

LING572 HW3
Mami Sasaki and Nat Byington
28/Jan/2010

Q1

training accuracy = 0.9425925925925925
test accuracy = 0.8833333333333333

Q3

Table 1 (Bernoulli NB model)

cond_prob_delta	Training accuracy	Test accuracy
0.1	0.93	0.88
0.5	0.91037037037	0.863333333333
1.0	0.897037037037	0.84
2.0	0.87962962963	0.823333333333

Q4

Table 2 (multinomial NB model)

cond_prob_delta	Training accuracy	Test accuracy
0.1	0.957037037037	0.913333333333
0.5	0.95037037037	0.906666666667
1.0	0.944814814815	0.9
2.0	0.94	0.896666666667

Q5

Table 3 (multinomial NB model with binary feature)

cond_prob_delta	Training accuracy	Test accuracy
0.1	0.95962962963	0.91
0.5	0.956296296296	0.903333333333
1.0	0.952962962963	0.896666666667
2.0	0.947777777778	0.896666666667

Q6

- (a) The multinomial model seems to give better results, and having a lower cond_delta helps too.
- (b) No, we don't think so. The individual probability doesn't tell you how that term relates to other classes, plus knowing the probability of more common terms (i.e. stop words) wouldn't help much either.
- (c) Same, mainly because the individual probability doesn't tell us anything about the number of occurrences.
- (d) Both were fast, but the multinomial model took about 3 times longer to run. The algorithm is slightly more complex.