Supplementary Material for

# *Semi-Supervised Mixture Models under the Concept of Missing at Random with Margin Confidence and Aranda–Ordaz Function*

## S1 Margin Confidence Derivation

In binary classification, the posterior probability for the positive class can be expressed using the logistic function as:

$$\tau_{1j} = \frac{1}{1 + e^{-d_j}}.$$

To derive the margin confidence, consider the transformation:

$$2\tau_{1j} - 1 = 2 \cdot \frac{1}{1 + e^{-d_j}} - 1 = \frac{1 - e^{-d_j}}{1 + e^{-d_j}}.$$

Multiplying the numerator and the denominator by $e^{d_j/2}$ yields

$$\frac{e^{d_j/2} - e^{-d_j/2}}{e^{d_j/2} + e^{-d_j/2}} = \tanh\left(\frac{d_j}{2}\right).$$

Therefore,

$$2\tau_{1j} - 1 = \tanh\left(\frac{d_j}{2}\right),$$

and the margin confidence is defined as

$$\mathrm{MC}_j = |2\tau_{1j} - 1| = \tanh\left(\frac{|d_j|}{2}\right).$$

## S2 Entropy Approximation

In the binary classification setting, let $\tau \in (0, 1)$ denote the posterior probability of the positive class. The **Shannon entropy** is defined as:

$$H(\tau) = -\tau \log \tau - (1 - \tau) \log(1 - \tau).$$

Let $m = |2\tau - 1|$ denote the **margin confidence**. Shannon entropy $H(\tau)$ can be expressed equivalently as a function of $m$:

$$H(m) = -\frac{1+m}{2} \log\left(\frac{1+m}{2}\right) - \frac{1-m}{2} \log\left(\frac{1-m}{2}\right), \quad m \in [0, 1].$$

The symmetric property of the entropy function permits its reparameterization in terms of margin confidence, enabling a margin-based representation that is analytically tractable.
A Taylor series expansion of the entropy function $H(m)$ around $m = 0$ can be used to describe its local behavior. Since $H(m)$ is smooth and symmetric, the first derivative of $m = 0$ is zero. Therefore, the second-order approximation is given by:

$$H(m) = H(0) + \frac{1}{2}H''(0)m^2 + O(m^4).$$

The first and second derivatives of $H(m)$ with respect to $m$ are given by:

$$H'(m) = \frac{1}{2}\log\left(\frac{1-m}{1+m}\right), \quad H''(m) = \frac{1}{m^2-1}.$$

Let $m = 0$, and substitute the resulting values into the previous second-order approximation. This yields the following.

$$H(m) \approx \log 2 - \frac{1}{2}m^2.$$

This result suggests a quadratic inverse relationship between entropy and the square of the margin confidence, indicating that the margin confidence may serve as a computationally tractable proxy for entropy in uncertainty estimation.

## S3 Log-Entropy Approximation

From the second-order approximation

$$H(m) \approx \log 2 - \frac{1}{2}m^2,$$

taking logarithms yields

$$\log H(m) \approx \log\left(\log 2 - \frac{1}{2}m^2\right).$$

This expression can be rewritten as

$$\begin{aligned}\log H(m) &\approx \log\left(\log 2\left(1 - \frac{1}{2\log 2}m^2\right)\right)\\ &= \log(\log 2) + \log\left(1 - \frac{1}{2\log 2}m^2\right).\end{aligned}$$

Applying the Taylor expansion of $\log\left(1 - \frac{1}{2\log 2}m^2\right)$ around zero, the following asymptotic approximation is obtained:

$$\log H(m) \approx \log(\log 2) - \frac{1}{2\log 2}m^2, \quad \text{as } m \to 0.$$

## S4 ECM Algorithm

We apply the Expectation–Conditional Maximization (ECM) algorithm to maximize the observed-data log-likelihood under the MAR mechanism.

**E-step:** Given current parameters $\boldsymbol{\Theta}^{(t)} = (\boldsymbol{\theta}^{(t)}, \boldsymbol{\alpha}^{(t)}, \lambda^{(t)})$, compute the posterior responsibilities

$$\tau_{ij}^{(t)} = \begin{cases} z_{ij}, & m_j = 0, \\ \dfrac{\pi_i^{(t)} \mathcal{N}(\boldsymbol{y}_j; \boldsymbol{\mu}_i^{(t)}, \boldsymbol{\Sigma}_i^{(t)})}{\sum_{\ell=1}^{K} \pi_\ell^{(t)} \mathcal{N}(\boldsymbol{y}_j; \boldsymbol{\mu}_\ell^{(t)}, \boldsymbol{\Sigma}_\ell^{(t)})}, & m_j = 1, \end{cases}$$

together with the margin confidence $\delta_j^{(t)} = \tau_{(1)j}^{(t)} - \tau_{(2)j}^{(t)}$. The missingness probability is then evaluated as

$$q_j^{(t)} = 1 - \big(1 + \lambda^{(t)} e^{\eta_j^{(t)}}\big)^{-1/\lambda^{(t)}}, \qquad \eta_j^{(t)} = \alpha_0^{(t)} + \alpha_1^{(t)} \delta_j^{(t)2}.$$

**CM-step 1 (update $\boldsymbol{\theta}$):** Fix $(\boldsymbol{\alpha}^{(t)}, \lambda^{(t)})$ and update

$$\begin{aligned} \boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} \Bigg\{ & \sum_{j=1}^{n} (1 - m_j) \sum_{i=1}^{K} z_{ij} \big[ \log \pi_i + \log \mathcal{N}(\boldsymbol{y}_j; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \big] \\ & + \sum_{j=1}^{n} m_j \sum_{i=1}^{K} \tau_{ij}^{(t)} \big[ \log \pi_i + \log \mathcal{N}(\boldsymbol{y}_j; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \big] \\ & + \sum_{j=1}^{n} \big[ (1 - m_j) \log(1 - q_j) + m_j \log q_j \big] \Bigg\}, \end{aligned}$$

where $q_j$ depends on $(\boldsymbol{\alpha}^{(t)}, \lambda^{(t)})$ and the current $\delta_j(\boldsymbol{\theta})$. Maximization can be performed using quasi-Newton methods.

**CM-step 2 (update $(\boldsymbol{\alpha}, \lambda)$):** Fix $\boldsymbol{\theta}^{(t+1)}$ and update

$$(\boldsymbol{\alpha}^{(t+1)}, \lambda^{(t+1)}) = \arg\max_{\boldsymbol{\alpha}, \lambda} \sum_{j=1}^{n} \big[ (1 - m_j) \log(1 - q_j) + m_j \log q_j \big],$$

with $q_j = 1 - (1 + \lambda e^{\eta_j})^{-1/\lambda}$ and $\eta_j = \alpha_0 + \alpha_1 \delta_j(\boldsymbol{\theta}^{(t+1)})^2$. Maximization can be performed with Newton or quasi-Newton methods; if $\lambda$ is estimated, use a one-dimensional line search.

# S5 Code

All `R` scripts used in the numerical analysis are publicly available. These include code for data generation under MAR mechanisms, implementation of ECM estimators with Aranda–Ordaz and logit links, logistic regression baselines, and all simulation and real-data experiments.

The full repository can be accessed at:

`https://github.com/LeomusUNSW/IJCNN.git`

The repository enables full reproduction of all reported results and facilitates further experimentation.