Deep Hazard Analysis

March 4, 2020

1 Proportional Hazards from scratch from an ML perspective

Event density is equal to the hazard times the survival function. Leveraging that the cumulative hazard is related to the survival function: $S(t,x) = e^{-\Lambda(t,x)}$, we have:

$f(t,x) = \lambda(t,x) S(t,x) = \lambda(t,x) e^{-\Lambda(t,x)} = \lambda_0(t) e^{x^\top \beta} e^{-\int_0^t \lambda(u) du} = \lambda_0(t) e^{x^\top \beta} e^{-e^{x^\top \beta} \int_0^t \lambda_0(u) du}$

For a dataset without censorship, we have the following log likelihood:

$\log f(t_{1:N}, x_{1:N}) = \sum_{i=1}^N \log \lambda_0(t_i) + x_i^\top \beta - e^{x_i^\top \beta} \int_0^{t_i} \lambda_0(u) du$

Generalizing beyond linear models we have:

$\log f(t_{1:N}, x_{1:N}) = \sum_{i=1}^N \log \lambda_0(t_i) + g(x_i) - e^{g(x_i)} \int_0^{t_i} \lambda_0(u) du$

For censored cases, we will need to maximize the survival function instead of maximizing the log likelihood of the censored times. Due to the functional form here, it's just the last term. Defining C as the set of sample indices that are censored, D as the set of sample indices that are uncensored, and I_A as the indicator function for inclusion in set A, we have:

$\log f(t_{1:N}, x_{1:N}) = \sum_{i=1}^N I_D(i) \log \lambda_0(t_i) + I_D(i) g(x_i) - e^{g(x_i)} \int_0^{t_i} \lambda_0(u) du$

1.1 Non-proportional hazards

For non-proportional hazards, all that you would need to do is the following:

$\log f(t_{1:N}, x_{1:N}) = \sum_{i=1}^N \log \lambda_0(t_i, x_i) + g(x_i) - e^{g(x_i)} \int_0^{t_i} \lambda_0(u, x_i) du$

Although the requirement of a start time to align on is problematic, the inclusion of x_i in the arguments of the hazard function allow for the hazard going forward from any point in time to depend heavily on the patient's covariates.

1.2 Cox Model

To recover the original non-parametric cox model you would just specify a piecewise constant baseline hazard.

1.3 Monte Carlo approximation

$\log f(t_{1:N}, x_{1:N}) = \sum_{i=1}^N \log \lambda_0(t_i) + g(x_i) - e^{g(x_i)} \int_0^{t_i} \lambda_0(u) du$

$\log f(t_{1:N}, x_{1:N}) = N \mathbb{E}_{x_i, t_i \sim \hat{p}_{data}} \left[\log \lambda_0(t_i) + g(x_i) - e^{g(x_i)} \int_0^{t_i} \frac{p_U(u;t_i)}{p_U(u;t_i)} \lambda_0(u) du \right]$

$= N \mathbb{E}_{x_i, t_i \sim \hat{p}_{data}} \left[\log \lambda_0(t_i) + g(x_i) - e^{g(x_i)} \int_0^{t_i} p_U(u;t_i) \frac{\lambda_0(u)}{p_U(u;t_i)} du \right]$

$= N \mathbb{E}_{x_i, t_i \sim \hat{p}_{data}} \left[\log \lambda_0(t_i) + g(x_i) - e^{g(x_i)} \mathbb{E}_{p_U(u;t_i)} \left[\frac{\lambda_0(u)}{p_U(u;t_i)} \right] \right]$

$= N \mathbb{E}_{x_i, t_i \sim \hat{p}_{data}} \left[\log \lambda_0(t_i) + g(x_i) - e^{g(x_i)} \mathbb{E}_{p_U(u;t_i)} \left[\frac{\lambda_0(u)}{\frac{1}{t_i}} \right] \right]$

$= N \mathbb{E}_{x_i, t_i \sim \hat{p}_{data}} \left[\log \lambda_0(t_i) + g(x_i) - e^{g(x_i)} t_i \mathbb{E}_{p_U(u;t_i)} \left[\lambda_0(u) \right] \right]$

$= N \mathbb{E}_{x_i, t_i \sim \hat{p}_{data}} \left[\log \lambda_0(t_i) + g(x_i) - e^{g(x_i)} t_i \mathbb{E}_{p_U(u;t_i)} \left[\lambda_0(u) \right] \right]$

For a given minibatch of m samples from $\hat{p}_{data}$ and a minibatch s samples from $p_U(u;t_i)$, this (and its gradients) can be estimated with Monte Carlo approaches.

1.4 Joint Survival Analysis and Competing Risks

If there are several survival problems being analyzed simultaneously, one could treat it is a multivariate regression problem. Taking the non-proportional hazards model from above:

$\log f(t_{1:N}, x_{1:N}) = \sum_{i=1}^N \log \lambda_0(t_i, x_i) + g(x_i) - e^{g(x_i)} \int_0^{t_i} \lambda_0(u, x_i) du$

We can share parameters and model the survival problems jointly as long as we can align them. For example, birth is a univeral aligner.

$\log \boldsymbol{f}(t_{1:N}, x_{1:N}) = \sum_{i=1}^N \log \boldsymbol{\lambda_0}(t_i, x_i) + \boldsymbol{g}(x_i) - e^{\boldsymbol{g}(x_i)} \int_0^{t_i} \boldsymbol{\lambda_0}(u, x_i) du$

Boldface functions are vector valued. This is a joint distribution over a set of survival times.

Competing risks are easily incorporated into such a framework through establishing several simultaneous survival analysis problems where the failure of one competing risk censors the others.

1.5 Deep Hazard Analysis with History

Again starting from the non-proportional hazards model as:

$\log f\left(t_{1:N}, x_{1:N}\right) = \sum_{i=1}^{N} \log \lambda_0\left(t_i, x_i\right) + g\left(x_i\right) - e^{g\left(x_i\right)} \int_0^{t_i} \lambda_0\left(u, x_i\right) du$

We can formulate a time series variant by making all functions dependent on some "filtration" function of the history:

$\log f\left(t_{1:N}, x_{1:N,-\infty:0}\right) = \sum_{i=1}^{N} \log \lambda_0\left(t_i, h\left(x_{i,-\infty:0}\right)\right) + g\left(h\left(x_{i,-\infty:0}\right)\right) - e^{g\left(h\left(x_{i,-\infty:0}\right)\right)} \int_0^{t_i} \lambda_0\left(u, h\left(x_{i,-\infty:0}\right)\right) du$

1.5.1 Anytime Deep Hazard Analysis with History

We would like our model to perform well regardless of the start time that is chosen for each patient and therefore optimize the parameters that perform well in expectation over models with different start times:

$\mathbb{E}_{p_{1:N}(t^*)}\left[\log \prod_{i=1}^{N} f\left(t_i, x_i\right)\right]$

$= \sum_{i=1}^{N} \mathbb{E}_{p_n(t^*)}\left[\log \lambda_0\left(t^* + t_i, h\left(x_{i,-\infty:t^*}\right)\right) + g\left(h\left(x_{i,-\infty:t^*}\right)\right) - e^{g\left(h\left(x_{i,-\infty:t^*}\right)\right)} \int_0^{t_i} \lambda_0\left(t^* + u, h\left(x_{i,-\infty:t^*}\right)\right) du\right]$