Introduction to Machine Learning Tech Report

Student Name Student ID: XXXX

May 20, 2025

Question 1

E-step

Introduce latent variables $z_{nk} \in \{0,1\}$ with $\sum_{k=1}^{K+1} z_{nk} = 1$. The posterior responsibilities are:

$$\gamma_{nk} := \mathbb{E}[z_{nk} \mid \mathbf{x}_n, \Theta^{\text{old}}] = \frac{w_k \mathcal{N}(\mathbf{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} w_j \mathcal{N}(\mathbf{x}_n; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) + w_{K+1} p_{\text{noise}}(\mathbf{x}_n)}, \quad k \leq K, \tag{1}$$

$$\gamma_{n,K+1} = \frac{w_{K+1} p_{\text{noise}}(\mathbf{x}_n)}{\sum_{j=1}^{K} w_j \mathcal{N}(\mathbf{x}_n; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) + w_{K+1} p_{\text{noise}}(\mathbf{x}_n)}. \tag{2}$$

M-step

Let $N_k = \sum_{n=1}^{N} \gamma_{nk}$. Maximizing the expected complete-data log-likelihood yields:

$$w_k^{\text{new}} = \frac{N_k}{N}, \quad k = 1, \ldots, K+1, \tag{3}$$

$$\boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk} \mathbf{x}_n, \quad k = 1, \ldots, K, \tag{4}$$

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk} (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})^\top, \quad k = 1, \ldots, K. \tag{5}$$

Algorithm 1 EM for GMM with Uniform Outliers

```
1: Initialise Θ^(0).
2: for t = 0, 1, ... until convergence do
3:     E-step: compute γ_nk^(t) using the formulas above.
4:     M-step: update parameters using the update rules.
5: end for
```

Question 2: Mean–Shift from an (M)MLE and MAP Perspective (4 Pts)

(a) Mean–Shift as Maximum Likelihood

MLE of mode of KDE solves $\nabla_{\mathbf{x}}\widehat{p}_h(\mathbf{x}) = 0$, leading to:

$$\begin{aligned}
\frac{\partial \hat{p}(x)}{\partial x} &= \frac{1}{N}\sum_{n=1}^{N}\frac{\partial K_h(x, x_n)}{\partial x} \\
&= -\frac{1}{Nh^2}\sum_{n=1}^{N} K_h(x, x_n)(x - x_n) \\
&\propto \sum_{n=1}^{N} K_h(x, x_n)(x_n - x) \\
&\propto \frac{\sum_{n=1}^{N} K_h(x, x_n)x_n}{\sum_{n=1}^{N} K_h(x, x_n)} - x
\end{aligned} \tag{6}$$

Thus,in every step we update $\mathbf{x}$ as $\frac{\sum_{n=1}^{N} K_h(x,x_n)x_n}{\sum_{n=1}^{N} K_h(x,x_n)}$

(b) MAP with a Gaussian Prior $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$

Maximising $\widehat{p}_h(\mathbf{x})\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \sigma^2\mathbf{I})$:

$$\begin{aligned}
\frac{\partial[\widehat{p}_h(\mathbf{x})\mathcal{N}(\mathbf{x})]}{\partial x} &= \frac{\partial[\widehat{p}_h(\mathbf{x})]}{\partial x}\mathcal{N}(\mathbf{x}) + \frac{\partial[\mathcal{N}(\mathbf{x})]}{\partial x}\widehat{p}_h(\mathbf{x}) \\
&= -\frac{1}{Nh^2}\mathcal{N}(\mathbf{x})\sum_{n=1}^{N} K_h(x, x_n)(x - x_n) - \frac{(x-\mu)}{N\sigma^2}\mathcal{N}(\mathbf{x})\sum_{n=1}^{N} K_h(x, x_n) \\
&\propto \sum_{n=1}^{N} K_h(x, x_n)(x_n - x + \frac{h^2(\mu - x)}{\sigma^2}) \\
&\propto \frac{\sum_{n=1}^{N} K_h(x, x_n)x_n}{\sum_{n=1}^{N} K_h(x, x_n)} + \frac{h^2\mu}{\sigma^2} - (1 + \frac{h^2}{\sigma^2})x
\end{aligned} \tag{7}$$

Thus, the regularised update is:

$$\mathbf{x}^{(t+1)} = \frac{\sigma^2}{h^2+\sigma^2}[\frac{\sum_{n=1}^{N} K_h(x^{(t)}, x_n)x_n}{\sum_{n=1}^{N} K_h(x^{(t)}, x_n)} + \frac{h^2\mu}{\sigma^2}] \tag{8}$$

(c) MAP with a GMM Prior

For isotropic GMM prior $p_{\text{prior}}(\mathbf{x}) = \frac{1}{K}\sum_{k=1}^{K}\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \sigma_k^2\mathbf{I})$, define $\alpha_k(\mathbf{x}) := \frac{\mathcal{N}(\mathbf{x};\boldsymbol{\mu}_k,\sigma_k^2\mathbf{I})}{\sum_j \mathcal{N}(\mathbf{x};\boldsymbol{\mu}_j,\sigma_j^2\mathbf{I})}$. Then:

$$\sum_n K_h(x)(\mathbf{x} - \mathbf{x}_n) + \sum_k \frac{h^2\alpha_k(\mathbf{x})}{\sigma_k^2}(\mathbf{x} - \boldsymbol{\mu}_k) = 0. \tag{9}$$

Hence the update is:

$$\mathbf{x}^{(t+1)} = \frac{\sum_n K_h(x^{(t)}, x_n)\mathbf{x}_n + h^2\sum_k \alpha_k(\mathbf{x}^{(t)})\sigma_k^{-2}\boldsymbol{\mu}_k}{\sum_n K_h(x^{(t)}, x_n) + h^2\sum_k \alpha_k(\mathbf{x}^{(t)})\sigma_k^{-2}}. \tag{10}$$