality [Kolmogorov and Fomin, 75]. Since the denominators v_i are also positive constants, equal for both sides, they do not alter the inequality.

3. $\rho(\lambda\vec{z}) = |\lambda|\rho(\vec{z})$. We have:

$$\rho(\lambda\vec{z}) = \left(\frac{1}{n'}\sum_{i=1}^{n}\left(\frac{|\lambda z_i|}{v_i}\right)^q\right)^{\frac{1}{q}}$$

($|\cdot|$ is a norm in $\mathbb{R}$)

$$= \left(\frac{1}{n'}\sum_{i=1}^{n}|\lambda|^q\left(\frac{|z_i|}{v_i}\right)^q\right)^{\frac{1}{q}}$$

$$= |\lambda|\left(\frac{1}{n'}\sum_{i=1}^{n}\left(\frac{|z_i|}{v_i}\right)^q\right)^{\frac{1}{q}} = |\lambda|\rho(\vec{z})$$

Proof of Proposition (4.6). By application of Propositions (4.5) and (A.1).

Proof of Proposition (4.7). First, note that an n-linear aggregation Θ fulfills the conditions of a general aggregation operator in (4.3). Second, since such an operator performs a linear combination of its arguments, and these are distances $d_i \in D(X^i)$, by Lemma (A.3), its n-linear aggregation Θ is a metric distance in $X = X^1 \times \ldots \times X^n$.

Proof of Proposition (4.8). It is clearly a n-linear operator (Definitions (4.5) and (4.6)).

Proof of Proposition (4.9). It is a particular case of Proposition (4.4), for $\vec{v} = \vec{1}, n' = n$.

Proof of Proposition (4.10). By Proposition (4.8) for $\vec{v} = (\frac{1}{n}, \ldots, \frac{1}{n})$. Alternatively, it can be seen as a particular case of (4.9) for $q = 1$.

Proof of Proposition (4.11). We prove it by fulfilment of the conditions in Definition (4.7).

1. Non-negativity. By definition of Θ_s.
2. Symmetry. By the symmetry of Θ_s.
3. Boundedness. By definition of Θ_s.
4. Minimality. This reads:

$$\Theta_s(\{s_1, \ldots, s_n\}) \Leftrightarrow \forall i : 1 \leq i \leq n : s_i = s_{max}$$

(a) ($\Leftarrow$) By the idempotency of Θ_s.

(b) ($\Rightarrow$) By contradiction. Suppose $\exists i : 1 \leq i \leq n : s_i < s_{max}$ such that $\Theta_s(\vec{s}) = s_{max}$. In this case, an $\vec{s}'$ such that $\vec{s}' = \{s_{max}, \ldots, s_{max}\}$ would yield, by the monotonicity property of Θ_s, $\Theta_s(\vec{s}') > \Theta_s(\vec{s}) = s_{max}$!

5. Semantics is expressed in conditions (v) to (vii).

Proof of Proposition (4.12). By fulfilment of the conditions in Definition (4.11). Let $\vec{s} = F_{\mathcal{X}}(\vec{s}_0)$, $\vec{s}_0$ of length n_0 and $\vec{s}$ of length n. All the properties but the last are to be valid for the present components in $\vec{s}_0$, that is to say, for all the components in $\vec{s}$. In all cases, the treatment of missing components is done by defining $\Theta_s(\vec{s}_0) = \Theta_s(\vec{s})$ (this fulfills property $[viii)]$). For the sake of clarity, we begin by the simplest of the three families.

The normalized modulus.

$$\Theta_s(\vec{s}) = \frac{1}{\sqrt[q]{n}}\|\vec{s}\|_q, \; q \geq 1 \in \mathbb{R}$$

1. Minimality. Since $\Theta_s^{q,n}(\vec{s}) = 0 \Rightarrow \forall i : 1 \leq i \leq n : s_i = 0$ both conditions are satisfied. The 0 is *not*, in this case, an absorbing element.
2. Symmetry. By the commutativity property of summation.
3. Monotonicity. The function $\Theta_s^{q,n}(\vec{s})$ is strictly monotonic w.r.t. any s_i.
4. Idempotency. For an arbitrary $s_k \in [0, s_{max}]$, let $\vec{s}_k = (s_k, \ldots, s_k)$.

$$s_k^n[\Theta_s] = \Theta_s^{q,n}(\vec{s}_k) = \frac{1}{\sqrt[q]{n}}\left(\sum_{i=1}^{n} s_k^q\right)^{\frac{1}{q}} = \frac{1}{\sqrt[q]{n}}(ns_k^q)^{\frac{1}{q}} = \frac{1}{\sqrt[q]{n}}n^{\frac{1}{q}}s_k = s_k$$

Note that this also ensures that $\Theta_s^{q,n} : [0, s_{max}]^n \to [0, s_{max}], \forall q \geq 1 \in \mathbb{R}, \forall n \in \mathbb{N}^+$.

5. Cancellation law.

$\frac{1}{\sqrt[q]{n}}\sqrt[q]{s_1^q + s_2^q} = \frac{1}{\sqrt[q]{n}}\sqrt[q]{s_1^q + s_3^q}$
$\equiv$
$\sqrt[q]{s_1^q + s_2^q} = \sqrt[q]{s_1^q + s_3^q}$
$s_1^q + s_2^q = s_1^q + s_3^q$
$\equiv$
$s_2^q = s_3^q$
$\equiv (s_2, s_3 \geq 0)$
$s_2 = s_3$ □

6. Continuity. $\Theta_s^{q,n}$ is a continuous function.
7. Compensativeness. $\min_i s_i \leq \Theta_s(\vec{s}) \leq \max_i s_i$

Let $\vec{s}_\mu(\vec{s}) = (\max_i s_i, \ldots, \max_i s_i)$. We have $\Theta_s^{q,n}(\vec{s}) \leq \Theta_s^{q,n}(\vec{s}_\mu(\vec{s}))$, by the monotonicity property, and $\Theta_s^{q,n}(\vec{s}_\mu(\vec{s})) = \max_i s_i$ by the idempotency property. Analogously for $\min_i s_i$.

Additive measures.

$$\Theta_s(\vec{s};\vec{v}) = f^{-1}\left(\sum_{i=1}^{n} v_i f(s_i)\right)$$

Where f is strictly increasing and continuous, $f(0) = 0$ and $f(s_{max}) = s_{max}$, and $\vec{v}$ is such that $\sum_{i=1}^{n} v_i = 1$, with $v_i > 0$.

1. Minimality. The first condition reads:
   $\Theta_s(\vec{s};\vec{v}) = 0$
   $\equiv$
   $f^{-1}\left(\sum_{i=1}^{n} v_i f(s_i)\right) = 0$
   $\equiv$
   $\sum_{i=1}^{n} v_i f(s_i) = 0$
   $\equiv (\forall i, v_i > 0)$
   $\forall i : 1 \leq i \leq n : f(s_i) = 0$
   $\equiv>$
   $\exists i : 1 \leq i \leq n : f(s_i) = 0$
   $\equiv$
   $\exists i : 1 \leq i \leq n : s_i = 0$ □
   The second condition is proven by reading the previous proof from the universal quantifier up.
2. Symmetry. By applying $\Theta_s(\vec{z};\vec{v}) = \Theta_s(\sigma(\vec{z});\sigma(\vec{v}))$, thanks to the commutativity property of summation.
3. Monotonicity. The functions f, f^{-1} are strictly increasing.
4. Idempotency. For an arbitrary $s_k \in [0, s_{max}]$, let $\vec{s}_k = (s_k, \ldots, s_k)$.

   $$s_k^n[\Theta_s] = \Theta_s(\vec{s}_k) = f^{-1}\left(\sum_{i=1}^{n} v_i f(s_k)\right) = f^{-1}\left(f(s_k)\sum_{i=1}^{n} v_i\right) = f^{-1}(f(s_k)) = s_k$$

   Again, this ensures that $\Theta_s^{q,n} : [0, s_{max}]^n \to [0, s_{max}], \forall q \geq 1 \in \mathbb{R}, \forall n \in \mathbb{N}^+$.
5. Cancellation law.
   $f^{-1}(v_1 f(s_1) + v_2 f(s_2)) = f^{-1}(v_1 f(s_1) + v_3 f(s_3))$
   $\equiv$
   $v_1 f(s_1) + v_2 f(s_2) = v_1 f(s_1) + v_3 f(s_3)$
   $\equiv$
   $v_2 f(s_2) = v_3 f(s_3)$
   $\equiv$ (in case all the v_i are equal)
   $f(s_2) = f(s_3)$
   $\equiv$
   $s_2 = s_3$ □
   This assumption appears in the common case of weightings as an averaging mechanism. If the v_i are different, then strict equality is not necessary, and the condition: $\frac{v_2}{v_3} = \frac{f(s_2)}{f(s_3)}$ has to be met.

6. Continuity. Θ_s is continuous because f, f^{-1} are.
7. Compensativeness. $\min_i s_i \leq \Theta_s(\vec{s};\vec{v}) \leq \max_i s_i$
   Let $\vec{s}_\mu(\vec{s};\vec{v}) = (\max_i s_i, \ldots, \max_i s_i)$. We have $\Theta_s(\vec{s};\vec{v}) \leq \Theta_s(\vec{s}_\mu(\vec{s});\vec{v})$, by the monotonicity property, and $\Theta_s(\vec{s}_\mu(\vec{s});\vec{v}) = \max_i s_i$ by the idempotency property. Analogously for $\min_i s_i$.

Multiplicative measures.

$$\Theta_s(\vec{s};\vec{v}) = f^{-1}\left(\prod_{i=1}^{n} f(s_i)^{v_i}\right)$$

with f and $\vec{v}$ in the same conditions as for additive measures.

1. Minimality. The first condition reads:
   $\Theta_s(\vec{s};\vec{v}) = 0$
   $\equiv$
   $f^{-1}\left(\prod_{i=1}^{n} f(s_i)^{v_i}\right) = 0$
   $\equiv$
   $\prod_{i=1}^{n} v_i f(s_i)^{v_i} = 0$
   $\equiv$
   $\exists i : 1 \leq i \leq n : f(s_i)^{v_i} = 0$
   $\equiv$
   $\exists i : 1 \leq i \leq n : f(s_i) = 0$
   $\equiv$
   $\exists i : 1 \leq i \leq n : s_i = 0 \quad \square$
   The second condition is proven by noting that:
   $$\forall i : 1 \leq i \leq n : s_i = 0 \Rightarrow \exists i : 1 \leq i \leq n : s_i = 0$$
   and reading the previous proof backwards.
2. Symmetry. For the same reasons than for additive measures.
3. Monotonicity. For the same reasons than for additive measures.
4. Idempotency. For an arbitrary $s_k \in [0, s_{max}]$, let $\vec{s}_k = (s_k, \ldots, s_k)$.
   $$s_k^n[\Theta_s] = \Theta_s(\vec{s}_k) = f^{-1}\left(\prod_{i=1}^{n} f(s_k)^{v_i}\right) = f^{-1}\left(f(s_k)^{\sum_{i=1}^{n} v_i}\right) = f^{-1}\left(f(s_k)\right) = s_k$$
   Again, this ensures that $\Theta_s^{q,n} : [0, s_{max}]^n \to [0, s_{max}], \forall q \geq 1 \in \mathbb{R}, \forall n \in \mathbb{N}^+$.
5. Cancellation law.
   An analogous derivation leads to: $f(s_2)^{v_2} = f(s_3)^{v_3}$.
6. Continuity. For the same reasons than for additive measures.

7. Compensativeness. $\min_i s_i \leq \Theta_s(\vec{s};\vec{v}) \leq \max_i s_i$
   Let $\vec{s}_\mu(\vec{s};\vec{v}) = (\max_i s_i, \ldots, \max_i s_i)$. We have $\Theta_s(\vec{s};\vec{v}) \leq \Theta_s(\vec{s}_\mu(\vec{s});\vec{v})$, by the monotonicity property, and $\Theta_s(\vec{s}_\mu(\vec{s});\vec{v}) = \max_i s_i$ by the idempotency property. Analogously for $\min_i s_i$.

Proof of Proposition (4.13). By fulfilment of the conditions in Definition (4.7).

1. Non-negativity. By definition of $\breve{s}$.
2. Symmetry. By the symmetry of s.
3. Boundedness. By definition of $\breve{s}$.
4. Minimality. This reads:
   $\breve{s}(s(x,y)) = \breve{s}_{max} \Leftrightarrow x = y$
   $\equiv$ ($\breve{s}(z) = \breve{s}_{max}$ only for $z = s_{max}$)
   $s(x,y) = s_{max} \Leftrightarrow x = y$
   $\equiv$ (s is a similarity)
   true □
5. The semantics of s is kept in $\breve{s} \circ s$ since $\breve{s}$ is a continuous and strictly increasing function.

Proof of Proposition (4.14). By fulfilment of the conditions in Definition (4.7).

1. Non-negativity. By definition of $\hat{s}$.
2. Symmetry. By symmetry of d.
   $s(x,y)$
   $\equiv$
   $\hat{s}(d(x,y))$
   $\equiv$ (d is a distance)
   $\hat{s}(d(y,x))$
   $\equiv$
   $s(y,x)$ □
3. Boundedness. By definition of $\hat{s}$.
4. Minimality. This reads:
   $s(x,y) = s_{max} \Leftrightarrow x = y$
   $\equiv$
   $\hat{s}(d(x,y)) = s_{max} \Leftrightarrow x = y$
   $\equiv$ ($\hat{s}(0) = s_{max}$, $\hat{s}$ is strictly decreasing)
   $d(x,y) = 0 \Leftrightarrow x = y$
   $\equiv$ (d is a distance)
   true □

5. The semantics of s is that of the inverse in $d(x,y)$ (the more the distance, the less the similarity, and vice versa), since $\hat{s}$ is a continuous and strictly decreasing function.

Proof of Proposition (4.15). Considering the properties fulfilled by $\breve{s}, \hat{s}$, we have:

- $\breve{s} \circ \hat{s}$ is strictly decreasing.
- $\breve{s}(\hat{s}(0)) = \breve{s}(s_{max}) = \breve{s}_{max}$
- $lim_{z\to\infty}\breve{s}(\hat{s}(z)) = lim_{z\to\infty}\breve{s}(0) = \breve{s}(0) = 0.$

Proof of Proposition (4.16). Since a normalized distance is a distance, the same proof as for Proposition (4.14) applies. The extra condition $\hat{s}(d_{max}) = 0$ is required to ensure that the formed similarity s covers $[0, s_{max}]$ (this is already guaranteed in Proposition (4.14), although not needed for the proof).

Proof of Proposition (4.17). It suffices to define:

$$\hat{s}'(z) = \frac{\hat{s}(z) - \hat{s}(d_{max})}{1 - \hat{s}(d_{max})}$$

Proof of Proposition (4.24). By fulfilment of the conditions in Definition (4.7).

1. Non-negativity. By definition of s.
2. Symmetry. By symmetry of s.
3. Boundedness. By definition of s.
4. Minimality. This reads:
   $s(x,y) = s_{max} \Leftrightarrow x = y$
   $\equiv (s_{max} = 1)$
   $s(x,y) = 1 \Leftrightarrow x = y$
   $\equiv$ (by def. of s)
   $true$ $\Box$
5. The semantics of s is binary, in that either two objects are similar or not. This is an equivalence relation.

Proof of Propositions (4.25, 4.27). They are discussed in the text, pages 94ss, and obtained by application of Proposition (4.16).

Proof of Proposition (4.26). By fulfilment of the conditions in Definition (4.7).

1. Non-negativity. By definition of s.
2. Symmetry. By symmetry of s.
3. Boundedness. By definition of s.
4. Minimality. This reads:
   $s(x, y) = s_{max} \Leftrightarrow x = y$
   $\equiv (s_{max} = 1)$
   $s(x, y) = 1 \Leftrightarrow x = y$
   $\equiv$ (by def. of s)
   $\frac{1}{m-1}\left(m\frac{min(\eta(x),\eta(y))}{max(\eta(x),\eta(y))} - 1\right) = 1 \Leftrightarrow x = y$
   $\equiv$
   $\left(m\frac{min(\eta(x),\eta(y))}{max(\eta(x),\eta(y))} - 1\right) = m - 1 \Leftrightarrow x = y$
   $\equiv$
   $\frac{min(\eta(x),\eta(y))}{max(\eta(x),\eta(y))} = 1 \Leftrightarrow x = y$
   $\equiv$
   $min(\eta(x), \eta(y)) = max(\eta(x), \eta(y)) \Leftrightarrow x = y$
   $\equiv$
   $true$ $\square$
5. This measure has a clear semantics, discussed in page (95).

Proof of Proposition (4.28). By fulfilment of the conditions in Definition (4.7).

1. Non-negativity. By definition of s.
2. Symmetry. By symmetry of s.
3. Boundedness. By definition of s.
4. Minimality. This reads:
   $s(x, y) = s_{max} \Leftrightarrow x = y$
   $\equiv (s_{max} = 1)$
   $s(x, y) = 1 \Leftrightarrow x = y$
   $\equiv$ (by def. of s)
   $\frac{\#x \cap y}{\#x \cup y} = 1 \Leftrightarrow x = y$
   $\equiv$
   $\#x \cap y = \#x \cup y \Leftrightarrow x = y$
   $\equiv$ (set properties)
   $true$ $\square$
5. This measure has a clear semantics: number of shared elements w.r.t. to the number of different elements apported between the two sets. Interestingly, the empty set means that the variable has no value and can thus be used to express missing information, with the semantics: $s(\emptyset, x) = 0,\ \forall x \in \mathcal{S}$, in accordance to its definition. Alternatively, $s(\emptyset, x) = \mathcal{X}$ can also be defined.

Appendix C

Notes on integrability

The following material can be found in textbooks on real analysis. Our references are [Jarauta, 93] for $\mathbb{R}$ and [de Burgos, 95] for $\mathbb{R}^n$. The propositions with proof are due to the author. We begin with a brief review on integrability in $\mathbb{R}$ and follow on to $\mathbb{R}^n$.

C.1 Integrability in $\mathbb{R}$

Let f a real function, integrable in any real interval $[a, x]$ for all $x > a$, that is:

$$F(x) = \int_a^x f(t)dt \qquad \text{exists and is finite } \forall x > a \in \mathbb{R}$$

Let us consider the limit of these integrals:

$$\lim_{x \to \infty} F(x) = \lim_{x \to +\infty} \int_a^x f(t)dt \tag{C.1}$$

If this limit exists, it is written:

$$\int_a^{+\infty} f(t)dt \tag{C.2}$$

and called **improper integral** (of the first kind) of f in $[a, +\infty]$. We say that (C.1) is *convergent* and the limit in (C.2) is the value of the integral. Otherwise, we say it is *divergent*. Analogously, whenever f is (Riemann) integrable in any real interval $[x, a]$ for all $x < a$, it can be defined:

$$\int_{-\infty}^a f(t)dt = \lim_{x \to -\infty} \int_x^a f(t)dt \tag{C.3}$$

Hence, if the function is integrable in any interval $[x, y] \subset \mathbb{R}$, then the improper integral can be defined:

$$\int_{-\infty}^{+\infty} f(t)dt = \lim_{x \to -\infty} \int_{x}^{a} f(t)dt + \lim_{y \to +\infty} \int_{a}^{y} f(t)dt, \qquad (a \in \mathbb{R}) \tag{C.4}$$

In any case, the improper integral is convergent if it exists and is finite; otherwise, we say it is divergent. From now on, we write:

$$\int_{-\infty}^{+\infty} f(t)dt = \int_{\mathbb{R}} f(t)dt \tag{C.5}$$

and consider $a = 0$. We say that f is *integrable* in Δ, $f : \Delta \to \mathbb{R}, \Delta \subseteq \mathbb{R}$ measurable if $\int_{\Delta} f(t)dt$ is convergent. Since the codomain of f is $\mathbb{R}$, the value of the integral, in case it exists, is a real number. We denote by $I(\Delta) = \{f : \Delta \to \mathbb{R} \mid f \text{ integrable in } \Delta\}$. We also denote by $C(\Delta)$ the set of continuous functions in Δ.

Proposition C.1 *Let f a positive function. If $f \in I(\Delta)$, then $\int_{\Delta} f(t)dt > 0$.*

Proposition C.2 *If a function $f \in I(\Delta)$, where Δ is the support of f (that is, f is null outside Δ), then $f \in I(\mathbb{R})$, and $\int_{\Delta} f(t)dt = \int_{\mathbb{R}} f(t)dt$.*

Proposition C.3 (Comparison criterion 1) *Let $f, g : \mathbb{R} \to \mathbb{R}$, such that $f, g \in I([a, x]), \forall x > a$, f is positive in $[a, +\infty)$ and $|g(x)| \leq f(x), \forall x \in [a, +\infty)$. If $\int_{a}^{\infty} f(t)dt$ is convergent, then $\int_{a}^{\infty} g(t)dt$ is convergent.*

Proposition C.4 (Comparison criterion 2) *Let $f, g : \mathbb{R} \to \mathbb{R}$, such that $f, g \in I([x, a]), \forall x < a$, f is positive in $(-\infty, a]$ and $|g(x)| \leq f(x), \forall x \in (-\infty, a]$. If $\int_{-\infty}^{a} f(t)dt$ is convergent, then $\int_{-\infty}^{a} g(t)dt$ is convergent.*

Proposition C.5 *The following families of functions $g(z) : [0, +\infty) \to \mathbb{R}$ are continuous and integrable in $[0, +\infty)$, and have a positive integral:*

1. $g_1(z) = \dfrac{1}{1 + e^{(az)^{\alpha}}}, \qquad a > 0, \alpha > 0$
2. $g_2(z) = \dfrac{1}{1 + (az)^{\alpha}}, \qquad a > 0, \alpha > 1$
3. $g_3(z) = e^{-(az)^{\alpha}}, \qquad a > 0, \alpha > 0$

As a consequence, the functions obtained replacing z by $|z|$ are defined $g(z) : \mathbb{R} \to \mathbb{R}$ and are integrable in $\mathbb{R}$.

Proposition C.6 *Let $f \in I(\mathbb{R})$. Let $h(x) = \alpha f(x), \alpha \in \mathbb{R}$. Then $h \in I(\mathbb{R})$.*

Proposition C.7 *Let $f, g \in I(\mathbb{R})$. Let $h(x) = f(x) + g(x)$. Then $h \in I(\mathbb{R})$.*

This can be extended to any number of functions. Let $\{f_i\}$ a collection of n integrable functions, $f_i \in I(\mathbf{R})$ such that $\int_{\mathbf{R}} f_i(x)dx = l_i \in \mathbf{R}$, for all $1 \leq i \leq n$. Then, the function $h(x) = \sum_{i=1}^{n} f_i(x) \in I(\mathbf{R})$ and $\int_{\mathbf{R}} h(x)dx = \sum_{i=1}^{n} l_i$.

C.2 Some notes on integrability in $\mathbb{R}^2$

Proposition C.8 *The function $h(x,y) = f(x) + g(y)$, $f(x) > 0, g(y) > 0, \forall x, y \in \mathbb{R}$ is* not *integrable in $\mathbf{R}^2$, even if $f, g \in I(\mathbf{R})$, except for the trivial case $f(x) = g(y) = 0,\ \forall x, y \in \mathbb{R}$.*

Proof. Since $f(x) > 0$ and $g(y) > 0$, $h(x,y) > f(x)$ and $h(x,y) > g(y)$. Let $F'(x) = f(x)$. Given x_0, consider a rectangular slice of x_0, $f(x_0) > 0$, and define: $A_{x_0,\epsilon} = \{(x,y) \mid |x - x_0| \leq \epsilon\}$.

We have:

$$\int_{A_{x_0,\epsilon}} h(x,y)dxdy > \int_{A_{x_0,\epsilon}} f(x)dxdy = \int_{\mathbb{R}} dy \int_{|x-x_0|\leq\epsilon} f(x)dx = \int_{\mathbb{R}} dy \left\{F(x)|_{x_0-\epsilon}^{x_0+\epsilon}\right\}$$
$$= \int_{\mathbf{R}} dy\{F(x_0+\epsilon) - F(x_0-\epsilon)\} = \{F(x_0+\epsilon) - F(x_0-\epsilon)\} \int_{\mathbb{R}} 1dy = \infty \quad \text{(C.6)}$$

Note that the Proposition is still valid for the particular case $f = g$.

Proposition C.9 *The function $h(x_1, \cdots, x_n) = \sum_{i=1}^{n} f_i(x_i)$, $f_i(x) > 0, \forall x, y \in \mathbb{R}$ is* not *integrable in $\mathbf{R}^n$, even if $f_i \in I(\mathbb{R})$, for all $1 \leq i \leq n$.*

Proof. Failure for the case $n = 2$ –Proposition (C.8)– invalidates any superior proof.

Proposition C.10 *The function $h(x,y) = f(x)g(y)$, $f(x) > 0, g(y) > 0, \forall x, y \in \mathbb{R}$ is integrable in $\mathbf{R}^2$, provided $f, g \in I(\mathbb{R})$.*

Proof. Let $\int_{\mathbf{R}} f(x)dx = l_f \in \mathbf{R}$, Let $\int_{\mathbf{R}} g(x)dx = l_g \in \mathbb{R}$. Then,

$$\int\int_{\mathbf{R}^2} f(x)g(y)dxdy = \int_{\mathbf{R}} dy \int_{\mathbf{R}} f(x)g(y)dx = \int_{\mathbb{R}} g(y)dy \left\{\int_{\mathbb{R}} f(x)dx\right\}$$
$$= \int_{\mathbf{R}} l_f g(y)dy = l_f \int_{\mathbf{R}} g(y)dy = l_f l_g \in \mathbb{R} \quad \text{(C.7)}$$

The extension to $\mathbf{R}^n$ is straightforward, just by iteration of the property for $n = 2$. Thus, the function $h(x_1, \cdots, x_n) = \prod_{i=1}^{n} f_i(x_i) \in I(\mathbf{R}^n)$ provided $f_i(x) \in I(\mathbb{R})$.

C.3 Integrability in $\mathbb{R}^n$

A set $I \subset \mathbb{R}^n$ has **zero measure** if $\forall\epsilon > 0$, there exists a covering (possibly finite) of I which is enumerated, formed by compact intervals, whose sum of measures is less than ϵ. Formally,

$$\forall\epsilon > 0, \exists I_1, \ldots, I_k, \ldots \text{ with } I_i \subset \mathbb{R}^n \text{ compact intervals, such that } I = \bigcup_{i=1}^{\infty} I_i \text{ and } \sum_{i=1}^{\infty} \mu(I_i) < \epsilon$$

In particular, any enumerable set has zero measure. The propositions and definitions about the Riemann integral in $\mathbb{R}$ can be extended to $\mathbb{R}^n$. The multiple integral is denoted:

$$\int_I f(\vec{x})d\vec{x} = \int_I \underbrace{\cdots}_{(n)} \int_I f(x_1, \ldots, x_n)dx_1, \ldots, dx_n$$

Theorem C.1 (Lebesgue's, about the Riemann integral) *Let $f : S \to \mathbb{R}$, with $S \subset \mathbb{R}^n$ compact interval and f bounded in S. Then $f \in I(S)$ if and only if the set of points $x \in S$ where $f(x)$ is discontinuous has zero measure.*

Many of the properties introduced for the real case (in an interval) are still valid. In particular, we remark the following. Let $S \subset \mathbb{R}^n$ be a compact interval:

1. Let $f : S \to \mathbb{R}$, $f \in C(S)$ implies $f \in I(S)$.
2. Let $f, g : S \to \mathbb{R}$, $f, g \in I(S)$. Then $\alpha f + \beta g, fg, f + g, |f| \in I(S)$, and also $\frac{f}{g}$ provided g is non-null in S.
3. Let $f(\vec{x}) \leq g(\vec{x})$ in S. Then, $\int_S f \leq \int_S g$.
4. Let $f : S \to \mathbb{R}$, $f \in I(S)$. Then, $\left|\int_S f\right| \leq \int_S |f|$.

For non-negative functions, the generic integrability conditions can be particularized in a useful way.

Definition C.1 *Let $f : S \to \mathbb{R}, f \in C(S)$, being $S \subseteq \mathbb{R}^n$ an arbitrary set. If there exists an increasing succession $\{S_i\}$ formed of compact and measurable subsets of S, such that $S_1 \cup \ldots \cup S_i \cup \ldots = S$, we say that such set S is σ-compact. In particular, any open set is σ-compact.*

Theorem C.2 *Let $S \in \mathbb{R}^n$ a σ-compact set, $f : S \to \mathbb{R}$ a function and let $\vec{x} = (x_1, \ldots, x_n)$. If $f(\vec{x}) \geq 0$, $\forall\vec{x} \in S$, then the integral $\int_S f$ exists and:*

$$\int_S f = sup \left\{ \int_{S'} f \;\mid\; S' \subset S, S' \text{ compact and measurable set} \right\}$$

This integral is then convergent or divergent depending only on the supremum being finite or infinite, respectively.

Theorem C.3 *In the same hypotheses of Theorem (C.2), a necessary and sufficient condition for the integral $\int_S f$ (guaranteed to exist by previous Theorem) to be convergent, is to find a succession $\{S_i\}$ making S an σ-compact set, such that the numeric succession $\{\int_{S_i} f\}$ is upper-bounded —that is to say, it has a finite limit, which will correspond to $\int_S f$.*

We note that S can be any σ-compact set. In particular, the entire set $\mathbb{R}^n$ is σ-compact in $\mathbb{R}^n$. In these conditions, the previously stated Propositions can be restated for S. We summarize the most relevant:

Proposition C.11 *Let $f, g : S \to \mathbb{R}, S \subseteq \mathbb{R}^n$, being S σ-compact in $\mathbb{R}^n$. Provided $\int_S f$, $\int_S g$ exist and are convergent:*

1. *Let $h = \alpha f + \beta g$. Then, $\int_S h$ exists and is convergent.*
2. *Let $f(\vec{x}) \leq g(\vec{x})$. Then, $\int_S f \leq \int_S g$.*
3. $\left|\int_S f\right| \leq \int_S |f|$.

As an example of use of Theorem (C.3), we develop the following integral of a non-negative function:

$$I = \int\int_{\mathbb{R}^2} e^{-x^2-y^2} dxdy$$

Let $B_n = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq n\}$. These are balls centered at $(0,0)$ and with radius $\sqrt{n}$. Note that $B_{n'} \subset B_n$, for $n' < n$, and $B_1 \cup \ldots \cup B_i \cup \ldots = \mathbb{R}^2$. Thus, $I = \lim_{n\to\infty} I_n$, with:

$$I_n = \int\int_{B_n} e^{-x^2-y^2} dxdy$$

Changing to polar coordinates: $x = \rho cos\theta, y = \rho sin\theta$, we have:

$$\begin{aligned} I_n &= \int_0^{2\pi} d\theta \int_0^n e^{-\rho^2} \rho d\rho \\ &= \int_0^{2\pi} \frac{1}{2}(1 - ne^{-n^2}) d\theta = \pi(1 - ne^{-n^2}) \end{aligned} \tag{C.8}$$

and $\lim_{n\to\infty} I_n = \pi = I$.

Proposition C.12 *Let $h(x,y) = f(\sqrt{x^2+y^2})$, with $f \in I(\mathbb{R})$. Then $h \in I(\mathbb{R}^2)$.*

Proof. By application of Theorem (C.3). Let $B_n = \{(x,y) \in \mathbb{R}^2 \mid x^2+y^2 \leq n\}$, and F a primitive of f, where $\int_{\mathbb{R}} f(x)dx = l_x \in \mathbb{R}$. Thus, $I = \lim_{n\to\infty} I_n$, with:

$$I_n = \int\int_{B_n} f(\sqrt{x^2+y^2})dxdy$$

Changing to polar coordinates: $x = \rho cos\theta, y = \rho sin\theta$, we have:

$$\begin{aligned} I_n &= \int_0^{2\pi} d\theta \int_0^n f(\rho)d\rho = \int_0^{2\pi} d\theta\{F(\rho)|_0^n\} = \int_0^{2\pi} \{F(n) - F(0)\}d\theta \\ &= \{F(n) - F(0)\} \int_0^{2\pi} 1d\theta = 2\pi\{F(n) - F(0)\} \end{aligned} \tag{C.9}$$

and $\lim_{n\to\infty} 2\pi\{F(n) - F(0)\} = 2\pi\{l_x - F(0)\} \in \mathbb{R}$.
This result can be extended to any number of coordinates.

Proposition C.13 *Let $f \in I(\mathbb{R})$, a positive monotonically decreasing function in $[0,+\infty)$. Then, given $h(\cdot) = f(\|\cdot\|_q)$, with $q \geq 1 \in \mathbb{R}$, $h \in I(\mathbb{R}^n)$.*

Proof. We make use of the equivalence of all norms in $\mathbb{R}^n$ –Proposition (A.2)– and the fact that the result holds for $q = 2$ –Proposition (C.12). Taking $b = 2$ in the Proposition, we have that:

$$\forall q \geq 1 \in \mathbb{R},\ \exists c_1, c_2 > 0 :\ c_1\|\vec{x}\|_q \leq \|\vec{x}\|_2 \leq c_2\|\vec{x}\|_q,\ \forall x \in \mathbb{R}^n$$

where c_1, c_2 only depend on q. Since f is monotonically decreasing in $[0,+\infty)$, it holds:

$$f(c_2\|\vec{x}\|_q) \leq f(\|\vec{x}\|_2), \qquad \forall x \in \mathbb{R}^n$$

Being $f(c_2\|\vec{x}\|_q)$ upper-bounded by an integrable function, it is integrable, by Proposition (C.3); note that the factor c_2 does not affect integrability.

Appendix D

Other Topics

Chebyshev inequality. Let X be a random variable with expected value μ and standard deviation σ. Then, $\forall t > 0$,

$$\Pr(|X - \mu| > t\sigma) \leq \frac{1}{t^2}$$

Jensen's theorem. Let I, J be two real intervals, and $f, g : I \to J$ two strictly increasing and continuous functions. Then, the following assertion holds :

$$f^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} f(x_i)\right) = g^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right)$$

if and only if there exist $\alpha, \beta \in \mathbb{R}$ such that $f(z) = \alpha g(z) + \beta,\ z \in I$. The proof of the assertion in page (319) is obtained by setting $f(z) = z^q, q \geq 1$, which is a strictly increasing and continuous function.

Proof of assertion in page (110). Let $D = \{x\}$ a real data set and $D' = \{x'\}$ the new (normalized) data set, with $x' = f(x)$ and $f(x) = ax + b$, a, b real constants obtained from D, $a > 0$. Assuming $D = [m, M]$, the similarity between any two $x, y \in D$ is given by (4.52):

$$s(x, y) = 1 - \frac{|x - y|}{M - m}$$

On the other hand, we have:

$1 - \frac{|x'-y'|}{M'-m'} = 1 - \frac{|f(x)-f(y)|}{f(M)-f(m)} = 1 - \frac{|ax+b-ay-b|}{aM+b-am-b} = 1 - \frac{|ax-ay|}{aM-am} = 1 - \frac{a|x-y|}{a(M-m)} = 1 - \frac{|x-y|}{M-m}$

Therefore $s(x, y) = s(x', y')$. This simple linear normalization scheme includes most of the commonly found methods: to $D' = [0, 1]$, by setting $a = \frac{1}{M-m}, b = \frac{m}{M-m}$; and to zero mean, unit standard deviation, by setting $a = \frac{1}{\sigma_x}, b = -\frac{\mu_x}{\sigma_x}$.

An analogous proof can be derived for ordinal types. For nominal types, the measure (4.45) is clearly not affected by any normalization. As for the fuzzy ones, their values are obtained from the continuous ones in such a way that scale is unimportant.