First of all...

...before we start

Bon St. Jordi!

NLP Block, 2–5 Lectures

Cristina España-Bonet

UdS & DFKI, Saarbrücken, Germany

Artificial Intelligence with Deep Learning

23rd April 2019

Seleted Topics on NLP

In 12 hours we'll talk about...

- Word Embeddings
- Text Classification
- Sentiment Analysis
- Sequence Modeling
- Language Modeling
- Machine Translation
- Dialogue Systems

Seleted Topics on NLP

In 12 hours we'll talk about...

- Introduction to the *language setting* in deep learning (with word embeddings and text classification)
- Selected NLP topics
- Lots of ideas, plots, and few equations (you know neural networks details already!)
- Brief introduction of classical architectures before going into deep learning

Seleted Topics on NLP

Distribution

Lectures: 7.5 hours on the previous topics

Labs: 3 hours

- Data Processing
- Word Embeddings
- Machine Translation

Research Talk: 1.5 hours

- Embeddings
- Self-supervised NMT
- Document-level NMT

Seleted Topics on NLP

Get prepared for the labs

Data processing and word embeddings

- python, jupyter-notebook
- nltk, gensim, matplotlib, sklearn
- word_tokenize, PorterStemmer, WordNetLemmatizer

Data processing and machine translation

- Marian NMT
- Moses scripts
- Sub-word NMT

Seleted Topics on NLP

Marian installation

```
Download Intel MKL (recomended by Marian developer)
https://software.intel.com/en-us/mkl
(needs a free register, downloads ~800 MB of data)

Make it visible:
export MKL_ROOT=$pathIntel/intel/mkl
export MKL_INCLUDE=$MKL_ROOT/include
export MKL_LIBRARY=$MKL_ROOT/lib/intel64
source $pathIntel/intel/mkl/bin/mklvars.sh intel64
source $pathIntel/intel/bin/compilervars.sh intel64
export CMAKE_INCLUDE_PATH=$MKL_INCLUDE:$CMAKE_INCLUDE_PATH
export CMAKE_LIBRARY_PATH=$MKL_LIBRARY:$CMAKE_LIBRARY_PATH

Install Boost if you don't have it:
sudo apt-get install libboost-all-dev

Install Marian. This assumes only CPU usage and builds binaries
that can be used in other machines with similar configurations
$ cmake .. -DUSE_STATIC_LIBS=on -DCOMPILE_CUDA=off
$ make -j8
```