Mathematical Summary of the Model and Gradient Computations

Sarah Urbut

Notation and Indices

- $i \in \{1, \ldots, N\}$: Index for individuals.
- $k \in \{1, \ldots, K\}$: Index for topics.
- $d \in \{1, \ldots, D\}$: Index for diseases.
- $t \in \{0, \ldots, T-1\}$: Index for time points.

Model Definition

Latent Variables

1. **Individual-Topic Latent Variables:**

$$\lambda_{ik}(t) \sim \mathcal{GP}\left(\Gamma_k^\top g_i, \Sigma_k\right)$$

where
- $\lambda_{ik}(t)$: Topic k's influence on individual i at time t.
- Γ_k: Coefficient vector for genetic covariates g_i.
- g_i: Genetic covariate vector for individual i.
- Σ_k: Covariance function (kernel) for topic k.
2. **Topic-Disease Latent Variables:**

$$\phi_{kd}(t) \sim \mathcal{GP}\left(\mu_d, \Omega_k\right)$$

where
- $\phi_{kd}(t)$: Relationship between topic k and disease d at time t.
- μ_d: Mean function for disease d.
- Ω_k: Covariance function (kernel) for topic k.

Transformation Functions

1. **Softmax Function for Topic Proportions:**

$$\theta_{ik}(t) = \text{softmax}\left(\lambda_{ik}(t)\right) = \frac{\exp\left(\lambda_{ik}(t)\right)}{\sum_{j=1}^{K} \exp\left(\lambda_{ij}(t)\right)}$$

ensuring $\theta_{ik}(t) \in (0,1)$ and $\sum_{k=1}^{K} \theta_{ik}(t) = 1$.

2. **Sigmoid Function for Topic-Disease Probabilities:**

$$\phi_{\text{prob}}(k,d,t) = \text{sigmoid}\left(\phi_{kd}(t)\right) = \frac{1}{1+\exp\left(-\phi_{kd}(t)\right)}$$

mapping $\phi_{kd}(t)$ to a probability in $(0,1)$.

Disease Occurrence Probability

The probability that individual i develops disease d at time t is:

$$\pi_{id}(t) = \sum_{k=1}^{K} \theta_{ik}(t) \cdot \phi_{\text{prob}}(k,d,t)$$

Likelihood Function

We use a discrete-time survival model for disease occurrence.

Negative Log-Likelihood

For individual i and disease d:
- **Event Occurs at Time t:**

$$\ell_{id} = -\sum_{s=0}^{t-1} \ln\left(1-\pi_{id}(s)\right) - \ln\left(\pi_{id}(t)\right)$$

- **Censored Observation at Time $t_c \geq t$:**

$$\ell_{id} = -\sum_{s=0}^{t_c} \ln\left(1-\pi_{id}(s)\right)$$

Gradient Computations

Derivative of Negative Log-Likelihood with Respect to $\pi_{id}(t)$

- **For $s < t$ (before event time):**

$$\frac{\partial \ell_{id}}{\partial \pi_{id}(s)} = \frac{1}{1-\pi_{id}(s)}$$

- **At Event Time $s = t$:**

$$\frac{\partial \ell_{id}}{\partial \pi_{id}(t)} = -\frac{1}{\pi_{id}(t)}$$

- **For Censored Observations:**

$$\frac{\partial \ell_{id}}{\partial \pi_{id}(s)} = \frac{1}{1-\pi_{id}(s)}, \quad \forall s \leq t_c$$

Gradient with Respect to $\lambda_{ik}(t)$

Applying the chain rule:

$$\frac{\partial \ell_{id}}{\partial \lambda_{ik}(t)} = \sum_{s=0}^{t_{\text{obs}}} \frac{\partial \ell_{id}}{\partial \pi_{id}(s)} \cdot \frac{\partial \pi_{id}(s)}{\partial \lambda_{ik}(t)}$$

However, since $\pi_{id}(s)$ depends on $\lambda_{ik}(s)$, the derivative $\frac{\partial \pi_{id}(s)}{\partial \lambda_{ik}(t)}$ is non-zero only when $s = t$:

$$\frac{\partial \pi_{id}(s)}{\partial \lambda_{ik}(t)} = \begin{cases} \theta_{ik}(t) \left(\phi_{\text{prob}}(k, d, t) - \pi_{id}(t) \right), & \text{if } s = t \\ 0, & \text{if } s \neq t \end{cases}$$

Therefore, the gradient simplifies to:

$$\frac{\partial \ell_{id}}{\partial \lambda_{ik}(t)} = \frac{\partial \ell_{id}}{\partial \pi_{id}(t)} \cdot \theta_{ik}(t) \left(\phi_{\text{prob}}(k, d, t) - \pi_{id}(t) \right)$$

Gradient with Respect to $\phi_{kd}(t)$

Similarly, applying the chain rule:

$$\frac{\partial \ell_{id}}{\partial \phi_{kd}(t)} = \sum_{s=0}^{t_{\text{obs}}} \frac{\partial \ell_{id}}{\partial \pi_{id}(s)} \cdot \frac{\partial \pi_{id}(s)}{\partial \phi_{kd}(t)}$$

Since $\pi_{id}(s)$ depends on $\phi_{kd}(s)$, the derivative $\frac{\partial \pi_{id}(s)}{\partial \phi_{kd}(t)}$ is non-zero only when $s = t$:

$$\frac{\partial \pi_{id}(s)}{\partial \phi_{kd}(t)} = \begin{cases} \theta_{ik}(t) \cdot \phi_{\text{prob}}(k, d, t) \left(1 - \phi_{\text{prob}}(k, d, t) \right), & \text{if } s = t \\ 0, & \text{if } s \neq t \end{cases}$$

Therefore, the gradient simplifies to:

$$\frac{\partial \ell_{id}}{\partial \phi_{kd}(t)} = \frac{\partial \ell_{id}}{\partial \pi_{id}(t)} \cdot \theta_{ik}(t) \cdot \phi_{\text{prob}}(k, d, t) \left(1 - \phi_{\text{prob}}(k, d, t) \right)$$

Incorporating the Gaussian Process Priors

1. **For $\lambda_{ik}(t)$:**

$$\frac{\partial \ln p(\lambda_{ik})}{\partial \lambda_{ik}(t)} = - \left[\Sigma_k^{-1} \left(\lambda_{ik} - \Gamma_k^\top g_i \right) \right]_t$$

2. **For $\phi_{kd}(t)$:**

$$\frac{\partial \ln p(\phi_{kd})}{\partial \phi_{kd}(t)} = - \left[\Omega_k^{-1} \left(\phi_{kd} - \mu_d \right) \right]_t$$

Total Gradients

Summing the contributions from the likelihood and the priors:
1. **Total Gradient for $\lambda_{ik}(t)$:**

$$\frac{\partial \mathcal{L}}{\partial \lambda_{ik}(t)} = \frac{\partial \ell_{id}}{\partial \lambda_{ik}(t)} + \frac{\partial \ln p(\lambda_{ik})}{\partial \lambda_{ik}(t)}$$

2. **Total Gradient for $\phi_{kd}(t)$:**

$$\frac{\partial \mathcal{L}}{\partial \phi_{kd}(t)} = \frac{\partial \ell_{id}}{\partial \phi_{kd}(t)} + \frac{\partial \ln p(\phi_{kd})}{\partial \phi_{kd}(t)}$$

Key Points

- **Softmax Function Gradient:**

$$\frac{\partial \theta_{ik}(t)}{\partial \lambda_{ij}(t)} = \theta_{ik}(t)\left(\delta_{kj} - \theta_{ij}(t)\right)$$

  where δ_{kj} is the Kronecker delta.
- **Sigmoid Function Gradient:**

$$\frac{\partial \phi_{\text{prob}}(k,d,t)}{\partial \phi_{kd}(t)} = \phi_{\text{prob}}(k,d,t)\left(1 - \phi_{\text{prob}}(k,d,t)\right)$$

- **Chain Rule Applications:** - Gradients with respect to $\lambda_{ik}(t)$ and $\phi_{kd}(t)$ involve only time t in the likelihood term.
- **Gaussian Process Priors:** - The GP priors introduce dependencies across time points, affecting the gradients through the covariance matrices.