CS 7650 Problem Set 1: Review of Probability

Chao Shi, 902604979

August 18, 2015

1 Zombie Bob

1.1 Bayes rules

Denote Bob's likelihood of saying *graagh* as $P(g|B) = 10^{-5}$, that of Zombie Bob as $P(g|Z) = 0.5$, and accordingly prior probability of Bob being a zombie was $P(Z) = 10^{-6}$.
According to Bayes rules, the posterior is:

$$P(Z|g) = \frac{P(Z)P(g|Z)}{P(Z)P(g|Z) + P(B)P(g|B)} = \frac{10^{-6} \times 0.5}{10^{-6} \times 0.5 + 10^{-5} \times (1 - 10^{-6})} = 0.0476 \tag{1}$$

1.2 Expected utility

The posterior probability that it is Bob rather than Zombie is $P(B|g) = 1 - P(Z|g) = 0.9524$.
If bob stays, the expected utility is:

$$u(1) = u(1, Bob) \times P(B|g) + u(1, Zombie) \times P(Z|g) = -20 \times 0.0476 = -0.952 \tag{2}$$

If bob runs, the expected utility is

$$u(2) = u(2, Bob) \times P(B|g) + u(2, Zombie) \times P(Z|g) = -1 \times 0.9524 + 3 \times 0.0476 = -0.8096 \tag{3}$$

Strategy 2 has higher expected utility, so Bob should run.

1.3 The chain rule and marginal probabilities

$$P(live) = P(live|Z, g)P(Z|g) + P(live|B, g)P(B|g) = 0.5 \times 0.0476 + 0.9524 \times 1.0 = 0.9762 \tag{4}$$

2 Necromantic Scrolls

2.1 Bayes rule

To simplify the notation, denote Anna as A, Barry as B, *abracadabra* as a, *gesundheit* as g.
According to the problem statement, the prior belief that Anna is the author is $P(A) = 0.6$, $P(a|A) = 0.005$, $P(a|b) = 0.01$, so we have:

$$P(A|a) = \frac{P(a|A)P(A)}{P(a|A)P(A) + P(a|B)P(B)} = \frac{0.005 \times 0.6}{0.005 \times 0.6 + 0.01 \times 0.4} = 0.4286 \tag{5}$$

2.2 Breakevenpoint

First we calculate posterior of author being Barry:

$$P(B|a) = \frac{P(a|B)P(B)}{P(a|B)P(B) + P(a|A)P(A)} \tag{6}$$

$$\Rightarrow \frac{P(A|a)}{P(B|a)} = \frac{P(a|A)P(A)}{P(a|B)P(B)} = \frac{P(a|A)P(A)}{P(a|B)(1-P(A))} = \frac{0.005 \times P(A)}{0.01 \times (1-P(A))} \tag{7}$$

What we want is the ratio to be 1, so let

$$\frac{0.005 \times P(A)}{0.01 \times (1-P(A))} = 1 \Rightarrow P(A) = \frac{1}{3} \tag{8}$$

2.3 Multiple words

- Denote the result in problem statement as event E. In order to calculate thte posterior belief P(A—E), we need to first calculate $P(E|A)$ and $P(E|B)$. For each word w, there're three cases: 1) w=a; 2) w=g; 3) otherwise. Denote the corresponding probabilities: 1) $P(a|A) = 0.005, P(a|B) = 0.01$; $P(g|A) = 0.006$, $P(g|B) = 0.001$; $P(o|A) = 1-P(a|A)-P(g|A) = 0.989$, $P(o|B) = 1-P(a|B)-P(g|B) = 0.989$. The first case happened twice, the second once, and thus the third 97 times. Then $P(E|A)$ can be described by multinomial distribution as following:

$$P(E|A) = \frac{100!}{2! \times 1! \times 97!} \times 0.005^2 \times 0.006 \times 0.989^{97} = 0.0249 \tag{9}$$

$$P(E|B) = \frac{100!}{2! \times 1! \times 97!} \times 0.01^2 \times 0.001 \times 0.989^{97} = 0.0166 \tag{10}$$

  Then the posterior can be calculated as:

$$P(A|E) = \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|B)P(B)} = \frac{0.0249 \times 0.5}{0.0249 \times 0.5 + 0.0166 \times 0.5} = 0.6 \tag{11}$$

- From the caculation above, we can see that we only need to know the probability of a word not being *abracadabra* or *gesundheit*, which can be calculated by the frequencies of these two words.
  So we **DO NOT** need any more information about the other 97 words.

3 Sentence lengths

3.1 Maximum likelihood estimation

Log likelihood:

$$\log P(l_{1:N}|\lambda) = \prod_{n=0}^{N} \lambda^n (1-\lambda) = \sum_{n=0}^{N} (\log \lambda^n + \log(1-\lambda)) = \frac{N(N+1)}{2} \log \lambda + N \log(1-\lambda) \tag{12}$$

Let the derivative of the log likelihood w.r.t λ be zero and we will have the value of λ:

$$\frac{\partial \log p(l_{1:N}|\lambda)}{\partial \lambda} = \frac{N(N+1)}{2\lambda} - \frac{N}{1-\lambda} = 0 \Rightarrow \lambda = \frac{N+1}{N+3} \tag{13}$$

3.2 Expectations

- Expectation of sentence length:

$$E[l] = \sum_{l=0}^{\infty} l\lambda^l(1-\lambda) = (1-\lambda)\sum_{l=1}^{\infty} l\lambda^l = \frac{\lambda}{1-\lambda} \tag{14}$$

- Modal sentence length.

$$\frac{\partial P(l)}{\partial l} = l\lambda^{n-1} - (l+1)\lambda^l = 0 \Rightarrow l = \frac{\lambda}{1-\lambda} \tag{15}$$

  So the modal sentence length is either

- Skewed distribution

4 Part-of-speech tagging accuracy

4.1

The probability of tagging a word correctly is $1 - 0.1 = 0.9$, and tagging each word correctly or not is IID. So for $n = 5$, the probability is simply $P = 0.9^5 = 0.59$

4.2

Since the number of verbs is less than or equal to the number of all words in a sentence(much less most of the time actually), the probability of Gregory's tagger to make errors is of course lower than that of Felicia's tagger. So Gregory's tagger will get more sentences completely correct.