LING 120: Language and Computers

Semester: FALL '17

Instructor: Sowmya Vajjala

Iowa State University, USA

25 October 2017

Outline

1. LING 410X advertisement
2. Question from last class
3. A text classification algorithm: Naive Bayes
4. Chatting with Eliza: Exercise

LING 410X: Language as Data

- Introductory course on data science - but specific to working with text
- Intended audience: people from different LAS backgrounds, and with no pre-reqs.
- Content: Methods of working with text data (collecting, pre-processing, storing, mining knowledge, doing text classification, visualizing text etc.
- Means: R programming language and associated libraries.
- First taught in Spring 2017. To know about syllabus, assignments, and projects - you can talk to me.

Question from last class

1. Let us say you are working on classifying webpages as "appropriate" and "inappropriate" for children and you developed two classifiers.
2. Now let us say you have a test set that has 500 texts labeled "appropriate", 250 texts labeled "inappropriate".
3. Here are the confusion matrices for Classifiers A and B:

(A) pred. $\rightarrow$	**App.**	**Inapp.**
App.	490	10
Inapp.	200	50

(B) pred. $\rightarrow$	**App.**	**Inapp.**
App.	400	100
Inapp.	50	200

Table: Confusion matrices for two scenarios

4. What is the classification accuracy for A and B respectively?
5. According to you, which one is doing better? A or B? Why?

Follow up

Which one is better:

(A) pred. $\rightarrow$	**App.**	**Inapp.**
App.	500	0
Inapp.	200	50

(B) pred. $\rightarrow$	**App.**	**Inapp.**
App.	300	200
Inapp.	20	230

Table: Confusion matrices for two scenarios

Note: Also wrote replies to comments posted on 23rd Oct forum explaining these. Check that out.

How do we evaluate?

- In general, overall accuracy is a good measure. However, it may not be the best one if you need one category to be more accurate than the other.
- In this case: Tagging appropriate ones as inappropriate ones is bad, but tolerable. But tagging inappropriate ones as appropriate is dangerous, because children will end up having access to inappropriate content.

Steps in Text classification?

- We need a collection of example texts with known categories (Training data)
- We need to extract "features" we want the machine to learn from these (feature extraction)
- **We should take these extracted features and give them to a "learning algorithm" (training/learning phase)**
- Evaluate if the "learned" classifier is doing well by "testing" it with a few more examples with known categories (test data, evaluation)
- If you are happy, start using in some real-world application!!

Naive Bayes Classifier

- Simplest, easy to understand method to do classification
- Primarily relies on probability and bayes theorem
- Although it is not the best algorithm around, it is commonly used to set a baseline whenever you see a new text classification problem.

Probability Primer

What is probability?

- In our class, there are about 20 people.
- If I have to randomly (and unbiasedly) pick one person to ask a question now, what is the probability that it is Emily?

What is probability?

- In our class, there are about 20 people.
- If I have to randomly (and unbiasedly) pick one person to ask a question now, what is the probability that it is Emily?
  Answer: $1/20$
- What is the probability that it will be Emily or Ethan?

What is probability?

- In our class, there are about 20 people.
- If I have to randomly (and unbiasedly) pick one person to ask a question now, what is the probability that it is Emily?
  Answer: $1/20$
- What is the probability that it will be Emily or Ethan?
  Answer: $1/20 + 1/20 = 2/20$ (it may turn out I can pick Emily again too).
- What is the probability that I pick either a Freshman or an International Student?

What is probability?

- In our class, there are about 20 people.
- If I have to randomly (and unbiasedly) pick one person to ask a question now, what is the probability that it is Emily?
  Answer: $1/20$
- What is the probability that it will be Emily or Ethan?
  Answer: $1/20 + 1/20 = 2/20$ (it may turn out I can pick Emily again too).
- What is the probability that I pick either a Freshman or an International Student?
  Formula: P(Freshmen) + P(Intl. student) - P(Freshmen who are Intl. students).

What is probability? - Examples

- Look at this age distribution for 10 students:

Name	Age
Dave	25
Pete	35
Ann	27
Chen	22
Blah	21
Clah	31
Meh	32
Neh	24
Cleh	30
Greg	29

- If I randomly pick one person, what is the probability that this person is below 30 years of age?

What is probability? - Examples

- Look at this age distribution for 10 students:

Name	Age
Dave	25
Pete	35
Ann	27
Chen	22
Blah	21
Clah	31
Meh	32
Neh	24
Cleh	30
Greg	29

- If I randomly pick one person, what is the probability that this person is below 30 years of age? Ans: 6/10
- If I randomly pick one person, what is the probability that it is Dave? what is the probability that this is not Dave?

What is probability? - Examples

- Look at this age distribution for 10 students:

Name	Age
Dave	25
Pete	35
Ann	27
Chen	22
Blah	21
Clah	31
Meh	32
Neh	24
Cleh	30
Greg	29

- If I randomly pick one person, what is the probability that this person is below 30 years of age? Ans: 6/10
- If I randomly pick one person, what is the probability that it is Dave? what is the probability that this is not Dave? Ans: 1/10 and 9/10.

Conditional probability

- Conditional probability is the probability of one event happening, when we know some other event has happened before.
- If one of my events is seeing 2 when I roll a die (E1), the other event is seeing an even number (E2), then, $P(E1|E2)$ is: $1/3$. Why??

Conditional probability

- Conditional probability is the probability of one event happening, when we know some other event has happened before.
- If one of my events is seeing 2 when I roll a die (E1), the other event is seeing an even number (E2), then, P(E1|E2) is: 1/3. Why??
- What is P(E1)?

Conditional probability

- Conditional probability is the probability of one event happening, when we know some other event has happened before.
- If one of my events is seeing 2 when I roll a die (E1), the other event is seeing an even number (E2), then, P(E1|E2) is: $1/3$. Why??
- What is P(E1)? $1/6$
- What is P(E2)?

Conditional probability

- Conditional probability is the probability of one event happening, when we know some other event has happened before.
- If one of my events is seeing 2 when I roll a die (E1), the other event is seeing an even number (E2), then, P(E1|E2) is: 1/3. Why??
- What is P(E1)? 1/6
- What is P(E2)? 3/6 i.e., 1/2
- What is P(E1, given E2)

Conditional probability

- Conditional probability is the probability of one event happening, when we know some other event has happened before.
- If one of my events is seeing 2 when I roll a die (E1), the other event is seeing an even number (E2), then, P(E1|E2) is: 1/3. Why??
- What is P(E1)? 1/6
- What is P(E2)? 3/6 i.e., 1/2
- What is P(E1, given E2) E2 has three possibilities (2,4,6). So, probability of getting 2 is 1/3.

Joint probability

- Probability that two events occur together. Represented as P(A,B) and is the same as P(B,A)
- So what is the difference between joint and conditional probability?

Joint probability

- Probability that two events occur together. Represented as P(A,B) and is the same as P(B,A)
- So what is the difference between joint and conditional probability?
- Useful example: `https://goo.gl/9MVM78`

Joint probability

- Probability that two events occur together. Represented as P(A,B) and is the same as P(B,A)
- So what is the difference between joint and conditional probability?
- Useful example: `https://goo.gl/9MVM78`
- Additional information:
  - Conditional probability is not commutative $P(A|B) \mathrel{!}= P(B|A)$
  - $P(A,B) = P(A|B)*P(B) = P(B|A)*P(A)$

Bayes Theorem

- $P(A|B) = \frac{P(B|A)*P(A)}{P(B)}$ - this is the theorem.
- Example when applied to Spam classification:
  1. P(Spam|Email) $= \frac{P(Email|Spam)*P(Spam)}{P(Email)}$
  2. P(Ham|Email) $= \frac{P(Email|Ham)*P(Ham)}{P(Email)}$
  3. Each time we see a new email, we calculate these two probabilities. If the first one is higher, we classify the email as spam. Else, as ham!
- P(Email) and P(Spam) are probabilities of seeing the Email and Probability of a spam email in your training data.
- How do we get P(Spam)?

Bayes Theorem

- $P(A|B) = \frac{P(B|A)*P(A)}{P(B)}$ - this is the theorem.
- Example when applied to Spam classification:
  1. $\text{P(Spam|Email)} = \frac{P(Email|Spam)*P(Spam)}{P(Email)}$
  2. $\text{P(Ham|Email)} = \frac{P(Email|Ham)*P(Ham)}{P(Email)}$
  3. Each time we see a new email, we calculate these two probabilities. If the first one is higher, we classify the email as spam. Else, as ham!
- P(Email) and P(Spam) are probabilities of seeing the Email and Probability of a spam email in your training data.
- How do we get P(Spam)?
- How do we get P(Email)?

Bayes Theorem

- $P(A|B) = \frac{P(B|A)*P(A)}{P(B)}$ - this is the theorem.
- Example when applied to Spam classification:
  1. P(Spam|Email) $= \frac{P(Email|Spam)*P(Spam)}{P(Email)}$
  2. P(Ham|Email) $= \frac{P(Email|Ham)*P(Ham)}{P(Email)}$
  3. Each time we see a new email, we calculate these two probabilities. If the first one is higher, we classify the email as spam. Else, as ham!
- P(Email) and P(Spam) are probabilities of seeing the Email and Probability of a spam email in your training data.
- How do we get P(Spam)?
- How do we get P(Email)? Do we even need it?

Bayes Theorem

- $P(A|B) = \frac{P(B|A)*P(A)}{P(B)}$ - this is the theorem.
- Example when applied to Spam classification:
  1. $\mathrm{P(Spam|Email)} = \frac{P(Email|Spam)*P(Spam)}{P(Email)}$
  2. $\mathrm{P(Ham|Email)} = \frac{P(Email|Ham)*P(Ham)}{P(Email)}$
  3. Each time we see a new email, we calculate these two probabilities. If the first one is higher, we classify the email as spam. Else, as ham!
- P(Email) and P(Spam) are probabilities of seeing the Email and Probability of a spam email in your training data.
- How do we get P(Spam)?
- How do we get P(Email)? Do we even need it?
- What is the difference between P(Spam|Email) and P(Email | Spam)?

Bayes Theorem

- $P(A|B) = \frac{P(B|A)*P(A)}{P(B)}$ - this is the theorem.
- Example when applied to Spam classification:
  1. $\text{P(Spam|Email)} = \frac{P(Email|Spam)*P(Spam)}{P(Email)}$
  2. $\text{P(Ham|Email)} = \frac{P(Email|Ham)*P(Ham)}{P(Email)}$
  3. Each time we see a new email, we calculate these two probabilities. If the first one is higher, we classify the email as spam. Else, as ham!
- P(Email) and P(Spam) are probabilities of seeing the Email and Probability of a spam email in your training data.
- How do we get P(Spam)?
- How do we get P(Email)? Do we even need it?
- What is the difference between P(Spam|Email) and P(Email | Spam)?

Reading recommendation:

`http://www.ling.upenn.edu/courses/cogs501/Bayes1.html`

How do we combine all evidence?

- We are operationalizing an email as a bunch of words.
- So ohow should we calculate P(Email|Spam) and P(Email|Ham)?

How do we combine all evidence?

- We are operationalizing an email as a bunch of words.
- So ohow should we calculate P(Email|Spam) and P(Email|Ham)?
- Product of individual word probabilities!
  $P(\text{Email}|\text{Spam}) = P(\text{Spam}) * \Pi_f P(f|\text{Spam})$
  $P(\text{Email}|\text{Ham}) = P(\text{Ham}) * \Pi_f P(f|\text{Ham})$
  Where f stands for "feature". In our example, we took words. But other features are: "is it all upper case", "is there large amounts of money mentioned" etc.

For further study

- We spoke briefly about others like - looking for nearest neighbors, creating a linear separator between classes, neural networks etc.
- There are 100s more.
- For more mathematical orientation on these, you should take a Machine Learning course
- For more practical applications, you should take courses on areas where machine learning is used to solve specific problems - such as natural language processing and computer vision.
- ISU offers a lot of these courses!
- Something to get started: `https://web.stanford.edu/~jurafsky/slp3/` (Chapter 6)

Preview to next topic: Attendance exercise

Post on Canvas.

- go to: `http://psych.fullerton.edu/mbirnbaum/psych101/Eliza.htm`
- Chat with Eliza for sometime and write your comments on the interaction addressing the below questions:
- Is it doing a good job of chatting? What is happening - how do you think is it able to understand what you say?
- Does it fail? In what cases?