

## References

- Francesca Albertini, Eduardo D. Sontag, and Vincent Maillot. Uniqueness of weights for neural networks. In *Artificial Neural Networks for Speech and Vision*, pages 113–125. Chapman & Hall, London, **1993**. Proceedings of a workshop held at Rutgers University in 1992. Access via [Eduardo D. Sontag](#). Cited on page 2.
- Shun-ichi Amari, Hyeyoung Park, and Tomoko Ozeki. Singularities affect dynamics of learning in neuromanifolds. *Neural Computation*, 18(5):1007–1065, **2006**. Access via [Crossref](#). Cited on page 2.
- Shun-ichi Amari, Tomoko Ozeki, Ryo Karakida, Yuki Yoshida, and Masato Okada. Dynamics of learning in MLP: Natural gradient and singularity revisited. *Neural Computation*, 30(1):1–33, **2018**. Access via [Crossref](#). Cited on page 2.
- Anonymous. Functional equivalence and path connectivity of reducible hyperbolic tangent networks. **2023**. Anonymised article included with supplementary material. Cited on pages 1, 2, 3, 8, and 28.
- Caglar Aytekin, Francesco Cricri, and Emre Aksu. Compressibility loss for neural network weights. **2019**. Preprint [arXiv:1905.01044](#) [cs.LG]. Cited on page 3.
- Piotr Berman, Alex D. Scott, and Marek Karpinski. Approximation hardness and satisfiability of bounded occurrence instances of SAT. Technical Report IHES/M/03/25, Institut des Hautes Études Scientifiques [Institute of Advanced Scientific Studies], **2003**. Access via [CERN](#). Cited on page 18.
- Cristian Buciluă, Rich Caruana, and Alexandru Niculescu-Mizil. Model compression. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 535–541. ACM, **2006**. Access via [Crossref](#). Cited on page 1.
- Stephen Casper, Xavier Boix, Vanessa D’Amario, Ling Guo, Martin Schrimpf, Kasper Vinken, and Gabriel Kreiman. Frivolous units: Wider networks are not really that wide. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence*, volume 8, pages 6921–6929. AAAI Press, **2021**. Access via [Crossref](#). Cited on page 3.
- Márcia R. Cerioli, Luerbio Faria, Talita O. Ferreira, and Fábio Protti. On minimum clique partition and maximum independent set on unit disk graphs and penny graphs: Complexity and approximation. *Electronic Notes in Discrete Mathematics*, 18:73–79, **2004**. Access via [Crossref](#). Cited on pages 6, 15, 18, and 25.
- Márcia R. Cerioli, Luerbio Faria, Talita O. Ferreira, and Fábio Protti. A note on maximum independent sets and minimum clique partitions in unit disk graphs and penny graphs: Complexity and approximation. *RAIRO: Theoretical Informatics and Applications*, 45(3):331–346, **2011**. Access via [Crossref](#). Cited on pages 6, 15, 18, and 25.
- An Mei Chen, Haw-minn Lu, and Robert Hecht-Nielsen. On the geometry of feedforward neural network error surfaces. *Neural Computation*, 5(6):910–927, **1993**. Access via [Crossref](#). Cited on pages 1 and 27.
- Yu Cheng, Duo Wang, Pan Zhou, and Tao Zhang. Model compression and acceleration for deep neural networks: The principles, progress, and challenges. *IEEE Signal Processing Magazine*, 35(1):126–136, **2018**. Access via [Crossref](#). Cited on pages 3 and 10.
- Yu Cheng, Duo Wang, Pan Zhou, and Tao Zhang. A survey of model compression and acceleration for deep neural networks. **2020**. Preprint [arXiv:1710.09282v9](#) [cs.LG]. Updated version of [Cheng et al. \(2018\)](#). Cited on page 3.
- Tejalal Choudhary, Vipul Mishra, Anurag Goswami, and Jagannathan Sarangapani. A comprehensive survey on model compression and acceleration. *Artificial Intelligence Review*, 53(7):5113–5155, **2020**. Access via [Crossref](#). Cited on page 3.
- Brent N. Clark, Charles J. Colbourn, and David S. Johnson. Unit disk graphs. *Discrete Mathematics*, 86(1-3):165–177, **1990**. Access via [Crossref](#). Cited on page 15.

- Stephen A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, pages 151–158. ACM, **1971**. Access via [Crossref](#). Cited on pages 3, 6, 17, and 18.
- Florent Cousseau, Tomoko Ozeki, and Shun-ichi Amari. Dynamics of learning in multilayer perceptrons near singularities. *IEEE Transactions on Neural Networks*, 19(8):1313–1328, **2008**. Access via [Crossref](#). Cited on page 2.
- Amit Daniely, Nati Linial, and Michael Saks. Clustering is difficult only when it does not matter. **2012**. Preprint [arXiv:1205.4891](#) [cs.LG]. Cited on page 9.
- Charles Fefferman. Reconstructing a neural net from its output. *Revista Matemática Iberoamericana*, 10(3):507–555, **1994**. Access via [Crossref](#). Cited on pages 1 and 2.
- Charles Fefferman and Scott Markel. Recovering a feed-forward net from its output. In *Advances in Neural Information Processing Systems 6*, pages 335–342. Morgan Kaufmann, **1993**. Access via [NeurIPS](#). Cited on page 2.
- Kenji Fukumizu. A regularity condition of the information matrix of a multilayer perceptron network. *Neural Networks*, 9(5):871–879, **1996**. Access via [Crossref](#). Cited on pages 1 and 28.
- Kenji Fukumizu and Shun-ichi Amari. Local minima and plateaus in hierarchical structures of multilayer perceptrons. *Neural Networks*, 13(3):317–327, **2000**. Access via [Crossref](#). Cited on page 2.
- Kenji Fukumizu, Shoichiro Yamaguchi, Yoh-ichi Mototake, and Mirai Tanaka. Semi-flat minima and saddle points by embedding neural networks to overparameterization. In *Advances in Neural Information Processing Systems 32*, pages 13868–13876. Curran Associates, **2019**. Access via [NeurIPS](#). Cited on page 2.
- Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, **1979**. Cited on page 3.
- S. L. Hakimi. Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research*, 12(3):450–459, **1964**. Access via [Crossref](#). Cited on page 15.
- S. L. Hakimi. Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations Research*, 13(3):462–475, **1965**. Access via [Crossref](#). Cited on page 15.
- Geoffrey E. Hinton and Drew van Camp. Keeping the neural networks simple by minimizing the description length of the weights. In *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, pages 5–13. ACM, **1993**. Access via [Crossref](#). Cited on page 3.
- Geoffrey E. Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. Presented at Twenty-eighth Conference on Neural Information Processing Systems, Deep Learning workshop, **2014**. Preprint [arXiv:1503.02531](#) [stat.ML]. Cited on page 1.
- Klaus Jansen and Haiko Müller. The minimum broadcast time problem for several processor networks. *Theoretical Computer Science*, 147(1-2):69–85, **1995**. Access via [Crossref](#). Cited on page 18.
- O. Kariv and S. L. Hakimi. An algorithmic approach to network location problems. I: the  $p$ -centers. *SIAM Journal on Applied Mathematics*, 37(3):513–538, **1979**. Access via [Crossref](#). Cited on page 15.
- Richard M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pages 85–103. Springer, **1972**. Access via [Crossref](#). Cited on page 15.
- Věra Kůrková and Paul C. Kainen. Functionally equivalent feedforward neural networks. *Neural Computation*, 6(3):543–558, **1994**. Access via [Crossref](#). Cited on page 2.
- Leonid A. Levin. Universal sequential search problems. *Problemy Peredachi Informatsii [Problems of Information Transmission]*, 9(3):115–116, **1973**. In Russian. Translated into English in [Trakhtenbrot \(1984\)](#). Cited on pages 3, 6, and 17.

David Lichtenstein. Planar formulae and their uses. *SIAM Journal on Computing*, 11(2):329–343, **1982**. Access via [Crossref](#). Cited on page 18.

Yanpei Liu, Aurora Morgana, and Bruno Simeone. A linear algorithm for 2-bend embeddings of planar graphs in the two-dimensional grid. *Discrete Applied Mathematics*, 81(1-3):69–91, **1998**. Access via [Crossref](#). Cited on pages 19 and 24.

Meena Mahajan, Prajakta Nimbhorkar, and Kasturi Varadarajan. The planar  $k$ -means problem is NP-hard. *Theoretical Computer Science*, 442:13–21, **2012**. Access via [Crossref](#). Cited on pages 6 and 15.

Nimrod Megiddo and Kenneth J. Supowit. On the complexity of some common geometric location problems. *SIAM Journal on Computing*, 13(1):182–196, **1984**. Access via [Crossref](#). Cited on page 15.

Mary Phuong and Christoph H. Lampert. Functional vs. parametric equivalence of ReLU networks. In *8th International Conference on Learning Representations*. OpenReview, **2020**. Access via [OpenReview](#). Cited on pages 1 and 2.

R. Piziak and P. L. Odell. Full rank factorization of matrices. *Mathematics Magazine*, 72(3):193–201, **1999**. Access via [Crossref](#). Cited on page 4.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. Presented at the Fifth Workshop on Energy Efficient Machine Learning and Cognitive Computing, **2019**. Preprint [arXiv:1910.01108](#) [cs.CL]. Cited on page 1.

Berfin Şimşek, François Ged, Arthur Jacot, Francesco Spadaro, Clément Hongler, Wulfram Gerstner, and Johanni Brea. Geometry of the loss landscape in overparameterized neural networks: Symmetries and invariances. In *Proceedings of the 38th International Conference on Machine Learning*, pages 9722–9732. PMLR, **2021**. Access via [PMLR](#). Cited on page 2.

Kenneth J. Supowit. *Topics in Computational Geometry*. Ph.D. thesis, University of Illinois at Urbana-Champaign, **1981**. Access via [ProQuest](#). Cited on pages 6 and 15.

Héctor J. Sussmann. Uniqueness of the weights for minimal feedforward nets with a given input-output map. *Neural Networks*, 5(4):589–593, **1992**. Access via [Crossref](#). Cited on pages 1, 2, 3, 4, 27, and 28.

Taiji Suzuki, Hiroshi Abe, Tomoya Murata, Shingo Horiuchi, Kotaro Ito, Tokuma Wachi, So Hirai, Masatoshi Yukishima, and Tomoaki Nishimura. Spectral pruning: Compressing deep neural networks via spectral analysis and its generalization error. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, pages 2839–2846. IJCAI, **2020a**. Access via [Crossref](#). Cited on page 3.

Taiji Suzuki, Hiroshi Abe, and Tomoaki Nishimura. Compression based bound for non-compressed network: Unified generalization error analysis of large compressible deep neural network. In *8th International Conference on Learning Representations*. OpenReview, **2020b**. Access via [OpenReview](#). Cited on page 3.

Craig A. Tovey. A simplified NP-complete satisfiability problem. *Discrete Applied Mathematics*, 8(1):85–89, **1984**. Access via [Crossref](#). Cited on page 18.

Boris A. Trakhtenbrot. A survey of Russian approaches to *perebor* (brute-force search) algorithms. *Annals of the History of Computing*, 6(4):384–400, **1984**. Access via [Crossref](#). Cited on page 11.

Leslie G. Valiant. Universality considerations in VLSI circuits. *IEEE Transactions on Computers*, 100(2):135–140, **1981**. Access via [Crossref](#). Cited on pages 6 and 19.

Verner Vlačić and Helmut Bölcskei. Affine symmetries and neural network identifiability. *Advances in Mathematics*, 376:107485, **2021**. Access via [Crossref](#). Cited on page 2.

Verner Vlačić and Helmut Bölcskei. Neural network identifiability for a family of sigmoidal nonlinearities. *Constructive Approximation*, 55(1):173–224, **2022**. Access via [Crossref](#). Cited on page 2.

- 528 Sumio Watanabe. *Algebraic Geometry and Statistical Learning Theory*. Cambridge University  
529 Press, **2009**. Cited on page 1.
- 530 Haikun Wei, Jun Zhang, Florent Cousseau, Tomoko Ozeki, and Shun-ichi Amari. Dynamics of learn-  
531 ing near singularities in layered networks. *Neural Computation*, 20(3):813–843, **2008**. Access  
532 via [Crossref](#). Cited on page 2.
- 533 Susan Wei, Daniel Murfet, Mingming Gong, Hui Li, Jesse Gell-Redman, and Thomas Quella. Deep  
534 learning is singular, and that’s good. *IEEE Transactions on Neural Networks and Learning Sys-*  
535 *tems*, **2022**. Access via [Crossref](#). To appear in an upcoming volume. Cited on page 1.
- 536 Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding  
537 deep learning requires rethinking generalization. In *5th International Conference on Learning*  
538 *Representations*. OpenReview, **2017**. Access via [OpenReview](#). Cited on pages 1 and 13.
- 539 Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding  
540 deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–  
541 115, **2021**. Access via [Crossref](#). Republication of [Zhang et al. \(2017\)](#). Cited on page 1.