End-to-End Bayesian Model with Attention Fusion

Let:

- M be the number of training subjects,
- N be the number of subgroups per subject,
- d be the dimension of each subgroup embedding,
- $\mathbf{x}_{ij} \in \mathbb{R}^d$ be the embedding for subject i and subgroup j, for $i = 1, \ldots, M$ and $j = 1, \ldots, N$.

1. Attention Weights

For each subject i, we model the attention weights over the N subgroups as:

$$\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iN}) \sim \text{Dirichlet}(\mathbf{1}),$$

where $\mathbf{1}$ is an N-dimensional vector of ones.

2. Fused Embedding

The fused embedding for subject i is computed as a weighted sum of the subgroup embeddings:

$$\tilde{\mathbf{x}}_i = \sum_{j=1}^{N} \alpha_{ij}\, \mathbf{x}_{ij}.$$

3. Classification Model (Logistic Regression)

We define a logistic regression model on the fused embeddings:

$$\text{logit}(\theta_i) = \beta^\top \tilde{\mathbf{x}}_i + b,$$

$$\theta_i = \sigma\left(\beta^\top \tilde{\mathbf{x}}_i + b\right) = \frac{1}{1 + \exp\left(-\left(\beta^\top \tilde{\mathbf{x}}_i + b\right)\right)},$$

with priors:

$$\beta \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad b \sim \mathcal{N}(0, 1).$$

4. Observation Model

The observed binary label y_i for subject i is modeled as:

$$y_i \sim \text{Bernoulli}(\theta_i).$$

Full Model

Putting it all together, for $i = 1, \ldots, M$:

$$\begin{aligned}
\boldsymbol{\alpha}_i &\sim \text{Dirichlet}(\mathbf{1}), \\
\tilde{\mathbf{x}}_i &= \sum_{j=1}^{N} \alpha_{ij}\, \mathbf{x}_{ij}, \\
\beta &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad b \sim \mathcal{N}(0, 1), \\
\text{logit}(\theta_i) &= \beta^\top \tilde{\mathbf{x}}_i + b, \quad \theta_i = \sigma\left(\beta^\top \tilde{\mathbf{x}}_i + b\right), \\
y_i &\sim \text{Bernoulli}(\theta_i).
\end{aligned}$$

This model jointly infers:

- The latent attention weights $\boldsymbol{\alpha}_i$ for each subject,
- The fused embedding $\tilde{\mathbf{x}}_i$ (a weighted sum of subgroup embeddings),
- The classification parameters β and b for logistic regression.