Question 1.

(i) $P(B_1 = 1) = 1/3$

(ii) $P(B_2 = 0|B_1 = 1) = 1$

(iii) $P(B_1 = 1|B_2 = 0) = \dfrac{P(B_1 = 1)}{P(B_2 = 0)} \cdot P(B_2 = 0|B_1 = 1) = \dfrac{1/3}{1} \cdot 1 = 1/3$

Note $P(B_2 = 0) = 1$ because the host always chooses the one that does not contain the prize.

(iv) We should change to B_3 because it has a higher probability of being the prize:

$$P(B_1 = 0|B_2 = 0) = 1 - P(B_1 = 1|B_2 = 0) = 1 - \frac{1}{3} = \frac{2}{3}.$$

Question 2.

(i) The maximum likelihood estimate of $P(y = k)$ is N_k/N_{doc}.

(ii) The maximum likelihood estimate of $P(w_i|y = k)$ is

$$\frac{\text{count}(w_i, k)}{\sum_{j=1}^{K} \text{count}(w_j, k)}.$$

(iii) We can perform Laplace smoothing, that estimates $P(w_i|y = k)$ with

$$\frac{\text{count}(w_i, k) + 1}{\sum_{j=1}^{K} \text{count}(w_j, k) + K}.$$

(iv) Infrequent words are likely not useful for classification since they are not as important and may likely act as noise instead, therefore those words can be neglected. Frequent words also do not contribute much to the meaning of text, therefore they also can be neglected.

Question 3.

(i) $P(y = +) = 5/8$; $P(y = -) = 3/8$. (iii) $P(y = +) = 0.6$; $P(y = -) = 0.4$.

(ii)

Vocabulary	+	−
annoying	0.0	0.125
awesome	0.08333	0.0
best	0.08333	0.0
easy	0.08333	0.0
good	0.08333	0.0
great	0.08333	0.0
is	0.08333	0.125
one	0.08333	0.0
rubbish	0.0	0.125
so	0.0	0.125
terrible	0.0	0.125
the	0.08333	0.0
this	0.08333	0.125
to	0.08333	0.0
use	0.08333	0.0
version	0.08333	0.125
very	0.0	0.125

Vocabulary	+	−
annoying	0.034483	0.08
awesome	0.068966	0.04
best	0.068966	0.04
easy	0.068966	0.04
good	0.068966	0.04
great	0.068966	0.04
is	0.068966	0.08
one	0.068966	0.04
rubbish	0.034483	0.08
so	0.034483	0.08
terrible	0.034483	0.08
the	0.068966	0.04
this	0.068966	0.08
to	0.068966	0.04
use	0.068966	0.04
version	0.068966	0.08
very	0.034483	0.08

(iv) $P(y = +|d) = 0.65776549$; $P(y = -|d) = 0.34223451$.
Hence, we conclude the text d is positive.