

1 Loss function in *cHawk*

1.1 Intensity function

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

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

$$L_d^i = f(t_{d,1}^i, t_{d,2}^i, \dots, t_{d,k}^i) = \prod_{j=1}^n f^*(t_{d,j}^i | \mathcal{H}(t_{j-1})) \quad (2)$$

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 .

$$f^* = \lambda^*(1 - F^*) \quad (3)$$

By solving the differential equation, we have

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

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

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

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

Taking logarithm on it, the result is

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

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

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

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

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

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

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

Therefore,

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

2.2 Derivatives of R

$$\begin{aligned}
\frac{\partial R}{\partial \alpha_{ij}} &= \lambda_1 \times sig(\alpha_{ij}) \\
\frac{\partial R}{\partial \boldsymbol{\mu}} &= \lambda_2 \boldsymbol{\mu}_d
\end{aligned} \tag{11}$$