Q2

(a):

x_i: Number of times word V_i appears in x

$n = 15$ total Vocabulary in V

x is of length 15. This is the input data

$y = 0$ when input data is NOT spam

$y = 1$ when input data is spam

$P(y=0) = P(\text{NOT Spam}) = \frac{4}{7}$ → Not spam / → total data

these are the prior

$P(y=1) = P(\text{Spam}) = \frac{3}{7}$

(b):

million dollar offer [0,1,0,0,0,0,0,1,1,0,0,0,0,0,0]

secret offer today [1,1,0,0,0,0,1,0,0,0,0,0,0,0,0]

secret is secret [2,0,0,0,0,0,0,0,0,0,1,0,0,0,0]

low price for valued customer [0,0,1,1,1,1,0,0,0,0,0,1,0,0,0]

Play secret sports today [1,0,0,0,0,0,1,0,0,1,0,0,1,0,0]

sports is healthy [0,0,0,0,0,0,0,0,0,1,1,0,0,1,0]

low price pizza [0,0,1,1,0,0,0,0,0,0,0,0,0,0,1]