Deep Learning for NLP
Lecture 2: Introduction to Theano

Nils Reimers

TECHNISCHE UNIVERSITÄT DARMSTADT

Course-Website: www.deeplearning4nlp.com

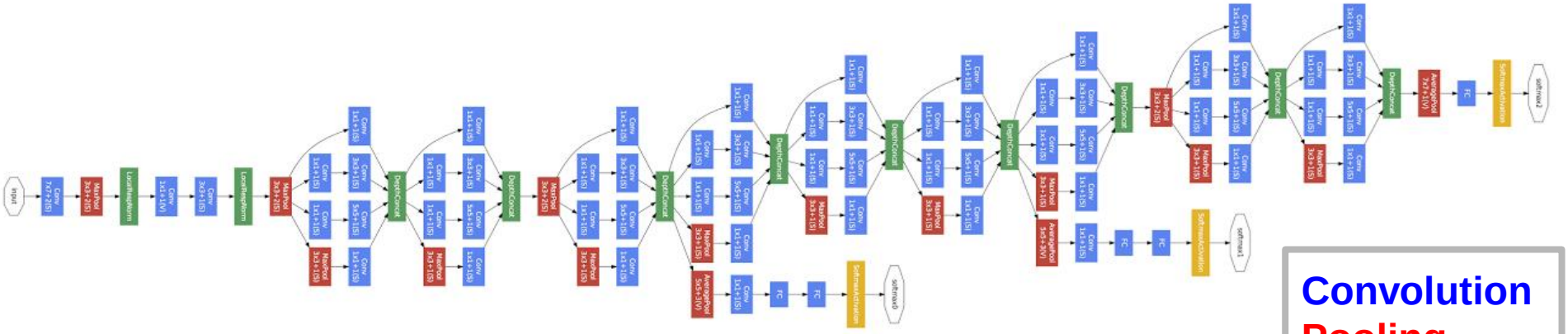
Please have your Python environment up and running

Recommended Readings

TECHNISCHE UNIVERSITÄT DARMSTADT

- Install Python (2.7), NumPy, SciPy and Theano. (Installing Theano for Ubuntu)
- Install Lasagne
- Refresh your knowledge on Python and Numpy:
  - Python and Numpy Tutorial
  - Python-Tutorial and Numpy refresher from the Theano website
- **Hint:** You can install Python, Theano etc. on you local desktop machine and log into it via SSH or via IPython Notebook during class

What does a Deep Network can look like?



Convolution
Pooling
Softmax
Other

- Google's Entry for the 2014 ImageNet Challenge
- "Just" 5 million parameters – 20MB model size
- Uses ReLu (sigmoid/tanh does not work in really deep networks)
  - See Glorot et al., 2011, *Deep Sparse Rectifier Neural Networks*

Source: Szegedy et al., 2014, Going Deeper with Convolutions

Requirements for a Framework

TECHNISCHE UNIVERSITÄT DARMSTADT

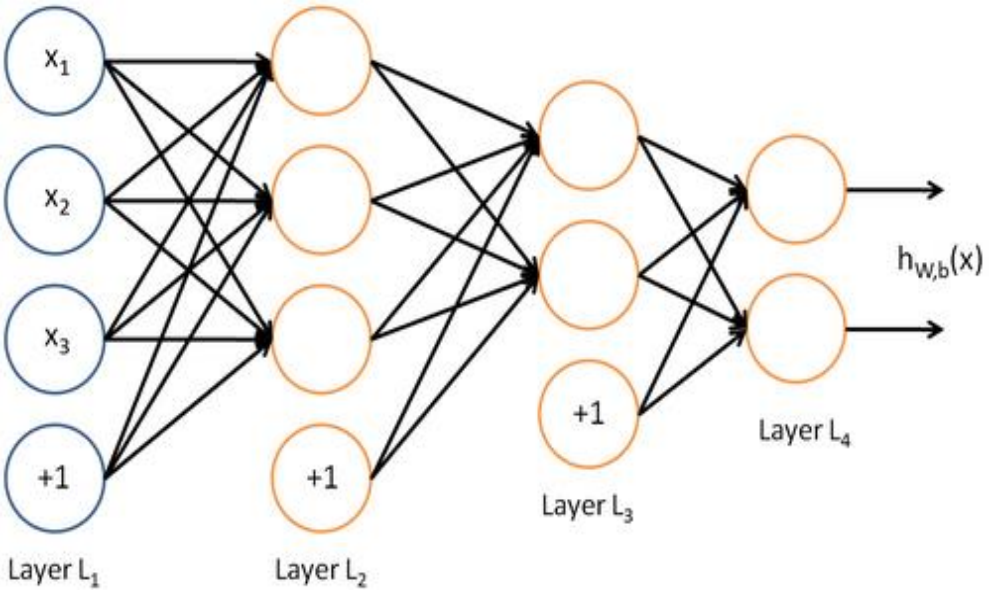
- To train Deep Neural Networks requires billions of operations
  - Google had trained some systems on up to 16.000 cores
- Performance in training time is **crucial**
  - More data = better results = your paper get published
  - Some times: Larger network = better results = your paper get published
  - Slow computation is not acceptable => Choose the framework that is the fastest
- Nearly all operations are matrix operations (multiplications, additions)
  - Easy syntax for matrices desired
  - Optimizing matrix multiplication for speed is hard, simple “two loop solution” is way too slow
- Must be runnable on a GPU
- Nice to have: Easy computation of gradients

Neural Network as One Long Function

- Neural Networks can be expressed as one long function of vector and matrix operations

$$output = \text{softmax}(b_3 + W_3 \tanh(b_2 + W_2 \tanh(b_1 + W_1 x)))$$

$$E(x, W, b) = -\log(\text{softmax}(b_3 + W_3 \tanh(b_2 + W_2 \tanh(b_1 + W_1 x)))_y)$$



Img-Source: http://ufldl.stanford.edu/wiki/

Common Frameworks

TECHNISCHE UNIVERSITÄT DARMSTADT

- C/C++:
  - If you need maximum performance, start from scratch
- Matlab
- Caffe
  - Ported Matlab's implementation of fast convolutional nets to C
  - Mainly used for machine-vision
- Torch:
  - Based on Lua,
  - Used by a lot of companies (Google Deep Mind, Facebook, IBM)
- Theano
  - Python based framework
  - Main framework used in the research community
  - For comparison: http://fastml.com/torch-vs-theano/

Introduction to Theano

TECHNISCHE UNIVERSITÄT DARMSTADT

Theano

TECHNISCHE UNIVERSITÄT DARMSTADT

Advantages

- Python library with tight integration of NumPy
  - Easy syntax for matrix operations ✓
- Transparent use of GPU (speed-up of up to 140x) ✓
- Efficient symbolic differentiation (Theano computes the gradient) ✓
- Speed and stability optimizations ✓
- Calculations are dynamically mapped to C code
  - We do our computations as fast as we would have written it in C
  - Great performance (>10 faster than Java in my experiments) ✓

Disadvantages

- Debugging is really hard

Some note on the installation of Theano

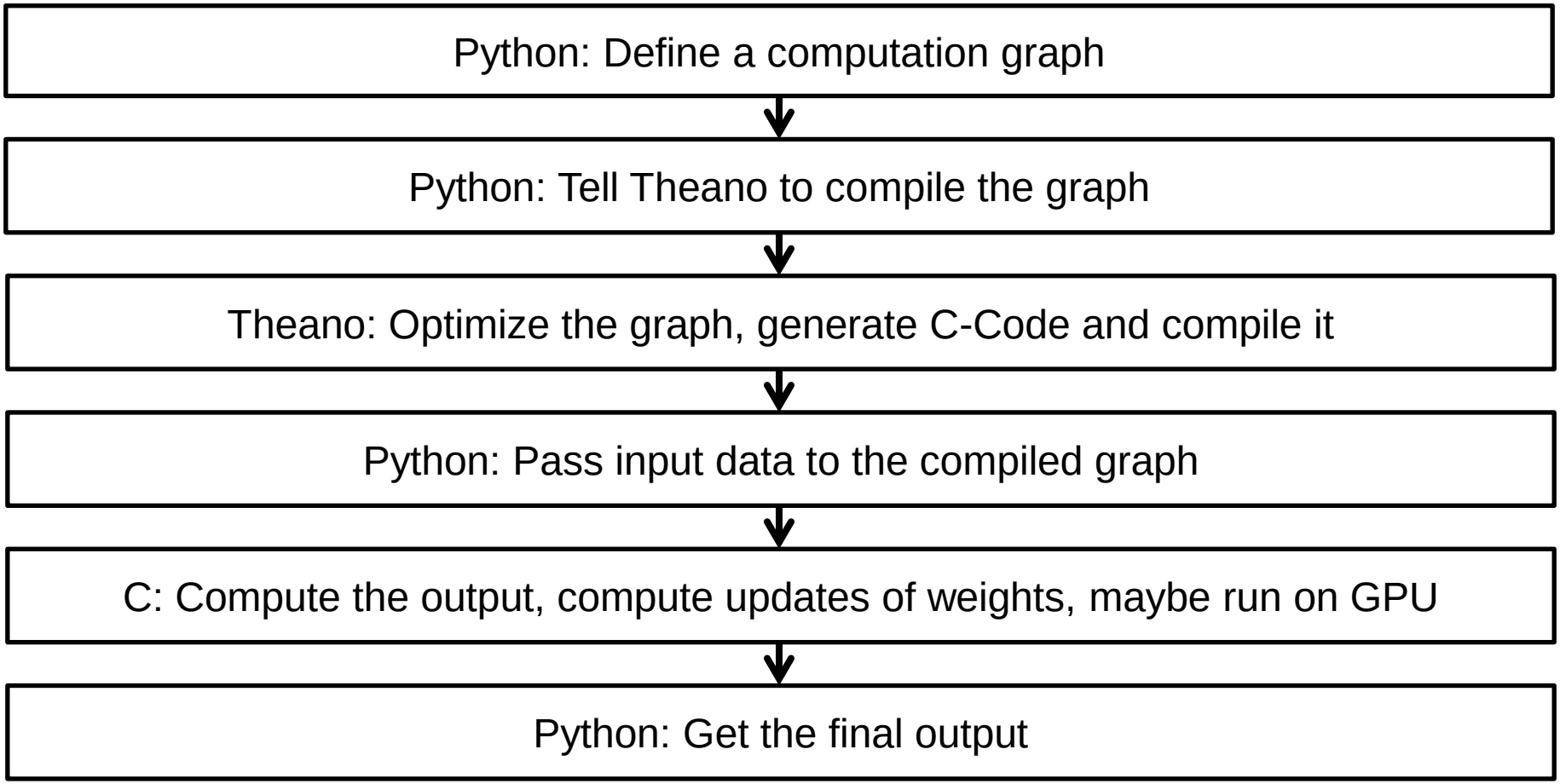
TECHNISCHE UNIVERSITÄT DARMSTADT

- Theano utilizes BLAS (Basic Linear Algebra Subprograms)
  - Building blocks for fast vector and matrix operations
  - Often written in Fortran, sometimes in Assembler
- For performance optimization install a BLAS package
- Benchmark different BLAS packages
- I use a manually compiled OpenBlas implementation
  - Installation notes: http://deeplearning.net/software/theano/install_ubuntu.html

Theano – Flow

TECHNISCHE UNIVERSITÄT DARMSTADT

- The execution of a Theano script is a bit different



UKP

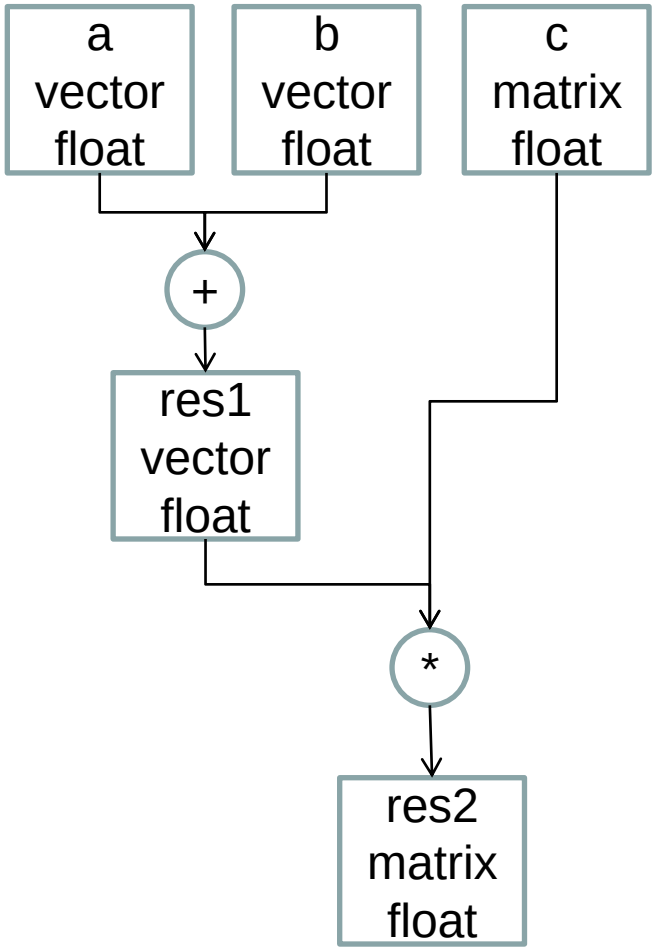
Theano – Computation Graph

- You write symbolic expression to define the computation graph

```
import theano
import theano.tensor as T


a = T.fvector()
b = T.fvector()
c = T.fmatrix()

res1 = a+b
res2 = T.dot(res1,c)
```



Next Lecture

TECHNISCHE UNIVERSITÄT DARMSTADT

- Preparation for the next tutorial
  - Please do the exercises mentioned on the website
  - Please watch the videos / read the papers

- **Next Lecture:** How use word2vec & word2vec hacks, Deep Feed Forward Network for NER
  - It is assumed that you know the basic theory behind word2vec (watch the videos)
  - We will discuss your questions regarding the videos/the theory next lecture
  - We will design and implement the network using Theano/Lasagne – please be familiar with the basics of these two frameworks