CS4248Assignment 1

U096883L Shawn Tan

1.

$$p(k) = \frac{\binom{n}{k}}{\binom{N}{n}} = \frac{n!}{k!(n-k)!}\frac{N!(N-n)!}{N!}$$

2.

$$P_{WB}(w \mid c_i) = \begin{cases} \dfrac{C(c_i, w)}{C(c_i) + T(c_i)} & \text{if } C(c_i, w) > 0 \\ \dfrac{T(c_i)}{Z(c_i)\,(C(c_i) + T(c_i))} & \text{if } C(c_i, w) = 0 \end{cases}$$

$$\begin{aligned} C(c_1) &= 7 \\ T(c_1) &= 6 \\ Z(c_1) &= 4 \\ P_{WB}(w \mid c_1) &= \frac{6}{4 \cdot (7+6)} \\ &\approx 0.1154 \end{aligned} \qquad \begin{aligned} C(c_2) &= 6 \\ T(c_2) &= 5 \\ Z(c_2) &= 5 \\ P_{WB}(w \mid c_2) &= \frac{5}{5 \cdot (6+5)} \\ &\approx 0.0909 \end{aligned}$$

	$P_{WB}(w\|c_1)$		$P_{WB}(w\|c_2)$	
Total	7		6	
	$C(c_1, w)$	$P_{WB}(w \mid c_1)$	$C(c_2, w)$	$P_{WB}(w \mid c_2)$
body	1	1/13	0	0.0909
fun	0	0.1154	1	1/11
is	0	0.1154	1	1/11
jogging	0	0.1154	2	2/11
John	1	1/13	0	0.0909
loves	1	1/13	1	1/11
Mary	0	0.1154	1	1/11
our	1	1/13	0	0.0909
strengthens	1	1/13	0	0.0909
swimming	2	2/13	0	0.0909

3. Table for edit distance.

p	5	4	3	4	3	4
a	4	3	2	3	4	5
e	3	2	1	2	3	4
h	2	1	2	3	4	5
c	1	2	3	4	5	6
	0	1	2	3	4	5
		h	e	l	p	s

4.

$$
\begin{aligned}
H(X,Y) &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log\, p(x,y) \\
&= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\,(\log\, p(y\,|\,x)\,p(x)) \\
&= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\,(\log\, p(y\,|\,x) + \log\, p(x)) \\
&= -\left(\sum_{x\in X}\sum_{y\in Y} p(x,y)\log\, p(y\,|\,x) + \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\, p(x)\right) \\
&= -\left(\sum_{x\in X}\sum_{y\in Y} p(x,y)\log\, p(y\,|\,x) + \sum_{x\in X}\log\, p(x)\sum_{y\in Y} p(x,y)\right)
\end{aligned}
$$

Marginalising out y,

$$
\begin{aligned}
&= -\left(\sum_{x\in X}\sum_{y\in Y} p(x,y)\log\, p(y\,|\,x) + \sum_{x\in X} p(x)\log\, p(x)\right) \\
&= \left(-\sum_{x\in X}\sum_{y\in Y} p(x,y)\log\, p(y\,|\,x)\right) + \left(-\sum_{x\in X} p(x)\log\, p(x)\right) \\
&= H(Y\,|\,X) + H(X)
\end{aligned}
$$