IN5550: Neural Methods in Natural Language Processing

Introduction

Andrey Kutuzov, Lilja Øvrelid,
Vinit Ravishankar, David Samuel, Erik Velldal, & You

University of Oslo

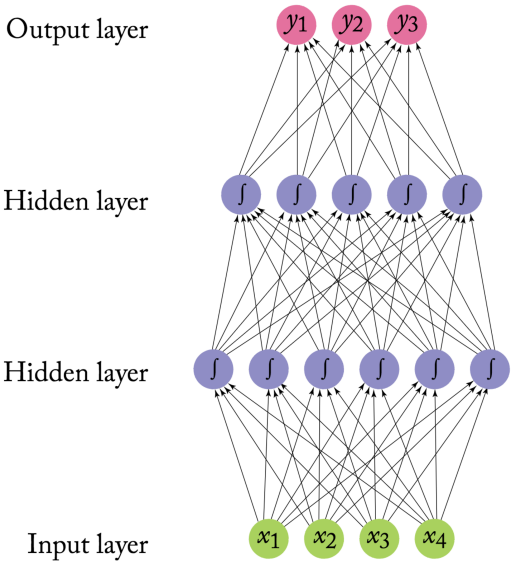
January 20, 2022



Today

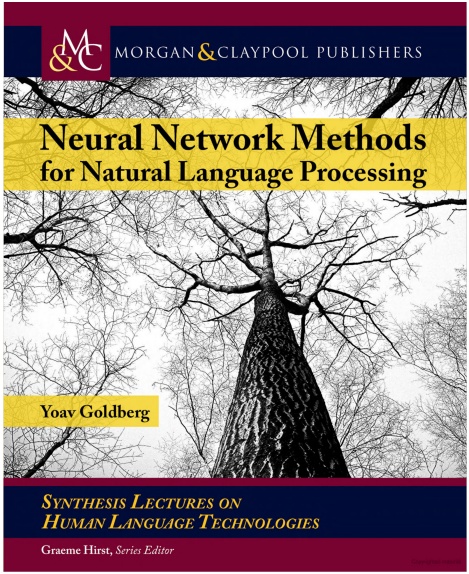
- Introduction
- Why a course on neural methods for NLP?
  - Motivation
  - Historical trends
  - Some contrasts between NNs and traditional ML
  - Success stories
- Course overview
  - Programming environment
  - Lab sessions and obligatory assignments
  - Home exam

What is a neural network (NN) model?



- Neural networks (NNs) are a family of powerful machine learning models.
- Weakly based on the metaphor of a neuron.
- Non-linear transformations of the input in the 'hidden layers'.
- Learns not only to make predictions, but how to represent the data
- 'Deep Learning': NNs with several hidden layers.

Textbook



- *Neural Network Methods for Natural Language Processing* by Yoav Goldberg (Morgan & Claypool Publishers, 2017).
- Free e-version available through UiO; `http://oria.no/`
- Supplementary research papers will be added.
- Check the schedule for the syllabus for each lecture!

Paradigm shifts in NLP (and AI at large)

- 50s–80s: mostly rule-based (symbolic / rationalist) approaches.
- Hand-crafted formal rules and manually encoded knowledge.
- (Though some AI research on neural networks in the 40s and 50s).

Paradigm shifts in NLP (and AI at large)

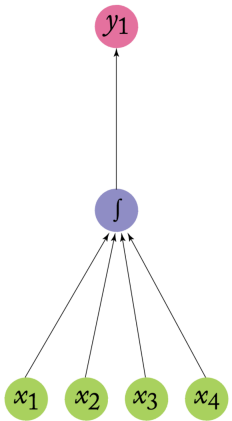
- 50s–80s: mostly rule-based (symbolic / rationalist) approaches.
- Hand-crafted formal rules and manually encoded knowledge.
- (Though some AI research on neural networks in the 40s and 50s).
- Late 80s: success with statistical ('empirical') methods in the fields of speech recognition and machine translation.
- Late 90s: massive shift towards machine-learning.
  - Based on automatically inferring statistical patterns from data.
- 00s: Machine-learning methods dominant.

Paradigm shifts in NLP (and AI at large)

- 50s–80s: mostly rule-based (symbolic / rationalist) approaches.
- Hand-crafted formal rules and manually encoded knowledge.
- (Though some AI research on neural networks in the 40s and 50s).
- Late 80s: success with statistical ('empirical') methods in the fields of speech recognition and machine translation.
- Late 90s: massive shift towards machine-learning.
  - Based on automatically inferring statistical patterns from data.
- 00s: Machine-learning methods dominant.
- 2010–: neural methods increasingly replacing traditional ML.
- A revival of techniques first considered in the 40s and 50s,
- but recent developments in computational power and availability of data have given great breakthroughs in scalability and accuracy.

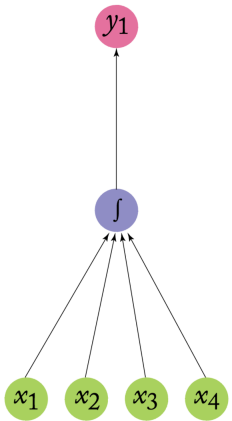
What is a classifier?

- Very high-level:
- A learned mapping from inputs to outputs.
- Learned from labeled examples; a set of objects with correct class labels.
- 1st step in creating a classifier; defining a representation of the input!
- Typically given as a feature vector.



What is a classifier?

- Very high-level:
- A learned mapping from inputs to outputs.
- Learned from labeled examples; a set of objects with correct class labels.
- 1st step in creating a classifier; defining a representation of the input!
- Typically given as a feature vector.



'*A good representation does most of the work.*'

– Anderson & Rosenfeld, 1988

Traditional ML; based on feature engineering

- The art of designing features for representing objects to a classifier.
- Typically also involves large-scale empirical tuning to identify the best-performing configuration.
- Although there is much overlap in the types of features used across tasks, performance is highly dependent on the specific task and dataset.
- Results in very high-dimensional and sparse feature vectors.
- Challenges; discreteness and lack of information sharing.

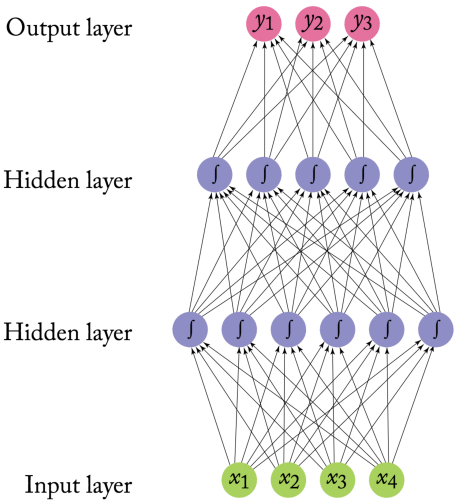
Representation learning

- One of the greatest strengths of neural models is their ability to learn their own internal task-specific representations of the data.

Representation learning

- One of the greatest strengths of neural models is their ability to learn their own internal task-specific representations of the data.
- One important example within NLP: word embeddings
  - Low-dimensional and dense vector representations of words.
  - Often pre-trained on unlabeled data for a text prediction task;
  - so-called language modeling;
  - the learned word representations can then be used to form the input representations for other down-stream supervised NN models.

Some pros and cons of NNs

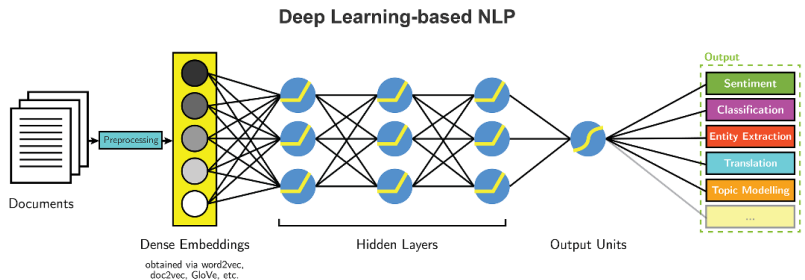


NNs hold promise to...

- reduce manual feature engineering,
- make better use of unlabeled data,
- through the use of distributed continuous input representations,
- and further learn more task-adapted internal representations of the data.
- Tends to scale better with more data.
- Though at the cost of being less interpretable and more complex with more parameters.

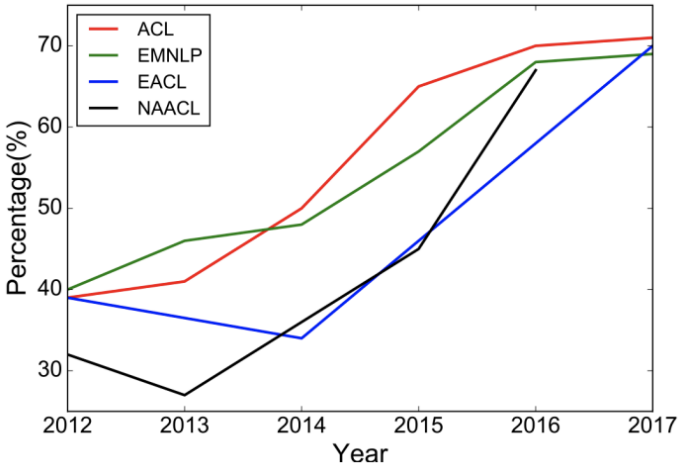
Neural methods in NLP

- Most NLP tasks approached as classification problems, using supervised machine learning
- Since at least 2014, neural networks completely dominating.
- Very successful on NLP benchmarks



https://ebookreading.net/view/book/EB9781787121423397.html

Neural methods in NLP



'Recent trends in deep learning based natural language processing' [Young et al., 2018]

Neural NLP tasks

Machine translation (Google Translate)

Neural NLP tasks

Machine translation (Google Translate)

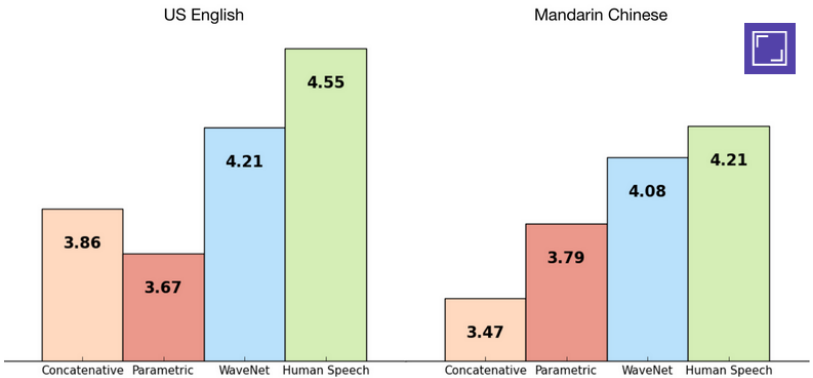
- No 1:
  Kilimanjaro is a snow-covered mountain 19,710 feet high, and is said to be the highest mountain in Africa. Its western summit is called the Masai “Ngaje Ngai,” the House of God. Close to the western summit there is the dried and frozen carcass of a leopard. No one has explained what the leopard was seeking at that altitude.
- No 2:
  Kilimanjaro is a mountain of 19,710 feet covered with snow and is said to be the highest mountain in Africa. The summit of the west is called “Ngaje Ngai” in Masai, the house of God. Near the top of the west there is a dry and frozen dead body of leopard. No one has ever explained what leopard wanted at that altitude.

Machine translation (Google Translate)

- No 3:
  Kilimanjaro is 19,710 feet of the mountain covered with snow, and it is said that the highest mountain in Africa. Top of the west, “Ngaje Ngai” in the Maasai language, has been referred to as the house of God. The top close to the west, there is a dry, frozen carcass of a leopard. Whether the leopard had what the demand at that altitude, there is no that nobody explained.

Neural NLP tasks

Text-to-Speech (van den Oord et al., 2016):



(https://deepmind.com/blog/wavenet-generative-model-raw-audio/)

Success stories

- Neural models have caused great advances in the field of image processing
- New tasks combining image and language are emerging

Success stories

- Neural models have caused great advances in the field of image processing
- New tasks combining image and language are emerging
- Visual Question Answering:

Who is wearing glasses?

man

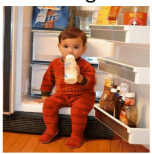


woman

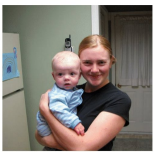


Where is the child sitting?

fridge

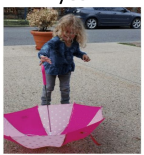


arms

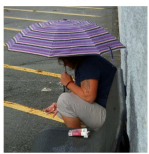


Is the umbrella upside down?

yes



no

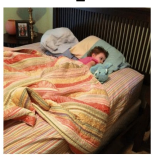


How many children are in the bed?

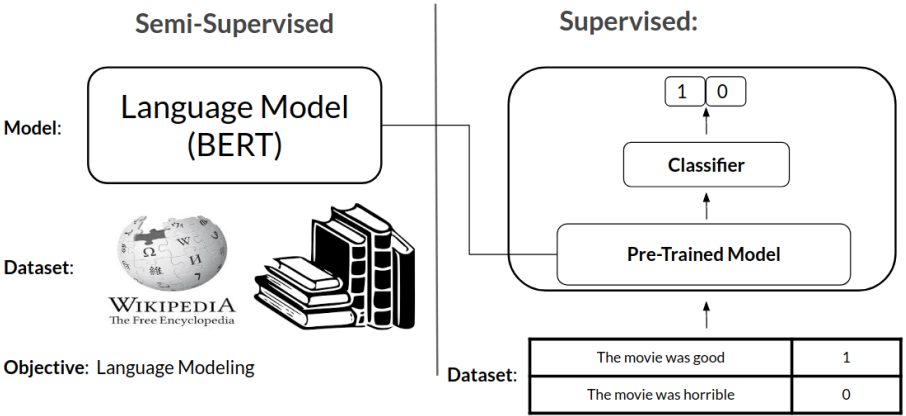
2



1

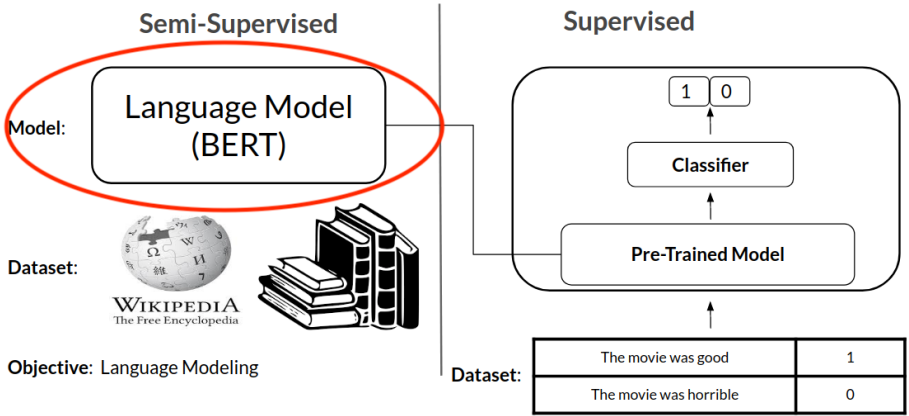


Typical neural NLP system



(https://blog.insightdatascience.com)

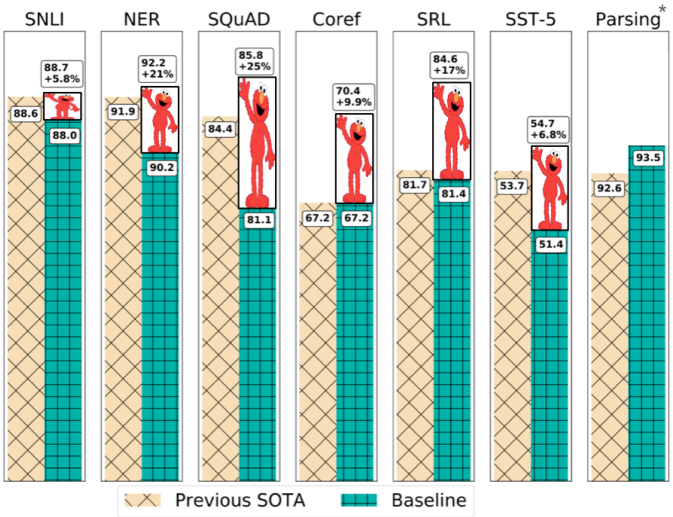
Typical neural NLP system



(https://blog.insightdatascience.com)

Language models

Pre-trained language models



*Kitaev and Klein, ACL 2018 (see also Joshi et al., ACL 2018)

Language models

- Given raw (i.e. unlabeled) text as input,
- Language Models (LMs) are trained to predict masked out parts of a sentence

It was a ~~dark~~ ??? and stormy ~~night~~ ???.

- 'Self-supervised'

Language models

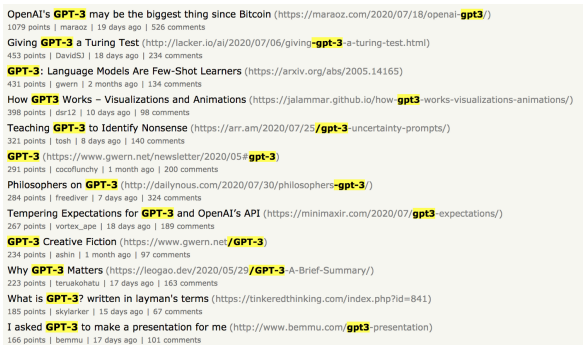
- Given raw (i.e. unlabeled) text as input,
- Language Models (LMs) are trained to predict masked out parts of a sentence

It was a ~~dark~~ ??? and stormy ~~night~~ ???.

- 'Self-supervised'
- Neural LMs are used as representation learners;

Language models

- Since then, a number of different models have been proposed: BERT, XLNet, GPT-2, GPT-3, etc.
- Extreme media attention:



(`https://towardsdatascience.com/why-are-you-seeing-gpt-3-everywhere-f156a71b77b0`)

- Limitations: [Bender and Koller, 2020] Best Theme Paper at ACL

Language models

- A year ago (`https://twitter.com/ltgoslo`):

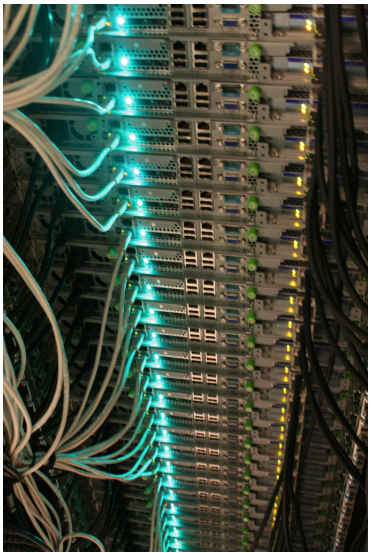


- New Norwegian language models (LMs) are coming...
- ...and Pre-trained LMs and Transformers lectures are coming later in the course

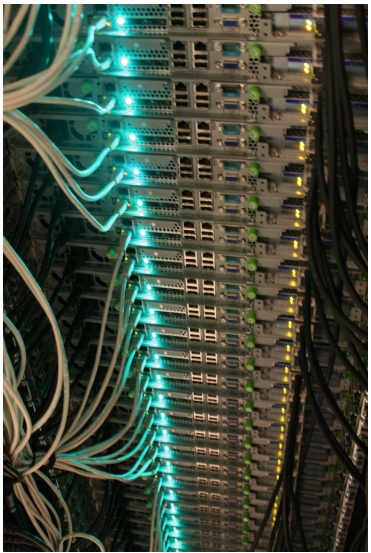
Getting to Know Each Other

A brief anonymous survey:

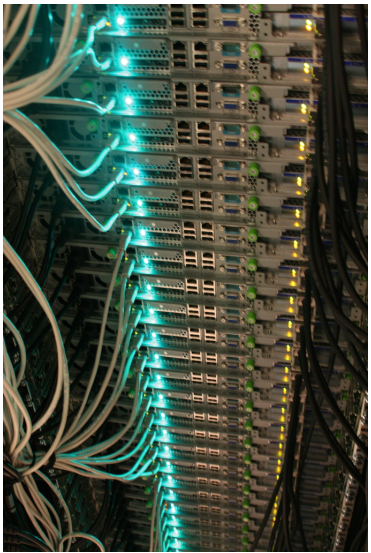
https://nettskjema.no/a/240888



- Strong practical and programming elements in this course;

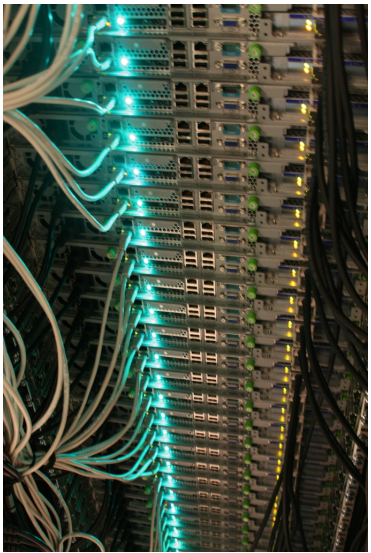


- Strong practical and programming elements in this course;
- three obligatory assignments; all predominantly ‘hands-on’;



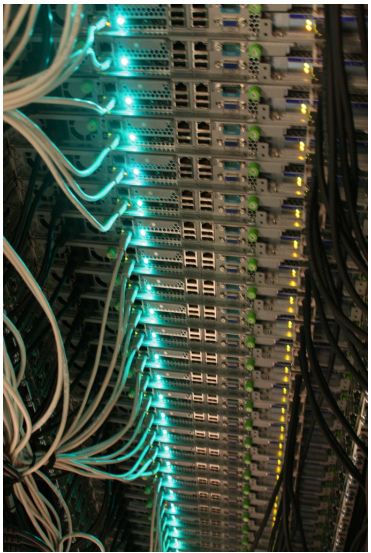
- Strong practical and programming elements in this course;
- three obligatory assignments; all predominantly ‘hands-on’;
- representative of sub-problems in typical MSc theses at LTG;

Scientific Programming in IN5550



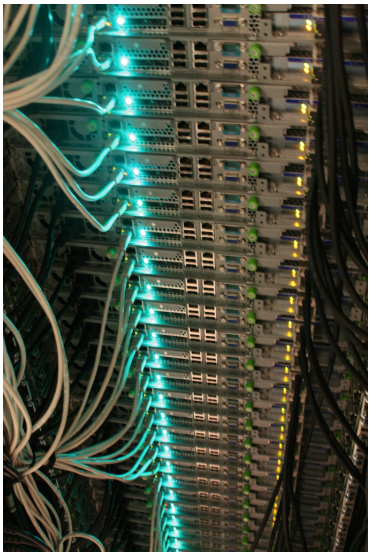
- Strong practical and programming elements in this course;
- three obligatory assignments; all predominantly 'hands-on';
- representative of sub-problems in typical MSc theses at LTG;
- fairly data- and compute-intensive throughout the semester;

Scientific Programming in IN5550



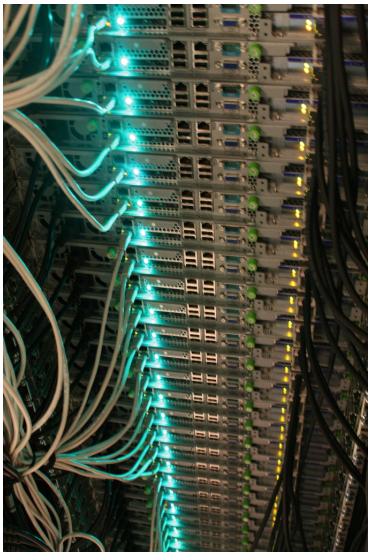
- Strong practical and programming elements in this course;
- three obligatory assignments; all predominantly 'hands-on';
- representative of sub-problems in typical MSc theses at LTG;
- fairly data- and compute-intensive throughout the semester;
- learn how to use the national Saga supercluster

Scientific Programming in IN5550



- Strong practical and programming elements in this course;
- three obligatory assignments; all predominantly 'hands-on';
- representative of sub-problems in typical MSc theses at LTG;
- fairly data- and compute-intensive throughout the semester;
- learn how to use the national Saga supercluster ($10{,}000^{+}$ CPUs and GPUs);

Scientific Programming in IN5550



- Strong practical and programming elements in this course;
- three obligatory assignments; all predominantly 'hands-on';
- representative of sub-problems in typical MSc theses at LTG;
- fairly data- and compute-intensive throughout the semester;
- learn how to use the national Saga supercluster (10,000$^+$ CPUs and GPUs);
- practical skills: Linux, batch jobs, experimentation, tuning, ...



- Python is a relatively simple programming language with great popularity for neural network–based machine learning;



- Python is a relatively simple programming language with great popularity for neural network–based machine learning;
- it provides a very convenient, high-level coding with a gentle learning curve; works easily across different platforms;



- Python is a relatively simple programming language with great popularity for neural network–based machine learning;
- it provides a very convenient, high-level coding with a gentle learning curve; works easily across different platforms;
- comprehensive standard library; ecosystem of community-maintained add-on modules with specialized (and optimized) functionality;



- Python is a relatively simple programming language with great popularity for neural network–based machine learning;
- it provides a very convenient, high-level coding with a gentle learning curve; works easily across different platforms;
- comprehensive standard library; ecosystem of community-maintained add-on modules with specialized (and optimized) functionality;
- pretty much everything open-source; we provide ready-to-use ‘virtual lab’ on Saga; in principle possible to install ‘at home’ (for development).

A Menagerie of Interoperable Modules









- NumPy for efficient multi-dimensional arrays (aka ‘tensors’) and general linear algebra;

A Menagerie of Interoperable Modules









- NumPy for efficient multi-dimensional arrays (aka ‘tensors’) and general linear algebra;
- PyTorch (Facebook) is a mature software infrastructure for deep learning (built natively in C^{++});
  - Any self-respecting technology giant today develops their own DL framework (Google has TensorFlow);
  - plus a few from university environments;
  - open source $\rightarrow$ community involvement;

A Menagerie of Interoperable Modules









- NumPy for efficient multi-dimensional arrays (aka 'tensors') and general linear algebra;
- PyTorch (Facebook) is a mature software infrastructure for deep learning (built natively in C^{++});
  - Any self-respecting technology giant today develops their own DL framework (Google has TensorFlow);
  - plus a few from university environments;
  - open source $\rightarrow$ community involvement;
- Gensim for (large-scale) distributional modeling $\rightarrow$ 'word embeddings';

A Menagerie of Interoperable Modules









- NumPy for efficient multi-dimensional arrays (aka 'tensors') and general linear algebra;
- PyTorch (Facebook) is a mature software infrastructure for deep learning (built natively in C^{++});
  - Any self-respecting technology giant today develops their own DL framework (Google has TensorFlow);
  - plus a few from university environments;
  - open source $\rightarrow$ community involvement;
- Gensim for (large-scale) distributional modeling $\rightarrow$ 'word embeddings';
- all integrated in standard modules on Saga (see the course page).

Lectures

No traditional lectures (except this one)

- Pre-recorded short videos;
- Published: every Tuesday afternoon (starting next week);
- watch whenever convenient;
- find them in the course schedule under 'Resources/curriculum';
- every Thursday at 12:15: interactive Q&A in Zoom.



Labs and Obligatory Assignments (*grupper & obliger*)

- Lab: Tuesday, 12:15–14:00 (starting next week);
- three obligatory assignments;
- rigid schedule; all assignments must be submitted;
- minimum 60% of points across all three required to qualify for exam;
- no re-submissions.



Labs and Obligatory Assignments (*grupper & obliger*)

- Lab: Tuesday, 12:15–14:00 (starting next week);
- three obligatory assignments;
- rigid schedule; all assignments must be submitted;
- minimum 60% of points across all three required to qualify for exam;
- no re-submissions.



Mechanics

- Schedule: `https://www.uio.no/studier/emner/matnat/ifi/IN5550/v22/obligatory-exercises/`
- Group work warmly encouraged (1–3 students); please give it a shot!
- Final exam as 'baby' MSc project in May; again, team work possible.

Home Exam

General Idea

- Our guiding metaphor: Preparing a scientific paper for publication.

Home Exam

General Idea

- Our guiding metaphor: Preparing a scientific paper for publication.

4th IN5550 Workshop on Neural NLP (WNNLP 2022)

Home Exam

General Idea

- Our guiding metaphor: Preparing a scientific paper for publication.

4th IN5550 Workshop on Neural NLP (WNNLP 2022)

Standard Process

(1) Experimentation

(2) Analysis

(3) Paper Submission

(4) Reviewing

(5) Camera-Ready Manuscript

(6) Presentation

For Example: The Annual ACL Conference

Important dates

Paper submission deadlines

- Deadline for submission to ACL Rolling Review: November 15, 2021
- Anonymity period for ACL 2022 starts: December 16, 2021
- Commitment deadline for ACL 2022: January 15, 2022
- Notification of acceptance: February 23, 2022
- Deadline to withdraw papers: February 23, 2022
- Camera-ready due: March 15, 2022

WNNLP 2022 (IN5550 Final Exam)

Three weeks for submissions; three weeks for reviewing and revision.

Home Exam

Our timeline

- Project presentations: April 21st
- Home Exam: April 28th – May 19th
- Paper submission May 19th
- Review period: May 23rd – May 30th
- Camera Ready: June 7th
- Oral presentations: June 9th

Course Communication

- Questions?

Course Communication

- Questions?
  - On-line chat room: https://mattermost.uio.no/ifi-undervisning/channels/in5550
  - GitHub issues: https://github.uio.no/in5550/2022/issues
  - email: in5550-help@ifi.uio.no reaches all course staff.
- submissions, solutions, etc. on the GitHub instance at UiO.

Course Communication

- Questions?
  - On-line chat room: https://mattermost.uio.no/ifi-undervisning/channels/in5550
  - GitHub issues: https://github.uio.no/in5550/2022/issues
  - email: in5550-help@ifi.uio.no reaches all course staff.
- submissions, solutions, etc. on the GitHub instance at UiO.
- Messages:
  - Read (or forward) your UiO email regularly;
  - Check the course page and schedule regularly;
  - Participate in the discussions on Mattermost and GitHub

Next week

Foundations

- Introducing mathematical notation for describing neural and not-neural classifiers.
- From linear regression to feed-forward networks and multi-layer perceptrons.
- Linear vs. non-linear decision boundaries

References I


Bender, E. M. and Koller, A. (2020).
Climbing towards NLU: On meaning, form, and understanding in the age of data.
In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 5185–5198, Online. Association for Computational Linguistics.


Young, T., Hazarika, D., Poria, S., and Cambria, E. (2018).
Recent trends in deep learning based natural language processing.
IEEE Computational intelligence magazine, 13(3):55–75.