CS4248
AY 2022/23 Semester 1
Tutorial 2 Solutions

1.

(a) $P(c_1) = \frac{2}{4} = \frac{1}{2}$ $P(c_2) = \frac{2}{4} = \frac{1}{2}$

(b)

$$P_{MLE}(w|c_i) = \frac{C(w)}{N}$$

$C(w)$: count of w in class c_i
N: count of all words in class c_i

(c)

$|V| = 11$

$$P_{AO}(w|c_i) = \frac{C(w)+1}{N+|V|}$$

w	$P(w\|c_1)$		$P(w\|c_2)$	
	MLE	add-one	MLE	add-one
brazil	0	$1/23$	$1/9$	$2/20$
champion	$1/12$	$2/23$	$1/9$	$2/20$
in	$2/12$	$3/23$	$1/9$	$2/20$
is	$2/12$	$3/23$	$2/9$	$3/20$
popular	0	$1/23$	$1/9$	$2/20$
soccer	0	$1/23$	$2/9$	$3/20$
tennis	$2/12$	$3/23$	0	$1/20$
the	$2/12$	$3/23$	$1/9$	$2/20$
ultimate	$1/12$	$2/23$	0	$1/20$
usa	$1/12$	$2/23$	0	$1/20$
wimbledon	$1/12$	$2/23$	0	$1/20$

(d)

$$\hat{c} = \underset{c}{\operatorname{argmax}} P(c|w_1, \dots, w_n)$$

$$= \underset{c}{\operatorname{argmax}} \frac{P(w_1, \dots, w_n|c)P(c)}{P(w_1, \dots, w_n)}$$

$$= \underset{c}{\operatorname{argmax}} P(w_1, \dots, w_n|c)P(c)$$

$$= \underset{c}{\operatorname{argmax}} P(w_1|c) \cdot \dots \cdot P(w_n|c) \cdot P(c)$$

(e)

For the text "germany is the champion in soccer":

$$P(\text{is}|c_1) \cdot P(\text{the}|c_1) \cdot P(\text{champion}|c_1) \cdot P(\text{in}|c_1) \cdot P(\text{soccer}|c_1) \cdot P(c_1)$$

$$= \frac{3}{23} \cdot \frac{3}{23} \cdot \frac{2}{23} \cdot \frac{3}{23} \cdot \frac{1}{23} \cdot \frac{1}{2} = 4.2 \times 10^{-6}$$

$$P(\text{is}|c_2) \cdot P(\text{the}|c_2) \cdot P(\text{champion}|c_2) \cdot P(\text{in}|c_2) \cdot P(\text{soccer}|c_2) \cdot P(c_2)$$

$$= \frac{3}{20} \cdot \frac{2}{20} \cdot \frac{2}{20} \cdot \frac{2}{20} \cdot \frac{3}{20} \cdot \frac{1}{2} = 1.1 \times 10^{-5}$$

Hence, this text is assigned to class c_2.

For the text "wimbledon is played in the uk":

$$P(\text{wimbledon}|c_1) \cdot P(\text{is}|c_1) \cdot P(\text{in}|c_1) \cdot P(\text{the}|c_1) \cdot P(c_1)$$

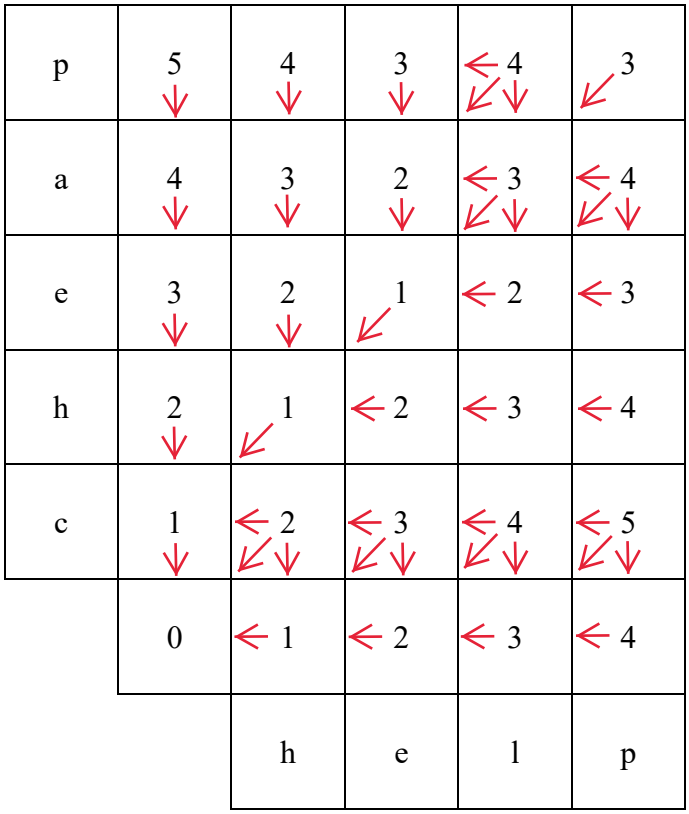
$$= \frac{2}{23} \cdot \frac{3}{23} \cdot \frac{3}{23} \cdot \frac{3}{23} \cdot \frac{1}{2} = 9.6 \times 10^{-5}$$

$$P(\text{wimbledon}|c_2) \cdot P(\text{is}|c_2) \cdot P(\text{in}|c_2) \cdot P(\text{the}|c_2) \cdot P(c_2)$$

$$= \frac{1}{20} \cdot \frac{3}{20} \cdot \frac{2}{20} \cdot \frac{2}{20} \cdot \frac{1}{2} = 3.8 \times 10^{-5}$$

Hence, this text is assigned to class c_1.

2.



3.

Application	Input to noisy channel	Output of noisy channel	$p(i)$	$p(o\|i)$
Spelling error correction	Correct (error-free) word	Misspelled word	Prob. of correct word	Prob. of misspelled word given correct word
Speech recognition	Word sequence	Speech signal	Prob. of word sequence (language model)	Prob. of speech signal given word sequence (acoustic model)