Applied Deep Learning for Sequences Class 1 -- Intro

Yoav Goldberg
Bar Ilan University

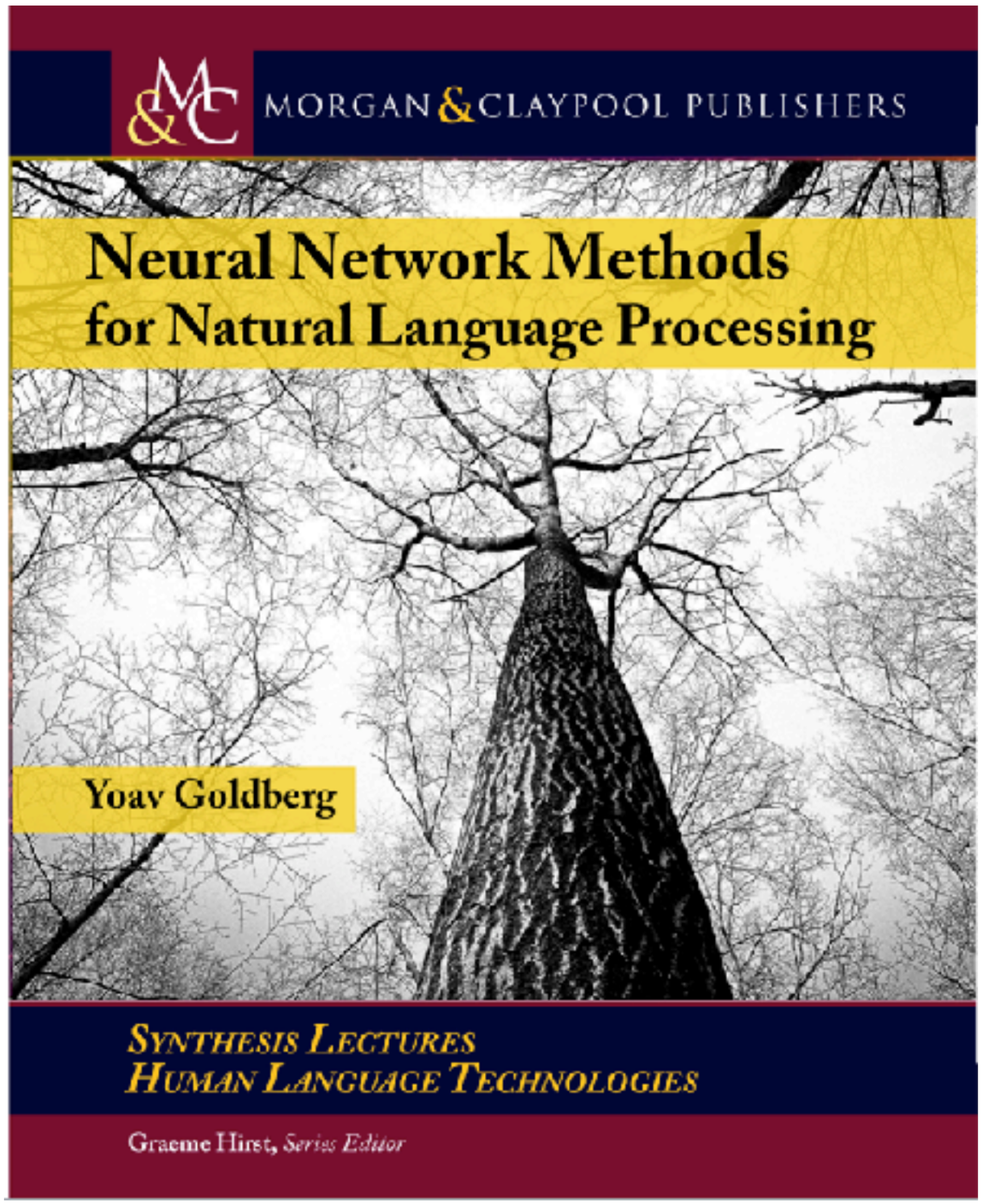
Course Info

- Website: http://www.cs.biu.ac.il/~89-687
  - Coming soon: link to piazza discussion forums
- Office hours: by appointment
- Requirements:
  - Small "exercises" (must submit, not graded)
  - Assignments (programming, math or both)
  - ~~Final Project~~ Final Exam

Course Info

- Background:
  - Linear Algebra
  - Calculus (differentiation)
  - A little bit of probability
  - Programming (python, some c++)
  - Machine Learning -- a plus

Course Info


MORGAN & CLAYPOOL PUBLISHERS
Neural Network Methods
for Natural Language Processing
Yoav Goldberg
SYNTHESIS LECTURES
HUMAN LANGUAGE TECHNOLOGIES
Graeme Hirst, Series Editor

Course Info

- This is not an easy course.
- It is also not very hard compared to similar ones.
- It is probably an important one.

http://www.phontron.com/class/nn4nlp2018/schedule.html

CMU CS 11-747, Fall 2017

Neural Networks for NLP

Course Schedule

What is Learning?

What is Learning?

- Making predictions
- Based on past observations
  - finding patterns in data
  - learning from data
  - generalization...

Types of Learning

- Supervised
  - Regression
  - Classification
    - Structured
- Unsupervised
  - Semi-supervised
- Reinforcement

Machine Learning

- Learning a mathematical function from x to y.
- Need to represent objects in the world as mathematical objects.
- Measurements / Features.
- Hypothesis class.
  - Inductive bias.

What is Deep Learning

What is Deep Learning

- Specific class of learning algorithms
- Rich hypothesis class (can represent borel measurable functions)
- Many "layers" of simple components.
- Find parameters with gradient-based methods (so everything needs to be differentiable)
- Minimal feature engineering (**learn** a good representation)

What are Sequences

- Sequential data is everywhere.

- Examples?

Types of Sequences

- Continuous / Numeric

- Discrete

Characteristics of Sequential Data

Characteristics of Sequential Data

- Sequential
- Items, "vocabulary", "alphabet"
- "time" dependence
- Locality
- Long-distance relations

Kinds of Problems

Kinds of Problems

- Classification
    - Of sequences
    - Of items within sequences
- Similarity
    - Of sequences
    - Of items within sequences
- Segmentation
- Mapping
- Sequence to Sequence
- Predict next item ("language modeling")
- ...

This course

- Linear models
- Gradient-based learning, losses, regularization
- Feed forward nets / Multi-layer perceptron
- Representation learning
- Language-models and word embeddings
- Multi-task learning
- 1D Convolutions
- Recurrent networks
    - Seq2seq. Attention. (Transformer?)
- Recursive networks.
- Adversarial training.
- Reinforcement learning.

This course

- Implementing neural networks:
    - by hand
    - with a toolkit
- Efficiency
    - GPU vs CPU
    - Batching

This course

- How to **model** with neural networks?
- How to **describe** neural networks?
- What are the **benefits** of different architectures?
- What are the **limitations** of different architectures?
- What is hard to learn? what is easy to learn?

Learning 101

- Building blocks -- Linear Model -- Perceptron

Linear Model

Housing Data

(whiteboard)

Language Classification

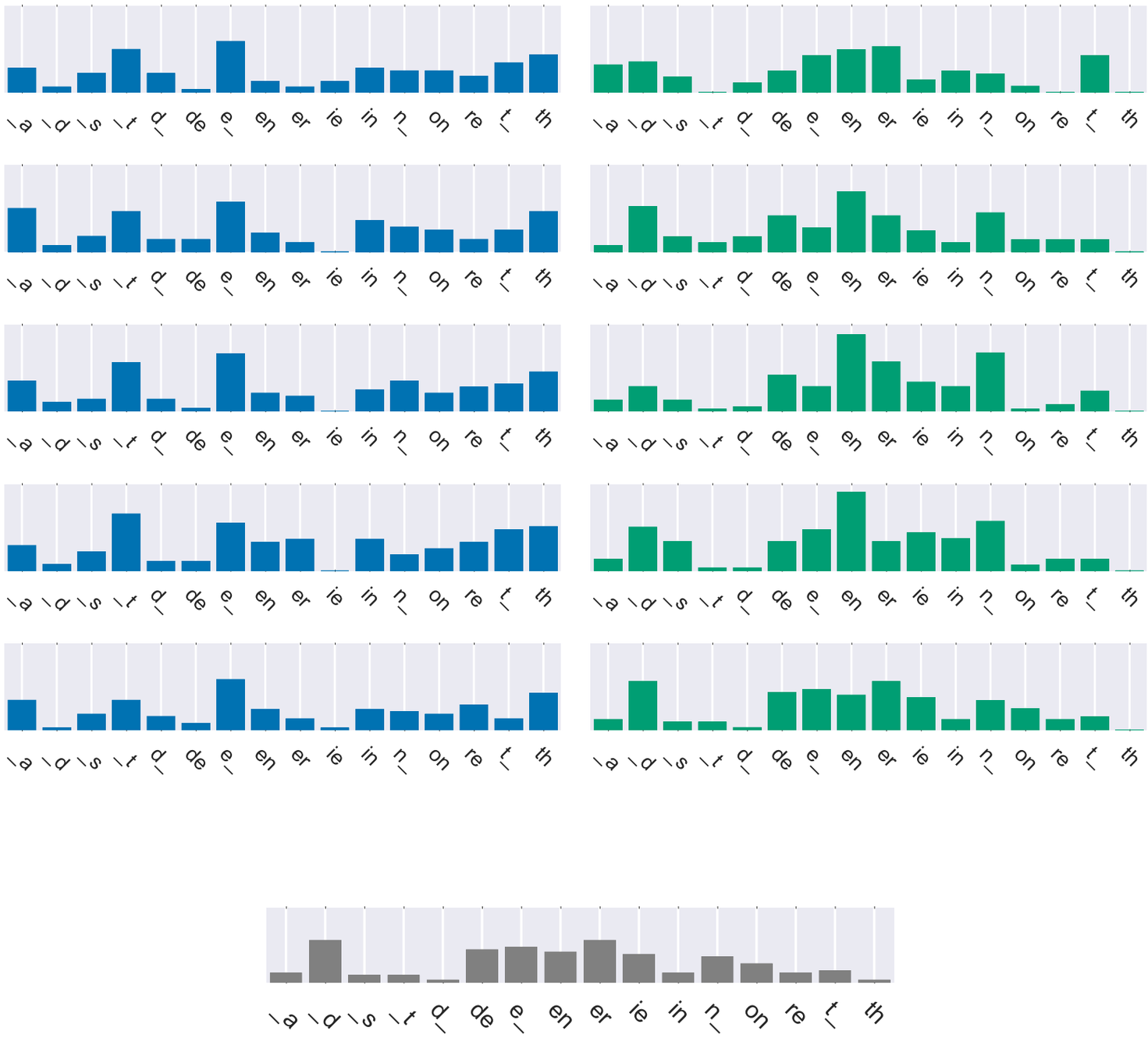
- Astrophysicists use the term "metal" to collectively describe all elements other than hydrogen and helium. In that sense, the metallicity of an object is the proportion of its matter made up of chemical elements other than hydrogen and helium.

- Ein einzelnes Atom dieser Elemente hat keine metallischen Eigenschaften; es ist kein Metall. Erst wenn mehrere solcher Atome miteinander wechselwirken und zwischen ihnen eine metallische Bindung besteht, zeigen solche Atomgruppen (cluster) metallische Eigenschaften.

Letter Bigrams

Letter Bigrams

Letter Bigrams



(whiteboard)

Summarize

- Simple hypothesis class -- linear model, perceptron

- See exercise 1 online.