Ingersol Rand Notes

Chris Comiskey, Open Data Group

July 13, 2018

Misc

- Model evaluation

  - Concordance index (C-index)
  - Brier score

- Packages

  - **rpart**, **survival**,

- Methods

  - Multi-Task Logistic Regression (MLTR)
  - Neural-MLTR (N-MLTR)

1 Articles

1.1 An Interval-Valued Neural Network Approach for Uncertainty Quantification in Short-Term Wind Speed Prediction

- From abstract: "We consider the task of performing prediction with neural networks on the basis of uncertain input data expressed in the form of intervals. We aim at quantifying the uncertainty in the prediction arising from both the input data and the prediction model. A multi-layer perceptron neural network (NN) is trained to map interval-valued input data into interval outputs, representing the prediction intervals (PIs) of the real target values."

- "Uncertainties can be classified in two distinct types: epistemic and aleatory.

  - epistemic – "imprecise model representation of the system behavior, in terms of uncertainty in both the hypotheses assumed (structural uncertainty) and the values of the parameters of the model (parameter uncertainty)"

  - aleatory – "inherent variability of the observed physical phenomenon, and it is therefore also named stochastic uncertainty, irreducible uncertainty, or inherent uncertainty"

- A. M. Roque, C. Mat, J. Arroyo, and . Sarabia, **iMLP: Applying Multi-Layer Perceptrons to Interval-Valued Data**, Neural Processing Letters, vol. 25, no. 2, pp. 157169, Apr. 2007.

- A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, **Lower Upper Bound Estimation Method for Construction of Neural Network- Based Prediction Intervals**, IEEE Transactions on Neural Networks, vol. 22, no. 3, pp. 337-346, March 2011. (ADDED TO DRIVE)

- "In this paper, we present an interval-valued prediction modeling framework based on a data-driven learning approach, more specifically a multi-layer perceptron neural network (NN)"

- Section 2: overview of "interval-valued NNs for PIs estimation"

1.2 Deep Learning for Patient-Specific Kidney Graft Survival Analysis

6 Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer. Random survival forests. The annals of applied statistics, pages 841860, 2008. (ADDED TO DRIVE)

1.3 Random Survival Forrests

- "As is well known, constructing ensembles from base learners, such as trees, can sub- stantially improve prediction performance."

- "ensemble learning can be improved further by inject- ing randomization into the base learning process"

- "In RF, randomization is introduced in two forms. First, a randomly drawn bootstrap sample of the data is used to grow a tree. Second, at each node of the tree, a randomly selected subset of vari- ables (covariates) is chosen as candidate variables for splitting. Averaging over trees, in combination with the randomization used in growing a tree, enables RF to approximate rich classes of functions while maintaining low generalization error."

- Bunch of limitations with traditional approaches–proportional hazards assumption, parametric, going beyond linearity problematic–"... these difficulties are handled automatically using forests."

- The setup:

  - h = terminal nodes
  - $l = 1, ..., N(h)$ = distinct event times at node h
  - $T_{l,h} = t_{1,h} < t_{2,h} < ... < t_{N(h),h}$ = survival time
  - $\delta_{l,h}$ =
    0 : censored
    1 : event
  - i = individual
  - $d_{l,h}$ = # deaths at event time l, node h
  - $Y_{l,h}$ = # individuals at risk at event time l, node h

- Hazard = $\lambda(t)$; cumulative Hazard = $\Lambda(t)$

- Cumulative Hazard function: think of it as "the sum of the risks you face going from duration 0 to t" (from Ch 7);