## 0.1 Regularized AFT models

We use the log-logistic model. The CDF is given as follows:

$$F(t;\alpha,\beta) = \frac{1}{1+(t/\alpha)^{-\beta}} \tag{1}$$

The probability density function is given as:

$$d(t;\alpha,\beta) = \frac{((t/\alpha)^{-\beta}}{1+(t/\alpha)^{-\beta}} \tag{2}$$

Survival and hazard functions can be calulated from the above:

$$s(t;\alpha,\beta) = \frac{1}{1+(t/\alpha)^{\beta}} \tag{3}$$

$$h(t;\alpha,\beta) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{1+(t/\alpha)^{\beta}} \tag{4}$$

For survival analysis, given time intervals $\{t_{l_i}, t_{r_i}\}$ and covariates $x_i$ for $i = 1 : n$, we define two indicator variables $\delta_l$ and $\delta_r$ with the following interpretation:

$$(\delta_l, \delta_r) = \begin{cases} (0,0) & : \text{Interval censoring} \\ (0,1) & : \text{Left censoring} \\ (1,0) & : \text{Right censoring} \\ (1,1) & : \text{No censoring} \end{cases} \tag{5}$$

In the observations with no censoring, the pdf is used, and in censored

observations, the cdf is used. Hence the likelihood can be written as:

$$L(w) = \prod_{i=1}^{n} [d(t_{r_i})]^{\delta_{l_i}\delta_{r_i}} [F(t_{l_i})]^{(1-\delta_{l_i})\delta_{r_i}} [s(t_{r_i})]^{\delta_{l_i}(1-\delta_{r_i})} [s(t_{l_i}) - s(t_{r_i})]^{(1-\delta_{l_i})(1-\delta_{r_i})} \tag{6}$$

where, $\alpha_i = e^{w^T x_i}$

Substituting from the previous equations and simplyfing, we obtain the following log-likelihood:

$$\begin{gathered} l(w) = \sum_{i=1}^{n} \{\delta_{l_i}\delta_{r_i} \log(\beta/t_{l_i}) - \delta_{r_i} \log(1+(t_{r_i}/\alpha_i)^{\beta}) \\ -\delta_{l_i} \log((\alpha_i^{\beta} + t_{r_i}^{\beta})/t_{l_i}^{\beta}) + \log(\alpha_i^{\beta} + t_{r_i}^{\beta}) - \log(\alpha_i^{\beta} + t_{l_i}^{\beta})\} \end{gathered} \tag{7}$$

The partial derivative of the log-likelihood wrt to $\alpha_i$ is given as:

$$\frac{\partial l(w)_i}{\partial \alpha_i} = \delta_{r_i}\frac{\beta/\alpha_i}{1+(t_{r_i}/\alpha_i)^\beta} - \delta_{l_i}\frac{\beta/\alpha_i}{1+(t_{r_i}/\alpha_i)^\beta} + \frac{\beta/\alpha_i}{1+(t_{r_i}/\alpha_i)^\beta} - \frac{\beta/\alpha_i}{1+(t_{l_i}/\alpha_i)^\beta} \tag{8}$$

Hence, partial derivative wrt $w$ is:

$$\frac{\partial l(w)}{\partial w_j} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial w_j} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \alpha_i} x_{ij} e^{w^T x_i} \tag{9}$$

Adding the elastic net (L1 + L2) penalty, we get the following optimization objective:

$$f(w) = -l(w) + (\lambda_1\|w\|_1 + 1/2\lambda_2\|w\|_2^2) \tag{10}$$

The subderivative of the optimization objective is given as:

$$\frac{\partial f}{\partial w_j} = \begin{cases} \{-\frac{\partial l(w)}{\partial w_j} + \lambda_2 w_j - \lambda_1\} & \text{if } w_j < 0 \\ [-\frac{\partial l(w)}{\partial w_j} - \lambda_1, -\frac{\partial l(w)}{\partial w_j} + \lambda_1] & \text{if } w_j = 0 \\ \{-\frac{\partial l(w)}{\partial w_j} + \lambda_2 w_j + \lambda_1\} & \text{if } w_j > 0 \end{cases} \tag{11}$$

Using the subderivative, upto three cases of solutions for $w_j$ may be obtained. The coordinate descent algorithm works by cycling through each $w_j$ in turn, keeping the others constant, and using the above estimate to calculate the optimal value $w_j^*$. This is repeated until convergence.

**Sources:**


- Machine Learning: A Probabilistic Perspective by Kevin Murphy
- AFT: TD Hocking