Midterm & Concept Review

CS-585

Natural Language Processing

Sonjia Waxmonsky

Transforming Lives. Inventing the Future. **www.iit.edu**

Midterm Exam info

- Exam locations and times
  - Midterm: Thurs, Oct. 5, 8:35-9:50am
  - Room: SH-118 (Siegel 118)

- Midterm details
  - Timed: 70 minutes
  - Multiple-choice or similar (bubble answer sheet)
  - Closed book, closed note, no electronics
  - Bring pencils with erasers

Mid-Term Logistics

Remote exam-takers

- Arrangements should already be finalized

- Instructor <u>cannot</u> proctor via zoom or video

In-person exam-takes

- Exam will be timed to 70 minutes (for fairness with online sections); most students will not need entire time

- Please arrive by 8:35am to Siegel 118.

- Late arrivers will be promptly seated by teaching staff, exam cannot be extended

- Bring pencils with erasers! (Does not need to be #2)

Topics – Up to Midterm

	Content	Reading
1	Welcome, linguistic concepts	MS 3
2	Math review 1: probability and linear algebra	MS 2.1
3	Math review 2: information theory	MS 2.2
4	Words and pattern matching	MS 4.2, E-NLP 4.3
5	Lexical representations for NLP	MS 1.4
6	Neural nets 1: neural word embeddings	E-NLP 14
7	Word sense disambiguation	MS 7.1-7.3
8	Text categorization and naive Bayes	E-NLP 2.1-2.2, 4.4
9	Generalized linear models	E-NLP 2.5-2.6
10	Neural nets 2: feedforward networks	E-NLP 3.1-3.3.3
11	Sentiment analysis	E-NLP 4.1
12	Unsupervised methods in NLP	E-NLP 5.1

SAMPLE QUESTIONS

Sample Question - Linguistics

You are interested in studying how word meaning is influenced by suffix endings, such as –ed and –ing. Which fields are BEST suited to inform your investigation?

A. Phonetics and phonology
B. Morphology and semantics
C. Historical Linguistics
D. Image processing

Sample Question – Information Theory

What is measured with **cross-entropy**?

A. The percentage of true positives that are detected in a binary classification model?

B. The distance between two probability distributions

C. The randomness or uncertainty of a random variable

D. Any of the above

Sample Question – Regular expressions

Which string is matched by the following regular expression?

`\bhedge-?hog.\b`

A. `hedge-hog`
B. `hedge-hogs`
C. `hedgehog`
D. All of the above

Sample Question – Neural Networks

You are building a text categorization model, and your teammate shares they were able to **reduce model complexity** by **adding a layer** to their feed forward neural network. What is unusual about this statement?

A. Neural networks cannot be feed-forward

B. Adding a layer would <u>generally</u> increase complexity, not decrease complexity

C. Complexity should always be increased, not decreased.

D. All of the above

Sample Question – Metrics

You have built a rules-based text classifier that uses regular expressions to flag text messages that include a telephone number. This tool outputs 1 when it finds any string of 10 digits, and 0 otherwise. What is a good metric to evaluate this text classification tool? Select the BEST answer

A. Precision
B. ROC-AUC
C. Cross-entropy
D. All of the above

Review question

- Which of the following are NLP tasks?
- Which correspond to algorithms (or families of algorithms? Which of those are <u>unsupervised</u>?

  - Word sense disambiguation

  - Sentiment analysis

  - Latent semantic analysis

  - word2vec