Assignment 3

Problem 1

1. Because $Y \sim \text{Multinomial}(1, p_1, \cdots, p_k)$

so that $P(Y=y_1, Y=y_2, \ldots Y_k=y_k) = \frac{N!}{y_1!y_2!\ldots y_k!} \cdot P_1^{y_1} \cdots P_k^{y_k}$

We can know that $N=1$ from the question,

so there is only a 1 in y_1 to y_k, and other factor are 0s.

Besides, we know that $X|Y=l \sim N(\mu_l, \sigma_l^2)$

so $P(X|Y=l) = \frac{1}{\sqrt{2\pi}\sigma_l} \exp\left(-\frac{(1-\mu_l^2)}{2\sigma_l^2}\right)$

It can be assumed that the probability at $Y=l$ is P_i

$P(X,Y) = P(X=x, Y=l)$

$= P(X=x|Y=l) \cdot P(Y=l)$

$= \frac{1}{\sqrt{2\pi}\sigma_l} \exp\left(-\frac{(1-\mu_l)^2}{2\sigma_l^2}\right) \cdot p_i^2$

$$= \frac{P_i^2}{\sqrt{2\pi}\,\sigma_\ell} \exp\left(-\frac{(1-\mu_\ell)^2}{2\sigma_\ell^2} \right)$$

From the information online, the exponential family is expressed as

$$f(y_i \mid \theta_i, \varphi_i) = \exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{\varphi_i^2} + C(y_i, \varphi_i) \right\}$$

and the above expression can be expressed as

$$f(x, y; \mu_\ell, \sigma_\ell) = C(\sigma_\ell)\, e^{\sum_{i=1}^{K} C_i(\mu_\ell) C_i(\sigma_\ell) T_i(x)} \cdot h(x)$$

Thus, it belongs to exponential family.

2. Because this is an i.i.d. sample,

so $F(X \mid Y) = \frac{F(XY)}{F(Y)} = \frac{F(X) \cdot F(Y)}{F(Y)} = F(X)$

thus the marginal density of X is

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma_\ell} \exp\left(-\frac{(x-\mu_\ell)^2}{2\sigma_\ell^2} \right)$$

3. Since the likelihood

$$L(\theta|X) = \prod_{i=1}^{n} f(x_i;\mu_l,\sigma_l^2)$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma_l}\right)^n e^{-\sum_{i=1}^{n}\frac{(x_i-\mu_l)^2}{2\sigma_l^2}}$$

$$= (2\pi\sigma_l)^{-\frac{n}{2}} e^{-\sum_{i=1}^{n}\frac{(x_i-\mu_l)^2}{2\sigma_l^2}}$$

If we take logarithm of $L(\theta|X)$, then it becomes

$$LL(\theta|X) = n\ln\left(\frac{1}{\sqrt{2\pi\sigma_l^2}}\right) - \frac{1}{2\sigma_l^2}\sum_{i=1}^{n}(x_i-\mu_l)^2$$

$$= -\frac{1}{2\sigma_l^2}\sum_{i=1}^{n}(x_i-\mu_i)^2 - \frac{n}{2}\ln\sigma_l^2 - \frac{n}{2}\ln 2\pi$$

The parameters are μ and σ_i

Problem 2

For each data point x_i, we introduce a latent variable $Y_i \in \{1, 2 \dots m\}$ denoting the component that points belongs to. For the E-step

$$\gamma_j(x_i) = P(y_i = j \mid x_i) \propto P(X_i \mid y_i = j) P(y_i = j)$$

$$= \pi_j f_L(x_i; \mu_j, \beta_j)$$

$$= \frac{\pi_j f_L(x_i; \mu_j, \beta_j)}{\sum_{l=1}^{m} \pi_l f_L(x_i; \mu_l, \beta_l)}$$

In M-step, we optimize

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_j(x_i) \log P(x_i, y_i = j) = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_j(x_i) \log \pi_j f_L(x_i; \mu_j, \beta_j)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_i(x_i) \left(\log \pi_j - \frac{1}{\beta_j} |x_i - \mu_j|\right) + \text{const}$$

We add Lagrange multiplier λ to make sure that $\sum_{j=1}^{m} \pi_j = 1$ and obtain the Lagrangian

$$\mathcal{L}(\pi,\mu,\lambda)=\sum_{i=1}^{n}\sum_{j=1}^{m} r_j(x_i)\left(\log\pi_j-\frac{1}{\beta_j}|x_i-\mu_j|\right)+\lambda\left(\sum_{j=1}^{m}\pi_j-1\right)$$

Setting the gradient with respect to π_j to zero,

$$\frac{\partial}{\partial\pi_j}\mathcal{L}(\pi,\mu,\lambda)=\sum_{i=1}^{n} r_i(x_i)/\pi_j+\lambda=0$$

$$\Rightarrow \quad \pi_j=\frac{\sum_{i=1}^{n} r_j(x_i)}{-\lambda}$$

So $\lambda=-n$ and if we want to maximize with respect to the variables μ_j, we have to solve m separate optimization problems, one for each μ_j. These m problems have following form

$$\underset{\mu_j}{\text{maximize}} -\sum_{i=1}^{n}\frac{r_j(x_i)}{\beta_j}|x_i-\mu_j|$$

These are one-dimensional convex optimization problems, and the negative of the objective is easily to be seen convex. A direct approach is possible if we observe that the function is piecewise linear and breakpoints are $x_1, x_2, \cdots, x_n$. Hence, the optimum must be attained at one of these n points and we can simply set μ_j to the point x_i with the largest objective value.