

## 参考文献

- 学習理論全般
  1. M. Mohri and A. Rostamizadeh and A. Talwalkar. *Foundations of Machine Learning*. The MIT Press, 2012.
  2. S. Shalev-Shwartz, and S. Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge University Press, 2014.
- 経験過程, Rademacher 複雑度, Dudley 積分
  1. A. W. van der Vaart, and J. A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media, 1996.
  2. R. M. Dudley: *Uniform Central Limit Theorems*. Cambridge University Press, 1999.
- Fast learning rate, 局所 Rademacher 複雑度, カーネル法
  1. I. Steinwart and A. Christmann. *Support Vector Machines*. Springer-Verlag New York, 2008.
- その他
  1. A. Rakhlin and K. Sridharan. *Statistical Learning and Sequential Prediction*. Lecture note, 2014.
  
- 万能近似能力・近似精度の理論
  1. G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989. (万能近似能力)
  2. H. N. Mhaskar. Neural networks for optimal approximation of smooth and analytic functions. *Neural Computation*, 8(1):164–177, 1996. (滑らかなシグモイド型活性化関数による Sobolev 空間における近似精度)
  3. A. R. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*. 39(3), 930–945, 1993. (Barron クラスと呼ばれる関数クラスを導入し, その近似精度を導出)
  4. S. Keiper, G. Kutyniok, and P. Petersen. DGD Approximation Theory Workshop. 2017. (深層 NN の近似理論関係の既存研究がまとめられている)
- Ridgelet 変換
  1. N. Murata. An integral representation of functions using three-layered networks and their approximation bounds. *Neural Networks*, 9(6):947–956, 1996.
  2. S. Kostadinova, S. Pilipović, K. Saneva, and J. Vindas. The ridgelet transform of distributions. *Integral Transforms and Special Functions*, 25(5):344–358, 2014.
  3. S. Sonoda and N. Murata. Neural network with unbounded activation functions is universal approximator. *Applied and Computational Harmonic Analysis*, 43(2):233–268, 2017.
- ReLU-NN の近似理論と統計的推定理論
  1. D. Yarotsky. Error bounds for approximations with deep relu networks. *Neural Networks*, 94:103–114, 2017. (近似理論)
  2. J. Schmidt-Hieber. Nonparametric regression using deep neural networks with ReLU activation function. ArXiv e-prints, Aug. 2017. (推定理論)
  3. T. Suzuki. Adaptivity of deep ReLU network for learning in Besov and mixed smooth Besov spaces: optimal rate and curse of dimensionality. ICLR2019, arXiv:1810.08033.
  
- 深層ニューラルネットワークの Rademacher 複雑度
  1. B. Neyshabur, R. Tomioka, and N. Srebro. Norm-based capacity control in neural networks. In *Conference on Learning Theory*, 1376–1401, 2015.
  2. P. L. Bartlett, D. J. Foster, and M. J. Telgarsky. Spectrally-normalized margin bounds for neural networks. In *Advances in Neural Information Processing Systems*, 6240–6249, 2017.
  3. N. Golowich, A. Rakhlin and O. Shamir. Size-independent sample complexity of neural networks. In *Conference On Learning Theory*, 297–299, 2018.
  
- 横幅の広いニューラルネットワークの勾配法による最適化の大域的最適性について。
  1. Du, S. S., Lee, J. D., Li, H., Wang, L., and Zhai, X. (2018). Gradient descent finds global minima of deep neural networks. arXiv preprint arXiv:1811.03804.
  2. Allen-Zhu, Z., Li, Y., and Song, Z. (2018). A convergence theory for deep learning via over-parameterization. arXiv preprint arXiv:1811.03962.
  3. Du, S. S., Zhai, X., Póczos, B., and Singh, A. (2018). Gradient descent provably optimizes over-parameterized neural networks. arXiv preprint arXiv:1810.02054.(これらは新しい論文であり, 間違いがありうることに注意されたい. また, 汎化性能については何も述

べていない.)