Post-Midterm Review

Adam Meyers
New York University
Outline

• Grading Policies
• Goals for Midterm Review
• Review
## Grade Curve

<table>
<thead>
<tr>
<th>Raw Score (out of 115)</th>
<th>Letter Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>105 and Above</td>
<td>A</td>
</tr>
<tr>
<td>101 and Above</td>
<td>A-</td>
</tr>
<tr>
<td>97 and Above</td>
<td>B+</td>
</tr>
<tr>
<td>92 and Above</td>
<td>B</td>
</tr>
<tr>
<td>87 and Above</td>
<td>B-</td>
</tr>
<tr>
<td>82 and Above</td>
<td>C+</td>
</tr>
<tr>
<td>77 and Above</td>
<td>C</td>
</tr>
<tr>
<td>72 and Above</td>
<td>C-</td>
</tr>
<tr>
<td>62 and Above</td>
<td>D</td>
</tr>
<tr>
<td>Below 55</td>
<td>F</td>
</tr>
</tbody>
</table>
Midterm Grade's Effect on Final Grade

• Grade = ¼ Mid + ¼ HW + ¼ Final Exam + ¼ Final Project
• Proviso
  – If HW, Final Exam and Final Project are all higher than Midterm, they count more
    • It is possible to do badly on the midterm, but get a good grade anyway.
Goals for Review

• **HW** – Too many Students are Currently Behind
  – There is a positive correlation between completion of homework and Midterm grade
  – Reviewing Sections of Midterm may make HW easier to complete

• **Final Exam**
  – Final Exam includes entire term of work
  – Reviewing Sections of Midterm helps with Final Exam

• **Final Project** – Reviewing Methodology and Correcting Misconceptions Should Lead to Better Final Projects
Question 1: Regular Expressions

- Example correct answer: \[0-9]+( ([A-Z][0-9]+)[a-z]*){1,3}( (Street|St\.| Ave))?, (New York|NY|NYC|New York City), (NY|New York)( [0-9]{5})?

- Limited use of very general expressions was OK, e.g., \[^\]+ or .* 

- Overuse meant that an expression matched both the set of addresses in Question 1 and lots of other stuff as well, e.g., a string of a certain length that starts with a number and has 2 commas in it.

- It seemed necessary to include at least a few anchor expressions that were not very general
  - (NY | NYC | New York)
  - (Street|Ave|Broadway|St\.)

- Careful about combinations of iteration and/or optionality
  - e.g., +? is the same thing as *

- Limit scope of disjunction
  - (NY | NYC | New York) not NY | NYC | New York
Question 2

- Include:
  - Phrase structure rules and Corresponding tree
  - The comma – it is a token, just like any word

- POS and Phrase Categories are based on distribution, not function

- Assume a Prepositional Phrase
  - *of* is a preposition (other preps: *with, for, in, …*)
  - PP → P NP       (IN and TO are Penn tags for P)

- **bagel** is a noun:
  - nouns can modify other nouns
  - If *bagel* was an adjective, than noun would be a subclass of adjective.

- The whole phrase is an NP, not an S

- Sample Tree (next slide)
Sample Answer for Question 2

- $NP \rightarrow NP, NP$
- $NP \rightarrow NNP NNP$
- $NP \rightarrow NP PP$

- $NP \rightarrow DT JJ NN NN$
- $PP \rightarrow IN NP$
Question 3 (Use all/only Rules)

- NP → Name
- Name → Name NNP
- Name → NNP
- NP → DT NOM
- NOM → JJ NOM
- NOM → NN NOM

- NOM → NN
- NN → red
- NN → robin
- DT → the
- JJ → red
- NNP → robin

<table>
<thead>
<tr>
<th>The</th>
<th>red</th>
<th>red</th>
<th>robin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>0</td>
<td>DT [0,1]</td>
<td>NP [0,2]</td>
<td>NP [0,3]</td>
</tr>
<tr>
<td>1</td>
<td>JJ, NN, NOM [1,2]</td>
<td>NOM [1,3]</td>
<td>NOM [1,4]</td>
</tr>
<tr>
<td>2</td>
<td>JJ, NN, NOM [2,3]</td>
<td>NOM [2,4]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>NN, NNP, NOM, NAME, NP [3,4]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Question 4

- Variation of tokenization of C.K. – OK
  - C.K. NNP B
  - C, ., K, . NNP,PU,NNP,PU B I I I
  - C., K. NNP, NNP B I

- his use PRP$ NN BI – possessive + nominalization

- is known – VBZ VBN O O

- observational, self-deprecating, dark and vulgar humor
  - JJ PU JJ PU JJ CC JJ NN
  - OOOOOOBI – only include consecutive JJ NN in noun group
  - BOBOBOBI – each adjective starts a noun group
  - BIIIIII – the conjoined adjectives are all left modifiers
    - This last one is what is assumed in HW 6
Question 5

- Vectors
  - query \([5.3, 8.1, 6.9, 3.5]\)
  - doc1 \([21.2, 16.2, 48.3, 17.5]\)
  - doc2 \([0.0, 16.2, 0.0, 3.5]\)
- Variation:
  - Natural Log vs Base 10 (doesn't matter)
  - Vector values are TF * IDF – some answers just used TF (this is incorrect)
  - missed 1 or 2 instances
  - counted \textit{wood} together with \textit{wooden}
- Similarity of Doc1 and Query: \[
\frac{638.1}{\sqrt{(153.56 \times 3351.02)}} = .9
\]
- Similarity of Doc2 and Query: \[
\frac{143.47}{\sqrt{(153.56 \times 274.69)}} = .7
\]
Question 6

- Probability of NNS VBP:
  - \(.33 \times 0.0007 \times 0.4 \times 0.035 \times 0.4 = 1.98 \times 10^{-6}\)
  - Start → NNS = .33, does|NNS = .007, NNS → VBP = .4, run|VBP = .035, VBP → End = .4

- Probability of VBZ VB
  - \(.42 \times 0.0035 \times 0.5 \times 0.006 \times 0.45 = 1.29 \times 10^{-6}\)
  - Start → VBZ = .43, does|VBZ = .035, VBZ → VB = .45, run|VBP = .006, VBP → End = .45
Does Run

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1: Does</th>
<th>2: Run</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNS</td>
<td></td>
<td>0.33 * 0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VB</td>
<td></td>
<td></td>
<td>0.5 * 0.006</td>
<td></td>
</tr>
<tr>
<td>VBP</td>
<td></td>
<td></td>
<td>0.4 * 0.035</td>
<td></td>
</tr>
<tr>
<td>VBZ</td>
<td></td>
<td>0.42 * 0.035</td>
<td></td>
<td></td>
</tr>
<tr>
<td>END</td>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
</tr>
</tbody>
</table>

Midterm Preparation 2016
Question 7

- The task: classify statements as Positive, Negative or Neutral, based on the intension of the speaker of the sentence

- Variation:
  - Knowledge of U.S. Politics
  - Annotator opinion about whether the idea of the sentence is Positive/Negative/Neutral
  - Lack of Context: Are isolated sentences enough?
  - Level of Assertion, i.e., anything is possible
Question 7: Controversial Sentences

• Donald Trump has “neutralized” the influence of well-funded super PACs.
  – Neutral: 30 Positive: 21 Negative: 1

• Donald Trump will protect Israel, and brutally and quickly cut the head off of ISIS.
  – Positive: 35 Neutral: 13 Negative: 4

• 5 sentences with 1 to 3 dissenting classifications
• 3 sentences with unanimous classifications
Question 7: Sentences with the biggest variation