Induction of Survival Trees by Quadric Splits and Dipolar Splitting Criteria

Aye Aye Maung[1,1*] and Drew Lazar[2,2]

[1*] Bioinformatics & Biostatistics, University of Louisville, 485 E. Gray St., Louisville, 40202, KY, USA.
[2]Mathematical Sciences, Ball State University, Robert Bell Building, Room 465, Muncie, 47306, IN, USA.

*Corresponding author(s). E-mail(s): a0maun01@louisville.edu; Contributing authors: dmlazar@bsu.edu;

Abstract

In recent decades, a variety of statistical models and supervised learning algorithms have been extended to predict survival outcomes. In particular, classification and regression trees (CART) have been modified to account for right-censored survival data. As in CART, sample data can be split univariately by considering covariates one at a time, obliquely with hyperplanes or non-linearly with curved surfaces. In this paper, we consider an existing non-parametric method which implements oblique splits and uses criteria based on pairs of covariates, and we extend this method to splits by quadratic surfaces. Data augmentation and optimization of combinations of piecewise-linear functions are used to find optimal quadric splits. Survival trees based on univariate splits and the log-rank statistic are also implemented. We compare the structure and predictive power of survival trees created by such trees on real and simulated data sets.

Keywords: survival analysis, regression trees, non-linear trees, right-censored data, dipolar criteria, non-parametric statistics

1 Introduction

Two seminal works are credited with forming the basis for the development of survival analysis in the 20th century. The first was the work of Kaplan and Meier [1], where they introduced the Kaplan-Meier estimates as a non-parametric method of estimating survival probabilities. The second was the proportional hazard model of Cox [2], a widely used semi-parametric method of estimating survival times assuming the proportional hazards assumption. Many more models and approaches from statistics have been adapted to accommodate survival analysis which often involves the presence of censored observations [3]. Survival analysis, in turn, has found wide application in finance, epidemiology, and insurance among many others.

In the last 40 years, tree-based survival analysis methods, which provide interpretable predictions without requiring the strong assumptions of parametric modeling techniques, have been developed. The first such model was introduced in [4] and further development was presented in [5] splitting nodes to attain the smallest amount of inter-node variability as measured by the Wasserstein metric. From there, many types of survival trees, mostly differing in the splitting criteria, were developed. These include univariate splits based on the log-rank statistic in [6], splits based on likelihood ratio statistic and the assumption of constant hazards of the exponential model [7], and completely non-parametric, oblique splits by hyperplanes in [8], among many others. This paper focuses on the improvement and extension of the methods introduced in [8] to accommodate non-linear splits.

2 Background

2.1 Survival Data

We assume that N observations, $(\mathbf{x}_i, t_i, \delta_i)_{i=1}^{N}$, are sampled from a random triple $(\mathbf{X}, T, \Delta)$. $\mathbf{X}$ is a D-dimensional random covariate vector,

$$T = \min(T_0, C)$$

where T_0 is a random variable indicating survival time, C is a random variable indicating right-censoring time and $\Delta = I(T_0 < C)$ is a censoring indicator.

2.2 Survival Dipolar Criterion

In [8], a non-parametric, tree-based method of separating the covariate space of right-censored data by hyperplanes was proposed which extended the approach of [9] to survival analysis. In both works, it is necessary to classify pairs of covariate vectors

$$\{(\mathbf{x}_i, \mathbf{x}_j)\}_{1 \leq i < j \leq N}$$

known as *dipoles* according to their time-difference information. Roughly speaking the aim is to label dipoles with small time differences as *pure* and to label dipoles with large time differences as *mixed*. We use the approach of [8] where, a vector of pairwise time differences ΔT is constructed as follows:

1. Initialize ΔT as the empty vector
2. For each pair of time-censorship tuples from $\left\{\left((t_i, \delta_i), (t_j, \delta_j)\right)\right\}_{1 \leq i < j \leq N}$ append $|t_i - t_j|$ to ΔT if $\delta_k = 1$, where $k = \arg\min(t_i, t_j)$

Next reals $0 < \eta < \zeta < 1$ are fixed as lower and upper percentile cutoffs of ΔT for determining *pure* and *mixed* dipoles. Letting $|\Delta T| = L$ the floored products $\lfloor \eta \cdot L \rfloor$ and $\lfloor \zeta \cdot L \rfloor$ approximate the η-th and ζ-th quantiles of ΔT. Finally the survival dipolar criterion for $\{(\mathbf{x}_i, \mathbf{x}_j)\}_{1 \leq i < j \leq N}$ is defined:

1. $(\mathbf{x}_i, \mathbf{x}_j)$ is *pure* if $\delta_i = \delta_j = 1$ and $|t_i - t_j| < \Delta T_{(\lfloor \eta \cdot L \rfloor)}$
2. $(\mathbf{x}_i, \mathbf{x}_j)$ is *mixed* if $|t_i - t_j| \geq \Delta T_{(\lfloor \zeta \cdot L \rfloor)}$ and $\delta_k = 1$, where $k = \arg\min(t_i, t_j)$
3. All other $(\mathbf{x}_i, \mathbf{x}_j)$ are *neither*

2.3 Piecewise-Linear Criterion Functions

With dipoles labeled [8] splits the covariate space at nodes by hyperplanes intended to split many mixed dipoles while splitting few pure dipoles. To achieve this [8] employs linear combinations of special piecewise-linear functions which are specified as follows.

We specify any hyperplane

$$H(\mathbf{v}) = \left\{(x_1, \ldots, x_D)^T \in \mathbb{R}^D;\ \mathbf{v}^T\mathbf{z} = 0 \text{ where } \mathbf{z} = (1, x_1, \ldots, x_D)^T\right\}$$

by its coefficient vector $\mathbf{v} = (-\theta, w_1, \ldots, w_D)^T \in \mathbb{R}^{D+1}$. The *augmented vector* of a covariate vector $\mathbf{x}_j$ is $\mathbf{z}_j = (1, \mathbf{x}_j^T)^T$ and the *augmented dipole* of a dipole $(\mathbf{x}_j, \mathbf{x}_k)$ is $(\mathbf{z}_j, \mathbf{z}_k)$.

Now consider the piecewise-linear functions $\varphi_j^+, \varphi_j^- : \mathbb{R}^{D+1} \to \mathbb{R}$:

$$\varphi_j^+(\mathbf{v}) = \max\{0, \varepsilon_j - \mathbf{v}^T\mathbf{z}_j\}, \quad \varphi_j^-(\mathbf{v}) = \max\{0, \varepsilon_j + \mathbf{v}^T\mathbf{z}_j\} \tag{1}$$

For $\mathbf{x}_j = (x_{1,j}, \ldots, x_{D,j})^T$, $\mathbf{z}_j = (1, \mathbf{x}_j^T)^T$, $\mathbf{w} = (w_1, \ldots, w_D)^T$, $\mathbf{v} = (-\theta, \mathbf{w}^T)^T$

$$\mathbf{v}^T\mathbf{z}_j = \|\mathbf{w}\|\|\mathbf{d}_j\|\text{sgn}(o_j)$$

where $\mathbf{d}_j = \frac{\mathbf{v}^T\mathbf{z}_j}{\|\mathbf{w}\|^2}\mathbf{w}$ is the shortest vector from $H(\mathbf{v})$ to $\mathbf{x}_j$ and o_j is 1 if $\mathbf{x}_j$ is orientated "above" $H(\mathbf{v})$ in the direction of $\mathbf{w}$ and o_j is -1 if $\mathbf{x}_j$ is orientated "below" $H(\mathbf{v})$ in the opposite direction of $\mathbf{w}$. Thus the ε_j's are called *margins* and following [9] we assume that

$$\varepsilon_j = \varepsilon \text{ uniformly for all } j = 1, \ldots, N \tag{2}$$

for some fixed $\varepsilon > 0$.

The piecewise-linear functions φ_j^+ and φ_j^- are shown in Figure 1.

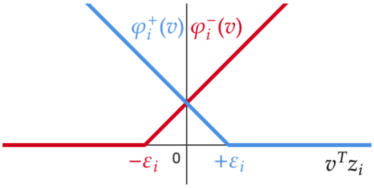

Fig. 1: Piecewise-Linear Criterion Functions

2.4 Dipole Penalty Functions

Next [8] combines the functions in (1) to define two pairs of penalty functions. The pair of functions

$$\varphi_{jk}^{m^+} = \varphi_j^+ + \varphi_k^- \text{ and } \varphi_{jk}^{m^-} = \varphi_j^- + \varphi_k^+ \tag{3}$$

penalize mixed dipoles that remain unsplit. Given a dipole $(\mathbf{x}_j, \mathbf{x}_k)$, each of these functions is minimized by a $\mathbf{v}$ which splits the dipole. On the other hand, the pair of functions

$$\varphi_{jk}^{p^+} = \varphi_j^+ + \varphi_k^+ \text{ and } \varphi_{jk}^{p^-} = \varphi_j^- + \varphi_k^- \tag{4}$$

penalize the splitting of pure dipoles. Given a dipole $(\mathbf{x}_j, \mathbf{x}_k)$, each of these functions is minimized by a $\mathbf{v}$ which does not split the dipole.

2.5 Dipolar Criterion Functions

Finally, using a weighted sum of penalty functions from (3) and (4), [8] constructs an objective function that is minimized at a $\mathbf{v}$ for which many mixed dipoles are split and many pure dipoles are not split by $H(\mathbf{v})$. In optimizing the objective function, $\varphi_{jk}^{m^+}$ or $\varphi_{jk}^{m^-}$ from (3) are assigned to each mixed dipole and $\varphi_{jk}^{p^+}$ or $\varphi_{jk}^{p^-}$ from (4) are assigned to each pure dipole.

While [8] states assignments are made according to the "orientation" of the dipoles, each dipole and its two elements do not intrinsically possess "orientation" without respect to a particular $\mathbf{v}$. At the same time [9] chooses $\varphi_{jk}^{m^+}$ for all mixed dipoles and chooses $\varphi_{jk}^{p^+}$ for all pure dipoles. In this paper, and as described in section 3.1, functions from (3) and (4) are assigned to dipoles in a geometrically reasonable manner with respect to initial $\mathbf{v}$'s at each step of the optimization algorithm and the dipolar criterion function is minimized with respect to those initial $\mathbf{v}$'s.

For now we introduce the dipolar criterion function presented in [8]. We can define dipoles as having "positive orientation" if $\varphi_{jk}^{m^+}$ or $\varphi_{jk}^{p^+}$ are assigned to them and as having "negative orientation" if $\varphi_{jk}^{m^-}$ or $\varphi_{jk}^{p^-}$ are assigned to them. Following [8] let $I^{p^+}, I^{p^-}, I^{m^+}, I^{m^-}$ be the disjoint sets of pairs of indices of dipoles that are respectively: pure with positive orientation, pure with negative orientation, mixed with positive orientation and mixed with negative orientation. Then the dipolar criterion function is

$$\Psi = \sum_{\substack{(j,k) \\ \in I^{p^+}}} \alpha_{jk}\varphi_{jk}^{p^+} + \sum_{\substack{(j,k) \\ \in I^{p^-}}} \alpha_{jk}\varphi_{jk}^{p^-} + \sum_{\substack{(j,k) \\ \in I^{m^+}}} \alpha_{jk}\varphi_{jk}^{m^+} + \sum_{\substack{(j,k) \\ \in I^{m^-}}} \alpha_{jk}\varphi_{jk}^{m^-} \tag{5}$$

The coefficients $\alpha_{jk} \geq 0$ are "price" factors of penalty functions. These can be fixed to prioritize splitting mixed dipoles or to prioritize not splitting pure dipoles depending on the relative sizes of the coefficients.

3 Methods

3.1 Orientation and Choice Dipole Penalty Functions

In this section, we specify how we assign "orientation" to dipoles with respect to a given hyperplane $\mathbf{v}^*$. Functions from (3) are assigned to mixed dipoles and functions from (4) are assigned to pure dipoles with respect to a $\mathbf{v}^*$ based on these orientations.

As in section 2.3, let $\mathbf{v}^*$ be a coefficient vector and $(\mathbf{z}_j, \mathbf{z}_k)$ be an augmented dipole.

1. Let $(\mathbf{z}_j, \mathbf{z}_k)$ be pure. It has *positive orientation* if ${\mathbf{v}^*}^T(\mathbf{z}_j + \mathbf{z}_k) \geq 0$ and in this case we assign $\varphi_{jk}^{p^+}$ to it. It has *negative orientation* if ${\mathbf{v}^*}^T(\mathbf{z}_j + \mathbf{z}_k) \leq 0$ and in this case we assign $\varphi_{jk}^{p^-}$ to it.
2. Let $(\mathbf{z}_j, \mathbf{z}_k)$ be mixed. It has a *positive orientation* if ${\mathbf{v}^*}^T(\mathbf{z}_j - \mathbf{z}_k) \geq 0$ and in this case we assign $\varphi_{jk}^{m^+}$ to it. It has a *negative orientation* if ${\mathbf{v}^*}^T(\mathbf{z}_j - \mathbf{z}_k) \leq 0$ and in this case we assign $\varphi_{jk}^{m^-}$ to it.

The index sets of the dipole criterion function (5) now depend on $\mathbf{v}^*$: $I_{\mathbf{v}^*}^{p^+}, I_{\mathbf{v}^*}^{p^-}, I_{\mathbf{v}^*}^{m^+}, I_{\mathbf{v}^*}^{m^-}$. In our algorithm to find a splitting hyperplane, we therefore optimize the function

$$\Psi_{\mathbf{v}^*} = \sum_{\substack{(j,k) \\ \in I_{\mathbf{v}^*}^{p^+}}} \alpha_{jk}\varphi_{jk}^{p^+} + \sum_{\substack{(j,k) \\ \in I_{\mathbf{v}^*}^{p^-}}} \alpha_{jk}\varphi_{jk}^{p^-} + \sum_{\substack{(j,k) \\ \in I_{\mathbf{v}^*}^{m^+}}} \alpha_{jk}\varphi_{jk}^{m^+} + \sum_{\substack{(j,k) \\ \in I_{\mathbf{v}^*}^{m^-}}} \alpha_{jk}\varphi_{jk}^{m^-} \tag{6}$$

Below we justify our assignments of penalty functions based on orientation with respect to a particular $\mathbf{v}^*$.

Proposition 3.1

(a) If ${\mathbf{v}^*}^T(\mathbf{z}_j + \mathbf{z}_k) \geq 0$*, then* $\varphi_{jk}^{p^-}(\mathbf{v}^*) \geq \varphi_{jk}^{p^+}(\mathbf{v}^*)$.

(b) If ${\mathbf{v}^*}^T(\mathbf{z}_j + \mathbf{z}_k) \leq 0$*, then* $\varphi_{jk}^{p^+}(\mathbf{v}^*) \geq \varphi_{jk}^{p^-}(\mathbf{v}^*)$.

(c) If ${\mathbf{v}^*}^T(\mathbf{z}_j - \mathbf{z}_k) \geq 0,$ *then* $\varphi_{jk}^{m^-}(\mathbf{v}^*) \geq \varphi_{jk}^{m^+}(\mathbf{v}^*)$.
(d) If ${\mathbf{v}^*}^T(\mathbf{z}_j - \mathbf{z}_k) \leq 0,$ *then* $\varphi_{jk}^{m^+}(\mathbf{v}^*) \geq \varphi_{jk}^{m^-}(\mathbf{v}^*)$.

Proof We show (a) with (b)-(d) shown similarly. Assuming $\varepsilon_j = \varepsilon_k = \varepsilon$ as in (2),

$$\begin{aligned}{\mathbf{v}^*}^T(\mathbf{z}_j + \mathbf{z}_k) \geq 0 &\implies \varepsilon + {\mathbf{v}^*}^T\mathbf{z}_j + {\mathbf{v}^*}^T\mathbf{z}_k \geq \varepsilon \\ &\implies \varepsilon_j + {\mathbf{v}^*}^T\mathbf{z}_j \geq \varepsilon_k - {\mathbf{v}^*}^T\mathbf{z}_k \text{ and } \varepsilon_k + {\mathbf{v}^*}^T\mathbf{z}_k \geq \varepsilon_j - {\mathbf{v}^*}^T\mathbf{z}_j\end{aligned}$$

Thus

$$\begin{aligned}\varphi_j^-(\mathbf{v}^*) &= \max\{0, \varepsilon_j + {\mathbf{v}^*}^T\mathbf{z}_j\} \geq \max\{0, \varepsilon_k - {\mathbf{v}^*}^T\mathbf{z}_k\} = \varphi_k^+(\mathbf{v}^*) \\ \varphi_k^-(\mathbf{v}^*) &= \max\{0, \varepsilon_k + {\mathbf{v}^*}^T\mathbf{z}_k\} \geq \max\{0, \varepsilon_j - {\mathbf{v}^*}^T\mathbf{z}_j\} = \varphi_j^+(\mathbf{v}^*)\end{aligned}$$

implying that

$$\varphi_{jk}^{p^-}(\mathbf{v}^*) = \varphi_j^-(\mathbf{v}^*) + \varphi_k^-(\mathbf{v}^*) \geq \varphi_k^+(\mathbf{v}^*) + \varphi_j^+(\mathbf{v}^*) = \varphi_j^+(\mathbf{v}^*) + \varphi_k^+(\mathbf{v}^*) = \varphi_{jk}^{p^+}(\mathbf{v}^*)$$

□

The sizes of the penalties $\varphi_{jk}^{p^+}(\mathbf{v}^*)$ and $\varphi_{jk}^{p^-}(\mathbf{v}^*)$ for a pure positively-oriented dipole are shown in Figure 2. $\|\mathbf{w}^*\| = 1$ is assumed for simplicity.

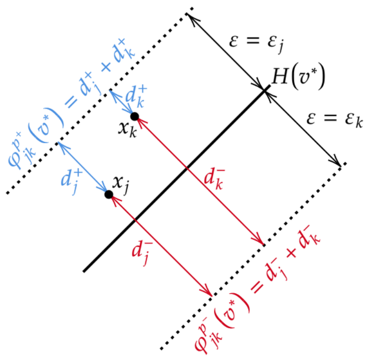

Fig. 2: Penalties for a Pure Positively-Oriented Dipole

In other words choosing $\varphi_{jk}^{p^+}$ over $\varphi_{jk}^{p^-}$, for a pure positively oriented dipole, penalizes our guess of the splitting hyperplane $\mathbf{v}^*$ less. This is what we want if $\mathbf{v}^*$ was close to the optimal to begin with. Of course in general the preliminary guess $\mathbf{v}^*$ would not be optimal. In section 3.3 we describe an algorithm to get rid of this dependence on a preliminary guess.

3.2 Optimization of Dipolar Criterion Functions

3.3 Re-Orientation Algorithm

3.4 Quadric Augmentation of Covariates

4 Data

5 Results

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

6 Conclusions

Declarations

Some journals require declarations to be submitted in a standardised format. Please check the Instructions for Authors of the journal to which you are submitting to see if you need to complete this section. If yes, your manuscript must contain the following sections under the heading 'Declarations':

- Funding
- Conflict of interest/Competing interests (check journal-specific guidelines for which heading to use)
- Ethics approval
- Consent to participate
- Consent for publication
- Availability of data and materials
- Code availability
- Authors' contributions

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

Editorial Policies for:

Springer journals and proceedings:
https://www.springer.com/gp/editorial-policies

Nature Portfolio journals:
https://www.nature.com/nature-research/editorial-policies

Scientific Reports:
https://www.nature.com/srep/journal-policies/editorial-policies

BMC journals:
https://www.biomedcentral.com/getpublished/editorial-policies

Springer Nature 2021 LaTeX template

Appendix A Section title of first appendix

An appendix contains supplementary information that is not an essential part of the text itself but which may be helpful in providing a more comprehensive understanding of the research problem or it is information that is too cumbersome to be included in the body of the paper.

References

[1] Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. Journal of the American Statistical Association **53**(282), 457–481 (1958) https://arxiv.org/abs/https://www.tandfonline.com/doi/pdf/10.1080/01621459.1958.10501452. https://doi.org/10.1080/01621459.1958.10501452

[2] Cox, D.R.: Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological) **34**(2), 187–220 (1972)

[3] Oakes, D.: Biometrika centenary: survival analysis. Biometrika **88**(1), 99–142 (2001)

[4] Ciampi, A., Bush, R.S., Gospodarowicz, M., Till, J.E.: An approach to classifying prognostic factors related to survival experience for non-hodgkin's lymphoma patients: Based on a series of 982 patients: 1967–1975. Cancer **47**(3), 621–627 (1981)

[5] Gordon, L., Olshen, R.A.: Tree-structured survival analysis. Cancer treatment reports **69**(10), 1065–1069 (1985)

[6] LeBlanc, M., Crowley, J.: Survival trees by goodness of split. Journal of the American Statistical Association **88**(422), 457–467 (1993)

[7] Bou-Hamad, I., Larocque, D., Ben-Ameur, H.: A review of survival trees. Statistics surveys **5**, 44–71 (2011)

[8] Kretowska, M.: Piecewise-linear criterion functions in oblique survival tree induction. Artificial Intelligence in Medicine **75**, 32–39 (2017). https://doi.org/10.1016/j.artmed.2016.12.004

[9] Bobrowski, L., Kretowski, M.: Induction of multivariate decision trees by using dipolar criteria. In: Lecture Notes in Computer Science, vol. 1910, pp. 331–336 (2000). https://doi.org/10.1007/3-540-45372-5_33