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:

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

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 $\{\underline{t}_i, \bar{t}_i\}$ and covariates x_i for $i = 1:n$, the likelihood is given as

$$L(w) = \prod_{i=1}^{n} \zeta(w, \underline{t}_i, \bar{t}_i) \tag{5}$$

where,

$$\zeta(w, \underline{t}_i, \bar{t}_i) = \begin{cases} d(t_i, \alpha_i, \beta) & \text{if: } -\infty < \underline{t}_i = \bar{t}_i < \infty \\ F(\bar{t}_i, \alpha_i, \beta) & \text{if: } -\infty = \underline{t}_i \ , \ \bar{t}_i < \infty \\ s(\underline{t}_i, \alpha_i, \beta) & \text{if: } -\infty < \underline{t}_i \ , \ \bar{t}_i = \infty \\ F(\bar{t}_i, \alpha_i, \beta) - F(\underline{t}_i, \alpha_i, \beta) & \text{if: } -\infty < \underline{t}_i \neq \bar{t}_i < \infty \end{cases} \tag{6}$$

where, $\alpha_i = e^{w^T x_i}$.
In the observations with no censoring, the pdf is used, and in censored observations, the cdf is used.
Hence the log likelihood is given as

$$l(w) = \sum_{i=1}^{n} \log \zeta(w, \underline{t}_i, \bar{t}_i) = \sum_{i=1}^{n} \ell_i \tag{7}$$

From 6 and 7, the partial derivative of the log-likelihood wrt to w_j is given as:

$$\frac{\partial l(w)}{\partial w_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial w_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \alpha_i} x_{ij} e^{w^T x_i} \tag{8}$$

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

$$f(w) = -l(w) + (\lambda_1 \|w\|_1 + 1/2\lambda_2 \|w\|_2^2) \tag{9}$$

The subderivative of the optimization objective is given as:

$$\frac{\partial f}{\partial w_j} = \begin{cases} \{-\frac{\partial l(w)}{\partial w_j} + \lambda_2 w_j - \lambda_1\} & \text{if } w_j < 0 \\ [-\frac{\partial l(w)}{\partial w_j} - \lambda_1, -\frac{\partial l(w)}{\partial w_j} + \lambda_1] & \text{if } w_j = 0 \\ \{-\frac{\partial l(w)}{\partial w_j} + \lambda_2 w_j + \lambda_1\} & \text{if } w_j > 0 \end{cases} \tag{10}$$

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