Korea Advanced Institute of Science and Technology
School of Electrical Engineering

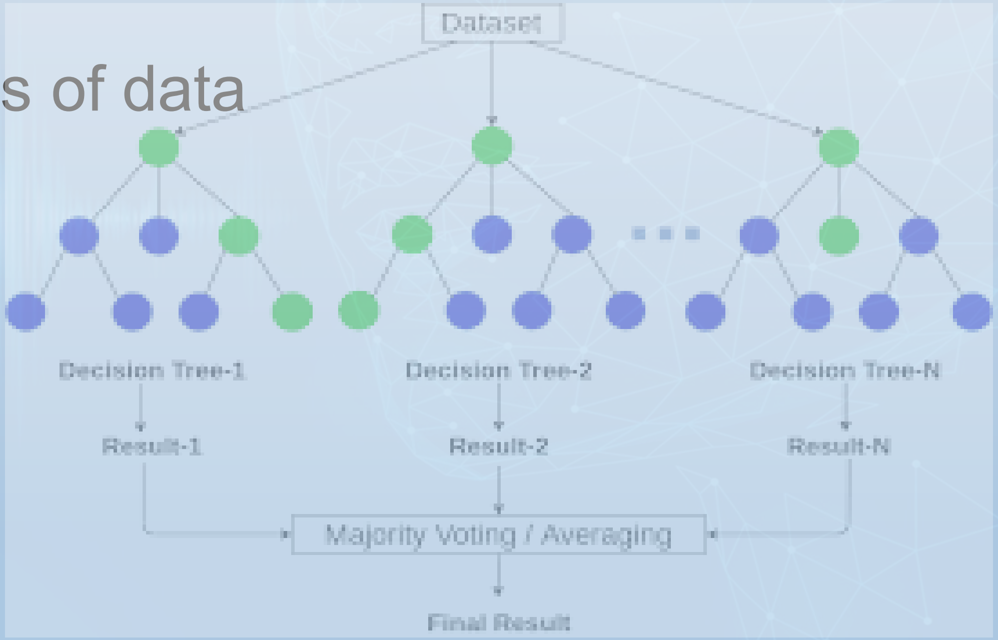
Machine Learning Technique for Survival Analysis

July 22, 2021

GWANGSU KIM

Contents

- Data, Model and Loss function
- Tree and Random Forests
- Survival Tree (Random Forests)
- XG Boosting
- Not providing examples of data analysis.



Data, Model and Loss function

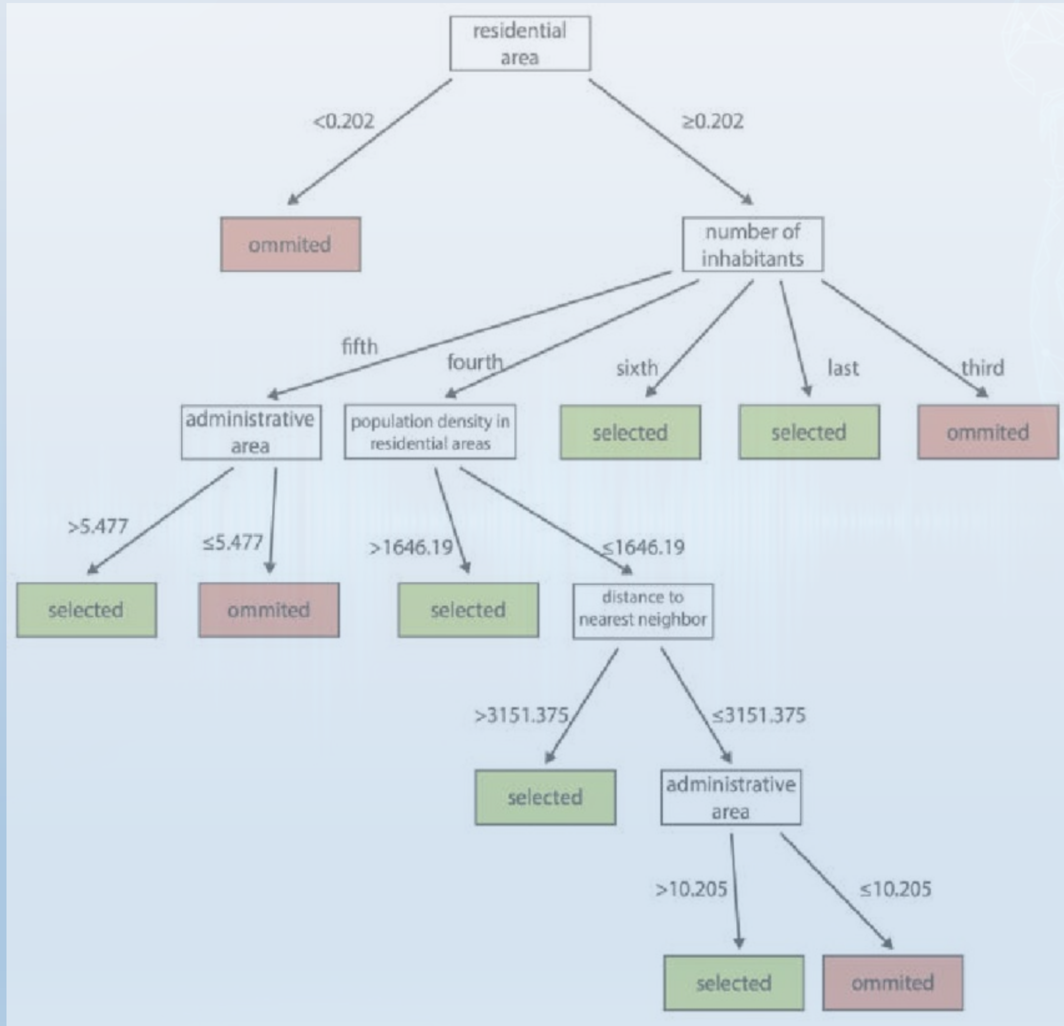
Data: $\{X_i, T_i, \delta_i\}_{i=1}^{N},\ T_i = \min(Y_i, C_i)$

Model: $\lambda(t \mid X) = \lim_{\delta \downarrow 0} P(t \leq Y < t + \delta \mid t \leq Y, X)/dt$

Loss function: Negative log likelihood

$$-\sum_{i=1}^{N} \left\{ \delta_i \log \lambda(T_i \mid X_i) - \int_0^{T_i} \lambda(s \mid X_i) ds \right\}$$

Tree

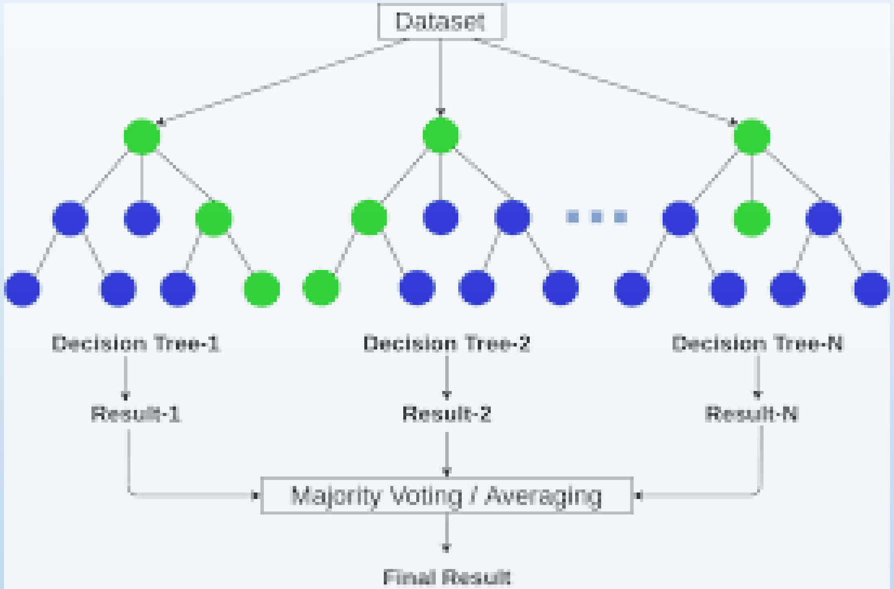


google

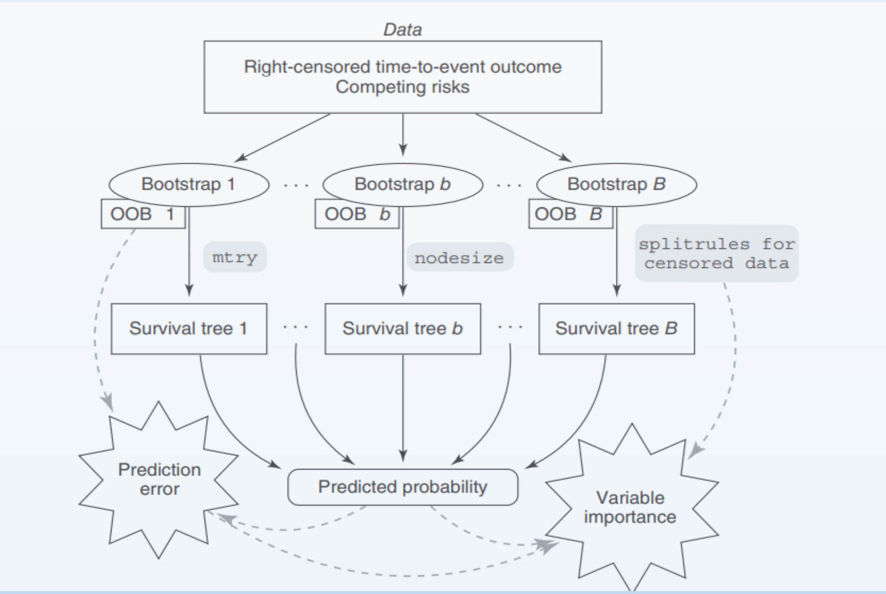
Node 1	Node 2	
	Node 3	Node 4

1. Sequentially selecting variable and splitting to minimize the criteria.
2. Applicable in high-dimensional features.
3. Most critical point is the rule of splitting.
4. Pruning (deleting nodes) to avoid the over-fitting.
5. Non-parametric prediction

Random Forests



google



Helen et al, 2018

- Making many trees (bootstrap sampling and subset selection of feature dimensions)
- As independent as possible
- Final prediction is a voting.

- Survival Tree: criteria rule
- Constraints: mtry-trying subset size of feature, nodesize, splitrules for censored data
- OOB: out-of-bag

Survival Trees

- Splitting rule

When $\mathcal{X}_l$ and $\mathcal{X}_r$ are supports for the left and right nodes respectively.
Log-rank test (equivalence of CHFs) statistic:

$$\frac{1}{\hat{\sigma}^{LR}} \int_0^t \frac{\bar{Y}_l(s)\bar{Y}_r(s)}{\bar{Y}_l(s)+\bar{Y}_r(s)} \left(\frac{d\bar{N}_l(s)}{\bar{Y}_l(s)} - \frac{d\bar{N}_r(s)}{\bar{Y}_r(s)} \right)$$

where
$$\begin{aligned} \bar{N}_r(s) &= \sum_{i=1}^{N} \mathbb{I}(X_i \in \mathcal{X}_r)\mathbb{I}(T_i \le t, \delta_i = 1) \\ \bar{Y}_r(t) &= \sum_{i=1}^{N} \mathbb{I}(X_i \in \mathcal{X}_r)\mathbb{I}(T_i > t) \end{aligned}$$

- Constraints

mtry: size of random subset in splitting
nodesize: number of nodes
split rules for censored data

Radom Forests for Survival Analysis

- How to estimate/predict the CHF using random forests (Ishwaran et al., 2008)

1. Combining N-A estimators $H_n(t)$ of all nodes.

   OOB ensemble $$H_e(t \mid X) = \frac{\sum_{n=1}^{ntree} H_n(t \mid X)(X \notin \mathcal{L}_n)}{\sum_{n=1}^{ntree} (X \notin \mathcal{L}_n)}$$

2. Prediction

   OOB ensemble for new X

3. Ensemble KM estimator (Hothorn, 2004)

$$\hat{S}(t \mid X) = \prod_{s \le t} \left(1 - \frac{\sum_{i=1}^{N} \sum_{n=1}^{ntrees} n_{i,n} L_{i,n}(X) N_i(ds)}{\sum_{i=1}^{N} \sum_{n=1}^{ntrees} n_{i,n} L_{i,n}(X) Y_i(s)} \right)$$

Prediction error and Importance of variables

- Prediction error

  Modification of the C-index,

  the C-index compare the pairs (small uncensored, large all)s.
  If the estimated hazard of large observation is larger than the other, then it is considered as error.

  OOB prediction error

  When $$\sum_{j=1}^{J} H_e(t_j^* \mid X_i) \geq \sum_{j=1}^{J} H_e(t_j^* \mid X_j) \text{ and } T_i \geq T_j$$

- Importance of variables (VIMP)

  For trees concerning a specific variable, convert the split by the variable to random splitting. Measure the ration of PE increasing

Issues

- Censoring ratio in each node
- Can variable selection be possible? In conventional random forests, some papers exist.
- Can high-dimensional analysis be available with certain levels?
- More explainable approach / extension such as cure model, some papers exist.

XG Boosting

- XG Boosting (Chen and Guestrin, 2016) is the best in many competitions for the prediction of structure datasets.

- Basic concept is to use the Talyor series expansions and gradient boosting.

$$\mathcal{L}^t = \sum_{i=1}^{N} \ell(y_i, \hat{y}_i^{t-1} + f_t(X_i)) + \Omega(f_t)$$

XG Boosting

$$
\begin{aligned}
\mathcal{L}^t &= \sum_{i=1}^{N} \ell(y_i, \hat{y}_i^{t-1} + f_t(X_i)) + \Omega(f_t) \\
&\approx \sum_{i=1}^{N} \left\{ \ell(y_i, \hat{y}_i^{t-1}) + g_i f_t(X_i) + \frac{1}{2} h_i f_t(x_i)^2 \right\} + \Omega(f_t) \\
&\approx \sum_{i=1}^{N} \left\{ \ell(y_i, \hat{y}_i^{t-1}) + g_i f_t(X_i) + \frac{1}{2} h_i f_t(x_i)^2 \right\} + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2 \\
&= \sum_{j=1}^{T} \left\{ w_j \sum_{i \in I_j} g_i + \frac{1}{2} w_j^2 \left(\sum_{i \in I_j} h_i + \lambda \right) \right\} + \gamma T \\
w_j^* &= -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}, \quad \mathcal{L}_{split} = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] \\
\mathcal{L}(w_j^*)^t &= -\frac{1}{2} \sum_{j=1}^{T} \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T
\end{aligned}
$$

XG Boosting

- Based on the previous additive model, simple tree (only one split) is added until at most several hundreds or thousands trees.

- Other Tricks:
  1) Greedy search for splitting point: using percentile of features.
  2) Weighted quantile (varying Hessian based)
  3) Sparsity-aware split: ignoring the missing points.
  4) Programming level techniques.

XG Boosting

- Gradient and Hessian are required.
- Partial likelihood and other likelihood can be candidates.
- Sparsity can be implemented with the boosting procedure.
- Feature extraction can be used for the conventional inference.

References

Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.

Genuer, Robin, Jean-Michel Poggi, and Christine Tuleau-Malot. "Variable selection using random forests." *Pattern recognition letters* 31.14 (2010): 2225-2236.

Ishwaran, Hemant, et al. "Random survival forests." *The annals of applied statistics* 2.3 (2008): 841-860.

Ma, Li, and Suohai Fan. "CURE-SMOTE algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests." *BMC bioinformatics* 18.1 (2017): 1-18.

Rytgaard, Helene C., and Thomas A. Gerds. "Random forests for survival analysis." *Wiley StatsRef: Statistics Reference Online* (2014): 1-8.