Foundations of Programming

Naïve Bayes revisited and Debugging Strategies

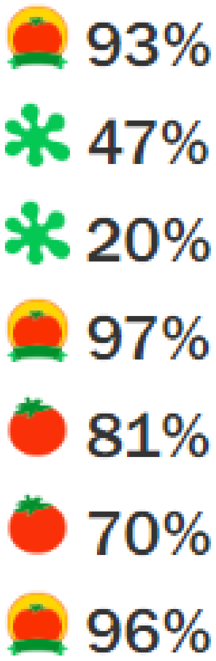
Announcements

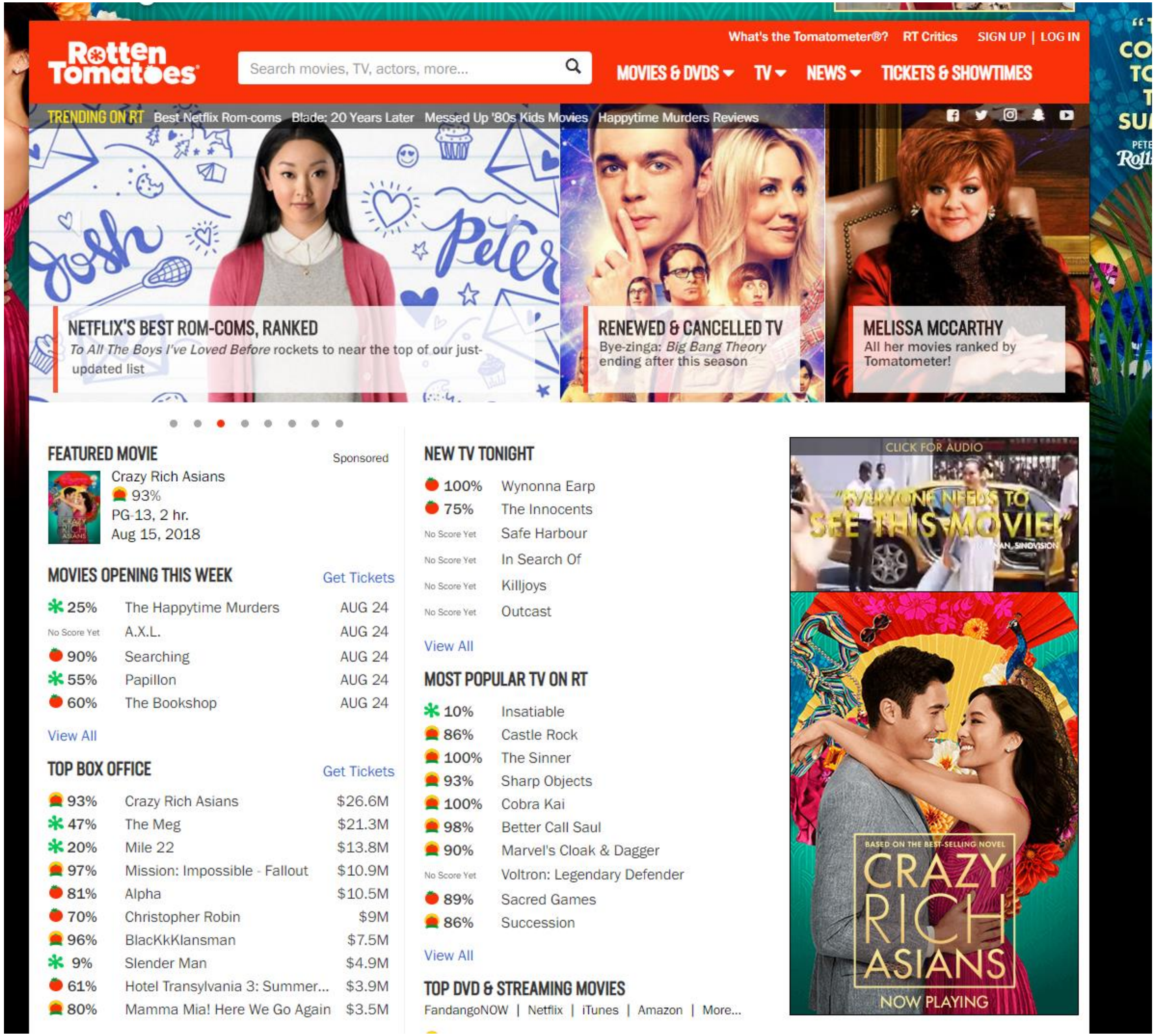
- Diversity in Computing Lunch: Meet here after next class

Learning outcomes/key ideas

- Probability and Naïve Bayes, revisited
- Debugging strategies

Text sentiment classification





Joint probability: Two or more events BOTH happening

Of 100 students completing a course, 20 were business (B) major. Ten students received As (A) in the course, and three of these were business majors.

What is the probability that a randomly selected student is a Business student and gets an A?
P(A = True, B = True) = ?

A. 0.03 B. 0.1 C. 0.3 D. 0.2 E. None of these

"the probability of A and B"

	B(usiness student) = True	B(usiness student) = False
A = True	0.03	0.07
A = False	0.17	0.73

P(A=True) = P(A=True, B=True) + P (A=True, B=False)

0.03 0.07

marginalization

sum gives Prior P(A)

Will the values in a joint probability table always sum to 1?

A. Yes B. No

But how does evidence change things?
Conditional probability

Of 100 students completing a course, 20 were business major. Ten students received As in the course, and three of these were business majors What is the probability of A after knowing B is true?

What is the probability that a randomly selected *business student* gets an A?

$3/20 = 0.15$

P(A = True | B = True) = ?   A. 0.03 B. 0.1 C. 0.3 D. 0.2 E. None of these

*"the probability of A **given** B"*

	B(usiness student) = True	B(usiness student) = False
A = True	0.03	0.07
A = False	0.17	0.73

$0.03 / (0.03 + 0.17)$

$\boxed{P(A=True \mid B = True) = P(A=True, B=True) / P(B=True)}$

$P(A,B) = P(A|B)\,P(B)$

$P(A,B) = P(B|A)\,P(A)$

Bayes Theorem

If $P(E2)>0$, then

$P(E1|E2) = P(E2|E1)P(E1) / P(E2)$

This can be derived from the definition of conditional probability.

Bayes Rule example II

Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding?

Let's start with this question: What if the weatherman hadn't made a prediction at all? What is the probability it will rain?

Returning to review sentiment classification…

Our evidence is the text in the review. We want to estimate

$$P(Sentiment = pos \mid text)$$

But we could never directly estimate this because we're unlikely to have ever seen this specific text before! How can Bayes rule help us?

$$P(Sentiment = pos \mid text) = \frac{P(text|Sentiment = pos)\, P(Sentiment = pos)}{P(text)}$$

Returning to review sentiment classification…

$$P(Sentiment = pos \mid text) \propto P(text|Sentiment = pos)\ P(Sentiment = pos)$$

Simplifying assumption 1: Represent text with a "bag of words" representation

$$P(w_1 = true, w_2 = true, w_3 = false, \ldots |Sentiment = pos)$$

But can we learn this??

Independence: Intuition

- Events are independent if one has nothing whatever to do with others. Therefore, for two independent events, knowing one happening does not change the probability of the other event happening.
  - one toss of coin is independent of another coin (assuming it is a regular coin).
  - price of tea in England is independent of the result of general election in Canada.

Independence: Definition

- Events A and B are independent iff:

$$P(A, B) = P(A) \times P(B)$$

which is equivalent to

$$P(A|B) = P(A) \text{ and}$$

$$P(B|A) = P(B)$$

when $P(A, B) > 0$.

T1: the first toss is a head.

T2: the second toss is a tail.

$$P(T2|T1) = P(T2)$$

Conditional Independence: Utility via Naïve Bayes

- Let E1 and E2 be two events, they are conditionally independent given E iff

  P(E1|E, E2)=P(E1|E)

  that is the probability of E1 is not changed after knowing E2, given E is true.
- Equivalent formulations:

  P(E1, E2|E)=P(E1|E) P(E2|E)

  P(E2|E, E1)=P(E2|E)

$$P(w_1 = true, w_2 = true, w_3 = false, \dots | Sent = pos)$$
$$= P(w_1 = true | Sent = pos)\, P(w_2 = true | Sent = pos) P(w_3 = false | Sent = pos) \dots$$

A ha! These we can learn from data!

Returning to review sentiment classification…

$$P(Sentiment = pos \mid text) \propto P(text|Sentiment = pos)\ P(Sentiment = pos)$$

Simplifying assumption 1: Represent text with a "bag of words" representation

$$P(w_1 = true, w_2 = true, w_3 = false, \ldots |Sentiment = pos)$$

Simplifying assumption 2: Words are conditionally independent given sentiment

$$P(w_1 = true, w_2 = true, w_3 = false, \ldots |Sent = pos)$$
$$= P(w_1 = true|Sent = pos)\ P(w_2 = true|Sent = pos) P(w_3 = false|Sent = pos) \ldots$$

Training a Naïve Bayes classifier

$$P(Sentiment = pos \mid text) \propto P(text|Sentiment = pos)\, P(Sentiment = pos)$$

Simplifying assumption 1: Represent text with a "bag of words" representation

$$P(w_1 = true, w_2 = true, w_3 = false, \ldots |Sentiment = pos)$$

Simplifying assumption 2: Words are conditionally independent given sentiment

$$P(w_1 = true, w_2 = true, w_3 = false, \ldots |Sent = pos)$$
$$= P(w_1 = true|Sent = pos)\, P(w_2 = true|Sent = pos) P(w_3 = false|Sent = pos) \ldots$$

Practice with Probability

- Which of the following statements are generally true? (If they are true only in certain conditions, state what the conditions are)

$P(A,B) = P(A)*P(B)$

$P(A,B) = P(A|B)$

$P(A,B) = P(A|B)P(B)$

$P(A|B) + P(A|\neg B) = 1$

$P(\neg A) + P(A) = 1$

$P(\neg A, B) + P(A, B) = P(B)$

$P(\neg A|B) + P(A|B) = P(B)$