Abstract

This paper introduces a new method which employs the concept of “Orientation Vectors” to train a feed forward neural network. It is shown that this method is suitable for problems where large dimensions are involved and the clusters are characteristically sparse. For such cases, the new method is not NP hard as the problem size increases. We ‘derive’ the present technique by starting from Kolmogrov’s method and then relax some of the stringent conditions. It is shown that for most classification problems three layers are sufficient and the number of processing elements in the first layer depends on the number of clusters in the feature space. This paper explicitly demonstrates that for large dimension space as the number of clusters increase from N to N+dN the number of processing elements in the first layer only increases by d(logN), and as the number of classes increase, the processing elements increase only proportionately, thus demonstrating that the method is not NP hard with increase in problem size. Many examples have been explicitly solved and it has been demonstrated through them that the method of Orientation Vectors requires much less computational effort than Radial Basis Function methods and other techniques wherein distance computations are required, in fact the present
method increases logarithmically with problem size compared to the Radial Basis Function method and the other methods which depend on distance computations e.g statistical methods where probabilistic distances are calculated. A practical method of applying the concept of Occum’s razor to choose between two architectures which solve the same classification problem has been described. The ramifications of the above findings on the field of Deep Learning have also been briefly investigated and we have found that it directly leads to the existence of certain types of NN architectures which can be used as a “mapping engine”, which has the property of “invertibility”, thus improving the prospect of their deployment for solving problems involving Deep Learning and hierarchical classification. The latter possibility has a lot of future scope in the areas of machine learning and cloud computing.

References

1. A. N. Kolmogorov: On the representation of continuous functions of many variables by superpositions of continuous functions of one variable and addition. Doklay Akademii Nauk USSR, 14(5):953 - 956, (1957). Translated in: Amer. Math Soc. Transl. 28, 55-59 (1963).
2. G.G. Lorentz: Approximation of functions. Athena Series, Selected Topics in Mathematics. Holt, Rinehart, Winston, Inc., New York (1966).
3. G. Lorentz, M. Golitschek, and Y. Makovoz: Constructive Approximation: Advanced Problems. Springer (1996).
4. D. A. Sprecher: On the structure of continuous functions of several variables. Transactions Amer. Math. Soc, 115(3):340 - 355 (1965).
5. D. A. Sprecher: An improvement in the superposition theorem of Kolmogorov. Journal of Mathematical Analysis and Applications, 38:208 - 213 (1972).
6. Bunpei Irie and Sei Miyake: Capabilities of Three-layered Perceptrons, IEEE International Conference on Neural Networks , pp641-648, Vol 1.24-27-July, (1988).
7. D. A. Sprecher: A numerical implementation of Kolmogorov’s superpositions. Neural Networks, 9(5):765 - 772 (1996).
8. D. A. Sprecher: A numerical implementation of Kolmogorov’s superpositions II. Neural Networks, 10(3):447 - 457 (1997).
9. Paul C. Kainen and Vera Kurkova: An Integral Upper Bound for Neural Network Approximation, Neural Computation, 21, 2970-2989 (2009).
10. Jrgen Braun, Michael Griebel: On a Constructive Proof of Kolmogorov’s Superposition Theorem, Constructive Approximation, Volume 30, Issue 3, pp 653-675 (2009).
11. David Sprecher: On computational algorithms for real-valued continuous functions of several variables, Neural Networks 59, 16-22(2014).
12. Vasco Brattka : From Hilbert’s 13th Problem to the theory of neural networks: constructive aspects of Kolmogorov’s Superposition Theorem, Kolmogorov’s Heritage in Mathematics, pp 273-274, Springer (2007).
13. Hecht-Nielsen, R.: Neurocomputing. Addison-Wesley, Reading (1990).
14. Hecht-Nielsen, R.: Kolmogorov’s mapping neural network existence theorem. In Proceedings IEEE International Conference On Neural Networks, volume II, pages 11-13, New York,IEEE Press (1987).
15. Rumelhart, D. E., Hinton, G. E., and R. J. Williams: Learning representations by
Some Theorems for Feed Forward Neural Networks

back-propagating errors. Nature, 323, 533–536 (1986).
17. Yoshua Bengio: Learning Deep Architectures for AI. Foundations and Trends in Machine Learning: Vol. 2: No. 1, pp 1-127 (2009).
18. J. Schmidhuber: Deep Learning in Neural Networks: An Overview. 75 pages, http://arxiv.org/abs/1404.7828,(2014).
19. D. George and J.C. Hawkins: Trainable hierarchical memory system and method, January 24 2012. URL https://www.google.com patents US8103603. US Patent 8,103,603.
20. Corinna Cortes and Vladimir Vapnik: Support-Vector Networks, Machine Learning, 20, 273-297 (1995).
21. William B. Johnson and Joram Lindenstrauss: Extensions of Lipschitz mappings on to a Hilbert Space, Contemporary Mathematics, 26, pp 189-206 (1984)
22. Sanjoy Dasgupta and Anupam Gupta: An Elementary Proof of a Theorem of Johnson and Lindenstrauss, Random Struct.Alg., 22: 60–65, 2002 Wiley Periodicals.
23. G.E. Hinton and R.R. Salkhutdinov: Reducing the Dimensionality with Neural Networks, v 313, Science, pp 504- 507 (2006)
24. Dasika Ratna Deepthi and K. Eswaran: A mirroring theorem and its application to a new method of unsupervised hierarchical pattern classification. International Journal of Computer Science and Information Security, pp. 016-025, vol 6, 2009.
25. Dasika Ratna Deepthi and K. Eswaran: Pattern recognition and memory mapping using mirroring neural networks. International Journal of Computer Applications 1(12):88-96, February 2010.
26. K Eswaran: Numenta lightning talk on dimension reduction and unsupervised learning. In Numenta HTM Workshop, Jun, pages 23-24, 2008a.
27. R.P. Lippmann: An introduction to computing with neural nets, IEEE,ASSP magazine, pp 4-22 (1987)
28. Ralph P. Boland and Jorge Urrutia: Separating Collection of points in Euclidean Spaces, Information Processing Letters, vol 53, no.4, pp. 177-183 (1995)
29. K.Eswaran:A system and method of classification etc. Patents filed IPO No.(a) 1256/CHE July 2006 and (b) 2669/CHE June 2015
30. K.Eswaran: A non iterative method of separation of points by planes and its application, Sent for publ. (2015) http://arxiv.org/abs/1509.08742
31. K.Eswaran and C. Chaitanya: Cloud based unsupervised learning architecture, Recent researches In AI and and Knowledge Engg. Data Bases, WSEAS Conf. at Cambridge Univ. U.K. ISBN 978- 960-474-273-8, 2011.

Index Terms
Computer Science Artificial Intelligence

Keywords
Neural Networks, Neural Architecture, Deep Learning, Orientation Vector, Kolmogorov
