1.1

$$P(x_1, x_2, \cdots, x_n) = \begin{cases} \frac{1}{\theta^n} & 0 < x_1 \leq \theta, x_2 \leq \theta, \cdots, x_n \leq \theta \\ 0 & else \end{cases} \tag{1}$$

So, to maximize the likelyhood , θ should be the min value that makes $P > 0$.
$\theta^* = max(x_1, x_2, \cdots, x_n)$
1.2
First:

$$P(k|x_n, \theta_1, \theta_2, w_1, w_2) = \frac{P(x_n, \theta_1, \theta_2, w_1, w_2|k)P(k)}{P(x_n, \theta_1, \theta_2, w_1, w_2)}$$

$$= \frac{\frac{1}{\theta_k} w_k 1[0 < x_n \leq \theta_k]}{\frac{1}{\theta_1} w_1 1[0 < x_n \leq \theta_1] + \frac{1}{\theta_2} w_2 1[0 < x_n \leq \theta_2]}$$

second:
Q=

$$\sum_n \frac{\frac{w_1}{\theta_1} 1[0 < x_n \leq \theta_1]}{\frac{w_1 1[0<x_n\leq\theta_1]}{\theta_1} + \frac{w_2 1[0<x_n\leq\theta_2]}{\theta_2}} log \frac{w_1 1[0 < x_n \leq \theta_1]}{\theta_1} + \frac{\frac{w_2}{\theta_2} 1[0 < x_n \leq \theta_2]}{\frac{w_1 1[0<x_n\leq\theta_1]}{\theta_1} + \frac{w_2 1[0<x_n\leq\theta_2]}{\theta_2}} log \frac{w_2 1[0 < x_n \leq \theta_2]}{\theta_2} \tag{2}$$

third:
First consider θ_1: the only part in $Q(\theta, \theta OLD)$ that is correlated to θ_1 is $\sum_{x_n \leq \theta_{1OLD}} log \frac{w_1 1[0<x_n\leq\theta_1]}{\theta_1}$
$\theta_1 = max(\forall x_n \leq \theta_{1OLD})$

second consider θ_2 the only part in $Q(\theta, \theta OLD)$ that is correlated to θ_2 is $\sum_{x_n \leq \theta_{2OLD}} log \frac{w_2 1[0<x_n\leq\theta_2]}{\theta_2}$
And $\theta_{2OLD} \geq max(x_1, x_2, \cdots, x_n)$ So $\theta_{2new} = max(x_1, x_2, \cdots, x_n)$

2.1
$P(X_b|X_a) = \frac{P(X)}{P(X_a)} = \frac{\sum_{k=1}^{K} \pi_k P(X|k)}{P(X_a)}$
$P(X|k) = P(x_b|x_a, k)P(x_a|k)$
$P(x_a) = \sum_k P(x_a|k)\pi_k$
So $P(X_b|X_a) = \frac{P(X)}{P(X_a)} = \sum_{k=1}^{K} \frac{\pi_k P(x_a|k)}{\sum_k \pi_k P(x_a|k)} P(x_b|x_a, k)$
$\lambda_k = \frac{\pi_k P(x_a|k)}{\sum_k \pi_k P(x_a|k)}$

3.1

$$\gamma(z_{nk}) = \frac{\pi_k exp(-(x_n-\mu_k)^2/(2\sigma^2))}{\sum_j \pi_j exp(-(x_n-\mu_k)^2+/(2\sigma^2))}$$

$$= \frac{\pi_k exp(-(x_n-\mu_k)^2/(2\sigma^2))}{1+\sum_{\forall j\neq k} \pi_j exp((x_n-\mu_k)^2-(x_n-\mu_j)^2/(2\sigma^2))}$$

$if\ k = argmin_{k'}||x_n-\mu_{k'}||^2:$
$then\ \forall j \neq k:$
$(x_n-\mu_k)^2-(x_n-\mu_j)^2 < 0$
$then \sigma \to 0, r(z_{nk}) \to \frac{\pi_k}{\pi_k} = 1 \qquad (0)$
else $\exists j(x_n-\mu_k)^2-(x_n-\mu_j) > 0$
$then \sigma \to 0, r(z_{nk}) \to \frac{\pi_k}{+\infty} = 0 \qquad (1)$

$$log\pi_k + logN(x_n|\mu_k,\sigma^2 I) = log\pi_k + log(\frac{1}{\sqrt{(2\pi)^k\sigma^{2k}}} * exp(-\frac{||x_n-\mu_k||^2}{2\sigma^{2k}}))$$

$$= C - \frac{||x_n-\mu_k||^2}{2\sigma^{2k}} \qquad (2)$$

in which C is the constant irrelivant to $||x_n-\mu_k||^2$.
And as $\sigma \to 0, C - \frac{||x_n-\mu_k||^2}{2\sigma^{2k}} \approx -\frac{||x_n-\mu_k||^2}{2\sigma^{2k}}$
From (0),(1),(2), we know that as $\sigma \to 0$, maximizing log-likelyhood equals to maxinmizing $-\gamma(z_{nk})\frac{||x_n-\mu_k||^2}{2\sigma^{2k}}$
which equals minimizing $\gamma(z_{nk})||x_n-\mu_k||^2$

4.1

$$P(x_n|Y=c;\mu,\sigma) = \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi\sigma_{cd}^2}} exp(-\frac{(x_d-\mu_{cd})^2}{2\sigma_{cd}^2})$$

$$P(X,Y) = \prod_{i=1}^{N} P(x=x_n, y=y_i) = \prod_{i=1}^{N} P(x=x_n|y=y_i)P(y=y_i)$$

$$= \prod_{i=1}^{N} \pi_{yi} \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi\sigma_{cd}^2}} exp(-\frac{(x_d - \mu_{cd})^2}{2\sigma_{cd}^2})$$

So log-likelyhood equals:

$$\sum_{i=1}^{N} (log\pi_{yi} + \sum_{d=1}^{D} (-\frac{(x_{id} - \mu_{y_i d})^2}{2\sigma_{y_i d}^2} + log\frac{1}{\sqrt{2\pi\sigma_{y_i d}^2}}))$$

4.2

$$\frac{\partial LL}{\partial \mu_{cd}} = \sum_{y_i = c} -\frac{x_{id} - \mu_{cd}}{\sigma_{y_i d}^2}$$

$$\mu_{cd}^* = \frac{\sum_{y_i=c} x_{id}}{number\ of\ data\ that\ y = c}$$

$$\frac{\partial LL}{\partial \sigma_{cd}} = \sum_{y_i = c} -\frac{(x_{id} - \mu_{cd})^2}{\sigma_{cd}^3} + \frac{1}{\sigma_{cd}}$$

plug $\mu_{cd}^* to \mu_{cd}$ and solve it we get

$$\sigma_{cd}^* = \frac{\sum_{y_i=c} (x_{id} - \frac{\sum_{y_i=c} x_{id}}{number\ of\ data\ that\ y=c})^2}{number\ of\ data\ that\ y = c}$$

In tuition,(as shown in lec 15, slide 49/58),

$$\pi_c = \frac{number\ of\ data\ labeled\ as\ c}{N}$$