University of Lethbridge
Department of Mathematics & Computer Science

CPSC 4310/5310/7310 – Spring 2020

Natural Language Processing (NLP)

Sample Solutions to Assignment 2 [60 points]

Due on February 7th, 2020

The problems are adopted from the textbook.

1. **[5 points]**

$$P(w_1, \ldots, w_n) = \Pi_{i=1}^{n} P(w_i | w_{i-2}, w_{i-1})$$

$$P(w_i | w_{i-1}, w_{i-2}) = \frac{C(w_{i-2}\ w_{i-1}\ w_i)}{C(w_{i-2}\ w_{i-1})}$$

2. **[10 points]**

$$\begin{aligned} P(c|d) &= argmax_{c \in C} P(d|c)P(c) \\ &= argmax_{c \in C} P(c)\Pi_{i=1}^{n} P(x_i|c) \end{aligned}$$

$$C = \{h, b\}$$

$$P(h) = \frac{4}{6}$$

$$P(b) = \frac{2}{6}$$

$$P(Flames|h) = \frac{4+1}{12+11}$$

$$P(Penguins|h) = \frac{1+1}{12+11}$$

$$P(Raptors|h) = \frac{1+1}{12+11}$$

$$P(Oilers|h) = \frac{2+1}{12+11}$$

$$P(Flames|b) = \frac{0+1}{6+11}$$

$$P(Penguins|b) = \frac{0+1}{6+11}$$

$$P(Raptors|b) = \frac{2+1}{6+11}$$

$$P(Oilers|b) = \frac{0+1}{6+11}$$

$$P(h|d_7) = \frac{4}{6} * \frac{5}{23} * \frac{2}{23} * \frac{3}{23} = 0.0016$$

$$P(b|d_7) = \frac{2}{6} * \frac{1}{17} * \frac{1}{17} * \frac{1}{17} = 0.00006$$

$$P(h|d_8) = \frac{4}{6} * \frac{2}{23} * \frac{3}{23} = 0.0075$$

$$P(b|d_8) = \frac{2}{6} * \frac{3}{17} * \frac{1}{17} = 0.0034$$

$$P(h|d_9) = \frac{4}{6} * \frac{5}{23} * \frac{2}{23} * \frac{2}{23} * \frac{3}{23} = 0.00014$$

$$P(b|d_9) = \frac{2}{6} * \frac{1}{17} * \frac{1}{17} * \frac{3}{17} * \frac{1}{17} = 0.000012$$