Task 1:

(1)

> $\{W_1 = a\} = \{aa\oplus, ab\oplus, ab\ominus\}$
Emails die als ersten Buchstaben „a" enthalten

> $\{W_2 = a\} = \{aa\oplus, ba\oplus\}$
Emails die als zweiten Buchstaben „a" enthalten

> $\{W_1 = b\} \cap \{W_2 = a\} = \{ba\oplus, bb\ominus\} \cap \{aa\oplus, ba\oplus\} = \{ba\oplus$
Emails die mit „ba" anfangen

> $\{W_2 \in \{a, b\}\} = \{aa\oplus, ba\oplus, ab\oplus, ab\ominus, bb\ominus\}$
Emails deren zweiter Buchstabe a oder b ist

> $\{Class = \ominus\} = \{ab\ominus, bb\ominus\}$
Emails die als „Spam" klassifiziert sind

> $\{W_2 = b\} \cap \{Class = \ominus\} = \{ab\ominus, bb\ominus\}$
Spam-Mails deren zweiter Buchstabe „b" ist.

(2)

> $P(W_1 = a, W_2 = b) = P(\{ab\oplus, ab\ominus\}) = \frac{3}{8}$
Relative Häufigkeit für Emails mit „ab"

> $P(Class = \ominus) = P(\{ab\ominus, bb\ominus\}) = \frac{3}{8}$
RH für Spam-Mails

> $P(Class = \oplus) = P(\{aa\oplus, ab\oplus, ba\oplus\}) = \frac{5}{8}$
RH für Ham-Mails

> $P(W_1 = a, W_2 = b \mid Class = \ominus) = \frac{P[(W_1 = a, W_2 = b) \cap Class = \ominus]}{P(Class = \ominus)}$

$$= \frac{1}{8} \cdot \frac{8}{3} = \frac{1}{3}$$

WSK das die Emails Spam sind, wenn sie als Spam klassifiziert wurden

> $P(W_1 = a, W_2 = b \mid Class = \oplus) = \frac{2}{8} \cdot \frac{8}{5} = \frac{2}{5}$

WSK das die Emails kein Spam sind, wenn sie als Ham klassifiziert wurden

> $P(Class = \ominus \mid W_1 = a, W_2 = b) = \frac{P[Class = \ominus \cap (W_1 = a, W_2 = b)]}{P(W_1 = a, W_2 = b)}$

$$= \frac{1}{8} \cdot \frac{8}{3} = \frac{1}{3}$$

WSK mit der Emails „ab" als Spam klassifiziert werden.

(3)

Da wir mit $P(A|B)$ im Wahrscheinlichkeitsraum von B sind, wo gilt $P(B|B) = 1$ und $P(B)$ die Anzahl aller möglichen Ereignisse beschreibt muss auch $P(A|B)P(B) + P(A|B^c)P(B^c) = P(B)$ gelten. Der Zähler beschreibt die Ereignisse von A und B -> $P(A \cap B) = P(A|B)P(B)$

Sei $A = \{W_1 = a, W_2 = b\}$, $B = \{Class = \Theta\}$. Dann ist:

$$\frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} = \frac{\frac{1}{8}}{\frac{1}{8} + \frac{2}{8}} = \frac{1}{8} \cdot \frac{8}{3} = \frac{1}{3}$$

(4)

$D = \{ab\oplus, ab\ominus\}$, $h^+ = \{ab\oplus\}$, $h^- = \{ab\ominus\}$

Dann ist:

$$P(h^-|D) = \frac{1}{8} \cdot \frac{8}{3} = \frac{1}{3}$$

und:

$$\frac{P(D|h^-)P(h^-)}{P(D|h^-)P(h^-) + P(D|h^+)P(h^+)} = \frac{\frac{1}{8}}{\frac{1}{8} + \frac{2}{8}} = \frac{1}{8} \cdot \frac{8}{3} = \frac{1}{3}$$

(5)

Es ist:

$$P(h|D) = \frac{2}{8} \cdot \frac{8}{3} = \frac{2}{3}$$

Und:

$$\frac{P(D|h)P(h)}{P(D)} = \frac{\frac{2}{8} \cdot \frac{8}{2} \cdot \frac{2}{8}}{\frac{3}{8}} = \frac{2}{8} \cdot \frac{8}{3} = \frac{2}{3}$$

Wenn wir z.B nur $P(h|D)$ wissen, also wie gut D als Spam klassifiziert wird. Aber daran interessiert sind, dass als Spam-Klassifizierte Mails auch Spam sind brauchen wir $P(D|h)$, was wir mit dieser Gleichung herrausfinden.