1 Loss function in *cHawk*

1.1 Intensity function

$$\lambda_d^i(t) = \boldsymbol{\mu}_d^\top \boldsymbol{f}_j^i + \sum_{t_j^i < t} \alpha_{d,d_j^i}\, g\left(t - t_j^i\right) \tag{1}$$

where f_j^i is the feature of the i-th patient at the j-th visit, α_{d,d_j^i} refers to the influence of disease d_j^i on disease d and we choose exponential kernel g as our decay kernel.

1.2 Log-likelihood

As a result, the probability of disease d in patient i is

$$L_d^i = f\left(t_{d,1}^i, t_{d,2}^i, \ldots, t_{d,k}^i\right) = \prod_{j=1}^{n} f^*\left(t_{d,j}^i \middle| \mathcal{H}\left(t_{j-1}\right)\right) \tag{2}$$

while we have the following equation as the probability of an event at time t is the intensity of the event times the probability of the event not happening before time t.

$$f^* = \lambda^*(1 - F^*) \tag{3}$$

By solving the differential equation, we have

$$f^*(t) = \lambda^*(t) \exp\left(-\int_{t_k}^{t} \lambda^*(u)\mathrm{d}u\right) \tag{4}$$

where t_k is the happening time of last event. Therefore the likelihood can be expressed as

$$L_d^i = \left[\prod_{j=1}^{n_d^i} \lambda^*\left(t_{d,j}^i\right)\right] \exp\left(-\int_{t_0}^{T} \lambda^*(u)\mathrm{d}u\right) \tag{5}$$

where t_0 and T represent the time span of whole process.

Taking logarithm on it, the result is

$$l_d^i = \log L_d^i = \sum_{j=1}^{n_d^i} \log \lambda^*\left(t_{d,j}^i\right) - \int_{t_0}^{T} \lambda^*\left(u\right)\mathrm{d}u \tag{6}$$

Sum up all i and d, the total log-likelihood is

$$L = \sum_i \sum_d \left\{\sum_{j=1}^{n_d^i} \log \lambda^*\left(t_{d,j}^i\right) - \int_{t_0}^{T} \lambda^*\left(u\right)\mathrm{d}u\right\} \tag{7}$$

With L^1 regularization on α and L^2 on μ, the loss function is written as

$$loss = L + R = L + \lambda_1 \|\boldsymbol{A}\|_1 + \frac{\lambda_2}{2} \sum_{d=1}^{D} \|\boldsymbol{\mu}_d\|_2^2 \tag{8}$$

where $\|\boldsymbol{A}\|_1$ refers to The sum of absolute values of each element of matrix $\boldsymbol{A}$.

2 Derivatives of loss function

2.1 Derivatives of L

Just consider derivatives of l_d^i and sum it up to get total derivatives.

$$\begin{aligned}
l_d^i =& \sum_{j=1}^{n_d^i} \log \lambda_d^i \left(t_{d,j}^i\right) - \int_{t_0}^{T} \lambda_d^i(t) \mathrm{d}t \\
=& \sum_{j=1}^{n_d^i} \log \lambda_d^i \left(t_{d,j}^i\right) - \int_{t_0}^{T} \left(\boldsymbol{\mu}_d^\top \boldsymbol{f}_j^i + \sum_{t_j^i < t} \alpha_{d,d_j^i} g\left(t - t_j^i\right) \right) \mathrm{d}t \\
=& \sum_{j=1}^{n_d^i} \log \lambda_d^i \left(t_{d,j}^i\right) - \boldsymbol{\mu}_d^\top \sum_{j=1}^{n^i} \boldsymbol{f_j^i} \left(t_j^i - t_{j-1}^i\right) - \int_{t_0}^{T} \sum_{t_j^i < t} \alpha_{d,d_j^i} g\left(t - t_j^i\right) \mathrm{d}t \\
=& \sum_{j=1}^{n_d^i} \log \lambda_d^i \left(t_{d,j}^i\right) - \boldsymbol{\mu}_d^\top \sum_{j=1}^{n^i} \boldsymbol{f_j^i} \left(t_j^i - t_{j-1}^i\right) - \sum_{j=1}^{n^i} \int_{t_{j-1}^i}^{t_j^i} \sum_{k=0}^{j-1} \alpha_{d,d_k^i} g\left(t - t_k^i\right) \mathrm{d}t \\
=& \sum_{j=1}^{n_d^i} \log \lambda_d^i \left(t_{d,j}^i\right) - \boldsymbol{\mu}_d^\top \sum_{j=1}^{n^i} \boldsymbol{f_j^i} \left(t_j^i - t_{j-1}^i\right) - \sum_{j=1}^{n^i} \sum_{k=0}^{j-1} \alpha_{d,d_k^i} \int_{t_{j-1}^i}^{t_j^i} g\left(t - t_k^i\right) \mathrm{d}t \\
=& \sum_{j=1}^{n_d^i} \log \lambda_d^i \left(t_{d,j}^i\right) - \boldsymbol{\mu}_d^\top \sum_{j=1}^{n^i} \boldsymbol{f_j^i} \left(t_j^i - t_{j-1}^i\right) - \sum_{j=1}^{n^i} \sum_{k=0}^{j-1} \alpha_{d,d_k^i} \left(G\left(t_j^i - t_k^i\right) - G\left(t_{j-1}^i - t_k^i\right)\right) \\
=& \sum_{j=1}^{n_d^i} \log \lambda_d^i \left(t_{d,j}^i\right) - \boldsymbol{\mu}_d^\top \sum_{j=1}^{n^i} \boldsymbol{f_j^i} \left(t_j^i - t_{j-1}^i\right) - \sum_{k=0}^{n^i-1} \sum_{j=k+1}^{n^i} \alpha_{d,d_k^i} \left(G\left(t_j^i - t_k^i\right) - G\left(t_{j-1}^i - t_k^i\right)\right) \\
=& \sum_{j=1}^{n_d^i} \log \lambda_d^i \left(t_{d,j}^i\right) - \boldsymbol{\mu}_d^\top \sum_{j=1}^{n^i} \boldsymbol{f_j^i} \left(t_j^i - t_{j-1}^i\right) - \sum_{k=0}^{n^i-1} \alpha_{d,d_k^i} \left(G\left(T - t_k^i\right) - G\left(0\right)\right) \\
=& \sum_{j=1}^{n_d^i} \log \lambda_d^i \left(t_{d,j}^i\right) - \boldsymbol{\mu}_d^\top \sum_{j=1}^{n^i} \boldsymbol{f_j^i} \left(t_j^i - t_{j-1}^i\right) - \sum_{k=0}^{n^i-1} \alpha_{d,d_k^i} G\left(T - t_k^i\right)
\end{aligned} \tag{9}$$

Therefore,

$$
\begin{aligned}
\frac{\partial L}{\partial \alpha_{ij}} &= \sum_i \sum_d \left\{ \frac{\sum_{t_k^i < t_j^i} g\left(t_j^i - t_k^i\right)}{\lambda_d^i\left(t_j^i\right)} + \sum_{d_k^i = d} G\left(T - t_k^i\right) \right\} \\
\frac{\partial L}{\partial \boldsymbol{\mu_d}} &= \sum_i \sum_d \left\{ \frac{\boldsymbol{f_j^i}}{\lambda_d^i(t_{d,j}^i)} - \sum_{j=1}^{n^i} \boldsymbol{f_j^i}\left(t_j^i - t_{j-1}^i\right) \right\}
\end{aligned} \tag{10}
$$

2.2 Derivatives of R

$$
\begin{aligned}
\frac{\partial R}{\partial \alpha_{ij}} &= \lambda 1 \times sig\left(\alpha_{ij}\right) \\
\frac{\partial R}{\partial \mu} &= \lambda_2 \boldsymbol{\mu}_d
\end{aligned} \tag{11}
$$