Exercise 1 Consider again the training data from slide 9: We have classes A and B and a training set of class-labeled documents:

<table>
<thead>
<tr>
<th>Training data:</th>
<th>d</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa A</td>
<td>d</td>
<td>c</td>
</tr>
<tr>
<td>ab A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ba A</td>
<td>d</td>
<td>c</td>
</tr>
<tr>
<td>bb B</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Calculate $P(A)$, $P(B)$, $P(a|A)$, $P(b|A)$, $P(a|B)$, $P(b|B)$ using Laplace smoothing for the conditional probabilities.

2. Now classify the following new data, deleting all unknown words:

<table>
<thead>
<tr>
<th>Documents:</th>
</tr>
</thead>
<tbody>
<tr>
<td>aaba</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>bbba</td>
</tr>
<tr>
<td>bccba</td>
</tr>
<tr>
<td>bbbb</td>
</tr>
</tbody>
</table>

Solution:

1. $P(A) = 0.75$, $P(B) = 0.25$.
   - $P(a|A) = \frac{3}{5}$, $P(b|A) = \frac{3}{5}$
   - $P(a|B) = \frac{3}{4}$, $P(b|B) = \frac{3}{4}$

2. 
   
   **aaba**
   
   A: $\frac{3}{4} \cdot \frac{5}{8} \cdot \frac{5}{8} \cdot \frac{3}{8} \cdot \frac{5}{8} = \frac{1125}{16384} = 0.07$
   
   B: $\frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{3}{1024} = 0.003$
   
   ⇒ class A is assigned to aaba

   **a**
   
   A: $\frac{3}{4} \cdot \frac{5}{8} = \frac{15}{32} = 0.47$
   
   B: $\frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16} = 0.06$
   
   ⇒ class A is assigned to aaba

   **bbba**
   
   A: $\frac{3}{4} \cdot \frac{3}{8} \cdot \frac{3}{8} \cdot \frac{3}{8} \cdot \frac{5}{8} = \frac{405}{16384} = 0.025$
   
   B: $\frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{27}{1024} = 0.026$
   
   ⇒ class B is assigned to bbba

   **bccba** same class as bbba

   **bbbb**
   
   A: $\frac{3}{4} \cdot \frac{3}{8} \cdot \frac{3}{8} \cdot \frac{3}{8} \cdot \frac{3}{8} = \frac{243}{16384} = 0.015$
   
   B: $\frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{81}{1024} = 0.08$
   
   ⇒ class B is assigned to bbbb

Exercise 2 Consider again the training data from the previous exercise and assume that we test on the following data:
<table>
<thead>
<tr>
<th>Documents</th>
<th>gold class</th>
<th>system class</th>
</tr>
</thead>
<tbody>
<tr>
<td>aaba</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>a</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>bbba</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>bceeba</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>bbbb</td>
<td>B</td>
<td>B</td>
</tr>
</tbody>
</table>

Compute precision, recall, accuracy and $F_1$ for the classification resulting from the training data in the previous exercise, for both classes $A$ and $B$.

Solution:

<table>
<thead>
<tr>
<th>Documents</th>
<th>gold class</th>
<th>system class</th>
</tr>
</thead>
<tbody>
<tr>
<td>aaba</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>a</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>bbba</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>bceeba</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>bbbb</td>
<td>B</td>
<td>B</td>
</tr>
</tbody>
</table>

Evaluation for $A$: $P = 1$, $R = \frac{1}{2}$, $F_1 = \frac{2}{3}$

Evaluation for $B$: $P = \frac{1}{3}$, $R = 1$, $F_1 = \frac{1}{2}$

Accuracy is in both cases $\frac{3}{5}$

**Exercise 3** Consider again the same example. Give the pooled confusion matrix for the results on the test set. Give the overall precision and recall by

1. macroaveraging, and
2. microaveraging.

Note that we have a special case here since the true positives of $A$ are the true negatives of $B$ and vice versa. Therefore we get a special confusion matrix.

Solution:

Pooled confusion matrix:

<table>
<thead>
<tr>
<th>system yes</th>
<th>gold yes</th>
<th>gold no</th>
</tr>
</thead>
<tbody>
<tr>
<td>system yes</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>system no</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

1. macroaveraging:

$P = \frac{1}{2}(1 + \frac{1}{3}) = \frac{2}{3}$

$R = \frac{1}{2}(\frac{1}{2} + 1) = \frac{3}{4}$

2. microaveraging:

$P = \frac{3}{5}$

$R = \frac{5}{6}$

($P$ and $R$ are the same in this special case where we have two classes excluding each other such that $\neg A = B$.)