Assignment 2A

Nolan McCafferty and Daniel Rosenbaum

February 7, 2019

1. Unigram probabilities:
   $P(a) = \frac{6}{12} = \frac{1}{2}$
   $P(b) = \frac{5}{12}$
   $P(c) = \frac{1}{12}$

2. Bigram probabilities:
   $P(a \mid a) = \frac{1}{5}$
   $P(b \mid a) = \frac{3}{5}$
   $P(c \mid a) = \frac{1}{5}$
   $P(a \mid b) = \frac{2}{3}$
   $P(b \mid b) = \frac{1}{3}$
   $P(c \mid b) = 0$
   $P(a \mid c) = 1$
   $P(b \mid c) = 0$
   $P(c \mid c) = 0$

3. Add-1 Smoothing:
   $P(a \mid a) = \frac{2}{8} = \frac{1}{4}$
   $P(b \mid a) = \frac{4}{8} = \frac{1}{2}$
   $P(c \mid a) = \frac{2}{8} = \frac{1}{4}$
   $P(a \mid b) = \frac{3}{6} = \frac{1}{2}$
   $P(b \mid b) = \frac{2}{6} = \frac{1}{3}$
   $P(c \mid b) = \frac{1}{6}$
   $P(a \mid c) = \frac{2}{4} = \frac{1}{2}$
   $P(b \mid c) = \frac{1}{4}$
   $P(c \mid c) = \frac{1}{4}$

4.

```
<UNK> a <UNK> b
a <UNK> a b
b a b a
```

5. New Corpus Analysis

   Unigram probabilites:

   $P(a) = \frac{5}{12}$

   $P(b) = \frac{4}{12} = \frac{1}{3}$

   $P(c) = \frac{3}{12} = \frac{1}{4}$

   Bigram probabilities:

   $P(a \mid a) = 0$

   $P(b \mid a) = \frac{2}{4} = \frac{1}{2}$

   $P(\texttt{<UNK>} \mid a) = \frac{2}{4} = \frac{1}{2}$

   $P(a \mid b) = \frac{2}{2} = 1$

   $P(b \mid b) = 0$

   $P(\texttt{<UNK>} \mid b) = 0$

   $P(a \mid \texttt{<UNK>}) = \frac{2}{3}$

   $P(b \mid \texttt{<UNK>}) = \frac{1}{3}$

   $P(c \mid \texttt{<UNK>}) = 0$

6. Absolute Discounting

   Calculating reserved mass:

   $\text{reserved\_mass}(a) = \frac{0.5*2}{4} = \frac{1}{4}$

   $\text{reserved\_mass}(b) = \frac{0.5*1}{2} = \frac{1}{4}$

   $\text{reserved\_mass}(\texttt{<UNK>}) = \frac{0.5*2}{3} = \frac{1}{3}$

   Calculating α:

   $\alpha(a) = \frac{reserved\_mass(a)}{1-(p(\texttt{<UNK>})+p(b))} = \frac{\frac{1}{4}}{1-\frac{3}{12}-\frac{4}{12}} = \frac{3}{5}$

   $\alpha(b) = \frac{reserved\_mass(b)}{1-p(a)} = \frac{\frac{1}{4}}{1-\frac{5}{12}} = \frac{3}{7}$

   $\alpha(\texttt{<UNK>}) = \frac{reserved\_mass(\texttt{<UNK>})}{1-(p(a)+p(b)} = \frac{\frac{1}{3}}{1-\frac{5}{12}-\frac{4}{12}} = \frac{4}{3}$

   Bigram probabilities:

   $p_{absolute}(a|a) = \alpha(a) * p(a) = \frac{3}{5} * \frac{5}{12} = \frac{1}{4}$

   $p_{absolute}(b|a) = \frac{C(ab)-D}{C(a)} = \frac{1.5}{4} = \frac{3}{8}$

   $p_{absolute}(\texttt{<UNK>}|a) = \frac{C(\texttt{a<UNK>})-D}{C(a)} = \frac{1.5}{4} = \frac{3}{8}$

   $p_{absolute}(a|b) = \frac{1.5}{2} = \frac{3}{4}$

$$p_{absolute}(b|b) = \frac{3}{7} * \frac{4}{12} = \frac{1}{7}$$
$$p_{absolute}(\texttt{<UNK>}|b) = \frac{3}{7} * \frac{3}{12} = \frac{3}{28}$$

$$p_{absolute}(a|\texttt{<UNK>}) = \frac{1.5}{3} = \frac{1}{2}$$
$$p_{absolute}(b|\texttt{<UNK>}) = \frac{0.5}{3} = \frac{1}{6}$$
$$p_{absolute}(\texttt{<UNK>}|\texttt{<UNK>}) = \frac{4}{3} * \frac{1}{4} = \frac{1}{3}$$