Q1.1)

$$L(\theta) = P(X_1 = x_1, X_2 = x_2, \ldots . X_n = x_n)$$
$$L(\theta) = P(X = x_1|\theta).(X = x_2|\theta) \ldots.$$
$$L(\theta) = \frac{1}{\theta}\mathbf{1}(x_1).\frac{1}{\theta}\mathbf{1}(x_2).\frac{1}{\theta}\mathbf{1}(x_3) \ldots.$$
$$L(\theta) = \prod_{i=1}^{N}\frac{1}{\theta}\mathbf{1}(x_i)$$

$$\frac{\partial \log L(\theta)}{\partial x} = 0$$
$$L(\theta) = \frac{1}{\theta^n}\mathbf{1}(x_1).\mathbf{1}(x_2) \ldots . \mathbf{1}(x_n)$$
$$\log L(\theta) = \log\frac{1}{\theta^n} + \log(\mathbf{1}(x_1).\mathbf{1}(x_2) \ldots) = 0$$
$$n \log\theta = \log\left(\prod_{i=1}^{N}\mathbf{1}(x_i)\right)$$
$$\theta = e^{\frac{1}{N}\log(\prod_{i=1}^{N}\mathbf{1}(x_i))}$$

Q1.2)

$$P\left(k\middle|x_n, \theta_1, \theta_{2,}w_1, w_2\right) = \frac{w_k U(X = x|\theta_k)}{w_1 U(X = x|\theta_1) + w_2 U(X = x|\theta_2)}$$
$$Q(\theta, \theta_{old}) = \sum_n\sum_k P\left(k\middle|x_n, \theta_1^{old}, \theta_2^{old}, w_1^{old}, w_2^{old}\right)\log\left(P\left(x_n, k\middle|\theta_1, \theta_{2,}w_1, w_2\right)\right)$$
$$= \sum_n\sum_{k=1,2}\frac{w_k^{old}U\left(X = x\middle|\theta_k^{old}\right)}{w_1^{old}\frac{1}{\theta_1^{old}}\mathbf{1}\left[0 < x \le \theta_1^{old}\right] + w_2^{old}\frac{1}{\theta_2^{old}}\mathbf{1}\left[0 < x \le \theta_2^{old}\right]}\times\log(w_k U[x = x_n\,|\theta_k])$$
$$= \sum_n\sum_k P_{old}(k|x_n).\log(w_k U[x = x_n|\,\theta_k])$$

Q2)

$$Given: P(x_b \cap x_a) = \pi_1\frac{P(x_a \cap x_b \cap 1)}{P(1)} + \pi_2\frac{P(x_a \cap x_b \cap 2)}{P(2)} + \cdots$$
$$Given: \frac{P(x_b \cap x_a)}{P(x_a)} = \pi_1\frac{P(x_a \cap x_b \cap 1)}{P(1)P(x_a)} + \pi_2\frac{P(x_a \cap x_b \cap 2)}{P(2)P(x_a)} + \cdots \qquad \ldots(1)$$
$$We\ have,\ P(x_b|x_a) = \lambda_1 P(x_b|x_a, 1) + \lambda_2 P(x_b|x_a, 2)$$

$$\Rightarrow P(x_b|x_a) = \lambda_1\frac{P(x_b \cap x_a \cap 1)}{P(x_a \cap 1)} + \lambda_2\frac{P(x_b \cap x_a \cap 2)}{P(x_a \cap 2)} + \cdots \qquad \ldots(2)$$
$$From\ (1)\&\ (2), \lambda_k = \frac{P(x_a \cap k)\ \mathrm{x}\ \pi_k}{P(k)P(x_a)} = \frac{P(x_a|\,k)\ \mathrm{x}\ \pi_k}{P(x_a)}$$

Q3)

$$Q = \sum_{n}^{N}\sum_{k}^{K} \gamma(z_{nk})[\log p\, \pi_k \; + \log \mathrm{N}\,(x_n|\mu_k, \sigma^2 I)]$$

$$\frac{\partial Q}{\partial \mu_k} = 0\,;\; We\ get,$$

$$= \sum_{n}^{N}\sum_{k}^{K} \gamma(z_{nk})\,.\frac{\partial}{\partial \mu_k}\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(x_n-\mu_k)^2}{2\sigma^2}}\right)$$

$$\mu_k = \frac{\sum_n^N \gamma(z_{nk})x_n}{\gamma(z_{nk})}$$

As σ→0, In $\gamma(z_{nk})$ we see that in the denominator, the form for which $\left\|x_n - \mu_j\right\|^2$ is smallest will go to zero most slowly, and hence the responsibilities $\gamma(z_{nk})$ for data point x_n all go to zero except for term j, for which the responsibility $\gamma(z_{nk})$will go to unity. Thus, in the limit, we obtain a hard assignment of data points to clusters, just as in the K-means, so that $\gamma(z_{nk}) \rightarrow\; r_{nk}$. Each data point is thereby assigned to the cluster having the closest mean.

(Reference: Pattern Recognition & Machine Learning – Bishop School)

$$J = \sum_{k}^{K}\sum_{n}^{N} r_{nk}\, \|x_n - \mu_k\|_2^2$$

$$\frac{\partial \mathrm{J}}{\partial \mathrm{x}} = \sum_{n}^{N}\sum_{k}^{K} r_{nk}\,(x_n - \mu_k)(-1) = 0$$

$$\mu_k = \frac{\sum_n^N r_{nk}x_n}{r_{nk}}$$

Hence proved.

Q4.1)

Parameters to be learned π, μ, σ

$$P(X = x \cap Y = c) = P(Y = c)P(X = x|Y = c)$$

$$log\prod_{n=1}^{N} P(X = x, Y = c) = log\prod_{n=1}^{N} P(Y = c)\prod_{d=1}^{D} P(X_d = x_d|Y = c)$$

$$= log\prod_{n=1}^{N} (\pi_c\prod_{d=1}^{D} P(X_d = x_d|Y = c))$$

$$We\ know, P(X_d = x_d|Y = c; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma_{cd}^2}} e^{\frac{-(x_d-\mu_{cd})^2}{2\sigma_{cd}^2}}$$

$$= \sum_n (log\pi_{y_n}) + \sum_n \sum_{d=1}^{D} \log\left(\frac{1}{\sqrt{2\pi\sigma_{cd}^2}} e^{\frac{-(x_d-\mu_{cd})^2}{2\sigma_{cd}^2}}\right)$$

$$= \sum_n log\pi_{y_n} + \sum_{n,d} -\frac{1}{2}\log(\ 2\pi\sigma_{cd}^2) - \frac{(x_d - \mu_{cd})^2}{2\sigma_{cd}^2}$$

Q4.2)

$$Q = \sum_n log\pi_{cn} + \sum_{n,d} -\frac{1}{2}\log(2\pi\sigma_{cd}^2) - \frac{(x_d - \mu_{cd})^2}{2\sigma_{cd}^2}$$

$$\frac{\partial Q}{\partial x} = 0, \sum_{n=1}^{N} -\frac{1}{4\pi\sigma_{cd}^2}.2\pi + \sum_{n=1}^{N} \frac{(x_d - \mu_{cd})^2}{2(\sigma_{cd}^2)^2} = 0$$

$$\sigma_{cd}^2(-N) + \sum_{n=1}^{N} (x_d - \mu_{cd})^2 = 0$$

$$\Rightarrow \sigma_{cd}^2 = \frac{\sum_{n=1}^{N} (x_d - \mu_{cd})^2}{N}$$

$$\frac{\partial Q}{\partial \mu_{cd}} = 0, \sum_{n=1}^{N} \frac{-2(x_d - \mu_{cd}).-1}{2\sigma_{cd}^2} = 0$$

$$\mu_{cd} = \frac{\sum_{n=1}^{N} x_d}{N}$$

$$\pi_c = \frac{|c|}{N}$$

Where $|c|$ is size of set [c] to which the dependent variable y belongs.