4.1

Using (9.10) and (9.11) we can get:

$$p(x) = \sum_z p(x|z)\,p(z) = \sum_z \prod_{k=1}^{K} (\pi_k \mathcal{N}(x|\mu_k, \Sigma_k))^{z_k}$$

Since z is a representation for 1 of K we can proceed with the following form:

$$\sum_{j=1}^{K} \prod_{k=1}^{K} (\pi_k \mathcal{N}(x|\mu_k, \Sigma_k))^{I_{kj}} = \sum_{j=1}^{K} \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)$$

where $I_{kj} = 1$ if $k = j$, or 0 otherwise.

4.2

If we explain k-means and the Gaussian mixture model in simple words, k-means is the hard assignment and Gaussian mixture is the soft assignment. Basically, while in k-means each data point can only belong to one cluster at each step of the algorithm, in Gaussian mixture each data point has a probability of belonging to each of the clusters, so the most straightforward way to modify the k-means algorithm is to make a list of probabilities for each data point representing the likelihood of belonging to a certain cluster. k-means has a great running time, it is easy to implement and interpret but it does not work with complex shapes (non-linear data). Gaussian mixture in its turn works with anything but it can be difficult to interpret and each be hard to initialize.

4.3

(a) $x_n = \sum_{i=1}^{D} \alpha_{ni} u_i = \alpha_{n1} u_1 + \alpha_{n2} u_2 + \dots + \alpha_{nD} u_D$

$x_n^T \cdot u_1 = (\alpha_{n1} u_1)^T u_1 + (\alpha_{n2} u_2)^T u_2 + \dots + (\alpha_{nD} u_D)^T u_1$

we know that the transpose of a scalar is just itself, so

$$\alpha_{nj} = x_n^T u_j$$

(1)

consider $x_n = \sum_{i=1}^{M} z_{ni} u_i + \sum_{i=M+1}^{D} b_i u_i$

and $J = \frac{1}{N} \sum_{n=1}^{N} \| x_n - \tilde{x}_n \|^2 = \frac{1}{N} \sum_{n=1}^{N} (x_n - \tilde{x}_n)^T (x_n - \tilde{x}_n)$

substitute for $\tilde{x}_n$ and derivate

$x_n^T \cdot \tilde{x}_n = (\sum \alpha_{ni} u_i)^T (\sum z_{ni} \cdot u_i + \sum_{n=1}^{D} b_i u_i)$

$= (\sum_{i=1}^{M} \alpha_{ni} z_{ni} + \dots)$

$\tilde{x}_n \cdot x_n^T = (\sum_{i=1}^{M} \alpha_{ni} z_{ni} + \dots)$

$\tilde{x}_n^T \tilde{x}_n = (\sum z_{ni} u_i + \sum b_i u_i)^T (\sum z_{ni} u_i + \sum b_i u_i)$

many terms will cancel out as $u_i^T u_j = \delta_{ij}$

So, $0 = -\frac{1}{N} \sum_{i=1}^{N} (2 \sum_{i=1}^{M} \alpha_{ni} - 2 \sum_{i=1}^{M} z_{ni})$

so we get that $z_{ni} = \alpha_{ni}$ and

$$z_{ni} = x_n^T u_j$$

(2)

$$x_n^T \tilde{x}_n = \sum \alpha_{ni} u_i^T \cdot \left(\sum_{i}^{D} z_{ni} u_i + \sum_{m+1}^{D} b_i u_i \right)$$

$$= \sum_{i=m+1}^{D} \alpha_{ni} b_i$$

$$\tilde{x}_n^T \tilde{x}_n = \left(\sum z_{ni} u_i^T + \sum b_i n_i^T \right) \left(\sum z_{ni} u_i + \sum b_i u_i \right) =$$

$$= \sum_{i=m+1}^{D} b_i^2$$

$$2 b_i = \frac{1}{N} \sum_{n=1}^{N} 2 \cdot \sum_{i=m+1}^{D} \alpha_{ni} = \frac{2}{N} \sum_{n=1}^{N} x_n^T \cdot u_i =$$

$$= 2 \left(\frac{1}{N} \sum x_n \right)^T u_i$$

$$\boxed{b_i = \bar{x}^T u_i}$$