Teaching the Basics of NLP and ML in an Introductory Course to Information Science

Apoorv Agarwal
Columbia University
COMS1001

- Introductory course on information science to undergraduates at Columbia University
COMS1001

• Introductory course on information science to undergraduates at Columbia University

• Mostly taken by freshmen and sophomores
COMS1001

- Introductory course on information science to undergraduates at Columbia University
- Mostly taken by freshmen and sophomores
- Assumes no prior programming or math background
COMS1001

• Introductory course on information science to undergraduates at Columbia University

• Mostly taken by freshmen and sophomores

• Assumes no prior programming or math background

• 10% : what’s a programming language?
Student demographics
Student demographics

- Economics: 16%
- Basic Sciences: 14%
- Arts and Humanities: 14%
- Social Science: 14%
- Religious Studies: 12%
- Political Science: 11%
- Math/Engineering: 11%
- Other: 9%
- Language: 6%
Student demographics

Math and Engineering majors
Student demographics

Challenge 1: Cannot use Math terminology: vector space, dot product, high-dimensional space etc.

![Pie chart showing student demographics]

Math and Engineering majors
Traditionally taught topics
Traditionally taught topics

- About thirty 75 min lectures
Traditionally taught topics

• About thirty 75 min lectures

• First half: Operating systems, WWW and the Internet, Binary and Machine Language, Spreadsheets, Database systems
Traditionally taught topics

• About thirty 75 min lectures

• First half: Operating systems, WWW and the Internet, Binary and Machine Language, Spreadsheets, Database systems

• Second half: Algorithms, Programming in Python
Traditionally taught topics

• About thirty 75 min lectures

• First half: Operating systems, WWW and the Internet, Binary and Machine Language, Spreadsheets, Database systems

• Second half: Algorithms, Programming in Python

Challenge 2: Introduce NLP/ML in one lecture
Overall Strategy
Overall Strategy

• Keep definitions simple
Overall Strategy

• Keep definitions simple

• Use analogies and concrete examples (also observed by Reva Freedman 2005)
Overall Strategy

• Keep definitions simple

• Use analogies and concrete examples (also observed by Reva Freedman 2005)

• Take baby steps -- incremental learning
Overall Strategy

- Keep definitions simple
- Use analogies and concrete examples (also observed by Reva Freedman 2005)
- Take baby steps -- incremental learning
- Introduce the core concepts in one lecture and build on them using homework and exam problems
Strategy

COMS1001-Spring2013
Jan 22, 2013 – May 2, 2013

ML HW
Feb 26, 2013 – Mar 12, 2013

NLP/ML Lecture
Feb 19, 2013

Mid-term Exam
Mar 14, 2013

Python Lectures
Apr 9, 2013 – Apr 23, 2013

ML in Python HW
Apr 18, 2013 – May 10, 2013

Feb 1, 2013  Mar 1, 2013  Apr 1, 2013  May 1, 2013
Strategy

COMS1001-Spring2013
Jan 22, 2013 – May 2, 2013

ML HW
Feb 26, 2013 – Mar 12, 2013

Sentiment analysis of movie reviews

NLP/ML Lecture
Feb 19, 2013

Sentiment analysis of tweets

Mid-term Exam
Mar 14, 2013

Python Lectures
Apr 9, 2013 – Apr 23, 2013

ML in Python HW
Apr 18, 2013 – May 10, 2013

Feb 1, 2013 Mar 1, 2013 Apr 1, 2013 May 1, 2013
Strategy

Sentiment analysis of movie reviews

Sentiment analysis of tweets

Email classification into Imp/Not-Imp

COMS1001-Spring2013
Jan 22, 2013 – May 2, 2013

ML HW
Feb 26, 2013 – Mar 12, 2013

NLP/ML Lecture
Feb 19, 2013

Sentiment analysis of tweets

Mid-term Exam
Mar 14, 2013

Python Lectures
Apr 9, 2013 – Apr 23, 2013

ML in Python HW
Apr 18, 2013 – May 10, 2013

Email classification into Imp/Not-Imp

Feb 1, 2013
Mar 1, 2013
Apr 1, 2013
May 1, 2013
Strategy

COMS1001-Spring2013
Jan 22, 2013 – May 2, 2013

ML HW
Feb 26, 2013 – Mar 12, 2013

NLP/ML Lecture
Feb 19, 2013

Sentiment analysis of movie reviews

Sentiment analysis of tweets

Gear towards text processing

Python Lectures
Apr 9, 2013 – Apr 23, 2013

Mid-term Exam
Mar 14, 2013

Email classification into Imp/Not-Imp

ML in Python HW
Apr 18, 2013 – May 10, 2013

Feb 1, 2013
Mar 1, 2013
Apr 1, 2013
May 1, 2013
Strategy

COMS1001-Spring 2013

Jan 22, 2013 – May 2, 2013

- **Sentiment analysis of movie reviews**
- **Sentiment analysis of tweets**
- **Email classification into Imp/Not-Imp**
- **Gear towards text processing**
- **Implement end-to-end SA pipeline**
Overview

• Lecture organization
• Questions asked in class
• Performance on the mid-term examination
• Final projects
• Conclusion
Lecture Organization
Lecture Organization

• General discussion on how to define intelligence
Lecture Organization

- General discussion on how to define *intelligence*

- Introduce a concrete application: sentiment analysis of *Twitter* data
Lecture Organization

• General discussion on how to define *intelligence*

• Introduce a concrete application: sentiment analysis of *Twitter* data

• Demonstrate annotation process
Lecture Organization

• General discussion on how to define *intelligence*

• Introduce a concrete application: sentiment analysis of *Twitter* data

• Demonstrate annotation process

• Demonstrate feature extraction
Lecture Organization

• General discussion on how to define *intelligence*

• Introduce a concrete application: sentiment analysis of *Twitter* data

• Demonstrate annotation process

• Demonstrate feature extraction

• Demonstrate a *basic* classification process
Points we drive home
Points we drive home

1. The machine automatically learns the connotation of words by looking at how often certain words appear in positive and negative tweets.
Points we drive home

1. The machine automatically learns the connotation of words by looking at how often certain words appear in positive and negative tweets.

2. The machine also learns more complex patterns that have to do with the conjunction and disjunction of features.
Points we drive home

1. The machine automatically learns the connotation of words by looking at how often certain words appear in positive and negative tweets.

2. The machine also learns more complex patterns that have to do with the conjunction and disjunction of features.

3. The quality and amount of training data is important – for if the training data fails to encode a substantial number of patterns important for classification, the machine is not going to learn well.
Questions asked in class by students
Questions asked in class by students

1. Could we create and use a dictionary that lists the prior polarity of commonly used words?
Questions asked in class by students

1. Could we create and use a dictionary that lists the prior polarity of commonly used words?

2. If the prediction score for the tweet is high, does that mean the machine is more confident about the prediction?
Questions asked in class by students

1. Could we create and use a dictionary that lists the prior polarity of commonly used words?

2. If the prediction score for the tweet is high, does that mean the machine is more confident about the prediction?

3. In the unigram approach, the sequence of words does not matter. But clearly, if “not” does not negate the words containing opinion, then won’t the machine learn a wrong pattern?
Questions asked in class by students

1. Could we create and use a dictionary that lists the prior polarity of commonly used words?

2. If the prediction score for the tweet is high, does that mean the machine is more confident about the prediction?

3. In the unigram approach, the sequence of words does not matter. But clearly, if “not” does not negate the words containing opinion, then won’t the machine learn a wrong pattern?

4. If we have too many negative tweets in our training data (as compared to the positive tweets), then would the machine not be predisposed to predict the polarity of an unseen tweet as negative?
Mid-term: Email classification

- 53 students
- Required to do only 2 out of the following 4 problems
# Mid-term: Email classification

- 53 students
- Required to do only 2 out of the following 4 problems

<table>
<thead>
<tr>
<th>Problem (25 points)</th>
<th>Average</th>
<th>Std-dev</th>
<th>Median</th>
<th># students attempted</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLP/ML</td>
<td>20.54</td>
<td>4.46</td>
<td>22</td>
<td>51</td>
</tr>
<tr>
<td>Logic Gates</td>
<td>16.94</td>
<td>6.48</td>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>Database design</td>
<td>13.63</td>
<td>6.48</td>
<td>14</td>
<td>42</td>
</tr>
<tr>
<td>Machine Instructions</td>
<td>12.8</td>
<td>6.81</td>
<td>14.5</td>
<td>30</td>
</tr>
</tbody>
</table>
# Mid-term: Email classification

- 53 students
- Required to do only 2 out of the following 4 problems

<table>
<thead>
<tr>
<th>Problem (25 points)</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLP/ML</td>
<td>20.54</td>
</tr>
<tr>
<td>Logic Gates</td>
<td>16.94</td>
</tr>
<tr>
<td>Database design</td>
<td>13.63</td>
</tr>
<tr>
<td>Machine Instructions</td>
<td>12.8</td>
</tr>
</tbody>
</table>
# Mid-term: Email classification

- 53 students
- Required to do only 2 out of the following 4 problems

<table>
<thead>
<tr>
<th>Problem (25 points)</th>
<th>Average</th>
<th>Std-dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLP/ML</td>
<td>20.54</td>
<td>4.46</td>
</tr>
<tr>
<td>Logic Gates</td>
<td>16.94</td>
<td>6.48</td>
</tr>
<tr>
<td>Database design</td>
<td>13.63</td>
<td>6.48</td>
</tr>
<tr>
<td>Machine Instructions</td>
<td>12.8</td>
<td>6.81</td>
</tr>
</tbody>
</table>

Sunday, September 8, 13
Mid-term: Email classification

• 53 students
• Required to do only 2 out of the following 4 problems

<table>
<thead>
<tr>
<th>Problem (25 points)</th>
<th>Average</th>
<th>Std-dev</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLP/ML</td>
<td>20.54</td>
<td>4.46</td>
<td>22</td>
</tr>
<tr>
<td>Logic Gates</td>
<td>16.94</td>
<td>6.48</td>
<td>20</td>
</tr>
<tr>
<td>Database design</td>
<td>13.63</td>
<td>6.48</td>
<td>14</td>
</tr>
<tr>
<td>Machine Instructions</td>
<td>12.8</td>
<td>6.81</td>
<td>14.5</td>
</tr>
</tbody>
</table>
## Mid-term: Email classification

- 53 students
- Required to do only 2 out of the following 4 problems

<table>
<thead>
<tr>
<th>Problem (25 points)</th>
<th>Average</th>
<th>Std-dev</th>
<th>Median</th>
<th># students attempted</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLP/ML</td>
<td>20.54</td>
<td>4.46</td>
<td>22</td>
<td>51</td>
</tr>
<tr>
<td>Logic Gates</td>
<td>16.94</td>
<td>6.48</td>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>Database design</td>
<td>13.63</td>
<td>6.48</td>
<td>14</td>
<td>42</td>
</tr>
<tr>
<td>Machine Instructions</td>
<td>12.8</td>
<td>6.81</td>
<td>14.5</td>
<td>30</td>
</tr>
</tbody>
</table>
Student projects
Student projects

• Formulate your own task
Student projects

• Formulate your own task

• Collect and annotate data
Student projects

- Formulate your own task
- Collect and annotate data
- Define the feature space
Student projects

- Formulate your own task
- Collect and annotate data
- Define the feature space
- Train and test
Student projects

• Formulate your own task

• Collect and annotate data

• Define the feature space

• Train and test

Incentive was low
Student projects

- Formulate your own task
- Collect and annotate data
- Define the feature space
- Train and test

Incentive was low

But still, 15/53 students decided to pursue the project and 11 actually managed to finish it
The Bechdel Test (bechdeltest.com)
The Bechdel Test (bechdeltest.com)

- A test for movies (Allison Bechdel, 1985)
The Bechdel Test (bechdeltest.com)

• A test for movies (Allison Bechdel, 1985)
• A movie passes this test if all 3 conditions are met:
The Bechdel Test (bechdeltest.com)

• A test for movies (Allison Bechdel, 1985)
• A movie passes this test if all 3 conditions are met:
  • There are at least 2 named female characters
The Bechdel Test (bechdeltest.com)

- A test for movies (Allison Bechdel, 1985)
- A movie passes this test if all 3 conditions are met:
  - There are at least 2 named female characters
  - They talk to each other
The Bechdel Test (bechdeltest.com)

- A test for movies (Allison Bechdel, 1985)
- A movie passes this test if all 3 conditions are met:
  - There are at least 2 named female characters
  - They talk to each other
  - They talk about something other than a man
The Bechdel Test (bechdeltest.com)

- A test for movies (Allison Bechdel, 1985)
- A movie passes this test if all 3 conditions are met:
  - There are at least 2 named female characters
  - They talk to each other
  - They talk about something other than a man
- Pass only 1 test: The Great Gatsby, Star trek into Darkness, Now you see me, The Internship
Conclusion
Conclusion

• We presented a strategy using which basic NLP/ML concepts may be taught in an introductory course, in one lecture (supported by HW and exam problems)
Conclusion

• We presented a strategy using which basic NLP/ML concepts may be taught in an introductory course, in one lecture (supported by HW and exam problems)

• The basics can be taught without using math terminology
Conclusion

• We presented a strategy using which basic NLP/ML concepts may be taught in an introductory course, in one lecture (supported by HW and exam problems)

• The basics can be taught without using math terminology

• Important outcome -- students find Watson playing Jeopardy! and Google’s self-driving car less “magical”
Thanks!

Questions?