1a)

objective function: $(e^{\beta_t} - e^{-\beta_t}) \sum_n w_t(n) \mathbb{I}[y_n \neq h_t(x_n)] + e^{-\beta_t} \sum_n w_t(n)$

$$\frac{\partial(\text{objective function})}{\partial \beta_t} = \left(e^{\beta_t} + e^{-\beta_t}\right) \sum_n w_t(n) \mathbb{I}[y_n \neq h_t(x_n)] - e^{-\beta_t} = 0$$

$$\left(e^{\beta_t} + e^{-\beta_t}\right)\left(e^{\beta_t}\right)(\varepsilon_t) - \left(e^{-\beta_t}\right)\left(e^{\beta_t}\right) = 0 \qquad \left(\sum_n w_t(n) \mathbb{I}(y_n \neq h_t(x_n)) = \varepsilon_t\right)$$

$$\left(e^{2\beta_t} + 1\right)(\varepsilon_t) - 1 = 0$$

$$e^{2\beta_t} = \frac{1}{\varepsilon_t} - 1 \longrightarrow \beta_t = \frac{\ln\left(\frac{1}{\varepsilon_t} - 1\right)}{2}$$

b) If the training data is linearly separable,

argmin $\varepsilon_t = 0$ at $t = 1$

$\therefore \beta_1 \to \infty$

This is because we are using a strong classifier instead of a weak one.

2)a) 3 clusters 4 data points

optimal clustering:

① : x_1, x_2 $M_1 = 1.5$

② : x_3 $N_2 = 5$

③ x_4 $N_3 = 7$

objective $= (0.5)^2 + (0.5)^2 + 0^2 + 0^2$

$= 0.5$

b) ① Suppose $N_1 = 1$, $N_2 = 2$, $N_3 = 6$

② First we will minimize over r_{nk}, so

① : x_1

② : x_2

③ : x_3, x_4

③ Then we will minimize over N_k, the initialized N values are already optimal

$\therefore$ final objective: 2 even though it is suboptimal (> 0.5)

$$\sum_k \left\{ \sum_n \gamma_{nk} \log N(x_n | N_k, \Sigma_k) \right\} = \sum_k \left\{ \sum_n \gamma_{nk} \left[\log\left(\frac{1}{\sqrt{2\pi |\Sigma_k|}} \right) + -\frac{1}{2}(x_n - N_k)^T \Sigma^{-1} (x_n - N_k) \right] \right\}$$

$$= \sum_k \left\{ \sum_n \gamma_{nk} \left(-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x_n - N_k)^T \Sigma^{-1}(x_n - N_k) \right) \right\}$$

3) a) $\nabla_{N_j} l(\theta) = 0 + \sum_n \gamma_{nj} \left(-\frac{1}{2}(x_n - N_j)\left((\Sigma^{-1})^T + \Sigma^{-1}\right)(-1) \right)$

b) $\frac{1}{2}\sum_n \gamma_{nj}(x_n - N_j)\left((\Sigma^{-1})^T + \Sigma^{-1}\right) = 0$

$$\sum_n (\gamma_{nj} x_n - \gamma_{nj} N_j) = 0$$

$$\sum_n \gamma_{nj} x_n = N_j \sum_n \gamma_{nj}$$

$$N_j = \frac{1}{\sum_n \gamma_{nj}} \sum_n \gamma_{nj} x_n \quad //$$

c) $x \in \{5, 15, 25, 30, 40\}$

$$w_1 = \frac{\sum_n \gamma_{n1}}{5} = \frac{0.2 + 0.2 + 0.8 + 0.9 + 0.9}{5} = \frac{3}{5} = 0.6$$

$$w_2 = \frac{\sum_n \gamma_{n2}}{5} = \frac{2}{3} = 0.4$$

$$N_1 = \frac{1}{3}\left(5(0.2) + 15(0.2) + 25(0.8) + 30(0.9) + 40(0.9)\right)$$

$$= 29$$

$$N_2 = \frac{1}{2}\left(5(0.8) + 15(0.8) + 25(0.2) + 30(0.1) + 40(0.1)\right)$$

$$= 14$$