Seminar: Theoretical Advances in Deep Learning

Debarghya Ghoshdastidar, Satyaki Mukherjee, Maximillian Fleissner, Mahalakshmi Sabanayagam

TU Munich, Department of Informatics

Winter Semester 2023

Course information

- Master seminar (IN2107, IN4409)
    - 5 ECTS, 2 SWS

Course information

- Master seminar (IN2107, IN4409)
  - 5 ECTS, 2 SWS
- Organisers:
  - Mahalakshmi Sabanayagam sabanaya@cit.tum.de (main coordinator of course)
  - Maximilian Fleissner fleissnm@cit.tum.de
  - Satyaki Mukherjee satyaki.mukherjee@cit.tum.de
  - Prof. Debarghya Ghoshdastidar ghoshdas@cit.tum.de

DL / ML papers

- New algorithms with some experiments showing their properties

DL / ML papers

- New algorithms with some experiments showing their properties
  - Provides some understanding (less common in ML than DL)

DL / ML papers

- New algorithms with some experiments showing their properties
  - Provides some understanding (less common in ML than DL)
- Empirical analysis of algorithmic properties

DL / ML papers

- New algorithms with some experiments showing their properties
  - Provides some understanding (less common in ML than DL)
- Empirical analysis of algorithmic properties
  - Important when algorithms are hard to analyse theoretically
  - Common in deep learning, non-convex optimisation

DL / ML papers

- New algorithms with some experiments showing their properties
  - Provides some understanding (less common in ML than DL)
- Empirical analysis of algorithmic properties
  - Important when algorithms are hard to analyse theoretically
  - Common in deep learning, non-convex optimisation
- Dedicated theory papers
  - Mathematically explain why DL / ML methods work (rare in DL)

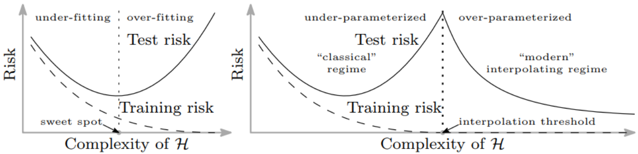
DL / ML papers

- New algorithms with some experiments showing their properties
  - Provides some understanding (less common in ML than DL)
- Empirical analysis of algorithmic properties
  - Important when algorithms are hard to analyse theoretically
  - Common in deep learning, non-convex optimisation
- Dedicated theory papers $\longleftarrow$ Focus of this seminar
  - Mathematically explain why DL / ML methods work (rare in DL)

Why do we need mathematical analysis of DL?

- Deep learning contradicts conventional wisdom

  Complex models generalise well

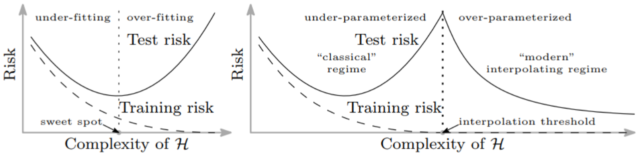


(a) U-shaped "bias-variance" risk curve (b) "double descent" risk curve

Why do we need mathematical analysis of DL?

- Deep learning contradicts conventional wisdom

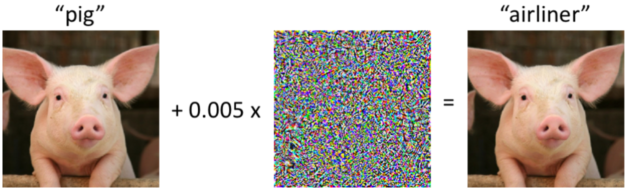
  Complex models generalise well



(a) U-shaped "bias-variance" risk curve (b) "double descent" risk curve

- Neural networks not robust

  Can be fooled to make error



Why do we need mathematical analysis of DL?

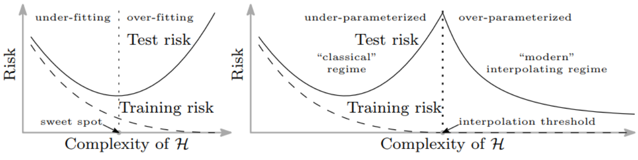
- Deep learning contradicts conventional wisdom

  Complex models generalise well

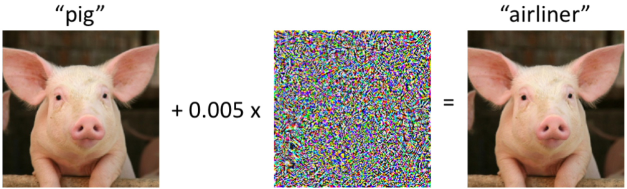
- Neural networks not robust

  Can be fooled to make error

- The output of deep networks lack explainability



(a) U-shaped "bias-variance" risk curve (b) "double descent" risk curve



Purpose of this seminar

- Theory in deep learning emerging

Purpose of this seminar

- Theory in deep learning emerging
  - What do we know so far?

Purpose of this seminar

- Theory in deep learning emerging
  - What do we know so far?
  - What are the limitations in theory, and gaps with practice?

Purpose of this seminar

- Theory in deep learning emerging
  - What do we know so far?
  - What are the limitations in theory, and gaps with practice?
- Familiarise with statistical foundations of learning (complements lecture IN2378)

Purpose of this seminar

- Theory in deep learning emerging
  - What do we know so far?
  - What are the limitations in theory, and gaps with practice?
- Familiarise with statistical foundations of learning (complements lecture IN2378)
- Familiarise with mathematical proof techniques
  - Considerable focus on math in this seminar

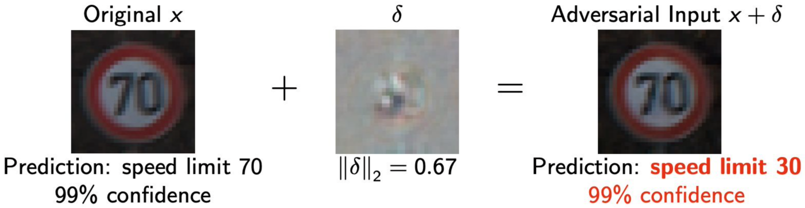
Purpose of this seminar

- Theory in deep learning emerging
  - What do we know so far?
  - What are the limitations in theory, and gaps with practice?
- Familiarise with statistical foundations of learning (complements lecture IN2378)
- Familiarise with mathematical proof techniques
  - Considerable focus on math in this seminar
- Familiarise with publication and review process in ML

Focus of this seminar
Possible Topics

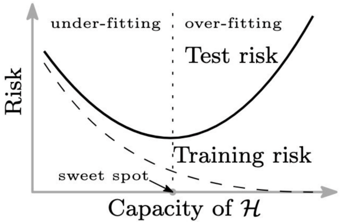
Adversarial ML / Robustness

- Performance of NNs significantly affected if data is slightly perturbed.
- Why? How can we build robust ML models / guarantee robustness?



Generalisation in neural networks

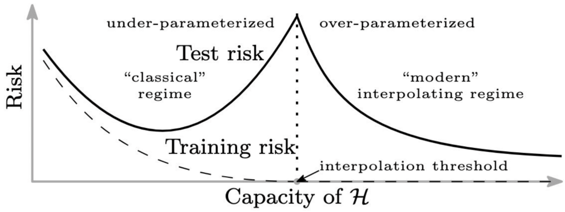
- Classical learning theory cannot explain generalisation in deep networks.
- Data-dependent generalisation error bounds more meaningful and practical.



Double-descent in bias-variance curve

- Over-parameterised NNs deviate from bias-variance trade-off - NNs may perform best in zero training loss / interpolating regime.

- Currently, this behaviour has been analytically derived in simpler settings.



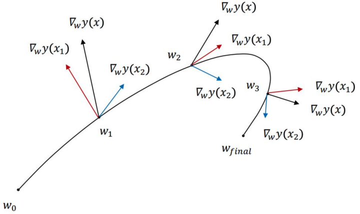
Over-parameterised NN (infinite width)

Analyse Over-parametrised NNs asymptotically as width goes to infinity

- Under small learning rate, (S)GD training $\equiv$ Neural Tangent Kernel (NTK), a dot product kernel in gradient space of the NN parameters

- Finite width networks can deviate from the kernel regime.

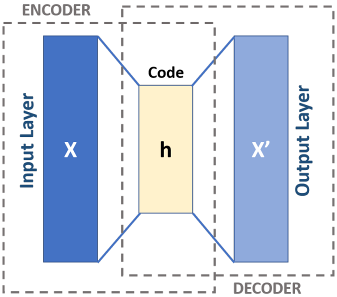
$$NTK(x_i, x_j) = \nabla_w y(x_i) \,.\, \nabla_w y(x_j)$$



Unsupervised Deep learning

Most of the current theoretical results in deep learning are in the supervised setting. What guarantees can we give in an unsupervised (e.g. clustering) setting.

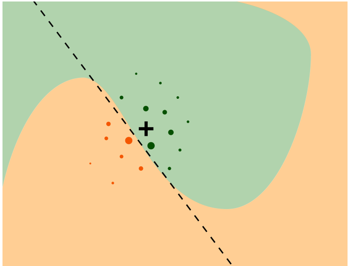
- Autoencoder
- Representation learning



Interpretability of DNN

Explain/interpret the prediction of the data using global or local schemes.

- LIME (Locally Interpretable Model-Agnostic Explanations) provides explanations for a datapoint by sampling data around it.
- SHAP (SHapley Additive exPlanations) can provide explanations on the model level by measuring the importance of every data feature.



Domain adaptation / Transfer learning

- What happens if the distribution at test time is not the same as during training?
- Can we still give generalization error bounds?

For a more in-depth look...

join the *recent advances in ML / DL* lecture as part of the *statistical foundations of deep learning* course

on 21.07.2023 (Friday) 16:00 - 18:00 (seminar room 00.13.009A)

Administration

Seminar details

- We will use Moodle for coordination

Seminar details

- We will use Moodle for coordination
- Desired number of participants $= 15$

Seminar details

- We will use Moodle for coordination
- Desired number of participants = 15
- Pre-requisites: Machine Learning (IN2064), Deep learning (IN2346)

Seminar details

- We will use Moodle for coordination
- Desired number of participants = 15
- Pre-requisites: Machine Learning (IN2064), Deep learning (IN2346)
- **Must be comfortable with mathematical techniques / proving results**
    - Taking Statistical foundations of learning (IN2378) would help

Assessment

- Everyone assigned one paper

Assessment

- Everyone assigned one paper
- Submit a report. Details will be provided in the introduction lecture.
    - summary of paper, explaining main results and their implications
    - review (we will discuss how to write reviews)
    - summary of proofs (main techniques, key lemmas and ideas)

Assessment

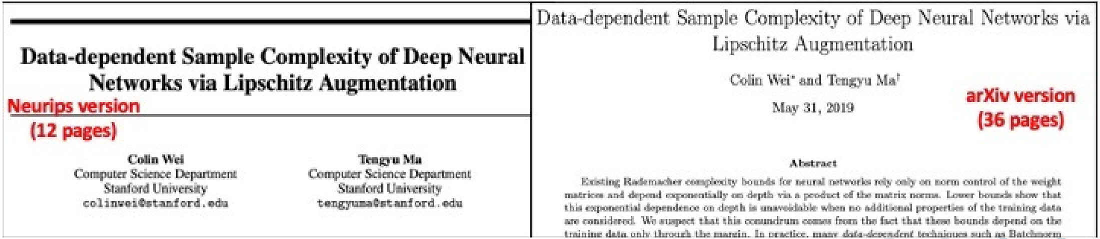
- Everyone assigned one paper
- Submit a report. Details will be provided in the introduction lecture.
  - summary of paper, explaining main results and their implications
  - review (we will discuss how to write reviews)
  - summary of proofs (main techniques, key lemmas and ideas)
- Present paper and your report
  - Block seminar; everyone needs to attend all talks

Assessment

- Everyone assigned one paper
- Submit a report. Details will be provided in the introduction lecture.
  - summary of paper, explaining main results and their implications
  - review (we will discuss how to write reviews)
  - summary of proofs (main techniques, key lemmas and ideas)
- Present paper and your report
  - Block seminar; everyone needs to attend all talks
- Grading: Report (40%) + Presentation (60%)
  - Bonus for asking interesting questions to other speakers

Report + Presentation of papers

- Mostly publications from recent ML conferences (ICML, ICLR, Neurips, COLT)
  - Conference papers are short (8 page, no proofs)
- **Report has to follow longer version on arXiv** (link will be provided)
  - Considerable focus on understanding mathematical results



Data-dependent Sample Complexity of Deep Neural Networks via Lipschitz Augmentation

Neurips version (12 pages)

Colin Wei
Computer Science Department
Stanford University
colinwei@stanford.edu

Tengyu Ma
Computer Science Department
Stanford University
tengyuma@stanford.edu

Data-dependent Sample Complexity of Deep Neural Networks via Lipschitz Augmentation

Colin Wei* and Tengyu Ma†

May 31, 2019

arXiv version (36 pages)

Abstract

Existing Rademacher complexity bounds for neural networks rely only on norm control of the weight matrices and depend exponentially on depth via a product of the matrix norms. Lower bounds show that this exponential dependence on depth is unavoidable when no additional properties of the training data are considered. We suspect that this conundrum comes from the fact that these bounds depend on the training data only through the margin. In practice, many data-dependent techniques such as Batchnorm

Timeline (tentative)

- First week of August: Get full paper list
- Last week of August: Provide preference for papers
- First week of lecture: First meeting (assignments, reports and organisation)
- November 01: Deadline for de-registration
- Mid January : Submit report and first version of slides (both as PDF)
- Mid February: Final presentation (block seminar, date to be finalised)
- Office hours: weekly 2h

Most important thing to do now...

Fill out the form to help us match you in the system
https://forms.gle/s7neSKFVL9iEToyG6

Slide deck and the form will be uploaded to the webpage