CS 771 Minor Assignment-2

Chadhurbala R V 251110022 **Nishu Kumar** 251110054 **Pranjal** 251110058

Department of Computer Science and Engineering
Indian Institute of Technology Kanpur, Kanpur, UP

Question 1:

Let z be a K-dimensional random variable, $z = (z_1, \ldots, z_K)$ with $z_k \in \{0, 1\}$ and $\sum_k z_k = 1$. Within class c,

$$p(z_k = 1 \mid y = c) = \pi_{c,k}, \qquad \sum_{k=1}^{K} \pi_{c,k} = 1, \;\; \pi_{c,k} \geq 0, \tag{1}$$

$$p(x \mid z_k = 1, y = c) = \mathcal{N}\big(x \mid \mu_{c,k}, \Sigma_{c,k}\big). \tag{2}$$

Marginalizing the latent z gives the class-conditional density:

$$p(x \mid y = c) \;=\; \sum_{z} p(z \mid y = c)\, p(x \mid z, y = c) \;=\; \sum_{k=1}^{K} \pi_{c,k}\, \mathcal{N}\big(x \mid \mu_{c,k}, \Sigma_{c,k}\big). \tag{3}$$

Training: Expectation–Maximization (EM) applied independently for each class [1]. Let $\mathcal{I}_c = \{n : y_n = c\}$ be the indices of training points of class c, and $N_c = |\mathcal{I}_c|$.

1. Initialization: Initialize the parameters $\{\pi_{c,k}, \mu_{c,k}, \Sigma_{c,k}\}$.

2. E-step: For $n \in \mathcal{I}_c$ and $k = 1, \ldots, K$,

$$q_{nk}^{(c)} \;\equiv\; p(z_k = 1 \mid x_n, y = c) = \frac{\pi_{c,k}\, \mathcal{N}(x_n \mid \mu_{c,k}, \Sigma_{c,k})}{\sum_{j=1}^{K} \pi_{c,j}\, \mathcal{N}(x_n \mid \mu_{c,j}, \Sigma_{c,j})}, \qquad N_{c,k} \;=\; \sum_{n \in \mathcal{I}_c} q_{nk}^{(c)}. \tag{4}$$

3. M-step: Using $N_{c,k}$ and $q_{nk}^{(c)}$,

$$\pi_{c,k}^{\text{new}} = \frac{N_{c,k}}{N_c}, \tag{5}$$

$$\mu_{c,k}^{\text{new}} = \frac{1}{N_{c,k}} \sum_{n \in \mathcal{I}_c} q_{nk}^{(c)}\, x_n, \tag{6}$$

$$\Sigma_{c,k}^{\text{new}} = \frac{1}{N_{c,k}} \sum_{n \in \mathcal{I}_c} q_{nk}^{(c)} \big(x_n - \mu_{c,k}^{\text{new}}\big)\big(x_n - \mu_{c,k}^{\text{new}}\big)^{\top}. \tag{7}$$

4. Check for convergence of the parameters. Go back to step 2 if the convergence criterion is not satisfied.

Repeat the same procedure independently for each class c to obtain all $\{\pi_{c,k}, \mu_{c,k}, \Sigma_{c,k}\}$.

Preprint. Under review.

Question 2:

After training, for each class $c \in [C]$ we have a mixture $\{(\pi_{c,k},\, \mu_{c,k},\, \Sigma_{c,k})\}_{k=1}^{K}$ and an estimated class prior $p(y = c)$. Given a test point $x \in \mathbb{R}^d$, we compute the posterior $p(y = c \mid x)$ and predict $\widehat{y} = \arg\max_c p(y = c \mid x)$.

$$p(y = c \mid x) = \frac{p(y = c)\, p(x \mid y = c)}{p(x)}$$

where,

$$p(x \mid y = c) \;=\; \sum_{k=1}^{K} \pi_{c,k}\, \mathcal{N}\big(x \mid \mu_{c,k}, \Sigma_{c,k}\big). \tag{8}$$

$$\text{Class prior: } p(y = c) \;=\; \frac{N_c}{N}. \tag{9}$$

The MAP classifier can ignore $p(x)$ because p(x) is the same for every class in the denominator.

$$\widehat{y} \;=\; \arg\max_{c} \;\Big\{ \log p(y = c) + \log p(x \mid y = c) \Big\}. \tag{10}$$

Question 3:

Changes Done in EM procedure:

E-step: we use the mixture prior $\pi_{c,k}$ and compute,

$$q_{nk}^{(c)} \;\propto\; \pi_{c,k}\; \exp\Big(-\tfrac{1}{2}\|x_n - \mu_{c,k}\|^2 \Big), \qquad \sum_{k=1}^{K} q_{nk}^{(c)} = 1.$$

M-step (update means and mixture weights). With $N_{c,k} = \sum_{n=1}^{N_c} q_{nk}^{(c)}$ and $N_c = |X_c|$,

$$\pi_{c,k}^{\text{new}} \;=\; \frac{N_{c,k}}{N_c}, \qquad \mu_{c,k}^{\text{new}} \;=\; \frac{1}{N_{c,k}} \sum_{n=1}^{N_c} q_{nk}^{(c)}\, x_n.$$

Covariances: computed once after EM with regularization. To avoid instability, we form the final covariances by the standard weighted estimator with a small diagonal term:

$$\Sigma_{c,k}^{\text{final}} = \frac{1}{N_{c,k}} \sum_{n=1}^{N_c} q_{nk}^{(c)} \big(x_n - \mu_{c,k}^{\text{new}}\big)\big(x_n - \mu_{c,k}^{\text{new}}\big)^{\top} + \varepsilon I, \quad \varepsilon = 10^{-5}.$$

Training: The above EM is run independently on each class subset X_c, producing $\{\mu_{c,k}, \Sigma_{c,k}, \pi_{c,k}\}_{k=1}^{K}$ and the class prior $p(y{=}c) = N_c/N$.

Prediction: For a test point x, we score each class by aggregating its components in the log domain:

$$\log p(x \mid y{=}c) = \mathrm{LSE}_{k=1}^{K}\Big(\log \pi_{c,k} + \log \mathcal{N}\big(x \mid \mu_{c,k}, \Sigma_{c,k}\big) \Big),$$

and take the MAP decision

$$\widehat{y} = \arg\max_{c} \Big[\log p(y{=}c) + \log p(x \mid y{=}c) \Big].$$

Here LSE is the log-sum-exp operator, which is numerically stable and implements the correct mixture aggregation; the common evidence $p(x)$ is constant across classes and is omitted for the $\arg\max$.

K	Train Accuracy (%)	Test Accuracy (%)
1	86.14	83.48
2	90.50	85.83
5	94.63	87.92
10	97.09	88.27
15	98.26	87.97
20	98.92	88.56

Table 1: Training and Testing Accuracy for Different Values of K

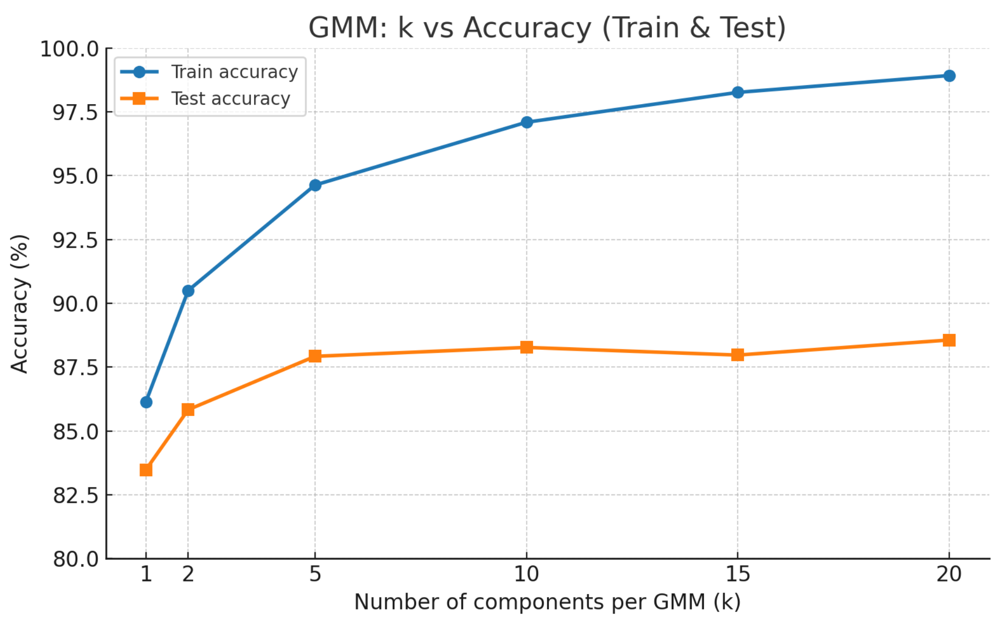


Figure 1: Train and Test Accuracy vs K plot

Question 4:

Classification: We use the already trained per-class GMM to predict the class of the image with missing pixels.

Reconstruction: For a test example with observed coordinates x_o and missing coordinates x_u, for each component k of predicted class c we compute the conditional mean (and use it in a mixture weighted by the posterior over components given x_o):

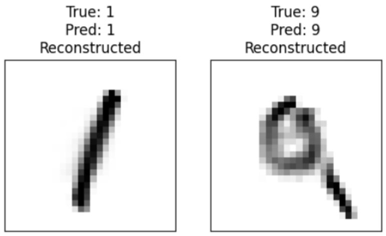
$$\mu^{(u)}_{c,k|o} = \mu^{(u)}_{c,k} + \Sigma^{(c,k)}_{u,o} \left(\Sigma^{(c,k)}_{o,o}\right)^{-1} \left(x_o - \mu^{(o)}_{c,k}\right).$$

The per-component log weight used to compute the posterior over components is:

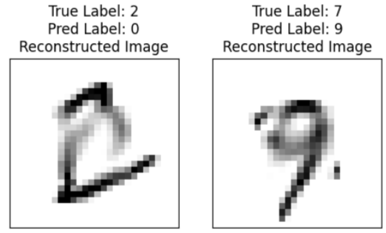
$$\log w_{c,k} = \log \pi_{c,k} + \log \mathcal{N}\left(x_o \mid \mu^{(o)}_{c,k}, \Sigma^{(c,k)}_{o,o}\right).$$

We compute $q_{c,k} = \mathrm{softmax}_k\left(\log w_{c,k}\right)$ and reconstruct the missing pixels by the mixture mean:

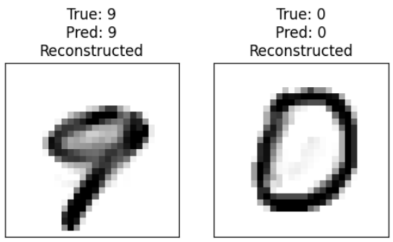
$$\widehat{x}_u = \sum_{k=1}^{K} q_{c,k}\, \mu^{(u)}_{c,k|o}.$$



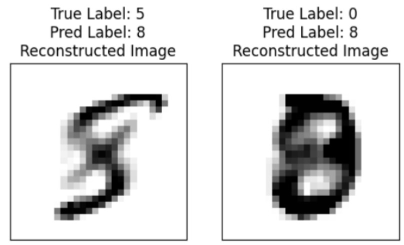
(a) K=1: Correct Classification



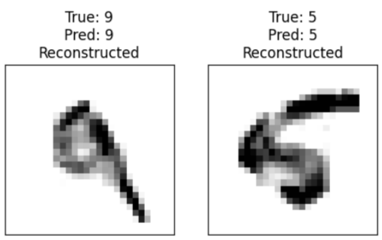
(b) K=1: Incorrect Classification



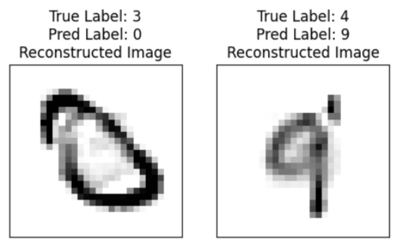
(c) K=2: Correct Classification



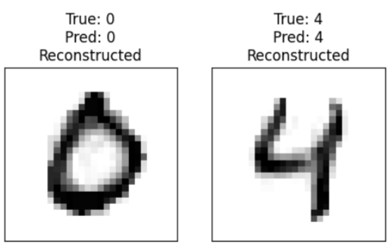
(d) K=2: Incorrect Classification



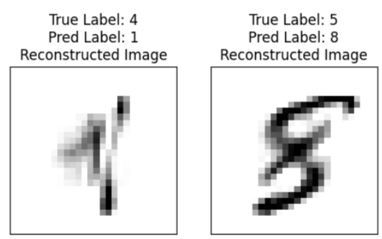
(e) K=5: Correct Classification



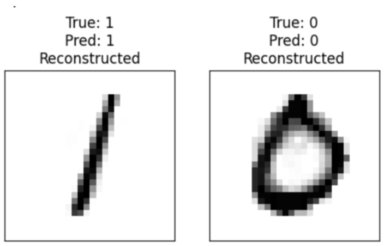
(f) K=5: Incorrect Classification



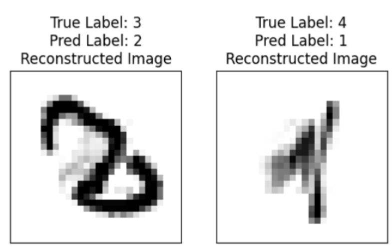
(g) K=10: Correct Classification



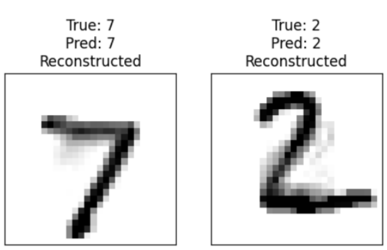
(h) K=10: Incorrect Classification



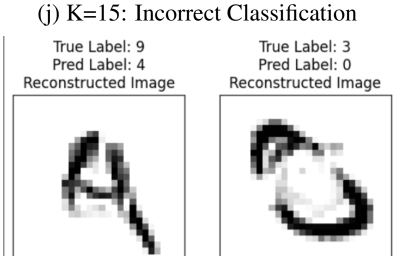
(i) K=15: Correct Classification



(j) K=15: Incorrect Classification



(k) K=20: Correct Classification



(l) K=20: Incorrect Classification

Figure 2: Comparison of Correct and Incorrect Classifications and Reconstruction for Censored images

K	**Test Accuracy (%)**
1	69.22
2	72.98
5	74.75
10	77.87
15	77.88
20	78.61

Table 2: Test Accuracy for Censored Images

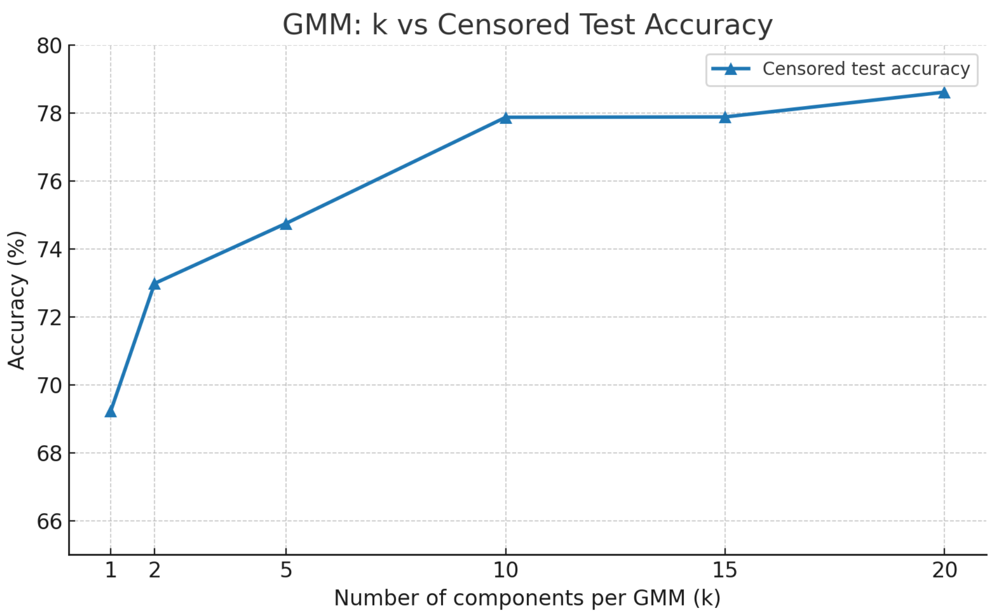


Figure 3: Test accuracy vs k plot in images with Missing Pixels

References

[1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[2] CS771 Lecture Material. Department of Computer Science and Engineering, IIT Kanpur, 2025. Available: `https://www.cse.iitk.ac.in/users/purushot/courses/ml/2025-26-a/lectures.html`