Chapter 4: Naive Bayes and Sentiment Classification

(Q4.1) S = "I always like foreign films"

$$P(neg|S) = \frac{P(S|neg)\,P(neg)}{P(S)}$$

$$P(pos|S) = \frac{P(S|pos)\,P(pos)}{P(S)}$$

ignore common base.

- $P(neg|S) = (0.16 \times 0.06^2 \times 0.15 \times 0.11) \times 0.5$

  $= 0.000004752$

- $P(pos|S) = (0.09 \times 0.07 \times 0.29 \times 0.04 \times 0.08) \times 0.5$

  $= 0.000002923$

The naive bayes will assign "neg" class to the sentence because $P(neg|S) > P(pos|S)$

(Q4.2) $P(\text{comedy}) = 2/5$ $\quad |V| = 7$

$P(\text{action}) = 3/5$

$$P(\text{fast} \mid \text{comedy}) = \frac{\text{count}(\text{fast}, \text{comedy}) + 1}{\sum_{w \in V} (\text{count}(w, \text{comedy}) + 1)}$$

$$= \frac{2}{9+7} = \frac{2}{16}$$

$$P(\text{fast} \mid \text{action}) = \frac{3}{11+7} = \frac{3}{18}$$

$P(\text{couple} \mid \text{comedy}) = \frac{3}{16}$ $\quad P(\text{shoot} \mid \text{comedy}) = \frac{1}{16}$

$P(\text{couple} \mid \text{action}) = \frac{1}{18}$ $\quad P(\text{shoot} \mid \text{action}) = \frac{5}{18}$

$P(\text{fly} \mid \text{comedy}) = \frac{2}{16}$

$P(\text{fly} \mid \text{action}) = \frac{2}{18}$

$$P(\text{comedy} \mid D) = P(D \mid \text{comedy}) P(\text{comedy})$$

$$= \frac{2}{16} \times \frac{3}{16} \times \frac{2}{16} \times \frac{1}{16} \times \frac{2}{5}$$

$$= 0.0000732$$

$$P(\text{action} \mid D) = 3/18 \times 1/18 \times 2/18 \times 5/18 \times 3/5$$

$$= 0.000171$$

D will be classified as "action".

(Q 4.3) • Binarized naive Bayes

$P(neg) = 0.6 \qquad P(pos) = 0.4$

$P(good \mid neg) = 3/9 \qquad P(good \mid pos) = 2/7$

$P(poor \mid neg) = 4/9 \qquad P(poor \mid pos) = 2/7$

$P(great \mid neg) = 2/9 \qquad P(great \mid pos) = 3/7$

$P(neg \mid D) = \frac{3}{9} \times \frac{4}{9} \times \frac{2}{9} \times 0.6 = 0.0197$

$P(pos \mid D) = \frac{2}{7} \times \frac{2}{7} \times \frac{3}{7} \times 0.4 = 0.0139$

Classified as "neg" by BNB.

• Multinomial Naive Bayes

$P(good \mid pos) = 4/12 \qquad P(good \mid neg) = 3/17$

$P(poor \mid pos) = 2/12 \qquad P(poor \mid neg) = 11/17$

$P(great \mid pos) = 6/12 \qquad P(great \mid neg) = 3/17$

$P(pos \mid D) = \left(\frac{4}{12}\right)^2 \times \frac{2}{12} \times \frac{6}{12} \times 0.6$

$= 0.0055$

$P(neg \mid D) = \left(\frac{3}{17}\right)^2 \times \left(\frac{11}{17}\right) \times \left(\frac{3}{17}\right) \times 0.4$

$= 0.0014$

Classified as "pos" by MNB.
Both models disagree.