Solutions for Questions in Chapter 6

Thien Huu Nguyen  
Compute Science Department  
New York University  
thien@cs.nyu.edu

1 Question 6.1

1.1 Solution
As there are 2 classes (pos and neg) in this problem and the prior probabilities for each class is equal, we have:

\[ P(\text{pos}) = P(\text{neg}) = 0.5 \]

For the given sentence \( S = \text{“I always like foreign films.”} \), the class that the Naive bayes classifier would assign to \( S \) is computed by:

\[
\hat{c} = \arg\max_{c \in \{\text{pos, neg}\}} P(c) \prod_{i \in \text{positions of } S} P(w_i|c)
\]

(1)

Note that we ignore the dot “.” in the end of the sentence \( S \) as it does not appear in the vocabulary in this case.

For \( c = \text{pos} \), we have:

\[
P(\text{pos})P(I|\text{pos})P(\text{always}|\text{pos})P(\text{like}|\text{pos})P(\text{foreign}|\text{pos})P(\text{films}|\text{pos})
\]

\[= 0.5 \times 0.09 \times 0.07 \times 0.29 \times 0.04 \times 0.08 \]

\[= 2.9232 \times 10^{-6} \]

(2)

For \( c = \text{neg} \), we have:

\[
P(\text{neg})P(I|\text{neg})P(\text{always}|\text{neg})P(\text{like}|\text{neg})P(\text{foreign}|\text{neg})P(\text{films}|\text{neg})
\]

\[= 0.5 \times 0.16 \times 0.06 \times 0.15 \times 0.11 \]

\[= 4.752 \times 10^{-6} \]

(3)

From equations (1), (2) and (3), we conclude that the Naive bayes classifier would assign the “neg” class to the given sentence.

2 Question 6.2

2.1 Solution
There are two classes for classification in this problem, i.e., “comedy” and “action” for movie reviews. In order to compute the most likely class for the new document \( D \) using the naive Bayes classifier, we
first build the vocabulary from the training documents and then compute likelihood for each word in the vocabulary (using the add-1 smoothing).

By taking the union of all the words occurring in the training documents, our vocabulary includes the 7 following words: 
*fun, couple, love, fast, furious, shoot, fly*

From the training documents, the frequencies of the vocabulary words with respect to different classes (\(\text{count}(\text{word}, \text{class})\)) are shown in the following table:

<table>
<thead>
<tr>
<th></th>
<th>count(*, comedy)</th>
<th>count(*, action)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fun</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>couple</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>love</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>fast</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>furious</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>shoot</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>fly</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>11</td>
</tr>
</tbody>
</table>

**Table 1: Word Frequencies with Respect to Different Classes**

Using the add-1 smoothing, the likelihoods of the words with respect to the classes are shown below:

\[
\hat{P}(\ast | \text{comedy}) \quad \hat{P}(\ast | \text{action})
\]

|        | \(\hat{P}(\ast | \text{comedy})\) | \(\hat{P}(\ast | \text{action})\) |
|--------|-------------------------------|-------------------------------|
| fun    | \(\frac{3+1}{9+7} = \frac{4}{16}\) | \(\frac{1+1}{11+7} = \frac{2}{18}\) |
| couple | \(\frac{2+1}{9+7} = \frac{3}{16}\) | \(\frac{0+1}{11+7} = \frac{1}{18}\) |
| love   | \(\frac{2+1}{9+7} = \frac{3}{16}\) | \(\frac{1+1}{11+7} = \frac{2}{18}\) |
| fast   | \(\frac{1+1}{9+7} = \frac{2}{16}\) | \(\frac{2+1}{11+7} = \frac{3}{18}\) |
| furious| \(\frac{0+1}{9+7} = \frac{1}{16}\) | \(\frac{2+1}{11+7} = \frac{3}{18}\) |
| shoot  | \(\frac{0+1}{9+7} = \frac{1}{16}\) | \(\frac{4+1}{11+7} = \frac{5}{18}\) |
| fly    | \(\frac{1+1}{9+7} = \frac{2}{16}\) | \(\frac{1+1}{11+7} = \frac{2}{18}\) |

**Table 2: Word Likelihoods with Respect to Different Classes**

Regarding the priors, there are 3 examples of “\text{action}” and 2 examples of “\text{comedy}” in the training data, so:

\(P(\text{action}) = \frac{3}{5}\) and \(P(\text{comedy}) = \frac{2}{5}\)

The most likely class for \(D = \text{“fast, couple, shoot, fly”}\) using the naive Bayes would be:

\[
\hat{c} = \arg\max_{c \in \{\text{action, comedy}\}} P(c) \prod_{i \in \text{positions of } D} P(w_i | c) \\
= \arg\max_{c \in \{\text{action, comedy}\}} P(c) P(\text{fast}|c)P(\text{couple}|c)P(\text{shoot}|c)P(\text{fly}|c)
\]

Now, for \(c = \text{action},\)
\begin{equation}
P(\text{action}) P(\text{fast}|\text{action}) P(\text{couple}|\text{action}) P(\text{shoot}|\text{action}) P(\text{fly}|\text{action})
\end{equation}
\begin{equation*}
= \frac{3}{5} \times \frac{3}{18} \times \frac{1}{18} \times \frac{5}{18} \times \frac{2}{18}
\approx 1.715 \times 10^{-4}
\end{equation*}

For \( c = \text{comedy} \), we have:

\begin{equation}
P(\text{comedy}) P(\text{fast}|\text{comedy}) P(\text{couple}|\text{comedy}) P(\text{shoot}|\text{comedy}) P(\text{fly}|\text{comedy})
\end{equation}
\begin{equation*}
= \frac{2}{5} \times \frac{2}{16} \times \frac{3}{16} \times \frac{1}{16} \times \frac{2}{16}
\approx 0.7324 \times 10^{-4}
\end{equation*}

From equations (4), (5) and (6), we conclude that the most likely class for \( D \) using the naive Bayes classifier is “\text{action}.”