ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

Statistical Physics of Learning and Inference

M. Biehl[1] and N. Caticha[2] and M. Opper[3] and T. Villmann[4] *

1- Univ. of Groningen, Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, Nijenborgh 9, NL-9747 AG Groningen, The Netherlands

2- Instituto de Física, Universidade de São Paulo Caixa Postal 66318, 05315-970, São Paulo, SP, Brazil

3- Technical University Berlin, Department of Electrical Engineering and Computer Science, D-10587 Berlin, Germany

4- University of Applied Sciences Mittweida, Computational Intelligence Group, Technikumplatz 17, D-09648 Mittweida, Germany

Abstract. The exchange of ideas between statistical physics and computer science has been very fruitful and is currently gaining momentum as a consequence of the revived interest in neural networks, machine learning and inference in general.

Statistical physics methods complement other approaches to the theoretical understanding of machine learning processes and inference in stochastic modeling. They facilitate, for instance, the study of dynamical and equilibrium properties of randomized training processes in model situations. At the same time, the approach inspires novel and efficient algorithms and facilitates interdisciplinary applications in a variety of scientific and technical disciplines.

1 Introduction

The regained popularity of machine learning in general and neural networks in particular [1–3] can be associated with at least two major trends: On the one hand, the ever-increasing amount of training data acquired in various domains facilitates the training of very powerful systems, *deep* neural networks being only the most prominent example [4–6]. On the other hand, the computational power needed for the data driven adaptation and optimization of such systems has become available quite broadly.

Both developments have made it possible to realize and deploy in practice several concepts that had been devised previously - some of them even decades ago, see [4–6] for examples and further references. In addition, and equally importantly, efficient computational techniques have been put forward, such as the use of pre-trained networks or sophisticated regularization techniques like drop-out or similar schemes [4–7]. Moreover, important modifications and conceptual extensions of the systems in use have contributed to the achieved progress significantly. With respect to the example of deep networks, this concerns, for instance, weight sharing in convolutional neural networks or the use of specific activation functions [4–6, 8].

*The authors thank the organizers of the ESANN 2019 conference for integrating this special session into the program. We are grateful to all authors for their contribution and the anonymous reviewers for their support.

Recently, several authors have argued that the level of theoretical understanding does not yet parallel the impressive practical success of machine learning techniques and that many heuristic and pragmatic concepts are not understood to a satisfactory degree, see for instance [9–13] in the context of deep learning.

While the partial lack of a solid theoretical background does not belittle the practical importance and success of the methods, it is certainly worthwhile to strengthen their theoretical foundations. Obviously, the optimization of existing tools and the development of novel concepts would benefit greatly from a deeper understanding of relevant phenomena for the design and training of adaptive systems. This concerns, for instance, their mathematical and statistical foundations, the dynamics of training dynamics and convergence behavior or the expected generalization ability.

2 Statistical physics and learning

Statistical mechanics based methods have been applied in several areas outside the traditional realms of physics. For instance, analytical and computational techniques from the statistical physics of disordered sytems have been applied in various areas of computer science and statistics, including inference, machine learning and optimization.

The wide-spread availability of powerful computational resources has facilitated the diffusion of these, often very involved, methods into neighboring fields. A superb example is the efficient use of Markov Chain Monte Carlo methods, which were developed to attack problems in Statistical mechanics in the middle of the last century [14]. Analytical methods, developed for the analysis of *disordered systems* with many degrees of freedom, constitute another important example [15]. They have been applied in a variety of problems on the basis of mathematical analogies, which appear to be purely formal, at a glance.

In fact it was such an analogy, pointed out by J. Hopfield [16], which triggered considerable interest in neural networks and similar systems within the physics community, originally: the conceptual similarity of simple models for dynamical neural networks and models of disordered magnetic materials [15]. Initially equilibrium and dynamical effects in so-called attractor neural networks such as the Little-Hopfield model had been addressed [17]. Later it was realized that the same or very similar theoretical concepts can be applied to analyse the weight space of neural networks. Inspired by the groundbreaking work of E. Grander [18, 19], a large variety of machine learning scenarios have been investigated, including the supervised training of feedforward neural networks and the unsupervised analysis of structured data sets, see [20–23] for reviews. In turn, the study of machine learning processes also triggered the development and better understanding of statistical physics tools and theories.

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

3 Current research questions and concrete problems

This special session brings together researchers who develop or apply statistical physics related methods in the context of machine learning, data analysis and inference.

The aim is to re-establish and intensify the fruitful interaction between statistical physics related research and the machine learning community. The organizers are convinced that statistical physics based approaches will be instrumental in obtaining the urgently needed insights for the design and further improvement of efficient machine learning techniques and algorithms.

Obviously, the special session and this tutorial paper can only address a small subset of the many challenges and research topics which are relevant in this area. Tools and concepts applied in this broad context cover a wide range of concepts and areas: information theory, the mathematical analysis of stochastic differential equations, the statistical mechanics of disordered systems, the theory of phase transitions, mean field theory, Monte Carlo simulations, variational calculus, renormalization group and a variety of other analytical and computational methods [7, 15, 24–27, 27–29].

Specific topics and questions of current interest include, but are by far not limited to the following list. Where available, we provide references to tutorial papers of relevant special sessions at recent ESANN conferences.

- The relation of statistical mechanics to information theoretical methods and other approaches to computational learning theory [25, 30]

  Information processing and statistical information theory are widely used in machine learning concepts. In particular the Boltzmann-Gibbs statistics is an essential tool in adaptive processes [25, 31–33]. The measuring of mutual information and the comparison of data in terms of divergences based on respective entropy concepts stimulated new approches in machine learning data analysis [34, 35]. For example, Tsallis entropy, known from non-extensive statistical physics [36,37], can be used to improve learning in decision trees [38] and kernel based learning [39]. Recent approaches relate the Tsallis entropy also to reinforcement and causal imitation learning [40, 41].

- Learning in deep layered networks and other complex architectures [42]

  Many tools and analytical methods have been developed and applied successfully to the analysis of relatively simple, mostly shallow neural networks [7, 20–22]. Currently, their application and significant conceptual extension is gaining momentum (pun intended) in the context of deep learning and other learning paradigms, see [7, 24, 43–47] for recent examples of these on-going efforts.

- Emergent behavior in societies of interacting agents

  Simple models of societies have been used to show that some social science problems are, at least in principle, not outside the reach of mathematical

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

modeling, see [48, 49] for examples and further references. To go beyond the analysis of simple two-state agents it seems reasonable to add more ingredients in the agent's model. These could include learning from the interaction with other agents and the capability of analyzing issues that can only be represented in multidimensional spaces. The modeling of societies of neural networks presents the type of problem that can be dealt with the methods and ideas of statistical mechanics.

- Symmetry breaking and transient dynamics in training processes
  Symmetry breaking phase transitions in neural networks and other learning systems have been a topic of great interest, see [7, 20–22, 51–53] for many examples and references. Their counterpart in off-equilibrium online learning scenarios are quasi-stationary plateau states in the learning curves [23, 50, 54–56]. The existence of these plateaux is in general a sign of symmetries that can often be only broken after the computational effort of including more data. Methods to analyse, identify, and possibly to partially alleviate these problems in simple feedforward networks have been presented in the context of statistical mechanics, see [50, 54–56] for some of the many examples. The problem of saddle-point plateau states has recently re-gained attention within the deep learning community, see e.g. [44].

- Equilibrium phenomena in vector quantization

  Phase transitions and equilibrium phenomena were intensively studied also in the context of self-organizing maps for unsupervised vector quantization and topographic vector quantization [57, 58]. Particularly, phase transitions in the context of violations of topology preservation in self-organizing maps (SOM) in dependence on the range of interacting neurons in the neural lattices were investigated applying Fokker- Planck-approaches [59, 60]. Moreover, energy function for those networks were considered in [61, 62] and [63]. Ordering processes and asymptotic behavior of SOMs were studied in terms of stationary states in particle systems of interacting particles delivering results for [61, 64, 65].

- Theoretical approaches to consciousness

  No agreement on what consciousness is seems to be around the corner [66]. However, some measures of casual relationships in complex systems, see e.g. [67], have been put forward as possible ways to discuss how to recognize when a certain degree of consciousness can be attributed to a system. Integrated information has been presented in several forms, including versions of Tononi's information integration [68, 69] based on information theory. Since the current state of the theory permits dealing with very few degrees of freedom, methods from the repertoire developed to study neural networks as versions of disordered systems, are a real possibility for advance our understanding in this field.

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

Without going into detail, we only mention some of the further topics of interest and on-going research:

- Design and anlysis of interpretable models and *white-box* systems [70–72]
- Probabilistic inference in stochastic systems and complex networks
- Learning in model space
- Transfer learning and lifelong learning in non-stationary environments [73]
- Complex optimization problems and related algorithmic approaches.

The diversity of methodological approaches inspired by statistical physics leads to a plethora of potential applications. The relevant scientific disciplines and application areas include neurosciences, systems biology and bioinformatics, environmental modelling, social sciences and signal processing, to name just very few examples. Methods borrowed from statistical physics continue to play an important role in the development all of these challenging areas.

4 Contributions to the ESANN 2019 special session on the "Statistical physics of learning and inference"

The three accepted contributions to the special session address a selection of diverse topics, which reflect the relevance of statistical physics ideas and concepts in a variety of areas.

Trust law and ideology in a NN agent model of the US Appellate Courts
In their contribution [74], N. Caticha and F. Alves employ systems of interacting neural networks as mathematica models of judicial panels. The authors investigate the the role of ideological bias, dampening and amplification effects in the decision process.

Noise helps optimization escape from saddle points in the neural dynamics
Synaptic plasticity is in the focus of a contribution by Y. Fang, Z. Yu and F. Chen [75]. The authors investigate the influence of saddle points and the role of noise in learning processes. Mathematical analysis and computer experiments demonstrate how noise can improve the performance of optimization strategies in this context.

On-line learning dynamics of ReLU neural networks using statistical physics techniques
The statistical physics of on-line learning is revisited in a contribution by M. Straat and M. Biehl [76]. They study the training of layered neural networks with rectified linear units (ReLU) from a stream of example data. Emphasis is put on the role of the specific activation function for the occurrance of sub-optimal quasi-stationary plateau states in the learning dynamics.

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

Statistical physics has contributed significantly to the investigation and understanding of relevant phenomena in machine learning and inference, and it continues to do so. We hope that the contributions to this special session on the "Statistical physics of learning and inference" helps to increase attention among active machine learning researchers.

References

[1] J. Hertz, A. Krogh, R.G. Palmer. *Introduction to the theory of neural computation,* Addison-Wesley, 1991.

[2] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* Springer, 2009.

[3] C. Bishop, *Pattern Recognition and Machine Learning,* Cambridge University Press, Cambridge, 2007.

[4] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning,* MIT Press, 2016.

[5] Y. LeCun, Y. Bengio, G. Hinton, Deep Learning, *Nature,* 521: 436-444, 2015.

[6] J. Schmidhuber. Deep Learning in Neural Networks: An Overview, *Neural Networks,* 61: 85-117, 2015.

[7] L. Saitta, A. Giordana, A. Cornuéjols. *Phase Transitions in Machine Learning,* Cambridge University Press, 383 pages, 2011.

[8] J. Rynkiewicz. Asymptotic statistics for multilayer perceptrons with ReLu hidden units. In: M. Verleysen (ed.), *Proc. European Symp. on Artificial Neural Networks (ESANN),* d-side publishing, 6 pages (2018)

[9] G. Marcus. Deep Learning: A Critical Appraisal. Available online: `http://arxiv.org/abs/1801.00631` (last accessed: April 23, 2018)

[10] C. Zhang, S. Bengio, M. Hardt, B. Recht, O. Vinyals. Understanding deep learning requires rethinking generalization. In: *Proc. of the 6th Intl. Conference on Learning Representations ICLR,* 2017.

[11] C.H. Martin and M.W. Mahoney. Rethinking generalization requires revisiting old ideas: statistical mechanics approaches and complex learning behavior. Computing Research Repository CoRR, eprint 1710.09553, 2017. Available online: `http://arxiv.org/abs/1710.09553`

[12] H.W. Lin, M. Tegmark, D. Rolnick. Why does deep and cheap learning work so well? *Journal of Statistical Physics* 168(6): 1223-1247, 2017.

[13] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent. Why does unsupervised pre-training help deep learning? *J. of Machine Learning Research* 11: 625-660, 2010.

[14] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller. Equations of State calculations by fast computing machines. *J. Chem. Phys.* 21: 1087, 1953.

[15] M. Mezard, G. Parisi, M. Virasoro. *Spin Glass Theory and Beyond,* World Scientific, 1986.

[16] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. of the National Academy of Sciences of the USA,* 79 (8): 2554-2558, 1982.

[17] D.J. Amit, H. Gutfreund, H. Sompolinsky. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters,* 55(14): 1530-1533, 1985

[18] E. Gardner. Maximum storage capacity in neural networks. *Europhysics Letters* 4(4): 481-486, 1988.

[19] E. Gardner. The space of interactions in neural network models. *J. of Physics A: Mathematical and General,* 21(1): 257-270, 1988.

[20] A. Engel, C. Van den Broeck. *Statistical Mechanics of Learning,* Cambridge University Press, 342 pages, 2001.

[21] T.L.H. Watkin, A. Rau, M. Biehl. The statistical mechanics of learning a rule. *Reviews of Modern Physics* 65(2): 499-556, 1993.

[22] H.S. Seung, H. Sompolinsky, N. Tishby. Statistical mechanics of learning from examples. *Physical Review A* 45: 6065-6091, 1992.

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

[23] D. Saad. *Online learning in neural networks*, Cambridge University Press, 1999.

[24] S. Cocco, R. Monasson, L. Posani, S. Rosay, J. Tubiana. Statistical physics and representations in real and artificial neural networks. *Physica A: Stat. Mech. and its Applications,* 504, 45-76, 2018.

[25] J.C. Principe. *Information Theoretic Learning,* Springer Information Science and Statistics, 448 pages, 2010.

[26] C.W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences,* Springer, 2004.

[27] M. Opper, D. Saad (editors). *Advanced Mean Field Methods: Theory and Practice.* MIT Press, 2001.

[28] L. Bachschmid-Romano, M. Opper. A statistical physics approach to learning curves for the inverse Ising problem. *J. of Statistical Mechanics: Theory and Experiment,* 2017 (6), 063406, 2017.

[29] G. Parisi. *Statistical Field Theory,* Addison-Wesley, 1988.

[30] T. Villmann, J.C. Principe, A. Cichocki. *Information theory related learning.* In: M. Verleysen, editor, *Proc. of the European Symposium on Artificial Neural Networks (ESANN 2011),* d-side pub. 1-10, 2011.

[31] G. Deco, D. Obradovic. *An Information-Theoretic Approach to Neural Computing.* Springer, 1997.

[32] F. Emmert-Streib, M. Dehmer. *Information Theory and Statistical Learning.* Springer Science and Business Media, 2009.

[33] D. Mackay. *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, 2003.

[34] A. Kraskov, H. Stögbauer, P. Grassberger. Estimating mutual information. *Physical Review E* 69(6):66–138, 2004.

[35] T. Villmann, S. Haase. Divergence based vector quantization. *Neural Computation* 23: 1343-1392, 2011.

[36] C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* 52: 479–487, 1988.

[37] C. Tsallis. *Introduction to nonextensive statistical mechanics : approaching a complex world.* Springer, 2009.

[38] T. Maszczyk, W. Duch. Comparison of Shannon, Rényi and Tsallis Entropy used in Decision Trees. In: L. Rutkowski, R. Tadeusiewicz, L. Zadeh, J. Zurada, editors, *Artificial Intelligence and Soft Computing - Proc. of the 9th International Conference Zakopane*, 643-651, 2008.

[39] D. Ghoshdastidar, A. Adsul, A. Dukkipati. Learning With Jensen-Tsallis Kernels. *IEEE Trans Neural Networks and Learning Systems* 10:2108–2119, 2016.

[40] K. Lee, S. Kim, S. Lim, S. Choi, S. Oh. Tsallis Reinforcement Learning: A Unified Framework for Maximum Entropy Reinforcement Learning. *arXiv:1902.00137v2,* 2019.

[41] K. Lee, S. Choi, S. Oh. Maximum Causal Tsallis Entropy Imitation Learning. *arXiv:1805.08336v2,* 2018.

[42] P. Angelov, A. Sperduti. *Challenges in Deep Learning.* In: M. Verleysen, editor, *Proc. of the European Symposium on Artificial Neural Networks (ESANN 2016),*i6doc.com, 489-495, 2016.

[43] J. Kadmon, H. Sompolinsky. Optimal Architectures in a Solvable Model of Deep Networks. In: D.D. Lee, M. Sugiyama, U.V. Luxburg, I. Guyon, R. Garnett (editors), *Advances in Neural Information Processing Systems (NIPS 29)*, Curran Associates Inc., 4781-4789, 2016.

[44] Y. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, Y. Bengio. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In: Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger (editors), *Advances in Neural Information Processing Systems (NIPS 27)*, Curran Associates Inc., 2933-2941, 2014.

[45] M. Pankaj, A.H. Lang, D. Schwab. An exact mapping from the Variational Renormalization Group to Deep Learning. *arXiv repository [stat.ML],* eprint 1410.3831v1, 2014. Available online: `https://arxiv.org/abs/1410.3831v1`

[46] A.M. Saxe, J.L. McClelland, S. Ganguli. Exact solutions to the non-linear dynamics of

learning in deep linear neural networks. In: Y. Bengio, Y. Le Cun (eds.), *Proc. Intl. Conf. on Learning Representations (ICLR)*, 2014.

[47] J. Sohl-Dickstein et al. Deep unsupervised learning using non-equilibrium thermodynamics. *Proc. of Machine Learning Research* 37, 2256-2265, 2016.

[48] N. Caticha, R. Calsaverini, R. Vicente. Phase transition from egalitarian to hierarchical societies driven between cognitive and social constraints. *arXiv:1608.03637*, available online: `http://arxiv.org/abs/1608.03637`, 2016.

[49] N. Caticha, R. Vicente. Agent-based social psychology: from neurocognitive processes to social data. *Advances in Complex Systems* 14 (05), 711-731, 2011.

[50] D. Saad, S.A. Solla. Exact Solution for On-Line Learning in Multilayer Neural Networks. *Phys. Rev. Lett.* 74, 4337-4340, 1995.

[51] W. Kinzel. Phase transitions of neural networks, Philosophical Magazine B, 77(5), 1455-1477, 1998.

[52] M. Opper. Learning and generalization in a two-layer neural network: The role of the Vapnik-Chervonenkis dimension. *Phys. Rev. Lett.,* 72, 2113, 1994.

[53] D. Herschkowitz, M. Opper. Retarded Learning: Rigorous Results from Statistical Mechanics. *Phys. Rev. Lett.,* 86, 2174, 2001.

[54] M. Biehl, P. Riegler, C. Wöhler. Transient dynamics of on-line learning in two-layered neural networks. *J. of Physics A: Math. and Gen.* 29, 4769-4780, 1996.

[55] R. Vicente. N. Caticha. Functional optimization of online algorithms in multilayer neural networks. *J. of Physics A: Math. and Gen.* 30 (17), L599, 1997.

[56] S. Amari, H. Park and T. Ozeki, Singularities affect dynamics of learning in neuromanifolds. *Neural Computation,* 18, 1007-1065, 2006.

[57] R. Der, M. Herrmann. Critical phenomena in self-organizing feature maps: A Ginzburg-Landau approach. *Physical Review E [Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics]* 49(6): 5840-5848, 1994.

[58] M. Biehl, B. Hammer, T. Villmann. Prototype-based models in machine learning. *WIREs Cogn. Sci.* 7, 92-111, 2016.

[59] H. Ritter, K. Schulten. On the Stationary State of Kohonen's Self-Organizing Sensory Mapping. *Biological Cybernetics* 54: 99–106, 1986.

[60] H. Ritter, K. Schulten. Convergence properties of Kohonen's topology preserving maps: fluctuations, stability, and dimension selection. *Biological Cybernetics* 60(1): 59–71, 1988.

[61] E. Erwin, K. Obermeyer, K. Schulten. Self-organizing maps: Ordering, convergence properties and energy functions. *Biological Cybernetics* 67(1): 47–55, 1992.

[62] E. Erwin, K. Obermeyer, K. Schulten. Self-organizing maps: Stationary states, metastability and convergence rate. *Biological Cybernetics* 67(1): 35–45, 1992.

[63] T. Heskes. Energy functions for self-organizing maps. In: E. Oja, S. Kaski, editors, *Kohonen Maps*, 303–316, Elsevier, 1999.

[64] H. Ritter. Asymptotic level density for a class of vector quantization processes. *IEEE Transactions on Neural Networks* 2(1):173–175, 1993.

[65] T. Martinetz, S. Berkovich, K. Schulten. 'Neural-Gas' Network for Vector Quantization and its Application to Time-Series Prediction. *IEEE Transactions on Neural Networks* 4(4):558–569, 1993.

[66] G. Tononi, C. Koch. Consciousness: here, there and everywhere? *Phil. Trans. of the R. Soc. B: Biological Sciences* 370: 20140167, 2015.

[67] J.A. Quinn, J. Mooij, T. Heskes, M. Biehl. Learning of causal relations. In: M. Verleysen, editor, *Proc. of the European Symposium on Artificial Neural Networks (ESANN 2011),* i6doc.com, 287-296, 2011.

[68] M. Oizumi, N. Tsuchiya, S. Amari. Unified framework for information integration based on information geometry. *Proc. of the National Academy of Sciences (PNAS)* 113 (51), 14817-14822, 2016.

[69] G. Tononi, M. Boly, M. Massimini, C. Koch. Integrated information theory: From consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17(7), 450-461, 2016.

[70] V. Van Belle, P. Lisboa. Research directions in interpretable machine learning models. In: M. Verleysen, editor, *Proc. of the European Symposium on Artificial Neural Networks (ESANN),* d-side pub. 533-541, 2013.

[71] A. Vellido, J.D. Martín-Guerro, P. Lisboa. Making machine learning models interpretable.

In: M. Verleysen, editor, *Proc. of the European Symposium on Artificial Neural Networks (ESANN),* d-side pub. 163-172, 2012.

[72] G. Bhanot, M. Biehl, T. Villmann, D. Zühlke. Biomedical data analysis in translational research: Integration of expert knowledge and interpretable models. In: M. Verleysen, editor, *Proc. of the European Symposium on Artificial Neural Networks (ESANN 2017),* i6doc.com, 177-186, 2017.

[73] A. Bifet, B. Hammer, F.-M. Schleif. Streaming data analysis, concept drift and analysis of dynamic data sets. In: M. Verleysen, editor, *Proc. of the European Symposium on Artificial Neural Networks (ESANN 2019),* i6doc.com, this volume, 2019.

[74] N. Caticha, F. Alves. Trust, law and ideology in a NN agent model of the US Appellate Courts. In: M. Verleysen, editor, *Proc. of the European Symposium on Artificial Neural Networks (ESANN 2019),* i6doc.com, this volume, 2019.

[75] Y. Fang, Z. Yu, F. Chen. Noise helps optimization escape from saddle points in the neural dynamics. In: M. Verleysen, editor, *Proc. of the European Symposium on Artificial Neural Networks (ESANN 2019),* i6doc.com, this volume, 2019.

[76] M. Straat, M. Biehl. On-line learning dynamics of ReLU neural networks using statistical physics techniques. In: M. Verleysen, editor, *Proc. of the European Symposium on Artificial Neural Networks (ESANN 2019),* i6doc.com, this volume, 2019.