Question 1.

For the mixture model the joint distribution can be written

$$p(x_a, x_b) = \sum_{k=1}^{K} \pi_k p(x_a, x_b | k)$$

We can find the conditional density $p(x_b | x_a)$ by making use of the relation

$$p(x_b | x_a) = \frac{p(x_a, x_b)}{p(x_a)}$$

The mixture model the marginal density of $x_a$ is give by

$$p(x_a) = \sum_{k=}^{K} \pi_k p(x_a | k)$$

where

$$p(x_a | k) = \int p(x_a, x_b | k) dx_b$$

Thus we can write the conditional density in the form

$$p(x_b | x_a) = \frac{\sum_{k=1}^{K} \pi_k p(x_a, x_b | k)}{\sum_{j=1}^{K} \pi_j p(x_a | j)}$$

Decompose the numerator using $p(x_a, x_b | k) = p(x_b | x_a, k) p(x_a | k)$

which allows us finally to write the conditional density as a mixture model of the form

$$p(x_b | x_a) = \sum_{k=1}^{K} \lambda_k p(x_b | x_a, k)$$

where the mixture coefficients are given by

$$\lambda_k \equiv p(k | x_a) = \frac{\pi_k p(x_a | k)}{\sum_j \pi_j p(x_a | j)}$$

and $p(x_b | x_a, k)$ is the conditional for component $k$.

Question 2

(a) $\hat{a} = \frac{\frac{1}{2}}{\frac{1}{2} + \mu} h \qquad \hat{b} = \frac{\mu}{\frac{1}{2} + \mu} h$

(b) $\hat{\mu} = \frac{h - a + c}{6(h - a + c + d)}$

Question 3

(a) $P(Z=1 | X=1) \propto P(X=1 | Z=1) P(Z=1) = \pi_1 e^{-1}$

$P(Z=2 | X=1) \propto P(X=1 | Z=2) P(Z=2) = \pi_2 4 e^{-2}$

$P(Z=3 | X=1) \propto P(X=1 | Z=3) P(Z=3) = \pi_3 16 e^{-4}$

$$P(Z=1 | X=1) = \frac{\pi_1 e^{-1}}{(\pi_1 e^{-1} + \pi_2 4 e^{-2} + \pi_3 16 e^{-4})}$$

(b) For each $X = x$ $\quad P(Z=k | X=x) = \frac{P(X=x | Z=k) P(Z=k)}{\sum_{k'} P(X=x | Z=k') P(Z=k')} = \frac{\beta_k^2 x e^{-\beta_k x} \pi_k}{\sum_{k'} \beta_{k'}^2 x e^{-\beta_{k'} x} \pi_{k'}}$

Question 4

$\sum_{k=1}^{K} \pi_k = 1 \qquad \tilde{Q} = \sum_{k=1}^{K} r(z_1 k) \ln \pi_k + \lambda \left( \sum_{k=1}^{K} \pi_k - 1 \right)$ Setting the derivative with respect to $\pi_k$ equal to zero

$$0 = r(z_1 k) \frac{1}{\pi_k} + \lambda \qquad \lambda = -\sum_{k=1}^{K} r(z_1 k) \qquad \pi_k = \frac{r(z_1 k)}{\sum_{j=1}^{K} r(z_1 j)}$$

maximization with respect to $A$ $\quad \sum_{k=1}^{K} A_{jk} = 1$ for $j = 1, \dots, K$

$$\hat{Q} = \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \ln A_{jk} + \sum_{j=1}^{K} \lambda_j \left( \sum_{k=1}^{K} A_{jk} - 1 \right)$$

Set the derivative of $Q$ with respect to $A_{jk}$ to zero

$$0 = \sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nk}) \frac{1}{A_{jk}} + \lambda_j$$

Multiply through by $A_{jk}$ and then sum over $k$, make use of summation constraint $\quad \lambda_j = -\sum_{n=2}^{N} \sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk})$

Substituting for $\lambda_j$ and solving for $A_{jk}$ $\Rightarrow A_{jk} = \frac{\sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nk})}{\sum_{l=1}^{K} \sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nl})}$

Questio 5

$$\sum_{n=1}^{N}\sum_{k=1}^{K} r(z_{nk}) \ln p(x_n|\phi_k) = \sum_{n=1}^{N}\sum_{k=1}^{K} r(z_{nk}) \sum_{i=1}^{D} x_{ni} \ln \mu_{ki}$$

Introduce Lagrange multipliers $\{\lambda_k\}$ and maximize the function given by

$$\sum_{n=1}^{N}\sum_{k=1}^{K} r(z_{nk}) \sum_{i=1}^{D} x_{ni} \ln \mu_{ki} + \sum_{k=1}^{K} \lambda_k \left(\sum_{i=1}^{D} \mu_{ki} - 1\right)$$

Setting the derivative with respect to $\mu_{ki}$ to zero we obtain

$$0 = \sum_{n=1}^{N} r(z_{nk}) \frac{x_{ni}}{\mu_{ki}} + \lambda_k$$

Multiplying through by $\mu_{ki}$, summing over $i$, and making use of the constraint on $\mu_{ki}$ together with result $\sum_i x_{ni} = 1$ we have

$$\lambda_k = -\sum_{n=1}^{N} r(z_{nk})$$

back-substituting for $\lambda_k$ and solving for $\mu_{ki}$

$$\mu_{ik} = \frac{\sum_{n=1}^{N} r(z_{nk}) x_{ni}}{\sum_{n=1}^{N} r(z_{nk})}$$

Similarly,

$$\sum_{n=1}^{N}\sum_{k=1}^{K} r(z_{nk}) \ln p(x_n|\phi_k) = \sum_{n=1}^{N}\sum_{k=1}^{K} r(z_{nk}) \sum_{i=1}^{D} \{x_{ni} \ln \mu_{ki} + (1-x_{ni}) \ln (1-\mu_{ki})\}$$

Maximizing with respect to $\mu_{ki}$ we obtain

$$\mu_{ki} = \frac{\sum_{n=1}^{N} r(z_{nk}) x_{ni}}{\sum_{n=1}^{N} r(z_{nk})}$$

Question 6

(a)

$$p(X, Z|\theta) = \prod_{r=1}^{R} p(X^{(r)}, Z^{(r)}|\theta)$$

$$p(Z|X,\theta) = \frac{p(X, Z|\theta)}{\sum_Z p(X, Z|\theta)}$$

$$= \frac{\prod_{r=1}^{R} p(X^{(r)}, Z^{(r)}|\theta)}{\sum_{Z^{(1)}} \cdots \sum_{Z^{(R)}} \prod_{r=1}^{R} p(X^{(r)}, Z^{(r)}|\theta)}$$

$$= \prod_{r=1}^{R} \left\{ \frac{p(X^{(r)}, Z^{(r)}|\theta)}{\sum_{Z^{(r)}} p(X^{(r)}, Z^{(r)}|\theta)} \right\}$$

$$= \prod_{r=1}^{R} p(Z^{(r)}|X^{(r)}, \theta)$$

(b) joint distribution

$$Q(\theta, \theta_{old}) = E_Z[\ln p(X, Z|\theta)] = E_Z\left[\sum_{r=1}^{R} \ln p(X^{(r)}, Z^{(r)}|\theta)\right]$$

$$= \sum_{r=1}^{R} p(Z^{(r)}|X^{(r)}, \theta_{old}) \ln p(X, Z|\theta)$$

$$= \sum_{r=1}^{R}\sum_{k=1}^{K} r(z_{1k}^{(r)}) \ln \pi_k + \sum_{r=1}^{R}\sum_{n=2}^{N}\sum_{j=1}^{K}\sum_{k=1}^{K} \xi(z_{n-1,j}^{(r)}, z_{n,k}^{(r)}) \ln A_{jk} + \sum_{r=1}^{R}\sum_{n=1}^{N}\sum_{k=1}^{K} r(z_{nk}^{(r)}) \ln p(x_n^{(r)}|\phi_k)$$

maximize the quantity with respect to $\pi$ and $A$ with Lagrange multiplier

Then we can get $\pi_k = \dfrac{\sum_{r=1}^{R} r(z_{1k}^{(r)})}{\sum_{r=1}^{R}\sum_{j=1}^{K} r(z_{1j}^{(r)})}$

$$A_{jk} = \frac{\sum_{r=1}^{R}\sum_{n=2}^{N} \xi(z_{n-1,j}^{(r)}, z_{n,k}^{(r)})}{\sum_{r=1}^{R}\sum_{l=1}^{K}\sum_{n=2}^{N} \xi(z_{n-1,j}^{(r)}, z_{n,l}^{(r)})}$$

(c) the parameters of the Gaussian emission densities appear only in the last term.

$$\sum_{n=1}^{N}\sum_{k=1}^{K} r(z_{nk}) \ln p(x_n|\phi_k) = \sum_{n=1}^{N}\sum_{k=1}^{K} r(z_{nk}) \ln N(x_n|\mu_k, \Sigma) = \sum_{n=1}^{N}\sum_{k=1}^{K} r(z_{nk}) \left\{-\frac{D}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}(x_n-\mu_k)^T \Sigma^{-1} (x_n-\mu_k)\right\}$$

maximize the quantity with respect to $\mu_k$ and $\Sigma_k$. Setting the derivative with respect to $\mu_k$ to zero and rearranging

$$\mu_k = \frac{\sum_{n=1}^{N} r(z_{nk}) x_n}{\sum_{n=1}^{N} r(z_{nk})}$$

Rewrite the final term in the form $\frac{N_k D \ln(2\pi)}{2} - \frac{N_k}{2}\ln|\Sigma_k| - \frac{1}{2}\mathrm{Tr}(\Sigma_k^{-1}\hat{S}_k)$

Differentiating w.r.t. $\Sigma_k^{-1}$

$$\Sigma_k = \frac{\sum_{n=1}^{N} r(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T}{\sum_{n=1}^{N} r(z_{nk})}$$