= previous work

= our contributions

Optimal Survival Trees With the Iterative Breslow Estimator and the Integrated Brier Score Objective

Background

1

We use dynamic programming to create the tree with the minimum training loss for a maximum depth

$$T(D,d) = \begin{cases} \min\limits_{\hat{\theta}} \mathfrak{L}(D,\hat{\theta}), & d = 0 \\ \min\limits_{f} [T(D(f), d-1) + T(D(\overline{f}, d-1)], & d > 0 \end{cases}$$

S(t) = P(T≥t), where T is true time-of-event

d(t) = deaths at time t

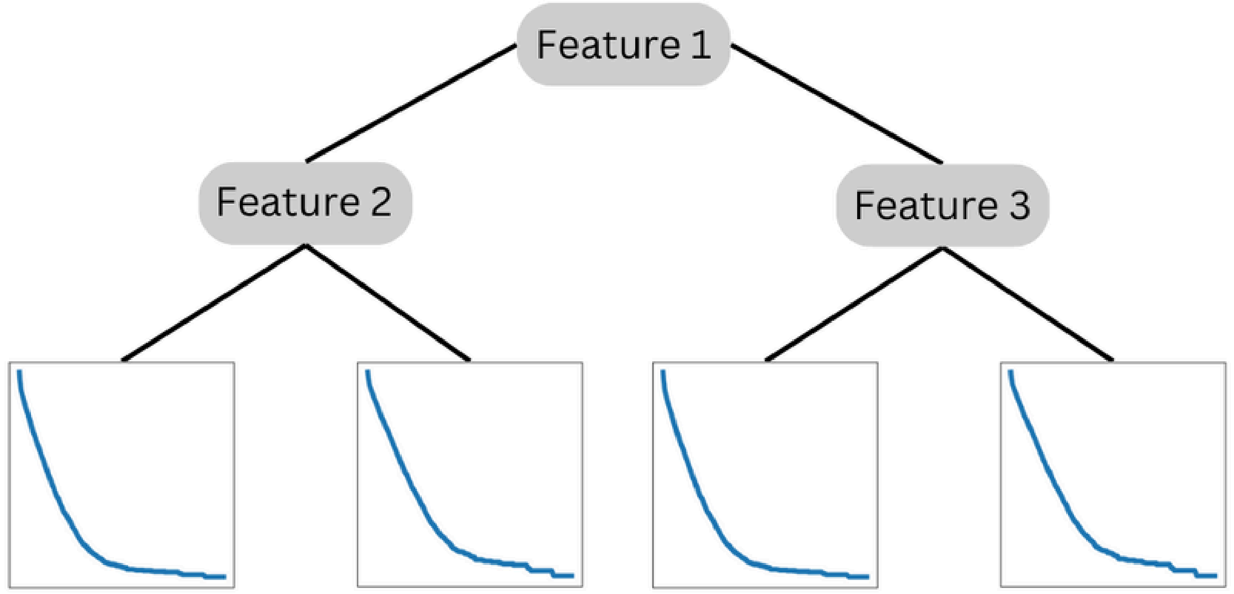
n(t) = number of survivors up to and until time t

Λ(t) = cumulative hazard function (CHF)

Λ_0(t) = baseline cumulative hazard function

θ = risk coefficient

IBS = Integrated Brier Score

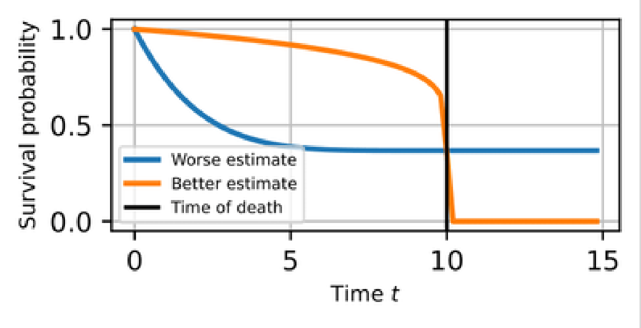


Partial likelihood objective

4

The partial likelihood objective function calculates the loss by taking the difference between the log likelihoods of the estimated risk coefficient and the saturated risk coefficient:

$$\theta_i^{sat} = \frac{\delta_i}{\hat{\Lambda}_0(t_i)}$$



The partial likelihood loss judges both survival functions equally on a dataset with instances that all die at the same time. This is a flaw.

Estimating the survival function

$$\hat{S}_i(t) = e^{-\hat{\Lambda}_i(t)}$$

$$\Lambda_i(t) = \theta_i \Lambda_0(t)$$

Risk coefficients are estimated in leaves. The baseline CHF is the same for entire tree

The Nelson-Aalen estimator estimates the CHF by weighing every instance equally:

$$\hat{\Lambda}_0(t) = \sum_{t' \leq t} \frac{d(t')}{n(t')}$$

2

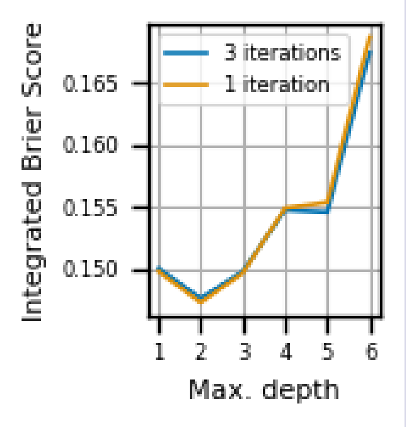
Iterative Breslow Estimator

3

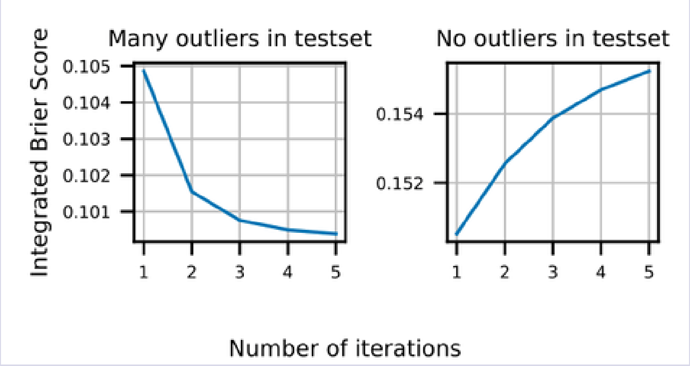
We estimate CHF iteratively with:

$$\hat{\Lambda}_0^k(t) = \sum_{t_j \leq t} \frac{d(t)}{\sum_{i \in D, t_i \geq t} \theta_i^k}$$

$$\theta_D^{k+1} = \frac{\sum_{i \in D} \delta_i}{\sum_{i \in D} \Lambda_0^k}$$



IBS for 1 and 3 Breslow iterations



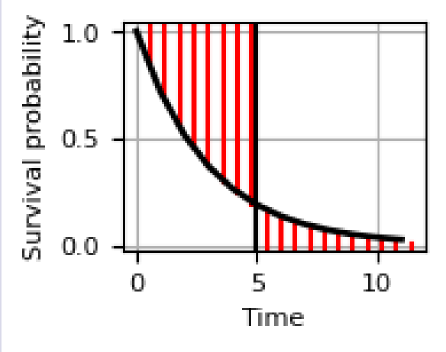
Iterative Breslow estimator on artificial data

Integrated Brier Score objective

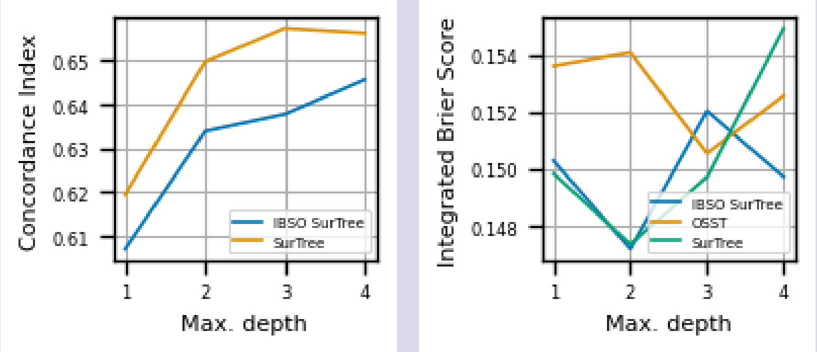
5

We find the tree with the optimal IBS on training data:

$$IBS = \frac{1}{t_{max}} \frac{1}{|D|} \sum_{i \in D} \int_0^{t_i} \frac{(1-\hat{S}_i(t))^2}{\hat{G}(t)} dt + \delta_i \int_{t_i}^{t_{max}} \frac{(\hat{S}_i(t))^2}{\hat{G}(t_i)} dt$$



Intuition of the area of loss for the IBS



Comparison of IBS and Concordance Index on test data