Cox SurTree: A Semi-Parametric Approach to Optimal Survival Trees

Author: Matei Mihai Mirica
(M.M.Mirica@student.tudelft.nl)
Supervisor: Jacobus G. M. van der Linden
Responsible Professor: Dr. Emir Demirović

TU Delft

1. Introduction

- **Survival Analysis**:
  - Find the expected time duration until one event occurs.
  - **Applications:** medicine, clinical research, engineering, economics, sociology.
  - **Objective**: fit survival function based on **historical data**, where the true time to event for some instances is **censored**.
- **Decision Trees**:
  - **Interpretable** model that can detect non-linear relations.
  - **Optimal decision trees**: fit optimal tree (for given size limits and train set) using **dynamic programming.**

2. Research question

Does fitting a Cox regression model [4] in the leaves of an optimal survival tree outperform the current state-of-the-art [1,2,3]?

3. Methodology

- **Cox Model**:
  - Semi-parametric model that specifies the effect of covariates without making assumptions about the overall shape of the survival function.
  - Fit parameters β maximising the likelihood.
  - Survival function for the Cox regression model: $S(t) = S0(t)^{e^{\beta^T * x}}$.
  - S0(t): baseline survival function. Same for all instances.
- **Regularised Cox models**:
  - Add elastic net penalty to likelihood: $\lambda * (\alpha * \sum_{i=1}^{p} |\beta_i| + (1-\alpha) * \sum_{i=1}^{p} \beta_i^2)$.
  - Simon et al. [5] proposed a way to fit a **path** of regularised solutions.
- **Cox SurTree**:
  - Fit optimal survival trees using a dynamic programming approach.
  - Generate a path of regularised Cox models per leaf and choose one:
    - *CoxSurTree LL*: likelihood optimisation.
    - CoxSurTree *CI*: C-Index optimisation [6].

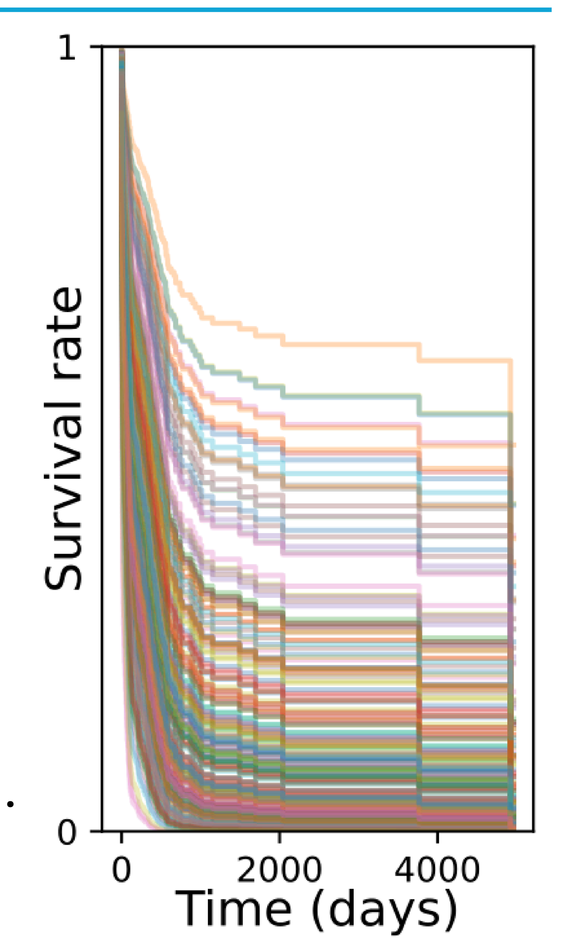

Figure 2: Example of estimated survival functions of all the instances.

4. Results and Findings

Survival metrics:
- **C-Index** [6]: proportion of correctly ordered comparable pairs of observations.
- **Integrated Brier score** [7]: deviation of predicted distributions from the actual ones.

	Data set characteristics				Harrell's C-Index					Integrated Brier Score				
Dataset	$\lvert\mathcal{D}\rvert$	Censoring (%)	$\lvert\mathcal{F}_{num}\rvert$	$\lvert\mathcal{F}\rvert$	CTree	OST	SurTree	Cox SurTree LL	Cox SurTree CI	CTree	OST	SurTree	Cox SurTree LL	Cox SurTree CI
Aids2	2839	38.0%	4	22	**0.54**	0.52	0.53	**0.54**	**0.54**	**0.01**	0.00	**0.01**	**0.01**	0.00
Dialysis	6805	76.4%	4	35	0.64	**0.67**	0.66	0.66	0.66	0.07	**0.12**	0.10	-0.03	-0.05
Framingham	4658	68.5%	7	60	0.68	0.67	0.67	**0.71**	**0.71**	0.11	0.10	0.10	**0.14**	**0.14**
Unempdur	3241	38.7%	6	45	0.69	0.69	0.69	0.69	**0.70**	**0.08**	**0.08**	0.07	0.03	0.03
Acath	2258	34.0%	3	21	0.59	0.59	0.59	0.59	**0.60**	**0.03**	**0.03**	**0.03**	**0.03**	0.01
Csl	2481	89.1%	6	42	**0.78**	0.76	0.77	**0.78**	**0.78**	0.12	0.14	0.13	**0.15**	0.14
Datadivat1	5943	83.6%	5	21	0.63	0.63	0.63	**0.64**	**0.64**	0.05	0.05	0.06	**0.09**	0.07
Datadivat3	4267	94.4%	7	30	0.66	0.65	0.66	**0.71**	**0.71**	-0.00	**0.03**	**0.03**	**0.03**	**0.03**
Divorce	3371	69.4%	3	5	**0.53**	**0.53**	**0.53**	**0.53**	**0.53**	**0.02**	**0.02**	**0.02**	**0.02**	**0.02**
Flchain	6524	69.9%	10	60	0.92	0.92	0.92	**0.93**	**0.93**	**0.64**	**0.64**	**0.64**	0.56	0.56
Hdfail	52422	94.5%	6	27	0.82	**0.86**	0.84	**0.86**	0.80	0.33	0.42	0.38	0.47	**0.50**
Nwtco	4028	85.8%	7	17	0.70	0.69	0.70	**0.72**	0.71	**0.13**	**0.13**	**0.13**	0.07	0.07
Oldmort	6495	69.7%	7	33	0.64	0.64	0.64	**0.66**	**0.66**	**0.07**	0.05	0.05	0.05	0.05
Prostatesurvival	14294	94.4%	3	8	**0.76**	**0.76**	**0.76**	**0.76**	**0.76**	**0.11**	**0.11**	**0.11**	-0.06	-0.05
Rott2	2982	57.3%	11	50	0.68	0.68	0.68	**0.71**	0.70	0.12	0.15	0.14	**0.16**	0.15
Wins per metric					4	4	2	**12**	11	**8**	**8**	7	**8**	4
Average rank					3.47	3.73	3.63	**2.07**	2.10	3.10	**2.77**	2.93	2.83	3.37

Table 1: Out-of-sample Harrell's C-Index [6] and integrated Brier score [7] for data sets from SurvSet [8]. CTree [3], OST [2], and SurTree [1] were tested on a maximum depth d = 4. CoxSurTree LL and Cox SurTree CI were tested on a fixed depth of d = 2.

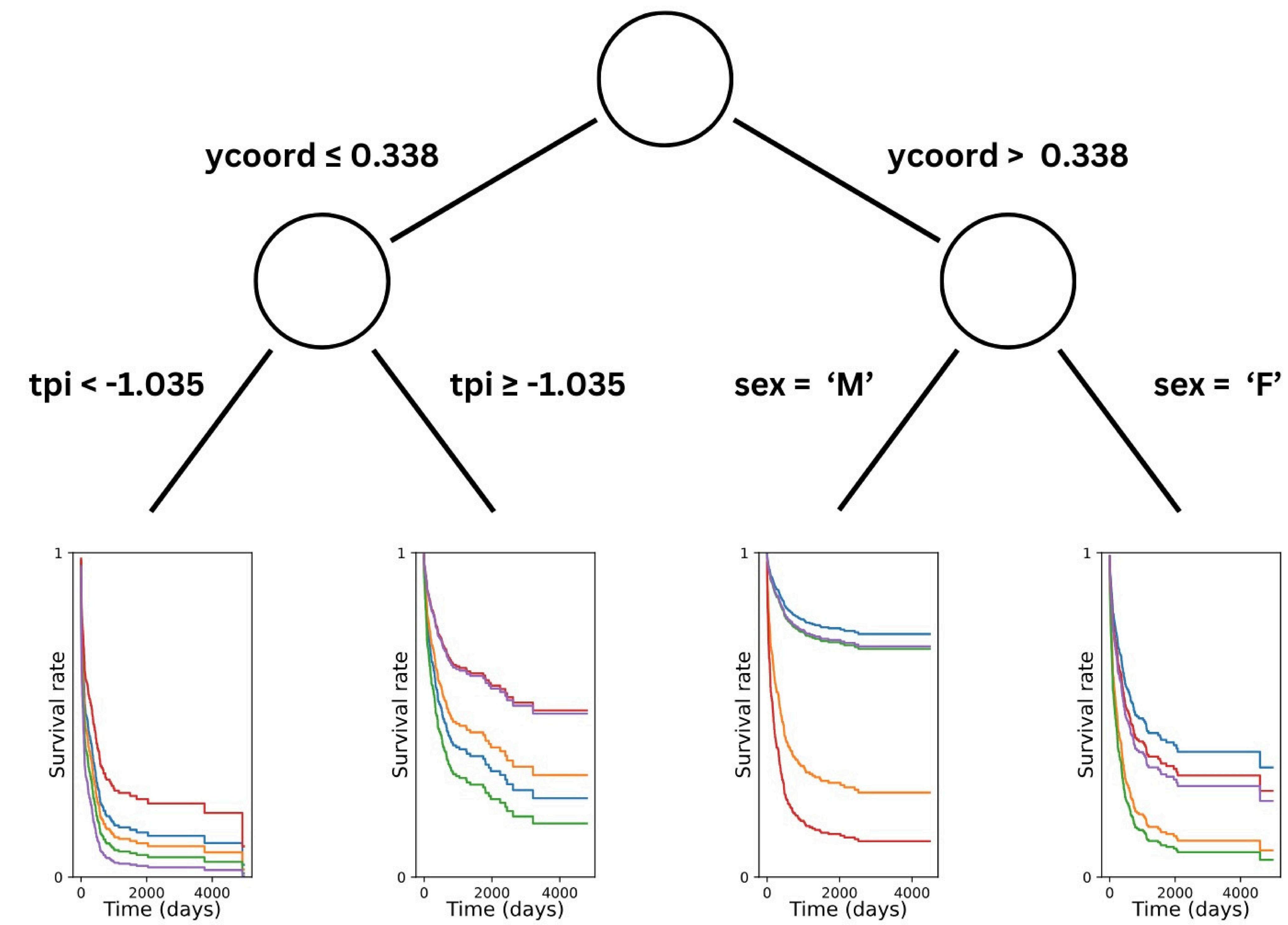

Figure 1: An example of a survival tree. Each leaf showcases five distributions, each corresponding to one instance in the respective leaf.

5. Limitations

- Cox models assume a specific form of the relationship between the covariates and the survival probability.
- Increased time complexity of leaf solvers (Regularisation Paths for Cox's Proportional Hazards Model via Coordinate Descent [5]).

6. Conclusion

Cox SurTree
- captures the relative risk better than the state-of-the-art methods.
- similarly estimates the overall distribution relative to the state-of-the-art methods.
- achieves these results for *smaller* trees.

References

[1] T. Huisman, J. G. M. van der Linden, and E. Demirović, "Optimal survival trees: A dynamic programming approach," in Proceedings of AAAI-24, 2024.
[2] D. Bertsimas, J. Dunn, E. Gibson, and A. Orfanoudaki, "Optimal survival trees," Machine Learning, vol. 111, no. 8, pp. 2951–3023, 2022.
[3] T. Hothorn, K. Hornik, and A. Zeileis, "Unbiased recursive partitioning: A conditional inference framework," Journal of Computational and Graphical Statistics, vol. 15, pp. 651–674, 2006.
[4] D. R. Cox, "Regression models and life-tables," Journal of the Royal Statistical Society: Series B (Methodological), vol. 34, no. 2, pp. 187–202, 1972.
[5] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for Cox's proportional hazards model via coordinate descent," Journal of statistical software, vol. 39, no. 5, p. 1, 2011.
[6] F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati, "Evaluating the yield of medical tests," Jama, vol. 247, no. 18, pp. 2543–2546, 1982.
[7] E. Graf, C. Schmoor, W. Sauerbrei, and M. Schumacher, "Assessment and comparison of prognostic classification schemes for survival data," Statistics in Medicine, vol. 18, no. 17-18, pp. 2529–2545, 1999.
[8] E. Drysdale, Survset: An open-source time-to-event dataset repository, 2022. arXiv: 2203.03094.