Question Answering System: Step by Step

Natural Language Processing: Jordan Boyd-Graber
University of Maryland
HOMEWORKS
Getting Started

- Sequence of small project to teach basics of question answering
- Introducing competition data, framework, useful tools
Information Retrieval

- Introduce dataset, Python
- Create document representations
Nearest-neighbor

- Actually produce a (simple) classifier
- Strong baseline to compare against
- Can submit to Codalab
Logistic Regression

- Using pytorch framework
- Optimizing an objective function
- Adding new labels to dataset
Deep Averaging Network

- Deep learning
- Question answering (same as nearest-neighbor)
- Real piece of (simple) research
Sequence Model: Networks

- More complicated architecture
- Learning *when* to answer questions
- Modifying objective functions
Reading Comprehension

- Modern NLP framework / model
- Combines IR, sequence models, deep learning
- Not confined to small answer set
- Alternative to previous approaches
Project

- Build on Homeworks: Answer questions!
- Focus on what you want:
  - Novels
  - Time periods
  - Multilingual data
- Live competitions
Advice

- Can start on competition immediately
- Work through subprojects: understand what’s important
- Designed to introduce elements systematically