Lab 4: Solutions

These solutions available in an html version[1] or a pdf version[2].

Pointwise mutual information

PMI using counts is:

$$\text{PMI}(x,y) = \log_2 \frac{N \cdot C(x,y)}{C(x)C(y)}$$

which can be derived from the fact that the MLE estimates are *P(x) = C(x)/N*, *P(y) = C(y)/N*, *P(x,y) = C(x,y)/N*.

Therefore:

$$\text{PMI}(x,y) = \log_2 \frac{12 \cdot 2}{6 \cdot 4} = \log_2 1 = 0$$

$$\text{PMI}(x,z) = \log_2 \frac{12 \cdot 1}{6 \cdot 3} = \log_2 \frac{2}{3} < 0$$

$$\text{PMI}(y,z) = \log_2 \frac{12 \cdot 2}{4 \cdot 3} = \log_2 2 = 1$$

- What do negative, zero and positive PMI values mean? (relate them to statistical independence).

  A PMI(x,y) = 0 means that the particular values of x and y are statistically independent; positive PMI means they co-occur *more* frequently than would be expected under an independence assumption, and negative PMI means they cooccur *less* frequently than would be expected.

  There is an important subtlety here: PMI is defined only over particular values of x and y, and can therefore be negative, zero, or positive; it considers only the independence of those two particular values. Normally when we talk about two random variables X and Y being independent, we actually mean that *P(X,Y) = P(X)P(Y)* for *all possible* values of X and Y. Similarly, the (non-pointwise) Mutual Information I between X and Y is defined as the expected value of PMI across all possible values of X and Y (here, all possible choices of word pairs):

  $$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) \log_2 \frac{P(x,y)}{P(x)P(y)}$$

  Unlike PMI, *I(X,Y)* cannot be a negative number, it always takes non-negative values. *I(X,Y) = b* can be interpreted to mean that knowing the value of X will, on average, reduce the uncertainty about the value of Y by b bits. This kind of interpretation doesn't make sense for PMI because negative PMI values still indicate important statistical correlations between x and y.

Examining and running the code

- How big is the `o_counts` dictionary?

  ```
  len(o_counts)
  ```

  gives 983308, the total number of distinct wordforms in the dataset.

- Which word occurs more often in this dataset, `like` or `love`?

  ```
  o_counts['like']
  ```

  `like` occurs 39941 times, while `love` only occurs 28956 times.

[1] http://homepages.inf.ed.ac.uk/sgwater/teaching/lsa2015/labs/lab4-sol.html

[2] http://homepages.inf.ed.ac.uk/sgwater/teaching/lsa2015/labs/lab4-sol.pdf

[3] http://homepages.inf.ed.ac.uk/sgwater/teaching/lsa2015/labs/lab4-sol.py

Do Twitter users like Justin Bieber?

See lab4-sol.py[3] for code. The PMI values for `love` and `hate` are 1.54 and 0.36 respectively, suggesting that Twitter users who mention Justin Bieber tend to like him, although the positive (but weak) PMI with `hate` suggests that at least some users feel negatively towards him. These scores are actually a bit unstable in this dataset if you start adding other positive/negative words, though in a larger dataset we found that overall Justin Bieber had a much higher postive than negative sentiment score.

Investigating other words

- Which of these two words (`husband` or `wife`) occurs more in this dataset. By how much? What are two possible explanations for this difference?

  `husband` occurs only about half as much as `wife` (562 vs 1074). I thought of two reasons: (1) people don't talk about husbands as much as they talk about wives, or (2) there are more men on Twitter than women, and everyone talks about their spouses equally often. Students thought of lots of other possibilities:

  - `husband`/`wife` are used self-referentially and there are more *women* than *men* on twitter, or women just refer to their role as wife more.
  - women tend to use other words for their husband (eg. `hubby`, or the husband's name).
  - men just tweet more than women (even if the number of users is similar).
  - Some popular movie, TV show or book with `wife` in the title came out at the time this data was collected.

- Are there noticeable differences in the sentiment of tweets in which people refer to husbands or wives? If so, what are they?

  If using just `love` and `hate` as the sentiment words, we can see that people use both words more often than by chance with `wife`, but the effect is stronger for `love`. For `husband`, people again use `love` more than by chance, but are *less* likely than chance to say `hate`.

  If you expand the list of sentiment words, you're likely to find similar results but less pronounced. If you use a co-occurrence frequency cutoff, however, you may find that `husband` no longer looks like it avoids negative words. This is likely an artifact: notice that `husband` is not that high frequency, and almost all the negative words only appear with it once or twice. Although any one of these probably isn't reliable, across a range of words the effect is more likely to be reliable. But if we filter out all low-frequency co-occurrences, we're left with only the negative words that happen to occur more frequently than expected with `husband`, so it looks like the negative-word avoidance pattern goes away.

- Pick a few words you looked at that you think are interesting and speak to some of the questions and issues above.

  See our code for some other words we tried. One issue in this dataset is that negative words in general are used less than positive ones, so it's hard to get enough cooccurrences with negative words to fully believe the PMI values if you add words besdies `hate`.

  (Also note that a few students confused the larger cooccurrence counts of some target words with `love` than with `hate` to mean that people felt positively towards those target words. However since `love` occurs more frequently in the data set overall, the higher co-occurrence counts don't necessarily mean that. In fact, that's exactly why we use PMI instead of raw cooccurrences: it tells us whether there are *surprisingly* many co-occurrences given the raw numbers of each word.)

  Despite these issues, our results suggest a few interesting things that could be explored further using larger datasets or more rigorous methods. For example,

  - Of the words I tested, the one people are most negative about is `myself`. Some words suggested by students that rival or even out-negative `myself` are `ex` and `men` (which is notably unlike `husband`, and also considerably more negative than `women`).
  - People are only mildly positive about Christmas, but they avoid being negative about it.
  - Unlike `bieber`, people don't seem to use emotional language when talking about `obama` (negative pmi for both sentiments). This was a bit surprising to me, but could be partly due to lots of official/news tweets, rather than personal ones. (One student found the same thing about `president`.)

- The most `hate`-avoiding word I found was `haiti`, which suffered an earthquake during the time these tweets were collected. It also seems to avoid negative sentiment in general, as well as positive sentiment. (Again, perhaps due to news items.)
- People `love` movies, but not `tv`. In fact, considering all sentiment words, `tv` has a mostly negative sentiment whereas `movie` has both, but more on the positive side.
- A lot of people `hate` Facebook. And not many people `love` it.
- Students found (among other things) that, unsurprisingly, people tend to `hate` Monday but not Friday; `hate` winter and college, and `love` dance. If you use more sentiment words, you'll probably find similar trends but weaker.