NLP needs ML

Advanced Machine Learning for NLP

Jordan Boyd-Graber

COURSE OVERVIEW AND LOGISTICS
What are ML and NLP?

• Machine Learning (Methods)
  ○ Linear classifiers
  ○ Logistic regression
  ○ Thinking about data in terms of feature vectors

• Natural Language Processing (Problems)
  ○ Summarization
  ○ POS Tagging
  ○ Question Answering
  ○ Parsing
  ○ Translation
What are ML and NLP?

- **Machine Learning (Methods)**
  - Linear classifiers
  - Logistic regression
  - Thinking about data in terms of feature vectors

- **Natural Language Processing (Problems)**
  - Summarization
  - POS Tagging
  - Question Answering
  - Parsing
  - Translation

- We’ll assume you know basics of both (5832/5622)
What will we be talking about?

- Objective functions
  - Deep learning
  - Bayesian approaches
- Structure
- Representation
- Algorithms
What will we be talking about?

- Objective functions
  - Deep learning
  - Bayesian approaches
- Structure
- Representation
- Algorithms
- Both critically influence by both ML and NLP
Philosophy of the Course

• Showing connections between methods
• Each week, go “in depth” into one specific method
• I don’t know everything well myself
• Organization: words → sentences → documents
Philosophy of Each Class

- Big picture: me
- Mathematical treatment: student(s)
- Hands on example/demo: students(s)
Philosophy of Each Class

- Big picture: me
- Mathematical treatment: student(s) (after first two weeks)
- Hands on example/demo: students(s) (after first two weeks)
Grade

- Mathematical Treatment 25%
- Detailed Demonstration 25%
- Final Project 30%
- Participation 20%
Timeline

- If you’re presenting on Monday of week $N$, you’ll need to
  - Present a five minute outline on Monday week $N-2$
  - Post a revised outline on Wednesday of week $N-2$ for general feedback on Piazza
  - Give a practice presentation before Wednesday of week $N-1$ to the course assistant
  - Submit materials to professor after class (via Github: https://github.com/Pinafore/adv-ml-nlp)
Timeline

- If you’re presenting on Monday of week $N$, you’ll need to
  - Present a five minute outline on Monday week $N-2$ (except for week 3)
  - Post a revised outline on Wednesday of week $N-2$ for general feedback on Piazza (except for week 3)
  - Give a practice presentation before Wednesday of week $N-1$ to the course assistant
  - Submit materials to professor after class (via Github: https://github.com/Pinafore/adv-ml-nlp)
Hands-on Demonstration

- Simple enough to be workable “by calculator”
- Through enough to understand algorithm (and implement)
- Provide scaffolding to work through the problem
Hands-on Demonstration

- Simple enough to be workable “by calculator”
- Through enough to understand algorithm (and implement)
- Provide scaffolding to work through the problem
- First two classes provide examples (but do better!)
Mathematical Treatment

- Provide enough background for the hands-on demonstration
- Don’t just give equations/algorithm, explain what’s going on
- Provide intuitions
- Work with person(s) doing the demonstration
Course Project

- Apply machine learning algorithm to language
- Proposal due around spring break
- Open-ended
- Stretch goal: NIPS submission (thus, needs ML novelty)
Administrivia

- Use Piazza
- Office Hours Monday 4-5
- Alvin is Course Assistant