

Zachet/exam syllabus. 6 points of your final grade

Theoretical Deep Learning #2, MIPT

All proofs are needed, if not stated otherwise.

1 Main part (4 points total).

1. Problem statement for bounding test-train risk difference. Worst-case bounds.
2. Worst-case bound for 0-1 loss.¹ VC-lemma (without proof). Growth function. VC-dimension. Sauer's lemma (without proof)
3. The reason to introduce γ -margin loss instead of 0-1 loss. Bounding $R - \hat{R}_{n,\gamma}$ with l_∞ -covering numbers (without proof; analogy with 0-1 loss case). Fat-shattering dimension. Connection of the latter with l_∞ -covering numbers (without proof; analogy with VC-dimension). Example of function class with finite fat-shattering dimension, but with infinite VC-dimension.²
4. McDiarmid's inequality (without proof). Rademacher complexity. Bounding test-train risk difference with Rademacher complexity.³
5. Bounding Rademacher complexity for 0-1 loss with growth function (Hoeffding's lemma — without proof).⁴
6. PAC-bayesian bound for at most countable hypothesis classes.
7. PAC-bayesian bound for uncountable hypothesis classes (in the form of McAllester)⁵.

¹In a similar manner as in Theorem 2 of Bartlett (1998): <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=661502>

²Based on Bartlett (1998)

³Have a look at lecture notes by Geoffrey Decrouez: https://1f912c10-a4be-4e1b-a1e2-6af556aeef2a.filesusr.com/ugd/dd0cbc_95c300090eb64378aaa1e0218987cbf9.pdf

⁴Again, have a look at lecture notes by Geoffrey Decrouez

⁵Theorem 2 of McAllester (1999): <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.21.1908&rep=rep1&type=pdf>

8. Failure of PAC-bayesian bounds for deterministic learning algorithms. Way to leverage: stochastization⁶.
9. PAC-bayesian bound for a deterministic learning algorithm (Neyshabur et al., 2018⁷).

2 Auxiliary part (2 points total).

1. Proof of bound for $R - \hat{R}_{n,\gamma}$ with l_∞ -covering numbers.⁸
2. Bounding $R - \hat{R}_{n,\gamma}$ for deep ReLU nets with spectral complexity (Dudley's integral and a covering number for a set of vectors — without proof)⁹.
3. Compression approach. Deriving a bound of Neyshabur et al. (2018) with compression approach (Arora et al., 2018¹⁰) (omit estimates for K in weight discretization step).

⁶Based on Dziugaite & Roy (2017): <https://arxiv.org/abs/1703.11008>

⁷https://openreview.net/forum?id=Skz_WfbCZ

⁸Theorem 2 in Bartlett (1998)

⁹Main result of Bartlett et al. (2017): <https://arxiv.org/abs/1706.08498>

¹⁰<http://proceedings.mlr.press/v80/arora18b.html>