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

Apoorv Agarwal
Columbia University

COMSI001

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

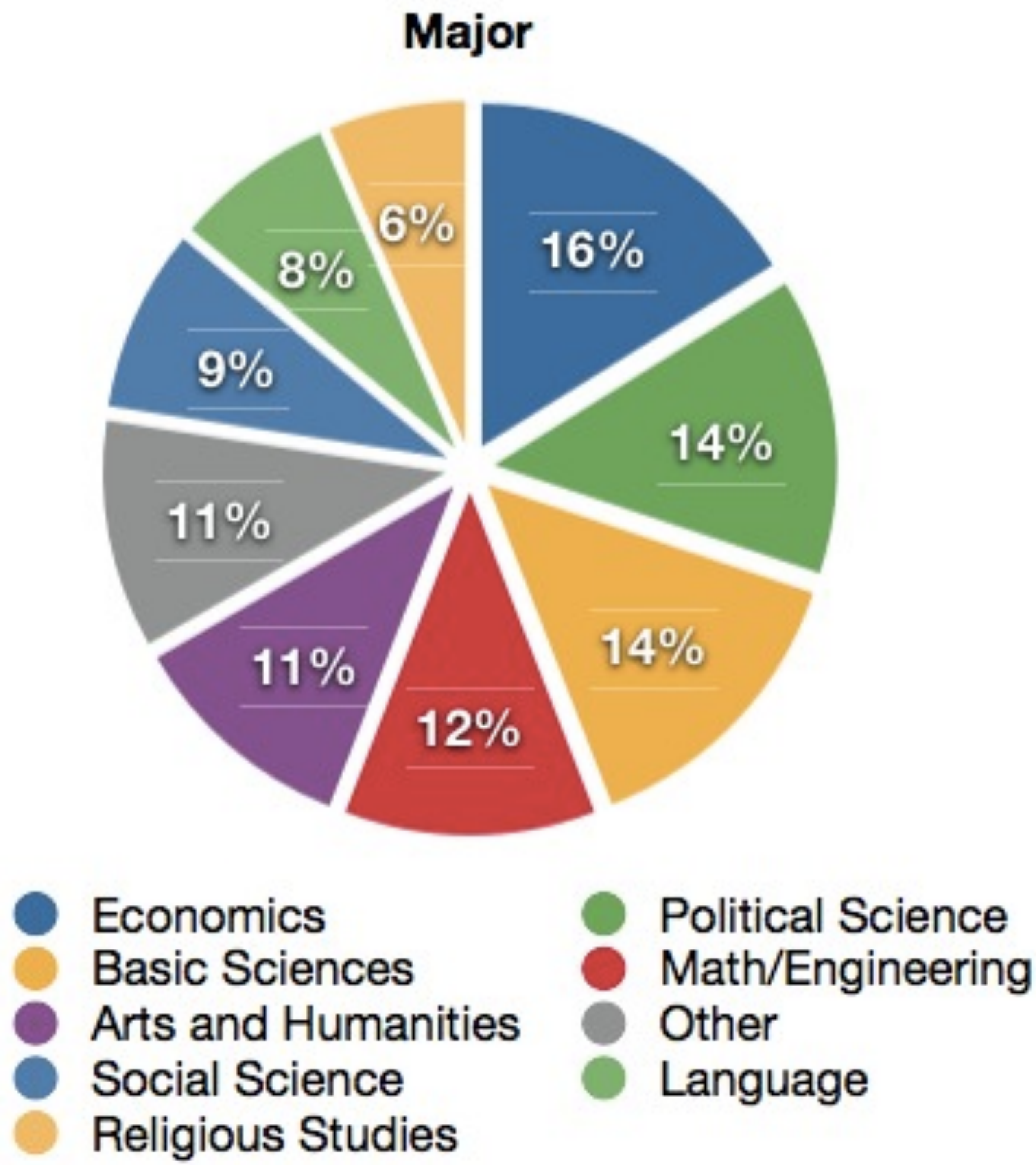
- Introductory course on information science to undergraduates at Columbia University
- Mostly taken by freshmen and sophomores
- Assumes no prior programming or math background

COMS1001

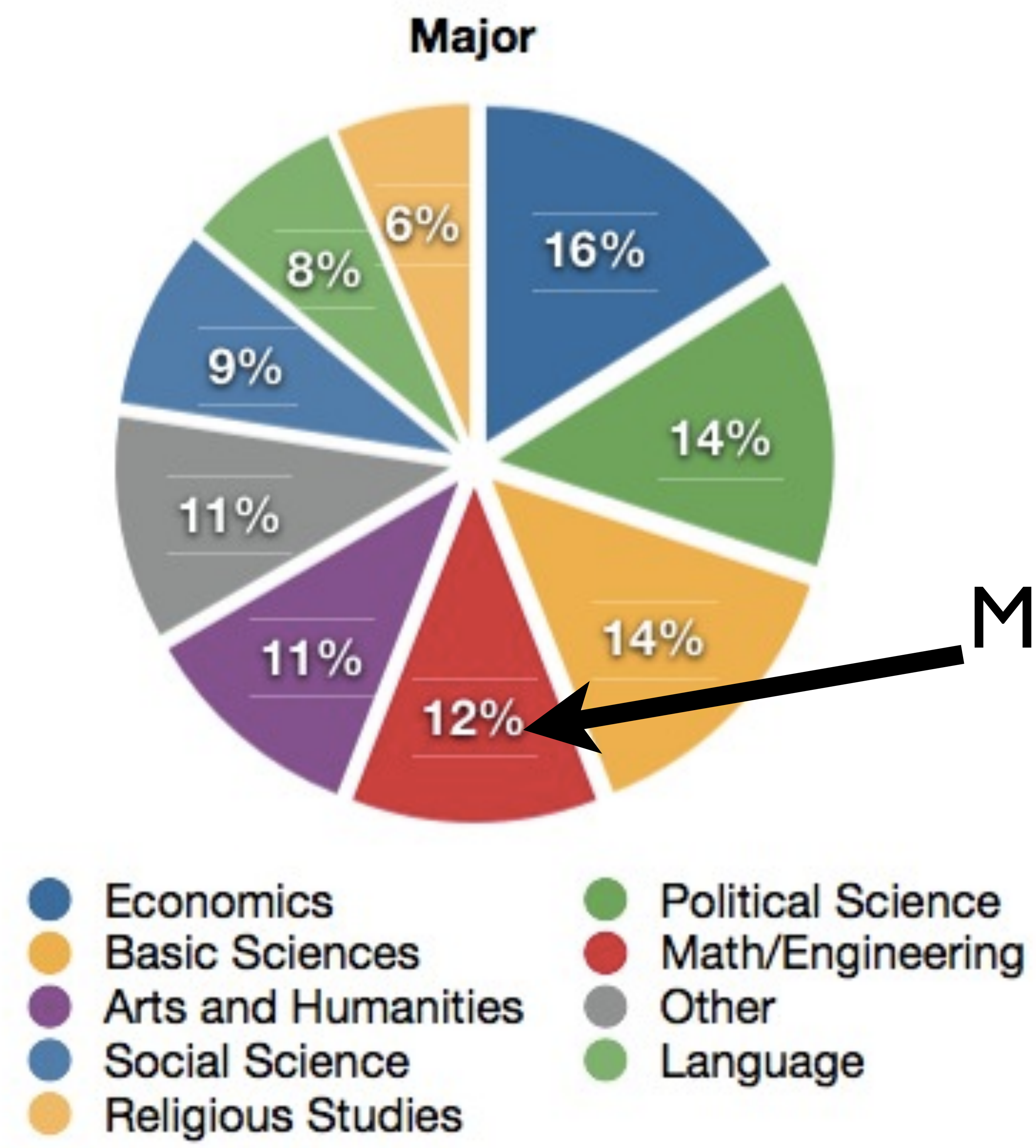
- Introductory course on information science to undergraduates at Columbia University
- Mostly taken by freshmen and sophomores
- Assumes no prior programming or math background
  - 10% : what's a programming language?

Student demographics

Student demographics

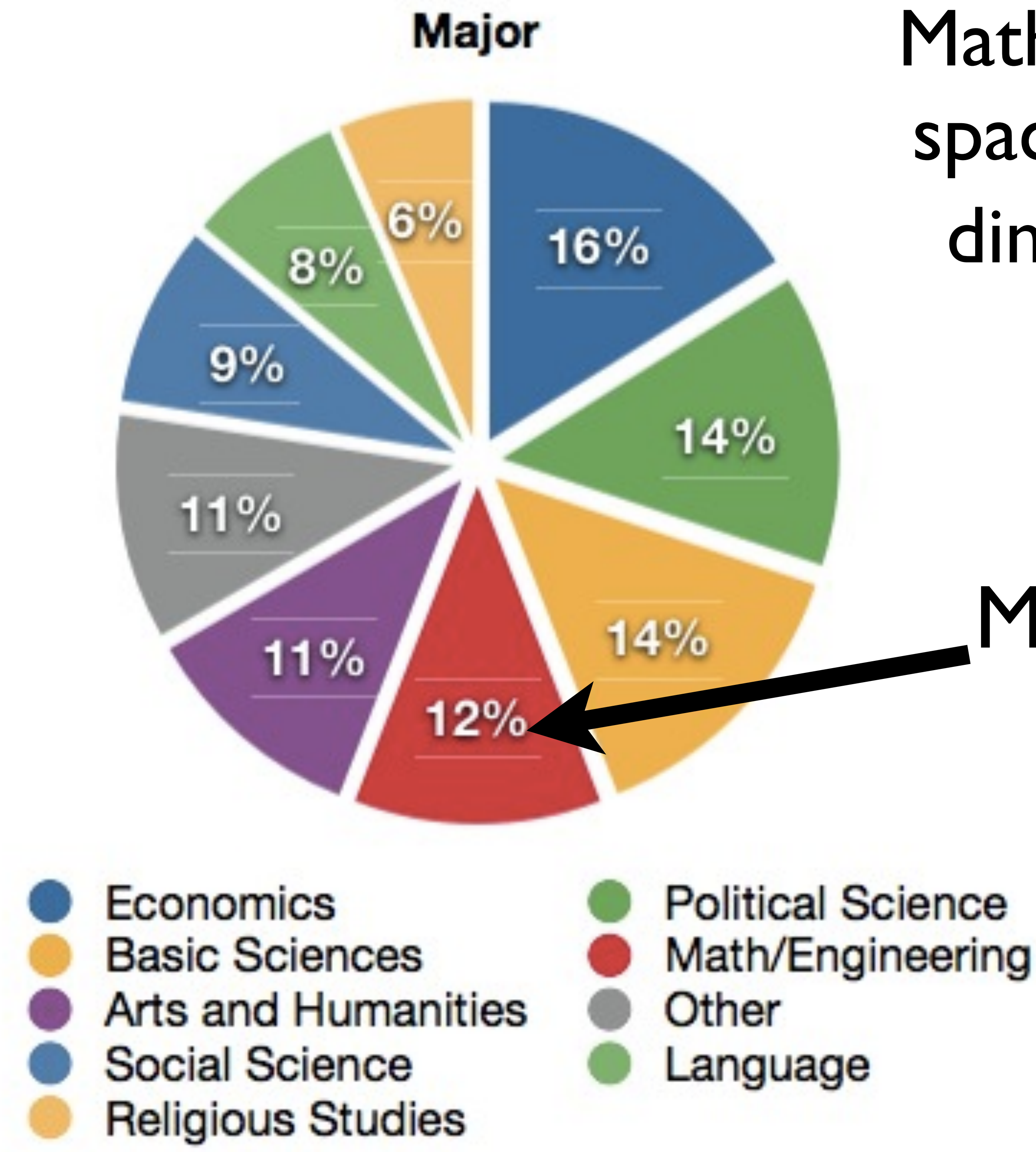


Student demographics



Math and Engineering majors

Student demographics



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

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)

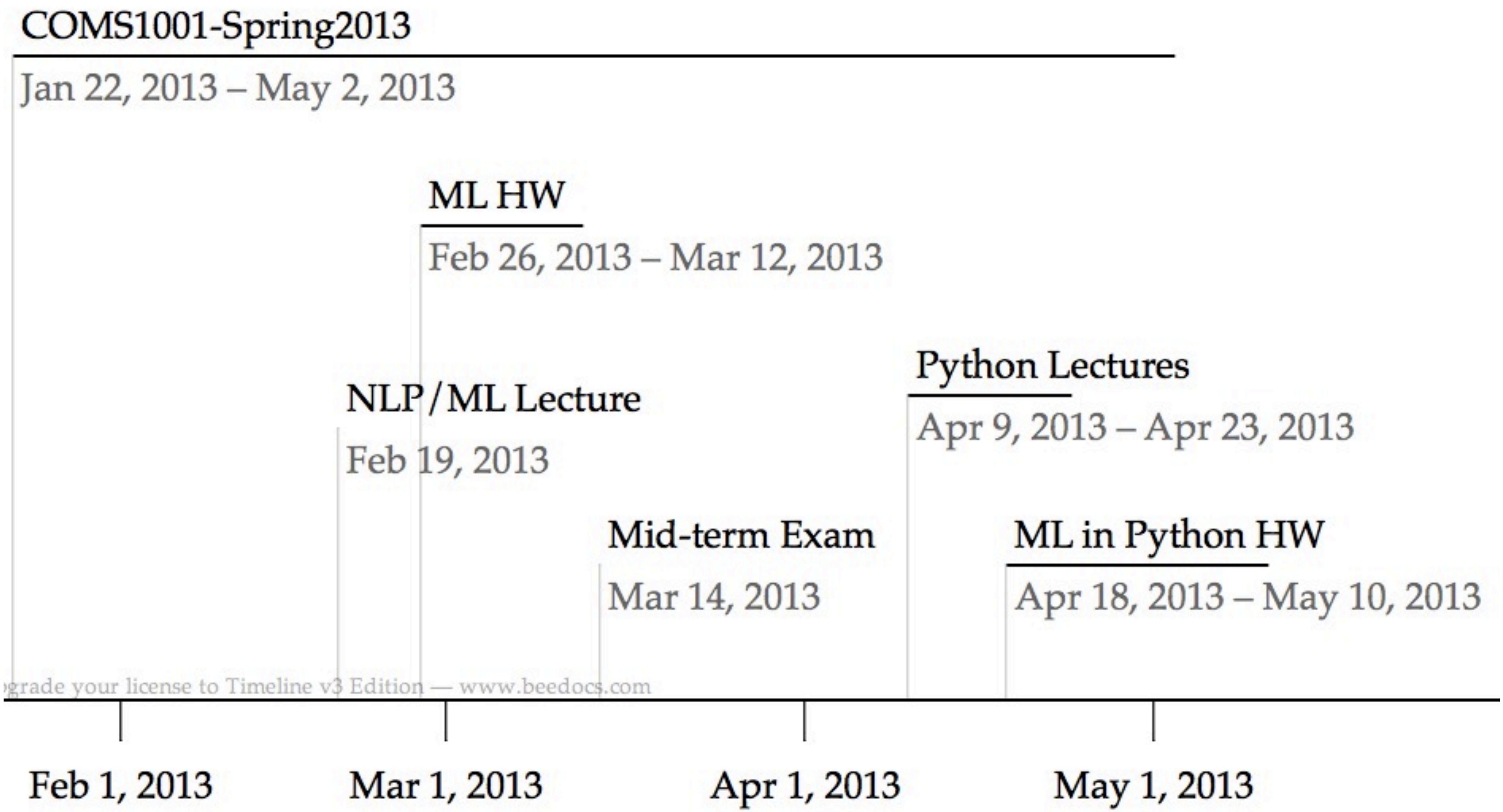
Overall Strategy

- Keep definitions simple
- Use analogies and concrete examples (also observed by Reva Freedman 2005)
- Take baby steps -- incremental learning

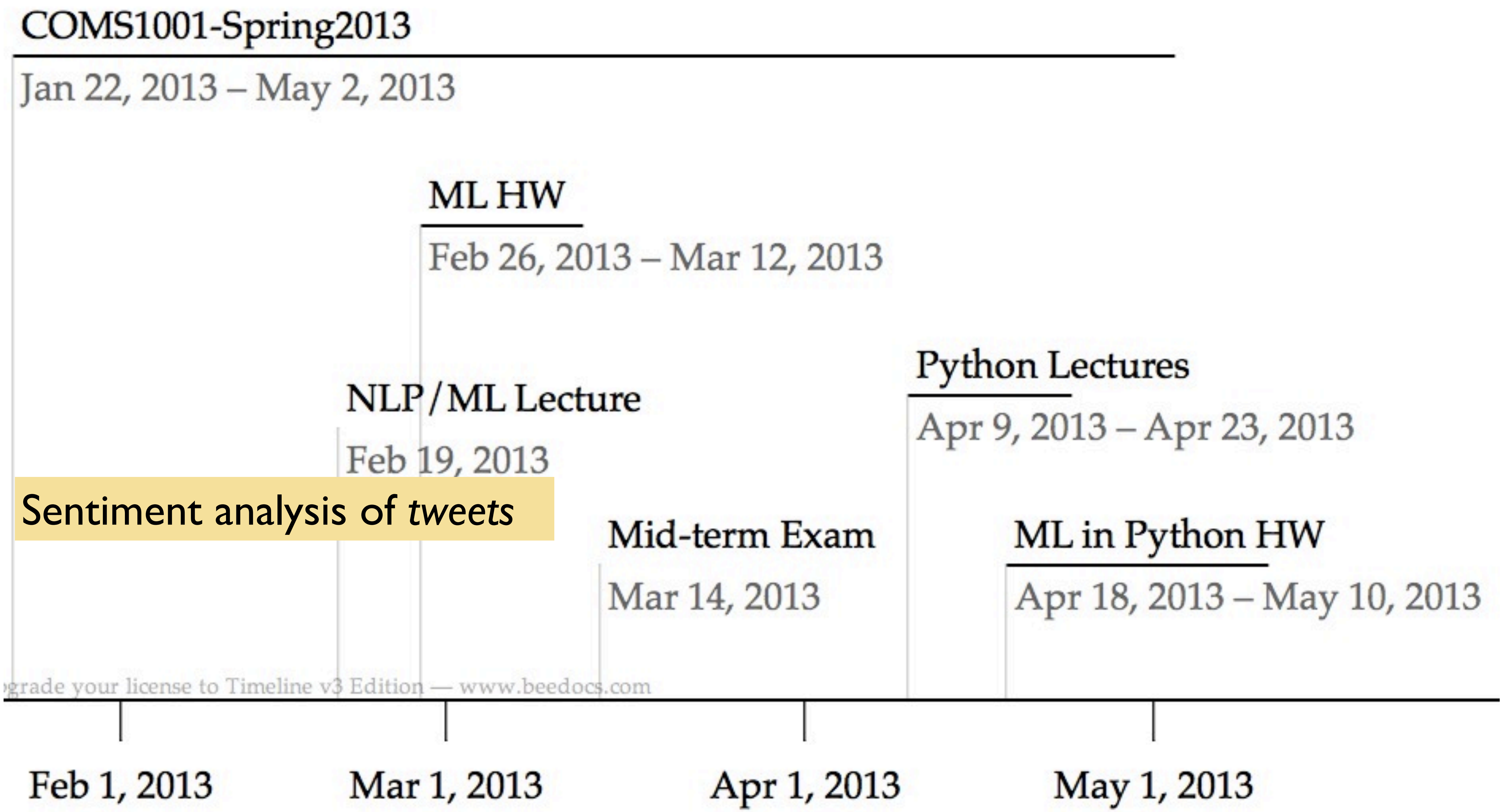
Overall Strategy

- 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

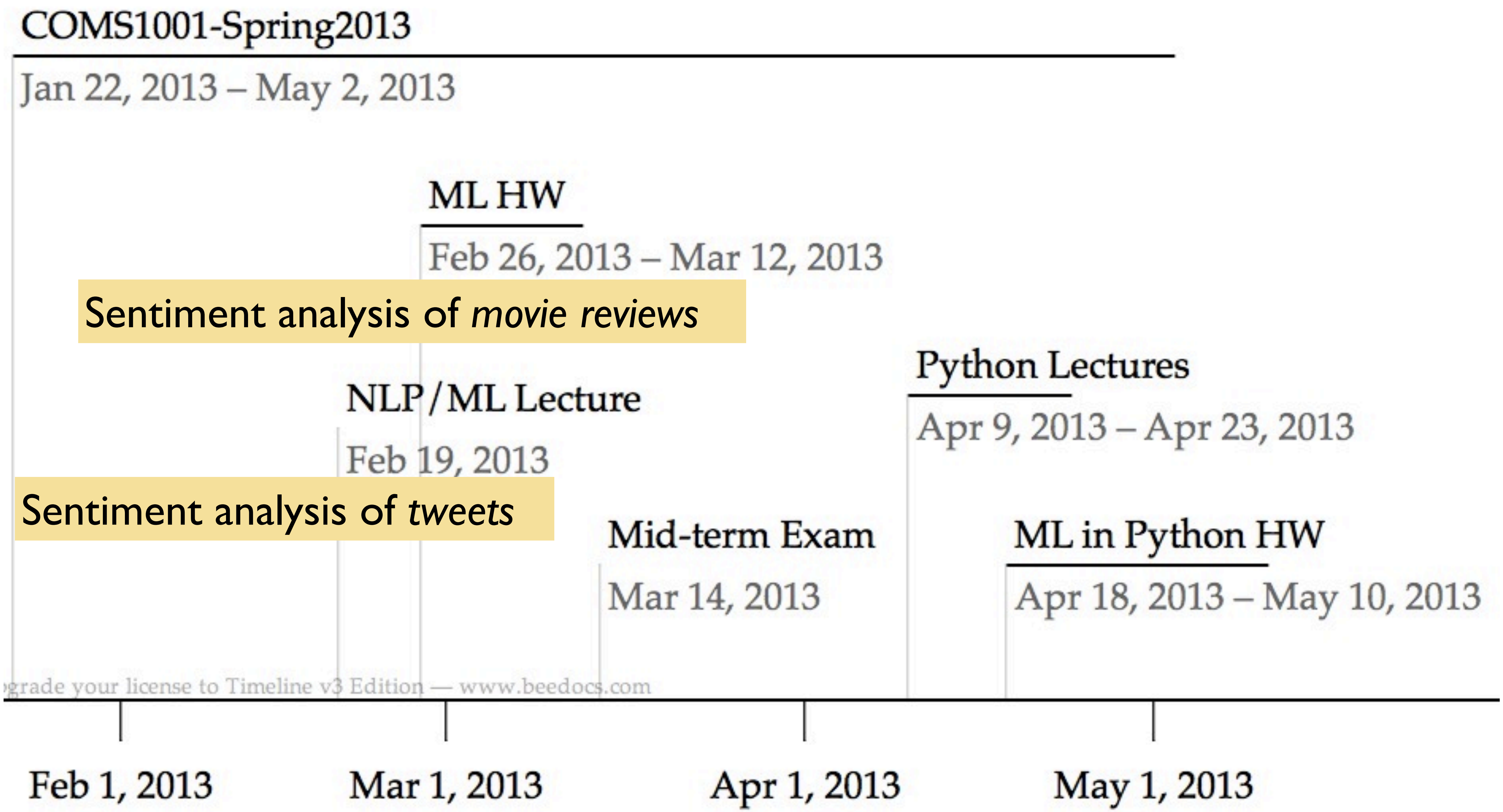

COMS1001-Spring2013
Jan 22, 2013 – May 2, 2013
ML HW
Feb 26, 2013 – Mar 12, 2013
NLP/ML Lecture
Feb 19, 2013
Python Lectures
Apr 9, 2013 – Apr 23, 2013
Mid-term Exam
Mar 14, 2013
ML in Python HW
Apr 18, 2013 – May 10, 2013
grade your license to Timeline v3 Edition — www.beedocs.com
Feb 1, 2013
Mar 1, 2013
Apr 1, 2013
May 1, 2013

Strategy



Sentiment analysis of *tweets*

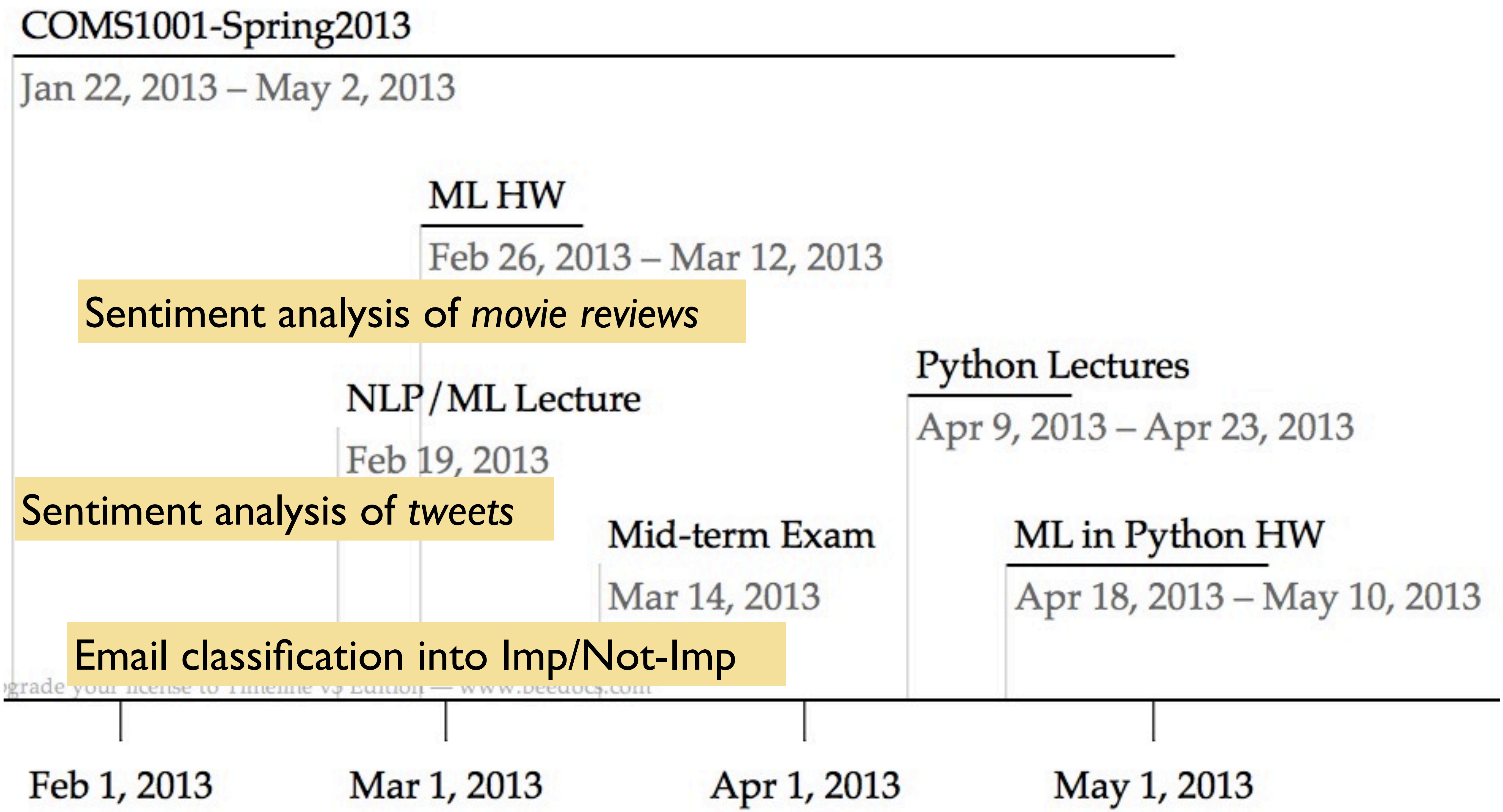
Strategy



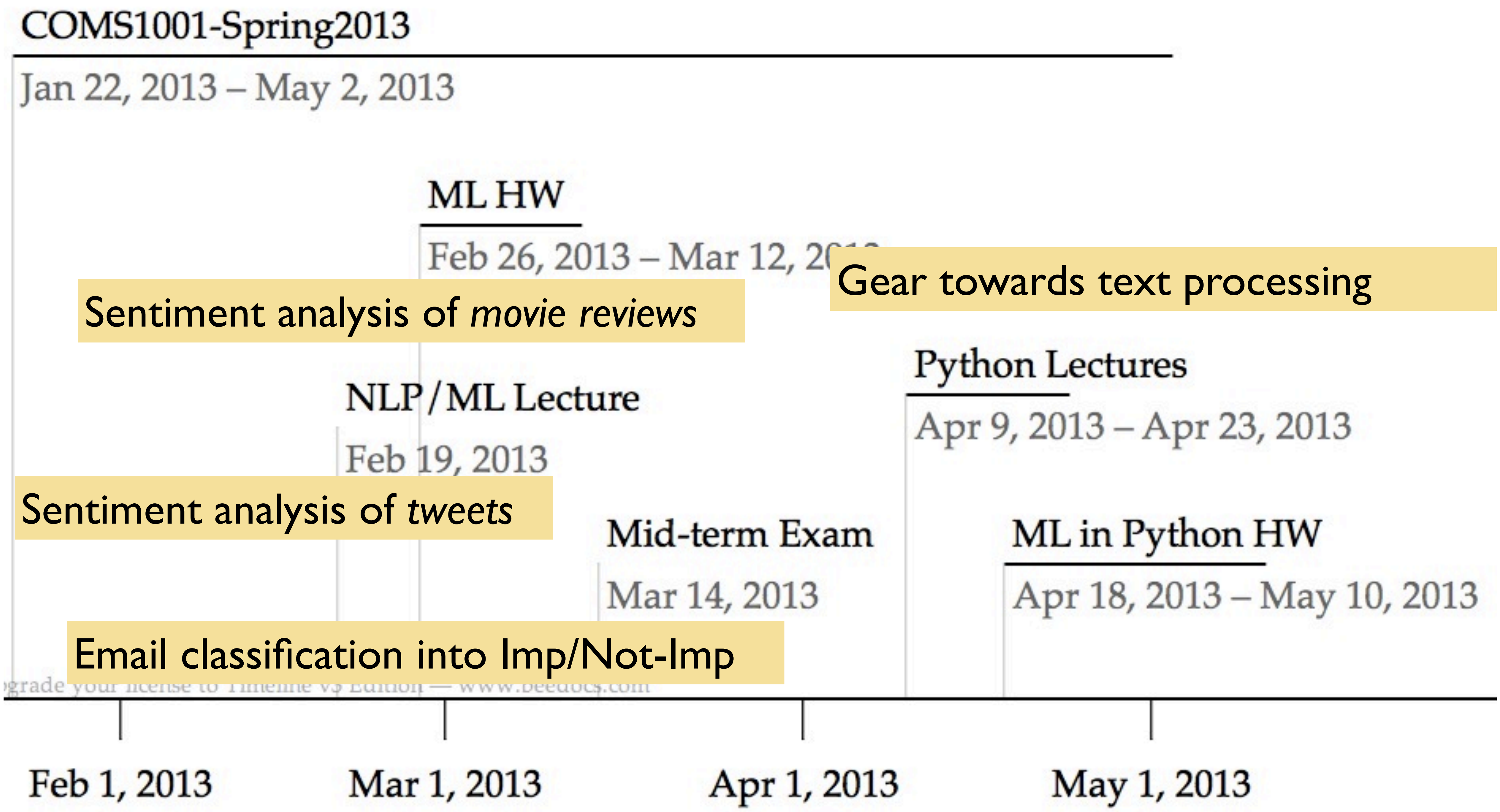
Sentiment analysis of *movie reviews*

Sentiment analysis of *tweets*

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
Python Lectures
Apr 9, 2013 – Apr 23, 2013
Sentiment analysis of *tweets*
Mid-term Exam
Mar 14, 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



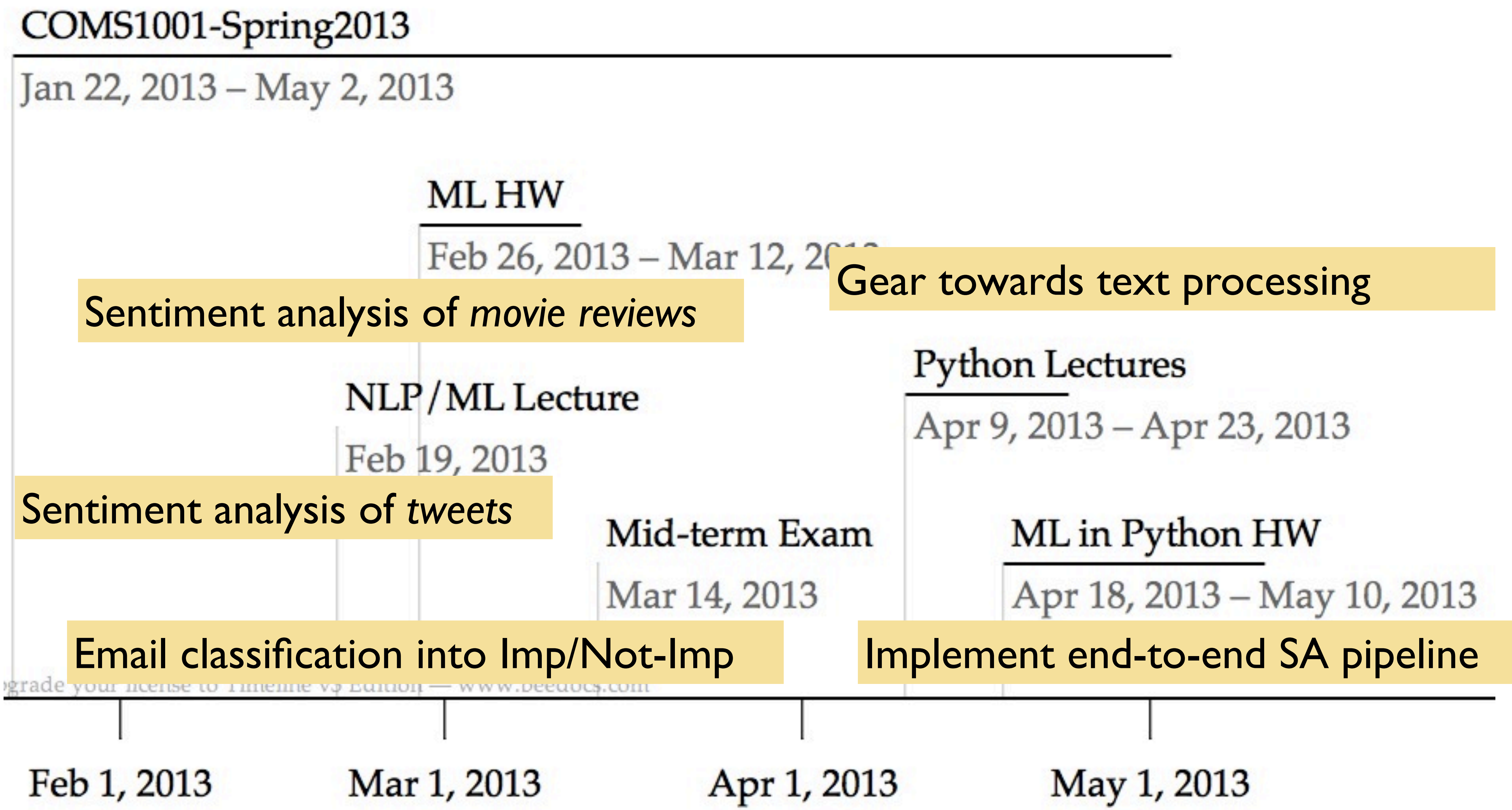
Sentiment analysis of *movie reviews*

Gear towards text processing

Sentiment analysis of *tweets*

Email classification into Imp/Not-Imp

Strategy


COMS1001-Spring2013
Jan 22, 2013 – May 2, 2013
ML HW
Feb 26, 2013 – Mar 12, 2013
Gear towards text processing
Sentiment analysis of *movie reviews*
Python Lectures
NLP / ML Lecture
Apr 9, 2013 – Apr 23, 2013
Feb 19, 2013
Sentiment analysis of *tweets*
Mid-term Exam
ML in Python HW
Mar 14, 2013
Apr 18, 2013 – May 10, 2013
Email classification into Imp/Not-Imp
Implement end-to-end SA pipeline
Feb 1, 2013
Mar 1, 2013
Apr 1, 2013
May 1, 2013

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

Problem (25 points)
NLP/ML
Logic Gates
Database design
Machine Instructions

Mid-term: Email classification

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

Problem (25 points)	Average
NLP/ML	20.54
Logic Gates	16.94
Database design	13.63
Machine Instructions	12.8

Mid-term: Email classification

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

Problem (25 points)	Average	Std-dev
NLP/ML	20.54	4.46
Logic Gates	16.94	6.48
Database design	13.63	6.48
Machine Instructions	12.8	6.81

Mid-term: Email classification

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

Problem (25 points)	Average	Std-dev	Median
NLP/ML	20.54	4.46	22
Logic Gates	16.94	6.48	20
Database design	13.63	6.48	14
Machine Instructions	12.8	6.81	14.5

Mid-term: Email classification

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

Problem (25 points)	Average	Std-dev	Median	# students attempted
NLP/ML	20.54	4.46	22	51
Logic Gates	16.94	6.48	20	36
Database design	13.63	6.48	14	42
Machine Instructions	12.8	6.81	14.5	30

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?