David Murray
dm966 for Sadiq Jaffer

Machine Learning Supo 1

Sentiment Lexicon

2.

No
Nothing
Nor
Neither
Hardly
Barely
But
However
Instead
Rarely

The sentiment of adjectives and adverbs can be flipped by adding a "not".
e.g. good -> not good
very -> not very

3. Yes you could use a lexicon that gives positive sentiment to words like "happy", "excited", "delighted" and negative sentiment to "sad", "depressed", "angry".

4. 0.796

5. If the probability of one class is 90% and the other is 10%, then the accuracy could be really high just by the lexicon always guessing class 1.

Extra Q.

~~Say the system has a probability P(A) of flagging a messange and P(C) of being correct.~~

~~Num of human reviews = Num messages · P(F)~~

Say the system has probability P(A) of correctly flagging a rude message, and P(B) of incorrectly flagging a non-rude message.

$$E[C] = 10 \cdot \left((N - N_r) \cdot P(B) + N_r \cdot P(A)\right)$$
$$+ 100{,}000 \cdot \left(N_r \cdot (1 - P(A))\right)$$

where N is the total number of messages and N_r is the number of rude messages

Naive Bayes

1. a) $P(A \mid F_1) = 1/2$ $\quad P(B \mid F_1) = 1/2$

$P(A \mid F_2) = 0$ $\quad P(B \mid F_2) = 1$

$P(A \mid F_3) = 1/10$ $\quad P(B \mid F_3) = 9/10$

b) ~~P(A) =~~

$$P(A \mid F_1, F_3) = P(A) \cdot P(F_1 \mid A) \cdot P(F_3 \mid A)$$
$$= \frac{1}{2} \cdot \frac{5}{8} \cdot \frac{3}{8}$$
$$= \frac{15}{128}$$

$$P(B \mid F_1, F_3) = P(B) \cdot P(F_1 \mid B) \cdot P(F_3 \mid B)$$
$$= \frac{1}{2} \cdot \frac{5}{42} \cdot \frac{27}{42}$$
$$= \frac{45}{1176}$$

so more likely to be A.

c. $P(A \mid F_1, F_3) = \frac{15}{128} \cdot \frac{1}{2} = \frac{15}{256}$

$P(B \mid F_1, F_3) = \frac{45}{1176} \cdot \frac{3}{2} = \frac{15}{262}$

so much closer but still A more likely.

d. F2 since $P(B \mid F_2) = 1$ and $P(A \mid F_2) = 0$

so F2 guarantees the class is B

e. Find the feature Fi such that $P(\text{class} \mid F_i)$ is highest for class 1 and lowest for class 2. The bigger the difference between $P(1 \mid F_i)$ and $P(2 \mid F_i)$ is the more useful the feature is. Calculate these probabilities with Bayes' Theorem.

2. Reusing the same word makes it less strong and impactful than using new words.
Only counting them once also lowers the estimate for high frequency words which increases the estimate for the lower frequency ones, giving a more even distribution.

Statistical properties of Language

1. They all have letters that do not commonly go together in English.

pferd -> pf
abtnuce -> btr
Ksc'a -> Ksc

Yes they are right these are not words.