Sample questions

Ekaterina Kochmar

May 27, 2017

Question 1

Suppose there is a movie rating website where $User_1$, $User_2$ and $User_3$ post their reviews. $User_1$ has written 30 positive (5-star reviews), 40 neutral (2 to 4 stars), and 30 negative (1-star) reviews in total; $User_2$ has written 10 positive, 10 neutral, and 0 negative reviews; and $User_3$ has written 30 positive, 30 neutral and 40 negative reviews.

The website uses a system that selects one of the reviews from this set to post on the main page every day.

(a) If a review posted today is chosen by such a system at random with probabilities $p(User_1) = 0.2$, $p(User_2) = 0.2$, and $p(User_3) = 0.6$, and then a review is selected at random (with equal probability of selecting any of the reviews from the chosen user), what is the probability that a *positive* review is posted on the main page?

(b) If we observe that the review posted on the main page today is in fact *neutral*, what is the probability that it was written by $User_1$?

Question 2

This question is on analysis and interpretation of the results of ML algorithms.

Suppose your inbox contains 100 emails, 19 of which are spam and the rest are emails from your friends, University, etc. (non-spam). You try out two spam filters:

Filter1:

	Non-spam (predicted)	Spam (predicted)	Total
Non-spam (actual)	65	16	81
Spam (actual)	4	15	19
Total	69	31	100

Filter2:

	Non-spam (predicted)	Spam (predicted)	Total
Non-spam (actual)	80	1	81
Spam (actual)	15	4	19
Total	95	5	100

What is the performance of each of the filters? If you only had a choice between these two filters in practice, which one would you choose and why?

Question 3

Questions on statistical laws of language

Briefly summarise Heaps' law and Zipf's law. What each of these tell you about the distribution of words in language? Why and how this can be taken into account in practical applications?

Question 4

(a) Describe an example of a graph where the diameter is more than three times as large as the average distance.

(b) Describe the concept of a triadic closure in social network theory.