

Predictive Modeling

Akhila Chowdary Kolla, Shadmaan Hye
Joint work with Dr.Toby Dylan Hocking

September 22, 2021

Our work so far

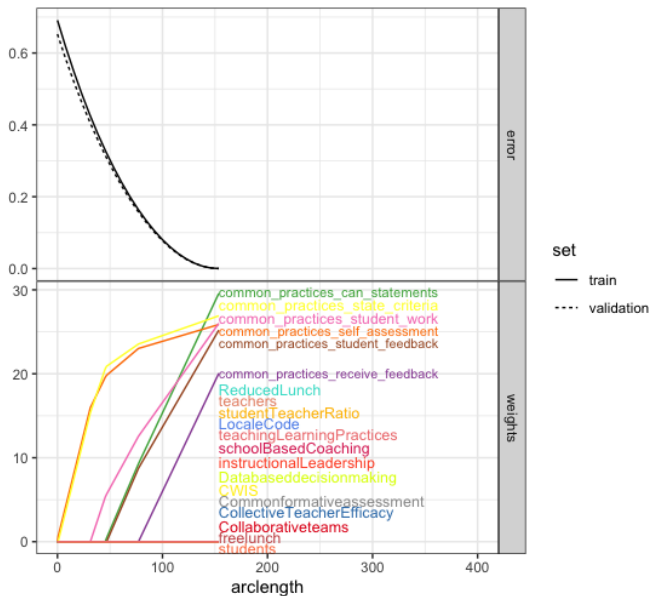
- Implementation of Lasso regression model
- Implementation of cvglmnet
- Implementation of Ordinal Forest
- Comparison of the models

DataSet Used

```
1 > str(input.data)
2 Classes      data      .table and 'data.frame':  25547 obs. of  37 variables:
3 $ Timestamp                                     : POSIXct, format: "
    2019-09-16 14:39:02" "2019-09-16 14:39:02" "2019-09-16 14:39:02" "
    2019-09-16 14:39:02" ...
4 $ Date of Event/Visit                           : POSIXct, format: "
    2019-09-12" "2019-09-12" "2019-09-12" "2019-09-12" ...
5 $ Duration of Event                             : chr   "2:30:00" "2:30:00
    " "2:30:00" "2:30:00" ...
6 $ Interaction Type                             : chr   "In-Person" "In-
    Person" "In-Person" "In-Person" ...
7 $ State District ID                             : chr   "MO-015002" "MO
    -015002" "MO-015002" "MO-015002" ...
8 $ DISTRICTS                                     : chr   "Camdenton R-III"
    "Camdenton R-III" "Camdenton R-III" "Camdenton R-III" ...
9 $ Collaborative teams                           : int    1 1 1 1 1 1 1 1 1
    1 ...
10 $ Common formative assessment                  : int    1 1 1 1 1 1 1 1 1
    1 ...
11 $ Data-based decision making                   : int    1 1 1 1 1 1 1 1 1
    1 ...
12 $ Effective teaching/learning practices        : int    1 1 1 1 1 1 1 1 1
    1 ...
13 $ Instructional Leadership                     : int    1 1 1 1 1 1 1 1 1
    1 ...
14 $ School-based implementation coaching          : int    1 1 1 1 1 1 1 1 1
    1 ...
15 $ Collective teacher efficacy                  : int    1 1 1 1 1 1 1 1 1
    1 ...
```

- LASSO (Least Absolute Shrinkage and Selection Operator) is a type of linear regression that uses shrinkage
- Shrinks Slope to zero meaning Excludes useless variables from features.
- Better at reducing variance that contains a lot of useless variables
- Returns a set of Lambda values

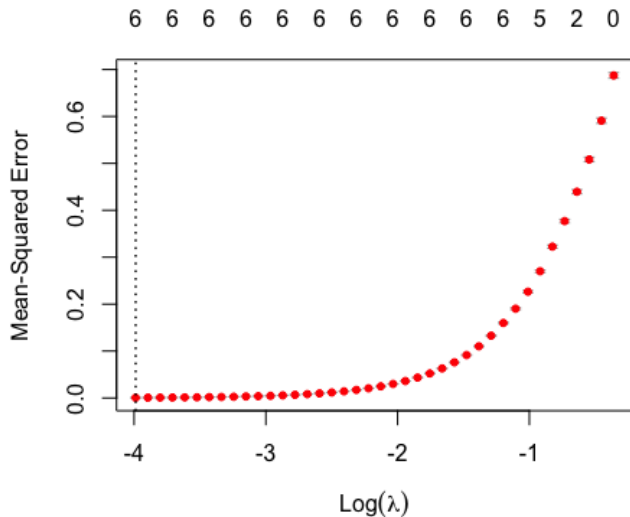
Results from Lasso & Subtrain vs Validation Loss



Why use cvglmnet?

- cvglmnet gives optimal Lambda value instead of a set of Lambda values
- easier to use the Lambda function

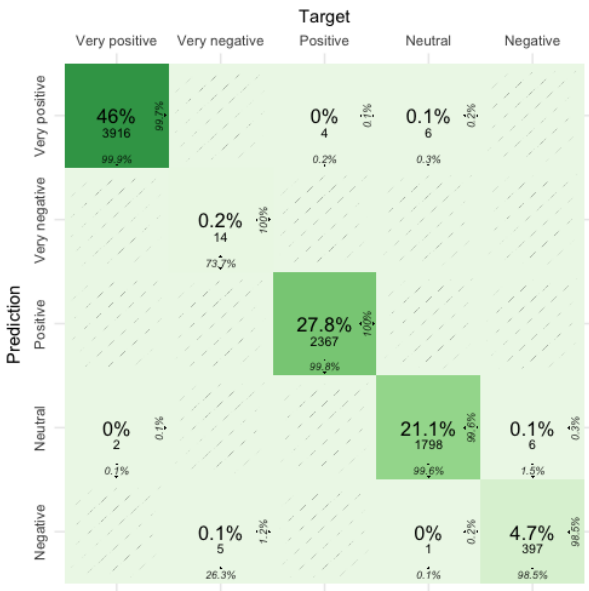
Results from cvglmnet



Why use Ordinal Forest?

- used a non linear model for performance measure
- Ordinal forest method is a random forest-based prediction method for ordinal variables
- used to rank covariates with respect to their importance for prediction.

Results from Ordinal Forest



- Comparisons between regular regression random forest and Ordinal Forest to understand which model works better

Plots

- Training the dataset with a) keeping missing rows, b) filling missing values
- Nonlinear Function (Random Forest) better than Linear Function (glmnet)? OrdinalNet vs Ordinal Forest
- Ordinal loss function better than Square loss function? ordinalNet vs glmnet (add baseline to all plots)