



NLP What Next?

Liad Magen

Learning Goals

Write

- Write efficient Regular Expressions, to solve (almost) any kind of text-extraction task.

Apply

- Apply Topic Modeling & Text-Distance algorithms to clustering and text-sequence problems.

Train

- Train a spelling / grammar corrector.

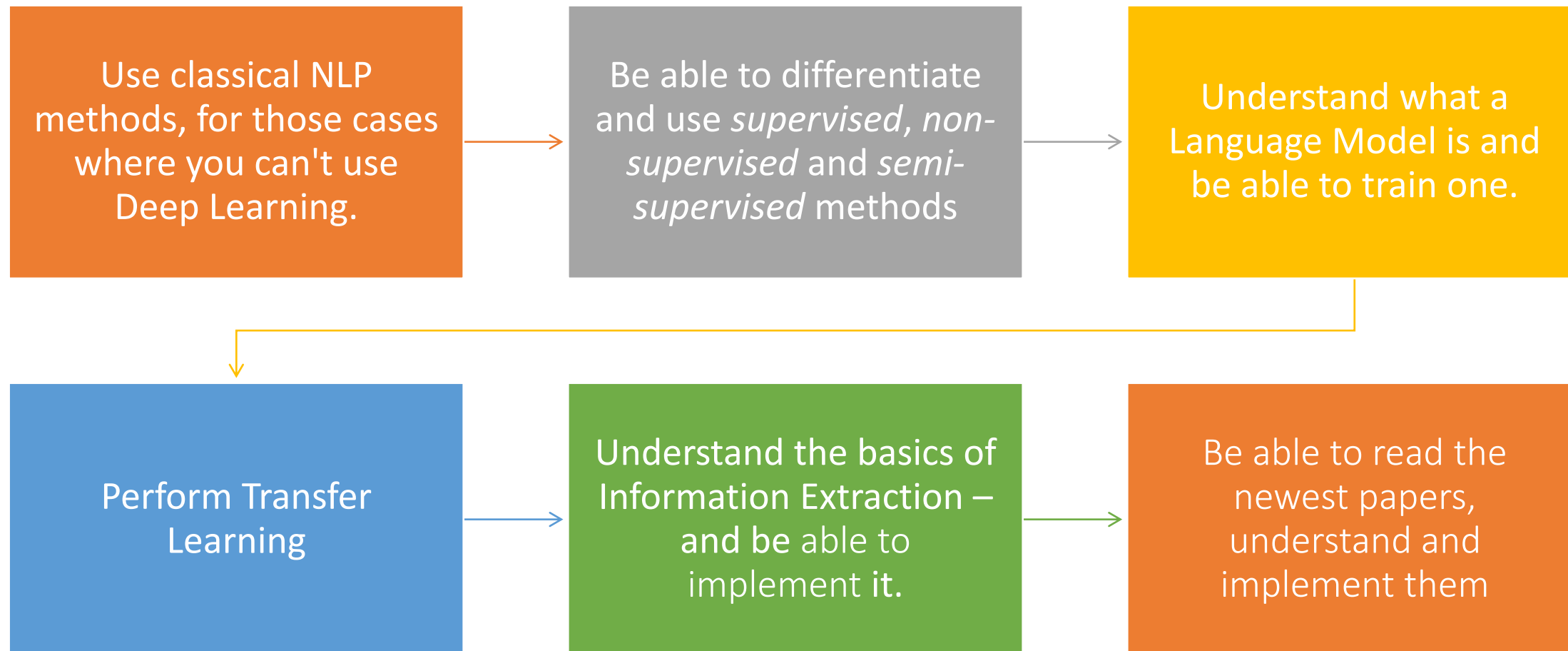
Classify

- Build a supervised classifier to solve problems such as Sentiment Analysis or a Spam-detector.

Build

- Build a recommendation engine.

Learning (hidden) Goals



What have
we learned so
far?





Day I

- What is NLP and where it is used
- 5 Levels of Linguistic Knowledge
- Why NLP is hard: Ambiguity, Variability + Zipf Law
- Basic Units of Processing
- Case Study: Sentence Boundary Detection
- Classification Types
- Pre-Processing: Tokenization, n-gram, stemming, lemmatizing
- TF/IDF, PMI

Day II

PP Attachment: Ambiguity Classification using Classical NLP

- Feature Extraction
- Collins and Brooks' Back Off Algorithm

Language Models

- Markov Assumption of n-grams
- Text Generation with a probabilistic model
- One-Hot Encoding
- Neural Language Model
- Embedding Layer

Distributional Semantics

- Word2Vec (== SVD)
- GloVe
- FastText

Pre-Training (Language Model)

- CBOW
- Skip-Gram

Semantic Distribution

- Word Similarity
- Sentence Representation (sum/avg)

RNN

- Seq2Seq (input: words, output: POS, NER, SRL)
- Intermediate vectors
- BiLSTM + deep RNN
- Multi-Task Learning
 - Architecture Creativity solves issues

Day III

Day IV (part 2/2)

- Contextual Word Embeddings
 - ELMo (AI2)
 - ULMFiT (fast.ai)
 - Attention
 - Transformers
- BERT
 - 3-layers of Input
 - SubWords + WordPiece
 - 12/24 layers of Word Embedding Vectors
 - BERTology Family
- Sustainability and Green Responsibility

Day V - Information Extraction

- Phrase Types
 - Chunking
- Named Entity Recognition (NER)
- Co-Reference Resolution
- Relation Extraction
 - Knowledge Graph Database
- Entity Linking
- Graphs
- Google PageRank
- ML-OPS for NLP



Day V


- Presentations:
 - Training Medical Images with Text annotations
 - Word-Order research in BERT
 - Sign Language & NLP
 - Hate Speech Detection
 - News Header Generation
- ... and many more.

What now?

- Dive deeper:
 - Neural Networks for NLP (2021) <http://www.phontron.com/class/nn4nlp2021/schedule.html> (All on [youtube](#))
 - Stanford: [Stanford CS 224N | Natural Language Processing with Deep Learning](#)
 - AI2 – Information Extraction software and talks: [SPIKE: Extractive Search from Allen Institute for AI \(allenai.org\)](#)
- Communication:
 - [\(21\) How to Speak – YouTube](#)



Additional Resources

- www.arxiv-sanity.com
 - SemanticScholar.org
 - Twitter – follow researchers
 - ACL - www.aclweb.org
 - NLPProgress.com
 - Ruder.io
 - PapersWithCode.com
 - Keep-Current.com
 - Listen to Podcasts
 - Attend (or organize) Meetups
- 

WIIIFM

On the first day, you've written down what you would like to do using NLP.

Please note down:

- Now that you know what you know, has your goal changed? If so, to what?
- How would you use the tools you've learned to achieve it?





Thank you!

It's been a true pleasure! :)