Q2

a) when $\Sigma_k = I$ and $\pi_k = \frac{1}{K}$,

$$z_i(EM) = \underset{k}{\operatorname{argmax}} \frac{\pi_k N(x_i | \mu_k, \Sigma_k)}{\sum_{k'} \pi_{k'} N(x_i | \mu_{k'}, \Sigma_{k'})} = \underset{k}{\operatorname{argmax}} \frac{N(x_i | \mu_k, I)}{\sum_{k'} N(x_i | \mu_{k'}, I)}$$

$\Rightarrow$ ML estimation

$$z_i(k\text{-Means}) = \underset{k}{\operatorname{argmin}} \|x - \mu_k\|_2^2 = (X^T X)^{-1} X^T y = w_{ML}.$$

$\therefore z_i(EM) = z_i(k\text{-Means})$.

we have $\mu_k(EM) = \frac{1}{\sum_i r_{ik}} \sum_i r_{ik} x_i$ and $\mu_k(k\text{-Means}) = \frac{1}{N_k} \sum_{i: z_i = k} x_i$

$\because \sum_i r_{ik} = \sum_i z_i$ = # of points assigned to each cluster center $= N_k$.

It follows that the 2 algorithms are equivalent given the conditions.

b) If the clustering is far apart, ie every point is responsible largly for only one center, then r_{ik} is sparse, ie

$$r_{ik} \approx \begin{cases} 1, & \text{for only } 1\ k \text{ for each } i \\ 0, & \text{otherwise} \end{cases}$$

$\therefore \pi_k$ is near diagonal and $\pi_k = \frac{1}{N} \sum_i r_{ik} = \frac{1}{K}$.

$\therefore$ We now satisfy the condition for (a)

$\therefore$ In this case, soft and hard assignments are nearly equivalent.

3

a) $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\,dy = \int_{-\infty}^{x} f(y) \cdot 1\,dy + \int_{x}^{\infty} f(y) \cdot 0\,dy$

$= E[\mathbb{I}\{X \leq x\}]$

b) $E[\hat{F}_n(x)] = E\left[\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{X_i \leq x\}\right] = \frac{1}{n} \sum_{i=1}^{n} E[\mathbb{I}\{X_i \leq x\}]$

$= \frac{1}{n} \sum_{i=1}^{n} F(X_i) = F(x)$

c) $var(\hat{F}_n(x)) = E[(\hat{F}_n(x) - F(x))^2] = E[(\hat{F}_n(x))^2] - E[\hat{F}_n(x)]^2$

$= \text{?}$

d) want to show $E[(\hat{F}_n(x) - F(x))^2] \le \frac{1}{4n}$

$\Leftrightarrow \quad \frac{F(x)(1-F(x))}{n} \le \frac{1}{4n} \quad \Leftrightarrow \quad F(x)(1-F(x)) \le \frac{1}{4}$.

We know that $F(x)(1-F(x)) = \frac{1}{4}$ when $F(x) = \frac{1}{2}$.

Any other x would make it $< \frac{1}{4}$.

$\therefore \quad F(x)(1-F(x)) \le \frac{1}{4}$ holds

$\therefore \quad E[(\hat{F}_n(x) - F(x))^2] \le \frac{1}{4n}$ holds ∎