Admin

Assignment 7 due next Monday

Will update schedule for the rest of the semester by Wed

Grading

---

Final project

1. Your project should relate to something involving NLP

2. Your project must include a solid experimental evaluation

3. Your project must have an ethical component

4. Your project should be in a group of 2-4. If you’d like to do it solo, please come talk to me.

---

Final project

<table>
<thead>
<tr>
<th>date</th>
<th>time</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/17</td>
<td>in-class</td>
<td>Project proposal presentation</td>
</tr>
<tr>
<td>4/19</td>
<td>11:59pm</td>
<td>Project proposal write-up</td>
</tr>
<tr>
<td>4/26</td>
<td>11:59pm</td>
<td>Status report</td>
</tr>
<tr>
<td>5/1</td>
<td>11:59pm</td>
<td>Paper draft</td>
</tr>
<tr>
<td>5/3</td>
<td>in-class</td>
<td>Presentation</td>
</tr>
<tr>
<td>5/3</td>
<td>11:59pm</td>
<td>Final paper and code</td>
</tr>
</tbody>
</table>

Read the final project handout ASAP!

Start forming groups and thinking about what you want to do
Final project ideas

- pick a text classification task
  - evaluate different machine learning methods
  - implement a machine learning method
  - analyze different feature categories

- n-gram language modeling
  - implement and compare other smoothing techniques
  - implement alternative models

- parsing
  - lexicalized PCFG (with smoothing)
  - n-best list generation
  - parse output reranking
  - implement another parsing approach and compare
  - parsing non-traditional domains (e.g., n-iter)

- EM
  - try and implement EM model 2
  - word-level translation models

Final project application areas

- spelling correction
- part of speech tagger
- text chunker
- dialogue generation
- pronoun resolution
- compare word similarity measures (more than the ones we looked at)
- word sense disambiguation
- machine translation
- information retrieval/search
- information extraction
- question answering
- summarization
- speech recognition

Final project application areas

- Ethical analysis of: Word2Vec, GPT-X, ChatGPT
- Utilize LLM (large language model), e.g., GPT
  - poetry
  - lyrics
  - code generation
  - ...
- Limited resource NLP (e.g., limited memory or computation)
- BART/BERT