

Part III: BERT



Outline

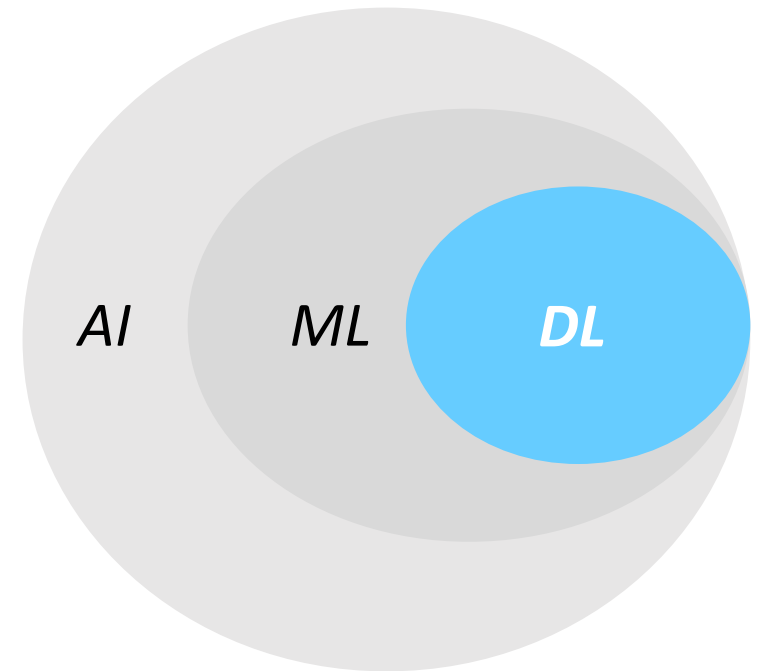
- i. Introduction to Deep Learning
- ii. BERT
 - i. Basic Application: Sentiment Analysis
 - i. Input Pre-Processing
 - ii. Pre-Training
 - iii. Fine-Tuning
 - ii. Advanced Application: Aspect-Based Sentiment Analysis

Part III: BERT

Introduction to Deep Learning

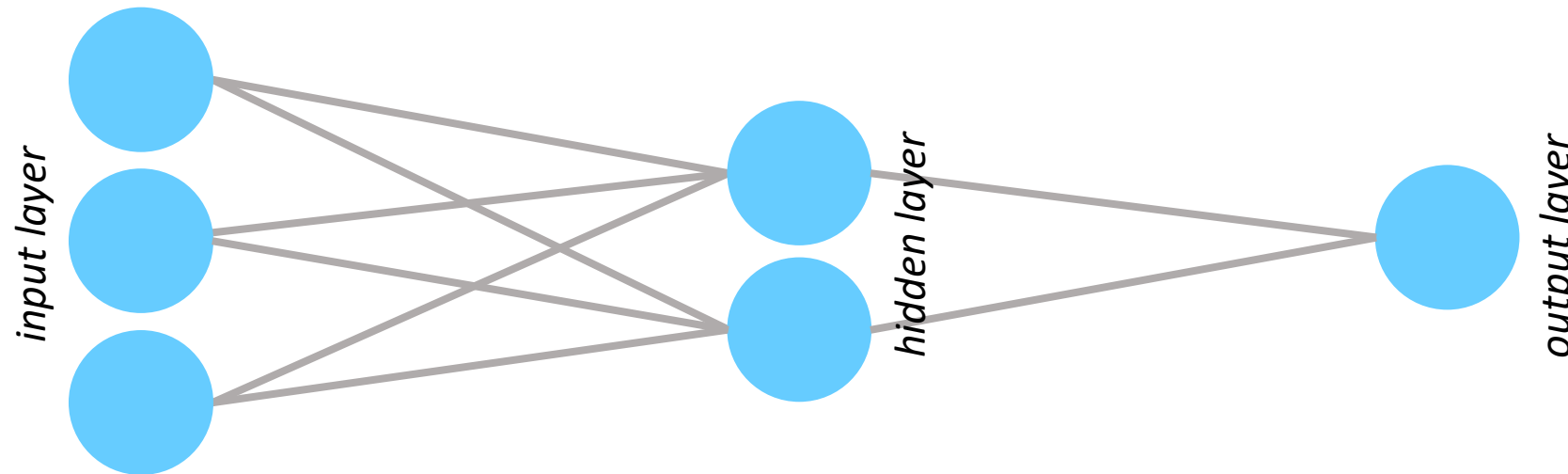
Intro DL Deep Learning

- Form of **artificial intelligence (AI)**
- Goal: simulate **human** intelligence processes by computer systems
- **Deep neural networks (DNN)**
inspired by organization of neurons in biological brains
- Applications of DL
 - Machine translation: Google Translate
 - Speech recognition and generation (OK, Google...)
 - Self-driving cars . . .



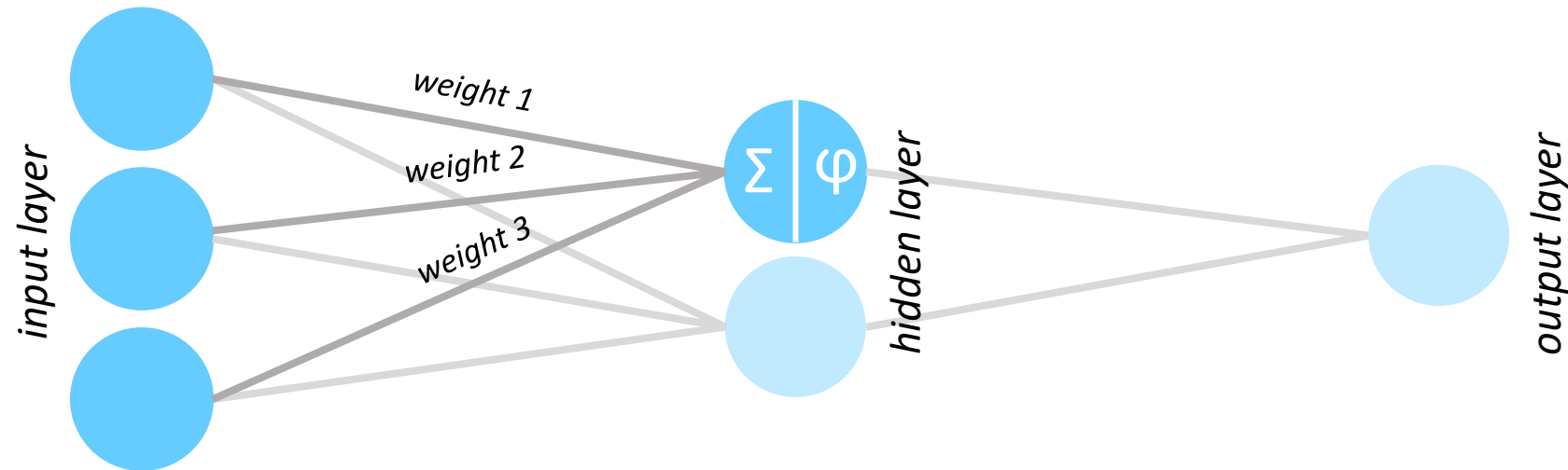
Intro DL Basic Network Structure

- **General goal:** learn complex functions, recognize patterns and generalize them to new data
- Core elements: artificial **neurons**



Intro DL Basic Network Structure

- Basic structure: **feed-forward** NN *aka* **multi-layer perceptron (MLP)**
- Optimization: **gradient descent** via **backpropagation**
- Hyperparameters: learning rate, batch size, # training epochs, ...



Intro DL Transfer Learning

- **Idea:** keep knowledge gained during solving a problem and later apply it to other (related) problems



why start from zero every time?

- Akin to human learning processes
- Particularly valuable in **NLP** – basic linguistic concepts required for all sorts of tasks

Part III: BERT

BERT

BERT Idea

- **Bi-directional Encoder Representations from Transformers**
- Modern, DL-based NLP model (2018)
- Architecture developed by Google researchers
 - Power of the transformer model: **attention** mechanism allows for much more parallelization
 - Power of transfer learning: **pre-train** language representations and **adapt** to a specific target task
- Implementation: Transformers library by Hugging Face (Python)



BERT Input Pre-Processing

- **Tokenization**: split strings in word or sub-word token strings
- Converting tokens strings to ids / integers
- Special tokens: [CLS], [SEP]
- Consumption limit: sequences of up to 512 tokens
→ **Padding** or **truncation**
- Token, sentence and positional **embeddings**

*[CLS] Die Ausgrenzung von MigrantInnen von der # EssenerTafel ist inakzeptabel und rassistisch.
Wir dürfen nicht zulassen, dass die Ärmsten gegeneinander ausgespielt werden. [SEP]*

BERT Pre-Training

- **Masked language modeling (MLM)**

1. Randomly sample 15% of tokens
2. Replace 80% of them by [MASK], 10% by a random token, keep 10% unchanged
3. Given a sequence, predict [MASK]-ed tokens

[CLS] Die Ausgrenzung von [MASK] von der # EssenerTafel ist inakzeptabel und [MASK]. Wir dürfen nicht zulassen, dass die [MASK] gegeneinander ausgespielt werden. [SEP]



targets: (MigrantInnen, rassistisch, Ärmsten)

BERT Pre-Training

- **Next sentence prediction (NSP)**

1. Randomly sample the next sentence for 50% of two-sentence pairs
2. Given two sentences, predict whether the second sentence follows the first one

[CLS] Die Ausgrenzung von [MASK] von der # EssenerTafel ist inakzeptabel und [MASK]. Wir dürfen nicht zulassen, dass die [MASK] gegeneinander ausgespielt werden. [SEP]

Wir dürfen nicht zulassen, dass die [MASK] gegeneinander ausgespielt werden. [SEP]



target label: IsNextSentence

BERT Fine-Tuning

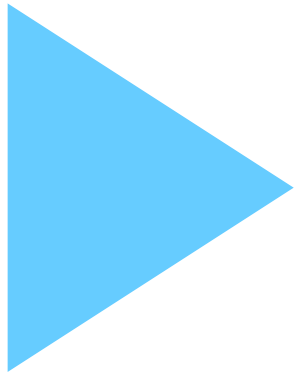
- Various language **tasks**: question answering, named entity recognition, sequence classification, etc.
- Exchange the **output layer** from pre-training & update parameters
 - Train and validation sets for model training procedure (80%)
 - Test set for evaluation (20%, unseen data)
- Set **hyperparameters** as recommended by the authors:
 - Learning rate $\in \{0.00005, 0.00003, 0.00002\}$
 - Batch size $\in \{16, 32\}$
 - Number of training epochs $\in \{2, 3, 4\}$

BERT Aspect-Based Sentiment Analysis

- **Goal:** classification of sentiment with respect to certain aspects of an entity
- **Idea:** sentiments potentially quite different across aspects
- General approach
 - Post-training
 - Aspect extraction
 - Aspect sentiment classification

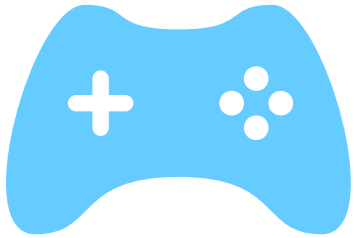
*[CLS] Die Ausgrenzung von **MigrantInnen** von der # EssenerTafel ist inakzeptabel und **rassistisch**.
Wir dürfen nicht zulassen, dass die Ärmsten gegeneinander ausgespielt werden. [SEP]*

BERT Sentiment Analysis with BERT



Demo 12: SA with BERT

BERT Sentiment Analysis with BERT



Exercise 6: SA with BERT

Part III: BERT

Literature and References

Chollet, F. (2017): Deep Learning with Python.

URL: <https://www.manning.com/books/deep-learning-with-python>

Chollet, F., and Allaire, J. (2018): Deep Learning with R.

URL: <https://www.manning.com/books/deep-learning-with-r>

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2019): BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv:1810.04805v2.

Goodfellow, I., Bengio, Y., and Courville, A. (2016): Deep Learning.

URL: <https://www.deeplearningbook.org/>.

Pan, S., and Yiang, Q. (2010): A Survey on Transfer Learning, *IEEE Transactions on Knowledge and Data Engineering* 22(10).

Ruder, S. (2019): Neural Transfer Learning for Natural Language Processing. PhD Thesis, National University of Ireland. URL: [https://ruder.io/thesis/neural transfer](https://ruder.io/thesis/neural%20transfer).

Vaswani, A., et al. (2017): Attention Is All You Need, arXiv:1706.03762v5.

Xu, H., Liu, B., Shu, L., and Yu, P. (2019): BERT Post-Training for Review Reading Comprehension and Aspect-Based Sentiment Analysis, *Proceedings of NAACL-HLT 2019*, pp. 2324–2335.

Zhang, A., Lipton, Z., Mu, L., and Smola, A.: Dive into Deep Learning. URL: <https://d2l.ai/>.

<https://deepset.ai/german-bert>

<https://huggingface.co/dbmdz/bert-base-german-cased>

https://huggingface.co/transformers/model_doc/bert.html#tfbertforsequenceclassification