NLP & ML

Assignment 2

Innopolis

Option one: PoS tagger

- Develop a PoS tagger using structured learning
- Try from CRF Suite
- Use the datasets on Moodle

- For B grade: implement the system with evaluation and a baseline
- For A grade: suggest, implement and evaluate improvements. They don't have to be successful!

Option two: Movie sentiment

- Develop a sentiment classifier
- Use LSTM
- Use the NLTK/Keras data

- For B grade: implement this, evaluate against a baseline
- For A grade: extend the approach and get a score over 92% - do something novel (perhaps find some new training data? Evaluate on product reviews? Try a new natural language?)

Global requirements

- Submit by end of March 31
- Hand in two separate items:
    - Code (ipynb OK, HTML not ok)
    - Essay
- Describe the system using 500-1000 words
- Tabulate data and make your baseline vs. system comparison clear