1 Semi-Supervised Learning:

1. $P(Label_5|Email_5)$:

   Given the Data set:

   Priors:

   P(Label=Spam) = 0.5, P(Label=Not Spam) = 0.5.

   Training :

   $$P(X_i = \text{word}) = \frac{\#\text{Observed Given Label} + \#\text{Hallucinated words} - 1}{\#\text{Total Observed words} + \#\text{Hallucianted words} + 11}$$

   The words present in this case are: '*linux*' and '*credit*'

   For example:

   $$P('linux'|Label = true) = \frac{0+1}{8+11} = \frac{1}{19} \tag{1}$$

   The following Table lists the probabilities:

   Table 1: Probability for Email 5

	Label=true	Label=false
'linux'	1/19	2/17
'credit'	1/19	2/17

   $$\begin{aligned}
   P(Label_5 = T|email_5) &= \frac{P(Label_5 = T) * P(email_5|Label_5)}{P(email_5)} \\
   &= \frac{P(Label_5 = T) * P(email_5|Label_5)}{P(email, Label_5 = T) + P(email, Label_5 = F)} \\
   &= \frac{P(Label_5 = T) * P(email_5|Label_5)}{P(Label_5 = T) * P(email_5|Label_5 = T) + P(Label_5 = F) * P(email_5|Label_5 = F)} \\
   &= \frac{0.5 * \frac{1}{19} * \frac{1}{19}}{0.5 * \frac{1}{19} * \frac{1}{19} + 0.5 * \frac{2}{17} * \frac{2}{17}} = 0.1668
   \end{aligned}$$

   $P(Label_5 = F|email_5) = 1 - P(Label_5 = T|email_5) = 0.8332$

   Hence for this case the label would be F.

2. $P(Label_6|Email_6)$:

   Words in this email are : *'reply'*,*'to'* and *'sale'*.

   The probability table for this case would be as follows:

   Table 2: Probability for Email 5

	Label=true	Label=false
'reply'	3/19	2/17
'to'	3/19	1/17
'sale'	2/19	1/17

$$
\begin{aligned}
P(Label_6 = T|email_6) &= \frac{P(Label_6 = T) * P(email_6|Label_6)}{P(email_6)} \\
&= \frac{P(Label_6 = T) * P(email_6|Label_6)}{P(email, Label_6 = T) + P(email, Label_6 = F)} \\
&= \frac{P(Label_6 = T) * P(email_6|Label_6)}{P(Label_6 = T) * P(email_6|Label_6 = T) + P(Label_6 = F) * P(email_5|Label_6 = F)} \\
&= \frac{0.5 * \frac{3}{19} * \frac{3}{19} * \frac{2}{19}}{0.5 * \frac{3}{19} * \frac{3}{19} * \frac{2}{19} + 0.5 * \frac{2}{17} * \frac{1}{17} * \frac{1}{17}} = 0.8657 \\
P(Label_6 = F|email_6) &= 1 - P(Label_6 = T|email_6) = 0.1343
\end{aligned}
$$

Hence the label for this email would be True.

3. M-step to retrain classifier, and classify *'to'* and *'credit'*. Recalculating the probabilities as follows:

$$
\begin{aligned}
P('to'|Label = true) &= \frac{2 + 0.8657 + 1}{8 + 3 * 0.8657 + 11 + 2 * (1 - 0.8332)} = 0.176 \\
P('to'|Label = false) &= \frac{0 + 1}{6 + 11 + 2 * 0.8332 + 3 * (1 - 0.8657)} = 0.052 \\
P('credit'|Label = true) &= \frac{1}{8 + 3 * 0.8657 + 11 + 2 * (1 - 0.8332)} = 0.046 \\
P('credit'|Label = false) &= \frac{2 + 0.8657 + 1}{6 + 11 + 2 * 0.8332 + 3 * (1 - 0.8657)} = 0.149
\end{aligned}
$$

Priors:

$$
\begin{aligned}
P(label = T) &= \frac{2 + 0.8657 + 1 - 0.8332}{6} = 0.505 \\
P(label = F) &= 1 - P(label = F) = 0.495
\end{aligned}
$$

Calculating the probabilities, in a similar way as above given the probabilities and priors:

$$
\begin{aligned}
P(Label = true|email) &= \frac{0.505 * 0.176 * 0.0046}{0.505 * 0.176 * 0.0046 + 0.495 * 0.0052 * 0.149} = 0.5131 \\
P(Label = false|email) &= 1 - P(Label = true|email) = 0.486
\end{aligned}
$$

Hence the email is classified as True.