Documentation on using Survival boosting model for predicting cirrhosis

Taking one lab, APRI score, for example. Assuming the raw data takes the format that

|  |  |  |
| --- | --- | --- |
| PatientID | DaysSinceEnrollment | Value |
| 1 | 1 | 1.8 |
| 1 | 30 | 1.5 |
| 1 | 80 | 1.7 |
| 1 | 120 | 1.4 |
| 1 | 260 | 1.5 |
| 1 | 400 | 1.6 |
| 1 | 500 | 1.4 |

Note that this data point is a made up one and does not reflect any real data in the VA cohort. Before applying the survival boosting models, the following steps need to be performed.

1. Obtain longitudinal summary statistics:
   1. Maximum (MAX) = 1.8
   2. Minimum (MIN) = 1.4

Calculate the slopes of these value, which is defined by the difference between two consecutive values divided by time difference, i.e., (1.5 – 1.8) /(30 – 1), (1.7 – 1.5)/(80 – 30), …, (1.4 – 1.6)/(500 – 400), then we take

* 1. Maximum of Slopes (DIFFMAX)
  2. Minimum of Slopes (DIFFMIN)
  3. Average of absolute value of slopes (TVAR)

1. Create a data frame, with each row being a patient, and the columns are:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PatientID | APRI\_MAX | APRI\_MIN | APRI\_DIFFMAX | APRI\_DIFFMIN | APRI\_TAVR |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |
| … |  |  |  |  |  |

1. Apply step 1 and 2 for other labs, with the lab variable name as specified in the [PLoS one paper](https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0208141&type=printable).
2. Merge all labs into a large data frame by PatientID
3. Merge TimeToOutcome and Outcome with the merged lab data frame.
4. Denote the merged data frame by datanew; apply the survival boosting model on the merged data frame by running the following codes:

load(‘fit\_boosting\_full\_lgt.RData’)

best.iter <- gbm.perf(fit1,method="OOB")

pred1 <- predict(fit1,newdata = datanew,n.trees = best.iter)

1. The model can be evaluated by:

CIndex1 <- survConcordance(Surv(TimeToOutcome, Outcome) ~ pred1, datanew)

CIndex1$concordance