

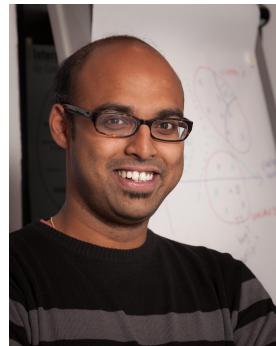
Deep Learning - 2019

Introduction

Prof. Avishek Anand

Who are we ?

- Knowledge-based Systems (KBS)
 - L3S Research Center
- Cutting edge research on
 - Information retrieval
 - Machine learning
 - Data mining
 - Deep learning
- We
 - Build systems and do large projects
 - Publish papers and travel the world
 - Are always looking for motivated students



Prof. Avishek Anand



Mateusz Malinowski



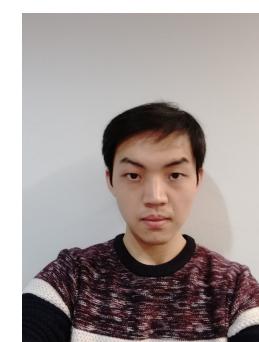
Dr. Megha Khosla



Teaching Assistants



Jurek Leonhardt



Zijian Zhang

Maximillian
Idahl

What is Deep Learning ?

Artificial Intelligence

Any technique that enables computers to mimic human behavior

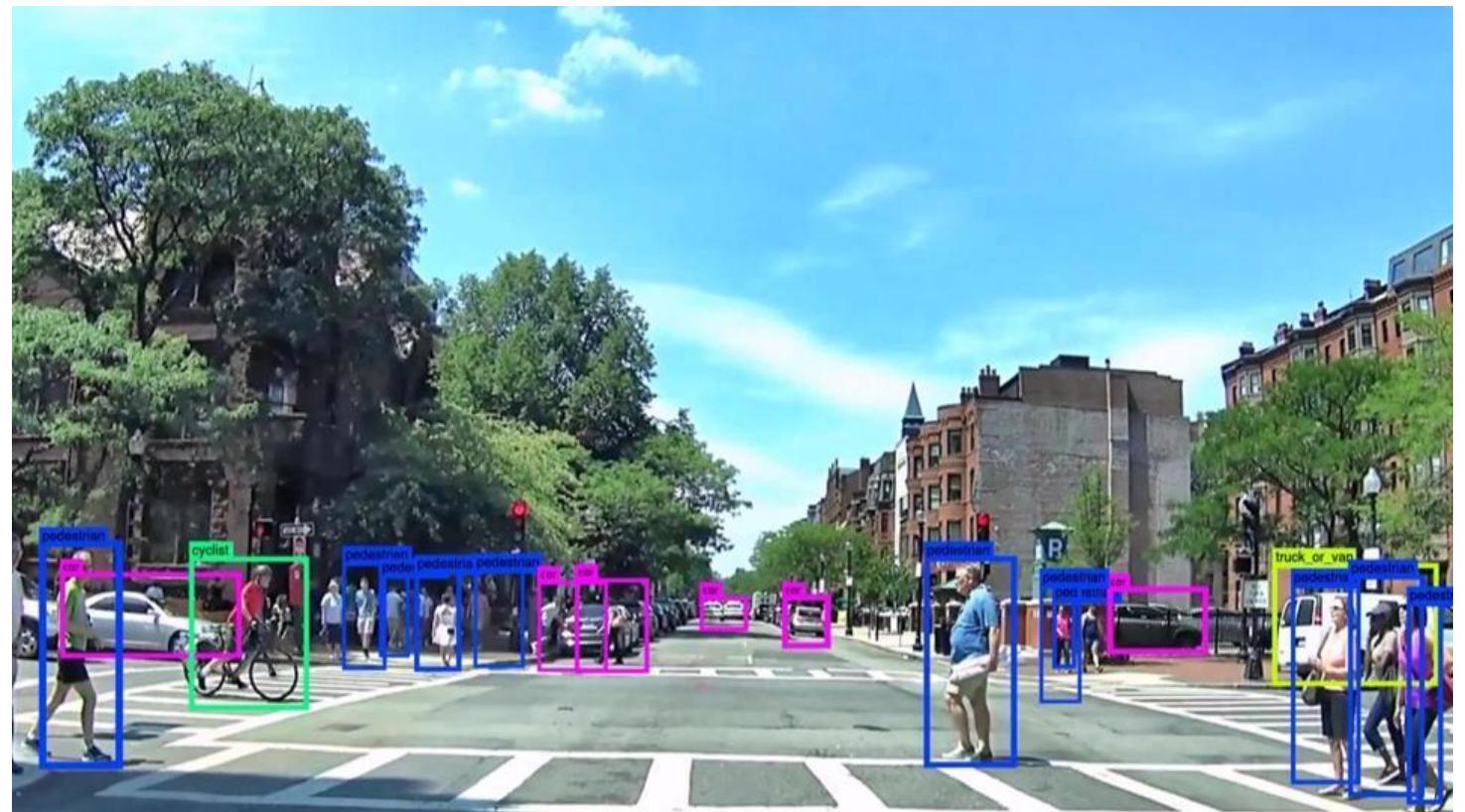
Machine Learning

Ability to learn without explicitly being programmed

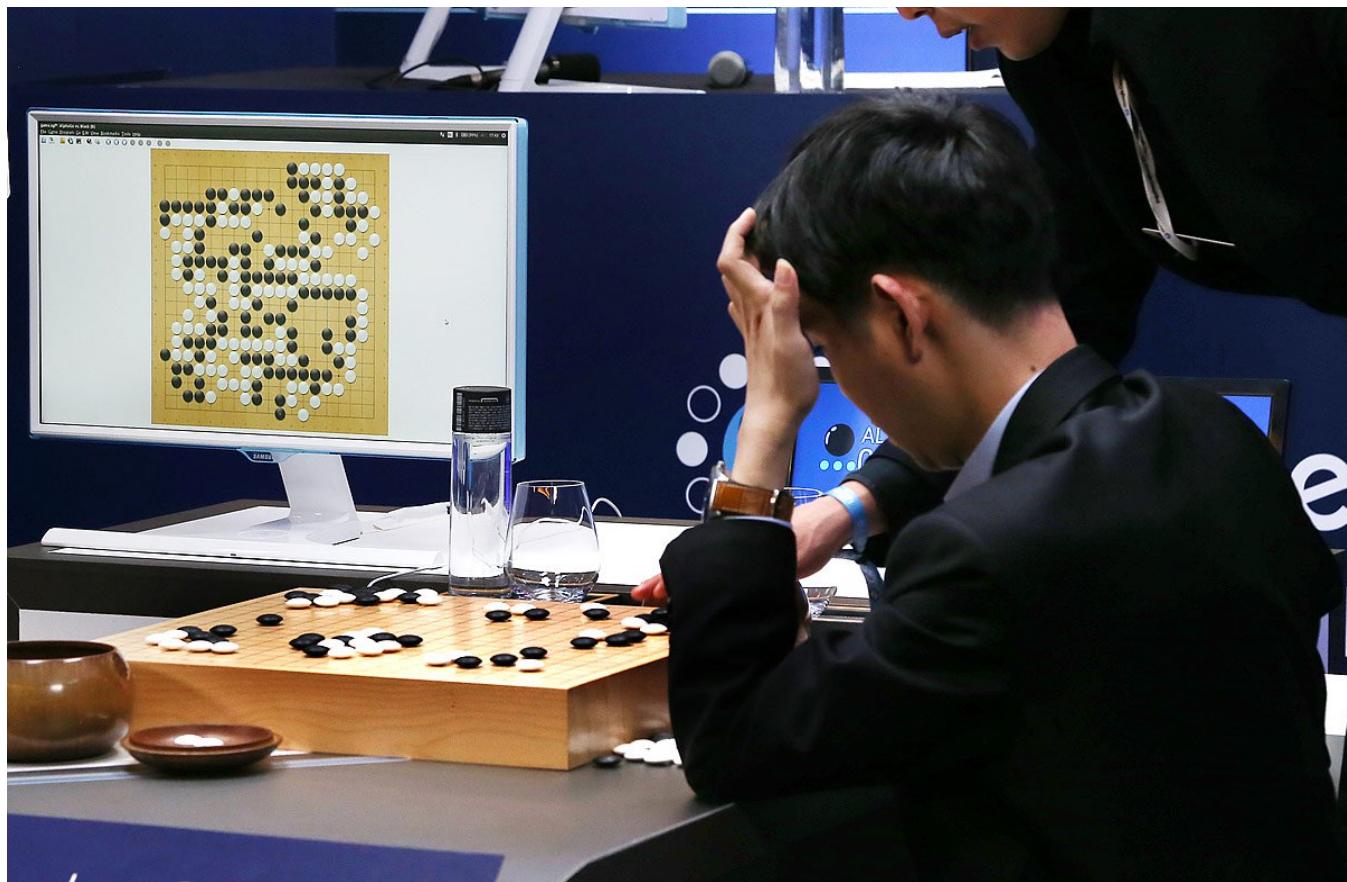
Deep Learning

Extract patterns from data using Neural Networks

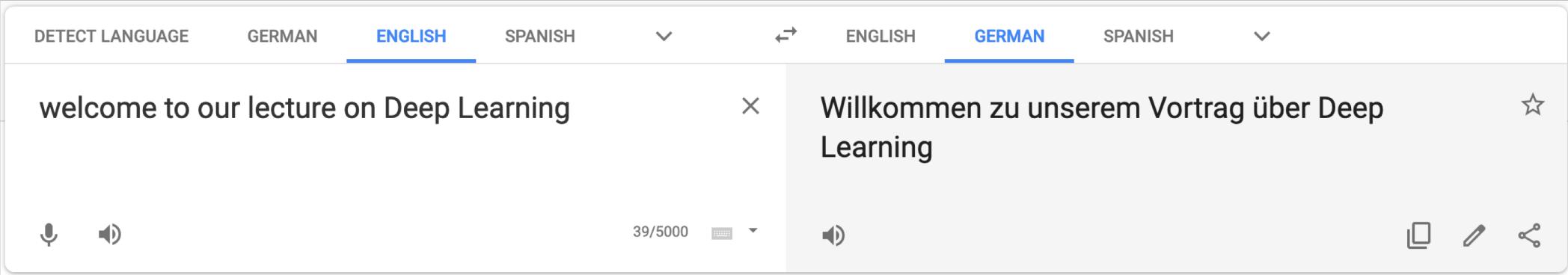
Achievements of DL



Achievements of DL



Achievements of DL



The screenshot shows the Google Translate interface. On the left, the input text is "welcome to our lecture on Deep Learning". Above it, the "DETECT LANGUAGE" dropdown is set to "GERMAN", and the "TRANSLATE TO" dropdown is set to "ENGLISH". Below the input, there are microphone and speaker icons, and a character count of "39/5000". On the right, the translated text is "Willkommen zu unserem Vortrag über Deep Learning". Above the output, the "TRANSLATE FROM" dropdown is set to "ENGLISH" and the "TRANSLATE TO" dropdown is set to "GERMAN". Below the output, there are edit and share icons. A star icon is also present in the top right corner of the output panel.

Google Translate



The Rise of Deep Learning

Using snippets of voices, Baidu's 'Deep Voice' can generate new speech, accents, and tones.



AI
N
pr
DEAN T

'Creative' AlphaZero leads way for chess computers and, maybe, science

Former chess world champion Garry Kasparov likes what he sees of computer that could be used to find cures for diseases



Stock Predictions Based On AI: Is the Market Truly Predictable?

By NICK BARTON, OUTLOOK AND CIO - INVESTOR JAN 1, 2018

How an A.I. 'Cat-and-Mouse Game' Generates Believable Fake Photos

By CADE METZ and KEITH COLLINS JAN 1, 2018

Complex of bacteria-infesting viral proteins modeled in CASP-13. The complex contains that were modeled individually. PROTEIN DATA BANK

Google's DeepMind aces protein folding

By Robert F. Service | Dec. 6, 2018, 12:05 PM

t with DEEPMIND'S STARCRAFT TRIUMPH FOR



Technology outpacing security measures

Facial recognition | Features and Interviews

Deep L

After Millions of Trials, These Simulated Humans Learned to Do Perfect Backflips and Cartwheels

George Dvorkin 477M 11.5k 11.5k · Post by AI ·

To create the final image in this set, the system generated 10 million revisions over 18 days.

Researchers introduce a deep learning method that converts mono audio recordings into 3D sounds using video scenes

By Michael Miller | December 28, 2018 10:21 am

AI beats docs in cancer spotting

A new study provides a fresh example of machine learning as an important diagnostic tool. Paul Biegler reports.



These faces show how far AI image generation has come in just four years

File photo on the right aren't real; they're the product of machine learning

Automation And Algorithms: De-Risking Manufacturing With Artificial Intelligence

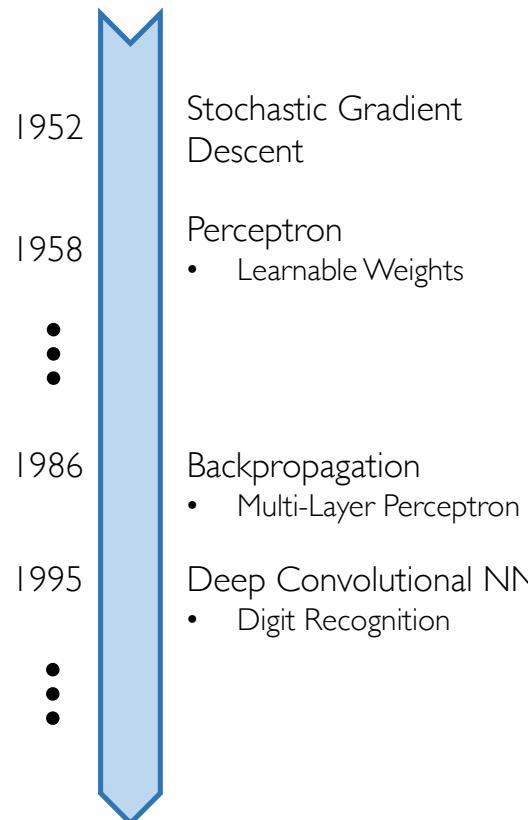
Sarah Goehrke Contributor Manufacturing Focus on the industrialization of additive manufacturing

TWEET THIS

The two key applications of AI in manufacturing are pricing and manufacturability feedback

Why now ?

Not an old idea, but the time has come now



I. Big Data

- Larger Datasets
- Easier Collection & Storage



WIKIPEDIA
The Free Encyclopedia



2. Hardware

- Graphics Processing Units (GPUs)
- Massively Parallelizable

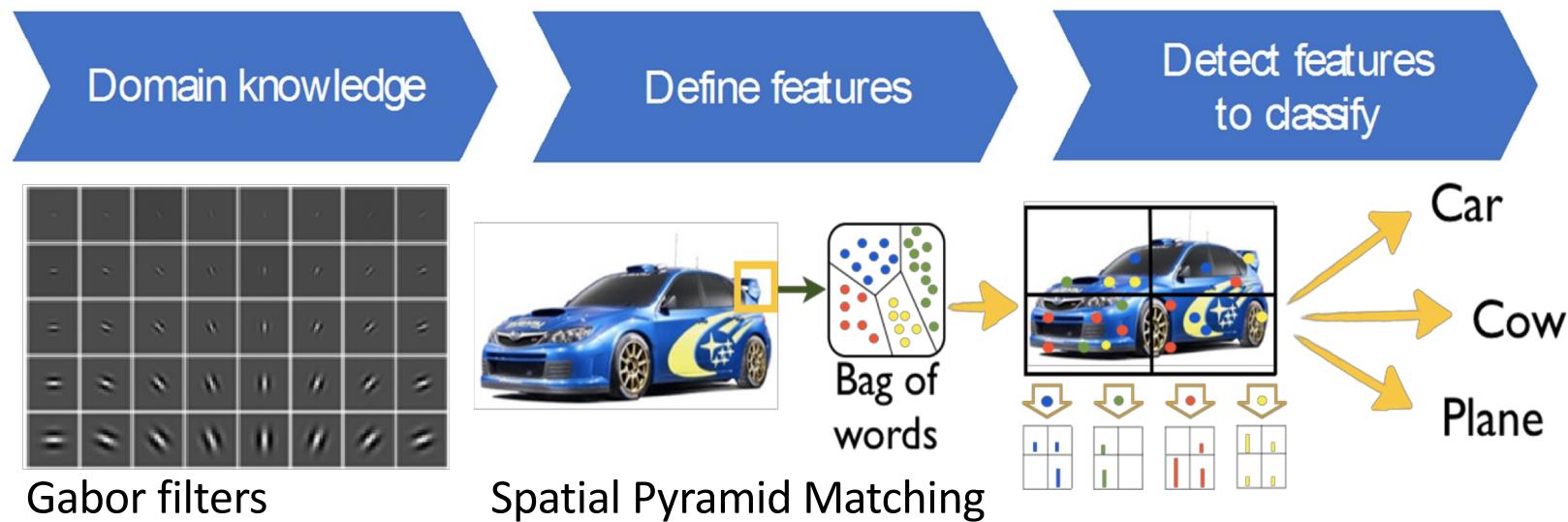


3. Software

- Improved Techniques
- New Models
- Toolboxes



Classical Machine Learning Pipeline



- Input Data → Engineer Features → Build Model
 - Needs domain knowledge
 - Time consuming
 - **Good Features make good models, but what are good features?**
- What if one could automatically learn features from the input data ?

Challenges in designing features

- Our perception works well in different circumstances
 - E.g., we can recognize our friends in the morning or in the afternoon



Scale variation



Deformation



Occlusion



Background clutter

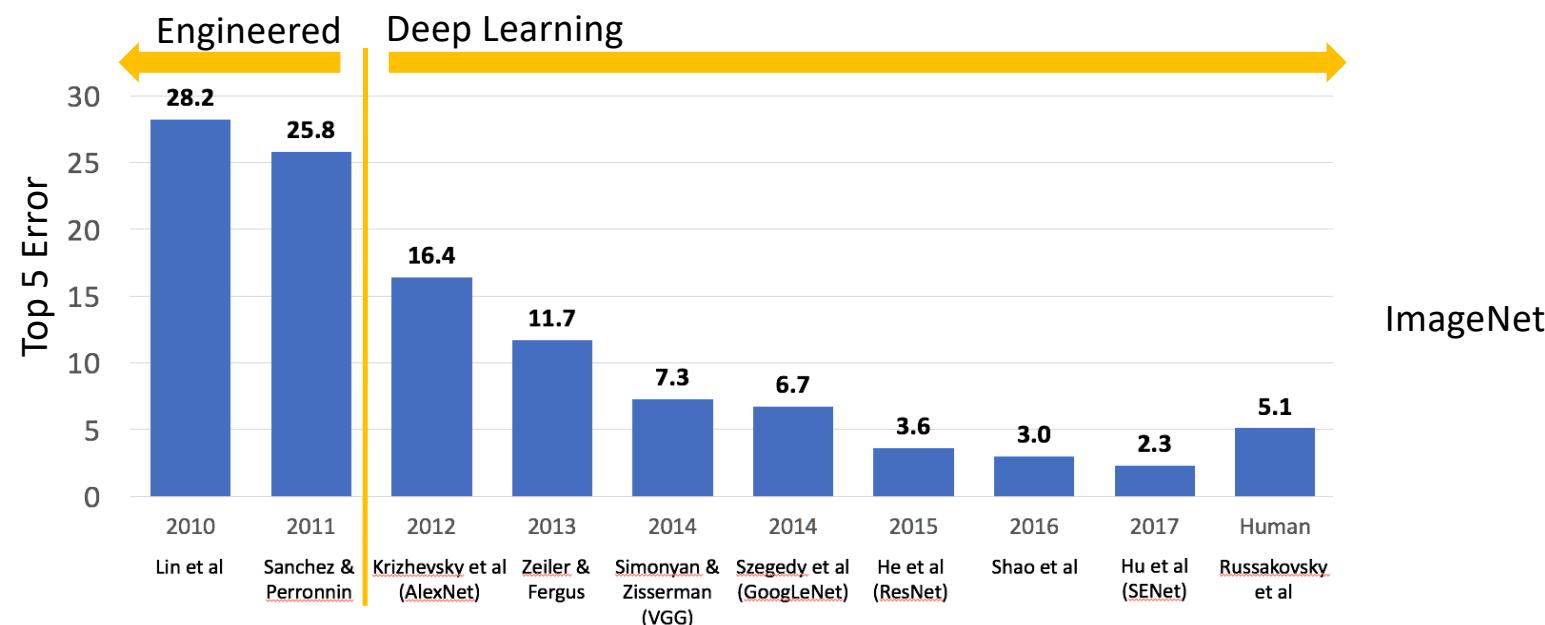


Intra-class variation



Deep Learning as Representation Learning

- New pipeline advertised by DL:
 - input data -> **learn representation** -> Build Models
 - learn representation together with classification (**end-to-end training**)



DL as Representation Learning

- A lot of time spent on creating new architectures
 - Still better performance than engineering features
 - Still less time than engineering features
 - there is still hope that with progress we will need less and less engineering of the architectures, and this will be either automated or we find the great ones
- Great progress on problems with unclear engineered features
 - Multimodal representation -- how to model language and visual signals at once ?

Course Logistics

Email for Assignment Submissions: pir-assignments@l3s.de

Email for Lecture related queries : leonhardt@l3s.de, zhang@l3s.de

Use [DL2019] in the subject line

Course Logistics

- Lecture Slides and Assignments will be available in **Stud.IP** a day before the lecture
- Mini-Projects (25% of the total credit)
 - Work in groups of max 3
 - Discuss progress with the tutors in the exercise sessions
 - Submit a writeup (Sample latex code will be provided)
 - Presentations in the last lecture + last tutorial
- Final written exam (75% of the total credit)
 - August 8
 - Topics released in May 7 :
 - DL for Language and Retrieval (Question Answering)
 - DL for Graphs (Social networks, knowledge graphs, Biological Networks)
 - Interpretability of DL approaches

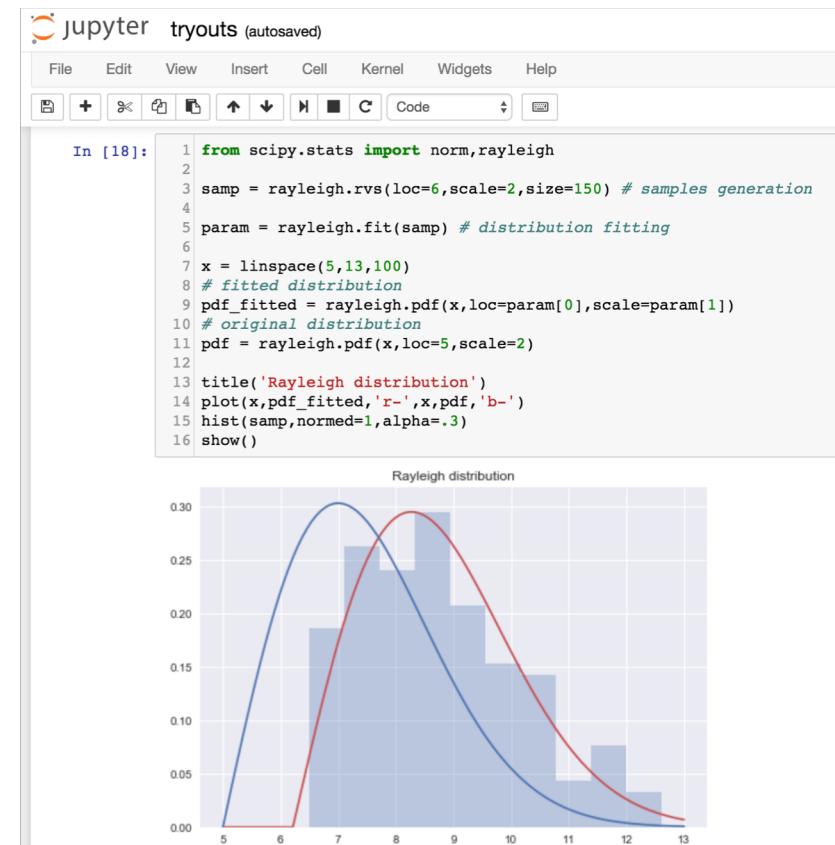


Tutorials and Exercises

- Tutorials and Recitations
 - 2 Recitations (In the exercise session)
 - Machine learning pipeline
 - Introduction to Tensor flow
 - 10 tutorials
- Exercises accompanied with each lecture to be discussed in 2 weeks from the release date
 - Tutorial Assignments to be turned in by 9 am Tuesday (In python notebooks)
- Composition of Exercises
 - Approx. 60% exercise questions based on formal matter
 - Approx. 40% exercise questions based on programming assignments

Python Notebooks

- Most/All lectures will be accompanied with python notebooks
 - Helps you see directly how the concepts translate into tangible value
 - Allows you play around for better understanding
 - Provides stub code for easier entry to code
- Programming Language
 - Python (v 3.0) – fast, easy, small code footprint, large online community
 - Tensorflow
 - Google/Stack Overflow is your friend but drive with caution



Bonus Structure

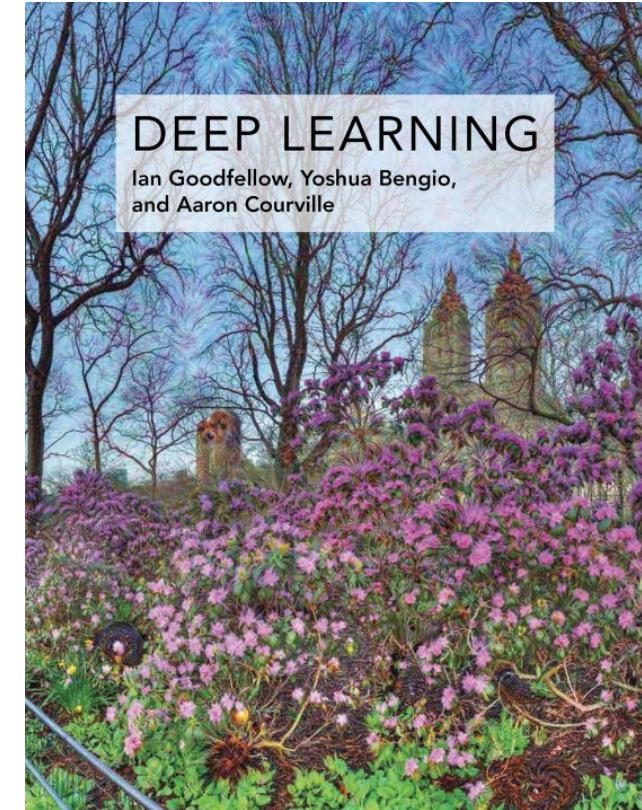
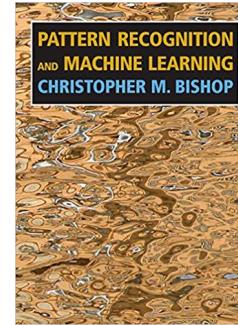
- Bonus to improve the final grade
- Students who submit correct solutions to the exercise are chosen by the tutors
 - Chosen students present their solution in the tutorials
- Each successful presentation = 1 badge
- 3 badges = 1 bonus point = 0.3 grade improvement on the final grade
 - We can decide to bump it up 1 badge = 0.5 improvement depending on the response
- Bonus in the Mini project (After reaching 100%)
 - Great job = 1 bonus point
 - Blow my mind = 2 bonus points (publication quality)



0.3333... Grade improvement

Literature and code

- Books to be followed
 - Deep Learning (Goodfellow et al.)
 - Pattern Recognition and machine learning
- References to papers to be included with the slides
- Programming language: Python + Tensorflow
- Google Colab : <https://colab.research.google.com/>



Pre-requisites

- Linear Algebra
 - Calculus (Practice your derivatives, needed in exercises)
 - Machine Learning basics
-
- Ability to work hard
 - Interest in setting up experiments
 - Being self critical

Goals of the lecture

- Get a solid understanding of the nuts and bolts of Supervised Neural Networks
 - Conv Nets, RNNs, GANs, Attention Mechanism
 - How they are applied to specific tasks
- A general understanding of optimization strategies to guide training Deep Architectures
 - The ability to design from scratch, and train novel deep architectures
 - Pick up the basics of a general purpose Neural Networks toolbox
- Ability to translate an architecture to code and into a fully functioning solution

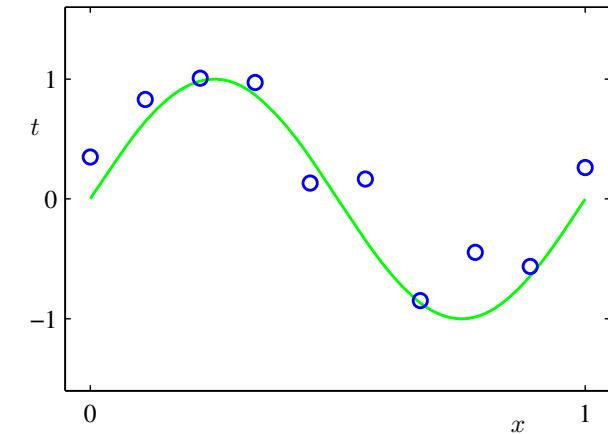
Is Machine Learning curve fitting ?

Simple Regression

- Assume that data is generated from a known function + random noise

$$\sin(2\pi x) + \epsilon$$

- You are only given some **sample of observations** (training data)
 - A finite sample of input and targets
- Goal:** to model the underlying distribution that generates data and to make predictions
- First thought: Lets try curve fitting



$$\mathbf{x} \equiv (x_1, \dots, x_N)^T$$
$$\mathbf{t} \equiv (t_1, \dots, t_N)^T$$

Polynomial curve fitting

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \dots + w_M x^M = \sum_{j=0}^M w_j x^j$$

Co-efficients , weights, parameters

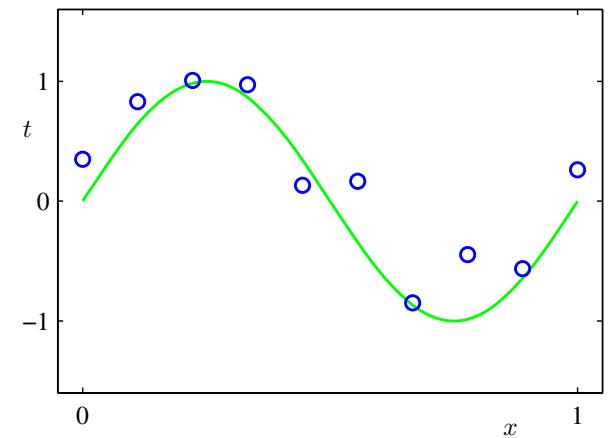
- Given a polynomial of degree M, goal is to find w that fits my data
- The best w is found with respect to some measure of goodness
 - When is one set of params better than another ?

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N (y(x_n, \mathbf{w}) - t_n)^2$$

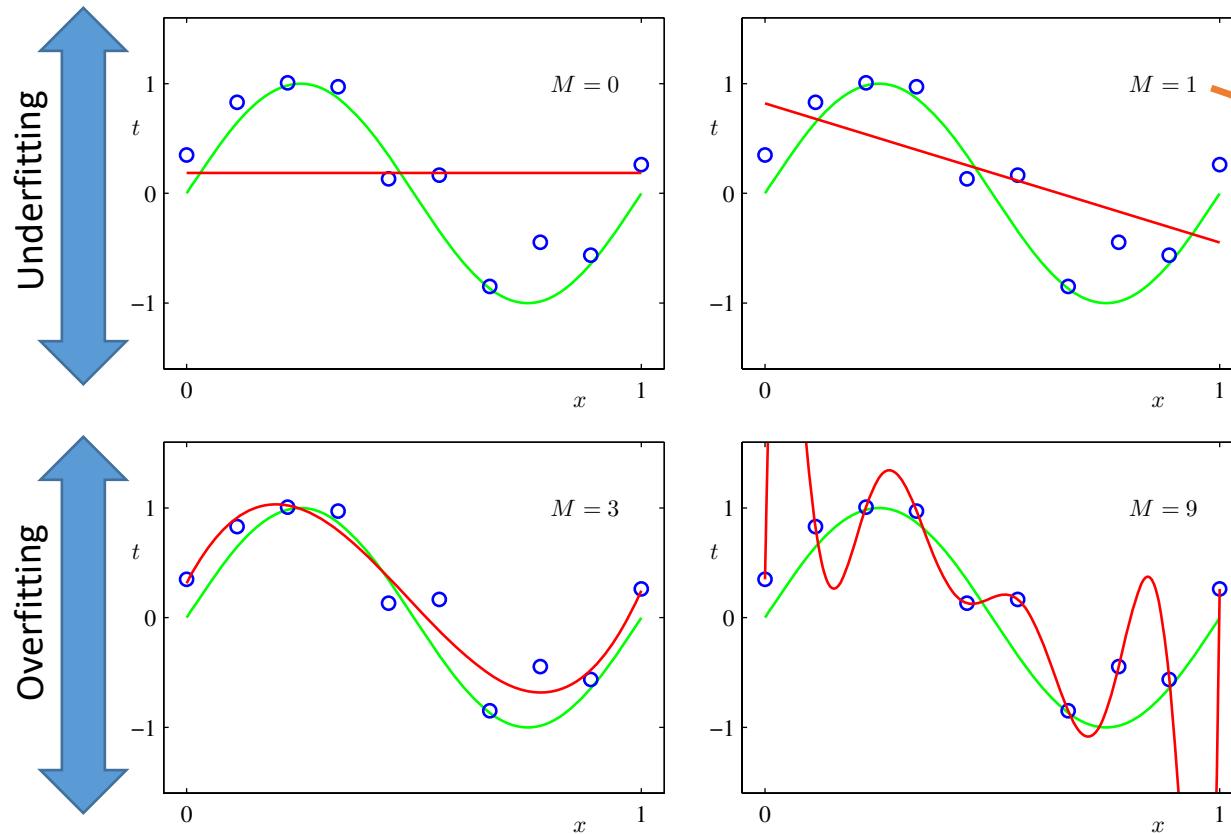
Loss function

- The best set of params are the ones with lowest loss; formally set-up as an optimization problem

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \mathcal{L}(\mathbf{w})$$



Which degree should you use ?



Hyperparameter
= degree of the
polynomial

Training, Validate and Test

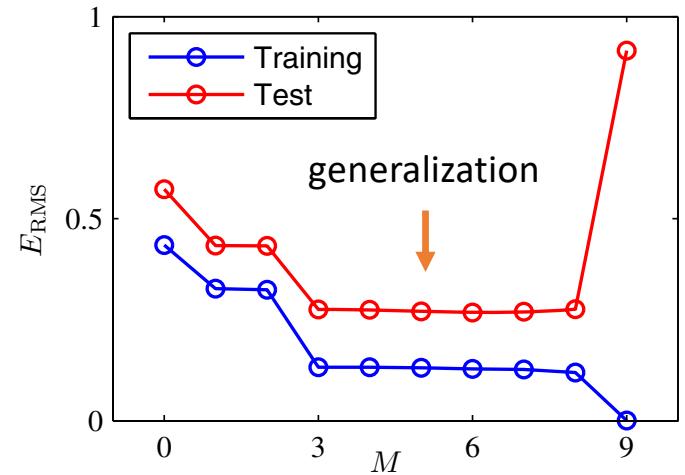
- **Generalization:** How well does your model perform on unseen data ?
- Set aside a test set to check for generalization
- In principle, we use a validation set/dev set for [model selection](#)

Generalization is the key to ML

Input Data

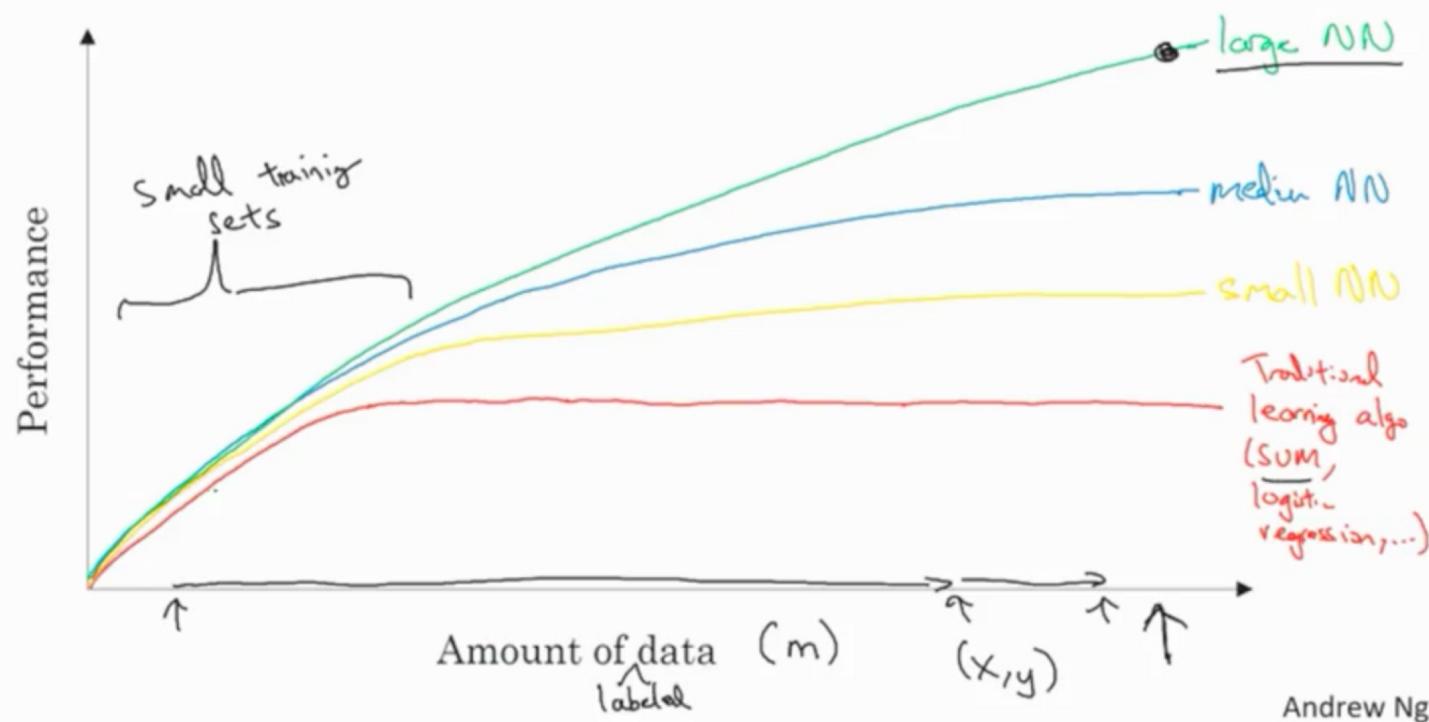
Training Data

Test Data



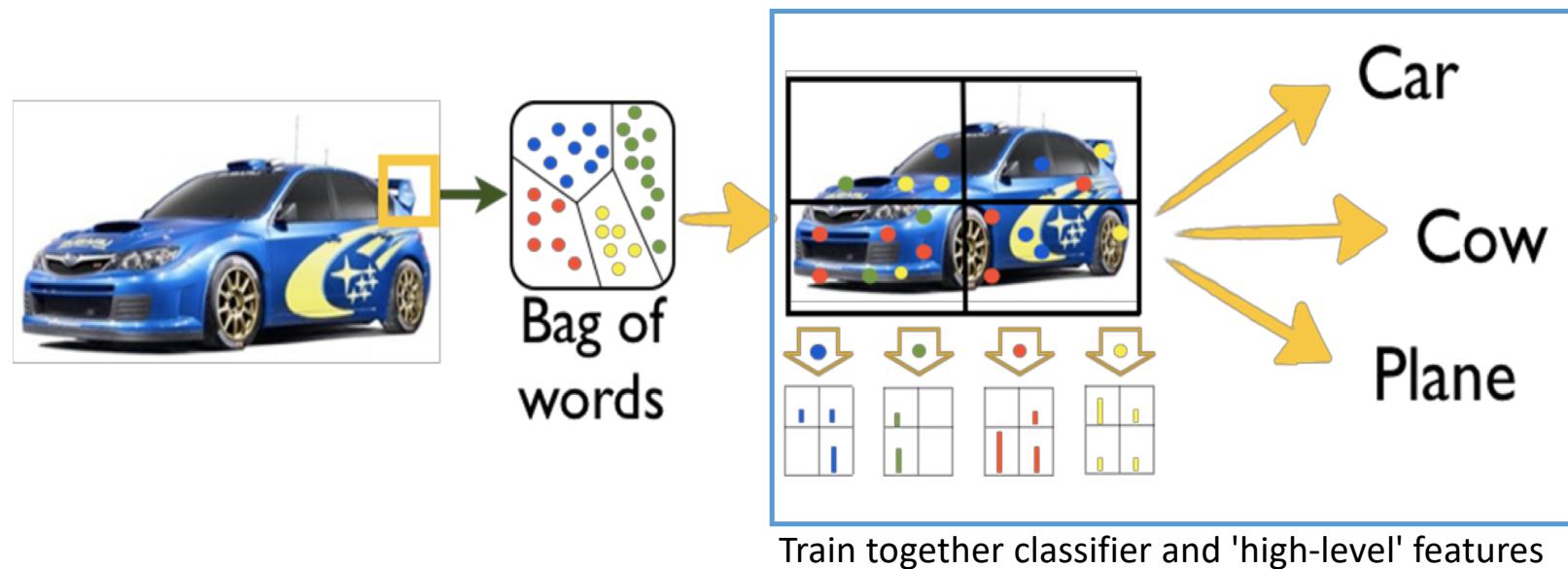
Why Deep Learning ? (revisited)

Scale drives deep learning progress



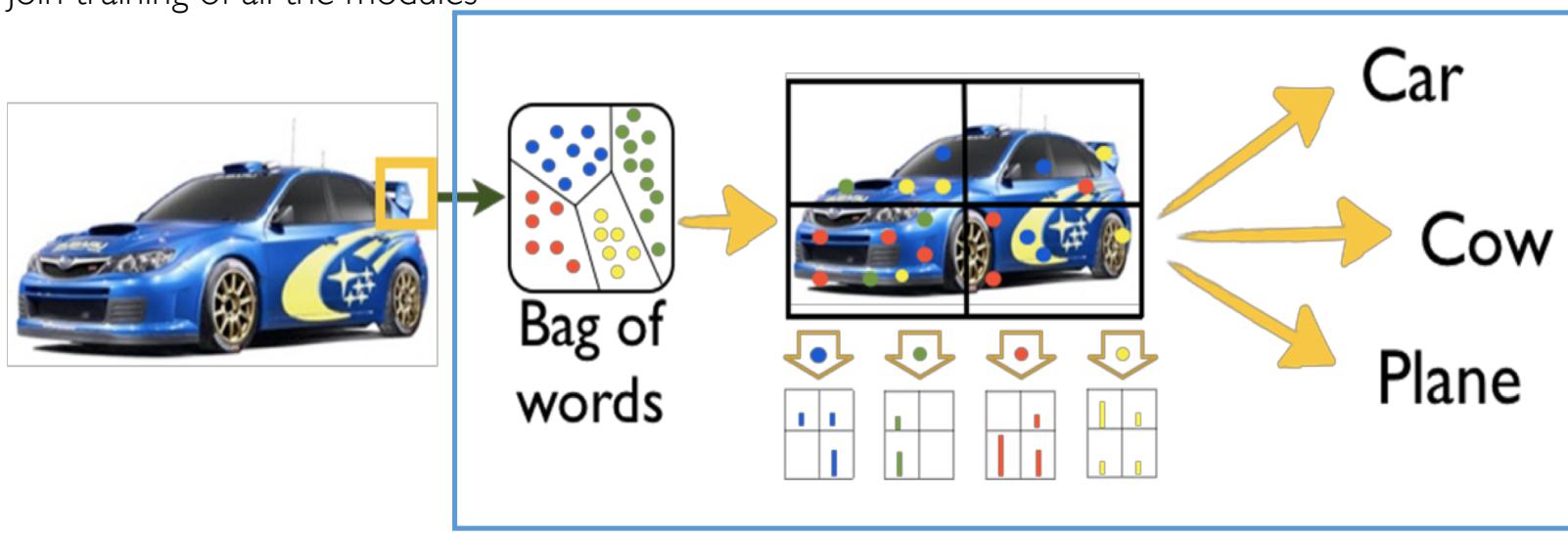
Basic Notions and Intuitions

- Joint training
 - Training at least two modules together
 - Pre-DL era: hand-engineer all the modules, train only the last one (e.g., classifier)



Basic Notions and Intuitions

- Joint training
 - Training at least two modules together
 - Pre-DL era: hand-engineer all the modules, train only the last one (e.g., classifier)
- End-to-end training (e2e training)
 - Join training of all the modules



Train together classifier and 'high-level' and 'low-level' features

Basic Notions and Intuitions

- Joint training
 - Training at least two modules together
 - Pre-DL era: hand-engineer all the modules, train only the last one (e.g., classifier)
- End-to-end training (e2e training)
 - Joint training of all the modules
- Back-propagation (Backprop)
 - Technique for training Neural Networks
 - Compositions of functions
 - Technically it computes a Jacobian matrix by using a chain rule
 - But it does it efficiently

$$\frac{\partial}{w} f(g(w)) = \frac{\partial}{g(w)} f(g(w)) \cdot \frac{\partial}{w} g(w)$$

Basic Notions and Intuitions

- Joint training
 - Training at least two modules together
 - Pre-DL era: hand-engineer all the modules, train only the last one (e.g., classifier)
- End-to-end training (e2e training)
 - Joint training of all the modules
- Back-propagation (Backprop)
 - Technique for training Neural Networks (compositions of functions)
 - Technically it computes a Jacobian matrix by using a chain rule
 - But it does it efficiently
- Optimization
 - We often try to find the best parameters that 'explains data' under the current model (e.g., neural network)
 - We did it in the curve-fitting example
 - There are, however, other alternatives like sampling

$$\boldsymbol{w}^* = \arg \min_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w})$$

Thank You (TA's Needed)

Email for Assignment Submissions: pir-assignments@l3s.de

Email for Lecture related queries : leonhardt@l3s.de, zhang@l3s.de

Use [DL2019] in the subject line