Committee-based Congressional Network Regression Analysis

Lynette Dang

2022-05-01

```
library(readr)
df <- read_csv("final_analysis.csv")
```

```
## Rows: 481 Columns: 11
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (6): Legislator name, Connections, Gender Estimates, Predicted Gender, E...
## dbl (5): Legislative Effectiveness Score, bc, ec, cc, dc
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
df$factor <- as.factor(sample(1:2, nrow(df), replace = TRUE))
head(df)
```

```
## # A tibble: 6 x 12
##   `Legislator name` Connections `Legislative E~`      bc        ec      cc    dc
##   <chr>             <chr>                  <dbl>   <dbl>     <dbl>   <dbl> <dbl>
## 1 Jack Reed         ['Kirsten ~            0.346 0.00178 -2.74e-18 0       0.117
## 2 Kirsten Gillibra~ ['Jack Ree~            0.339 0.00178 -3.85e-18 0.00181 0.108
## 3 Tammy Duckworth   ['Jack Ree~            1.53  0.00178  1.94e-18 0.00361 0.116
## 4 Gary Peters       ['Jack Ree~            5.02  0.00225  1.74e-17 0.00722 0.119
## 5 Elizabeth Warren  ['Jack Ree~            0.725 0.00180  2.01e-17 0.00903 0.112
## 6 Martin Heinrich   ['Jack Ree~            0.781 0.00239  1.96e-17 0.0108  0.125
## # ... with 5 more variables: `Gender Estimates` <chr>,
## #   `Predicted Gender` <chr>, `Ethnicity Estimates` <chr>,
## #   `Predicted Ethnicity` <chr>, factor <fct>
```

```
library(glmnet)
```

```
## Loading required package: Matrix
## Loaded glmnet 4.1-3
```

```
# the original data frame and formula
set.seed(23)
f <- as.formula(df$`Legislative Effectiveness Score` ~ (df$bc+ df$ec+ df$cc+ df$dc + df$`Predicted Gende
# transform dataframe to matrices as required by glmnet
x <- model.matrix(f, df)
y <- as.matrix(df$`Legislative Effectiveness Score`, ncol=1)
```

```r
# fit ridge regression model with a wide range of penalty parameter
lambdas <- 10^seq(3, -2, by = -.1)
cv_fit <- cv.glmnet(x,y, lambda=lambdas, alpha = 0)
fit <- cv_fit$glmnet.fit
summary(fit)
```

```
##           Length Class     Mode
## a0          51   -none-    numeric
## beta      1734   dgCMatrix S4
## df          51   -none-    numeric
## dim          2   -none-    numeric
## lambda      51   -none-    numeric
## dev.ratio   51   -none-    numeric
## nulldev      1   -none-    numeric
## npasses      1   -none-    numeric
## jerr         1   -none-    numeric
## offset       1   -none-    logical
## call         5   -none-    call
## nobs         1   -none-    numeric
```

```r
# get BIC value
tLL <- fit$nulldev - deviance(fit)
k <- fit$df
n <- fit$nobs

BIC<-log(n)*k - tLL
BIC
```

```
##  [1] 203.7148 203.6922 203.6636 203.6279 203.5832 203.5275 203.4580 203.3717
##  [9] 203.2649 203.1332 202.9715 202.7790 202.5427 202.2597 201.9245 201.5326
## [17] 201.0809 200.5688 199.9986 199.3758 198.7085 198.0079 197.2856 196.5534
## [25] 195.8211 195.0955 194.3794 193.6714 192.9683 192.2578 191.5308 190.7818
## [33] 189.9976 189.1778 188.3069 187.3999 186.4561 185.4871 184.5072 183.5323
## [41] 182.5796 181.6579 180.7752 179.9356 179.1394 178.3807 177.6596 176.9605
## [49] 176.2710 175.6056 174.9604
```

```r
opt_lambda = lambdas[which(BIC == min(BIC))]
# Choose best model
best_model <- glmnet(x,y, lambda=opt_lambda, alpha = 0)
print(coef(best_model))
```

```
## 35 x 1 sparse Matrix of class "dgCMatrix"
##                                                                    s0
## (Intercept)                                                1.497341e+00
## (Intercept)                                                .
## df$bc                                                     -1.130367e+02
## df$ec                                                     -2.108612e-01
## df$cc                                                      4.343528e-01
## df$dc                                                     -2.443166e+00
## df$`Predicted Gender`male                                  4.855377e-01
## df$`Predicted Ethnicity`black                              1.842198e-01
```

```
## df$`Predicted Ethnicity`hispanic                                 5.753947e-01
## df$`Predicted Ethnicity`white                                   -2.049838e-01
## df$bc:df$ec                                                     -3.244254e+03
## df$bc:df$cc                                                      3.456645e+02
## df$bc:df$dc                                                      3.972438e+02
## df$bc:df$`Predicted Gender`male                                 -2.179875e+01
## df$bc:df$`Predicted Ethnicity`black                              8.332574e+02
## df$bc:df$`Predicted Ethnicity`hispanic                          -1.106393e+02
## df$bc:df$`Predicted Ethnicity`white                              3.791436e+01
## df$ec:df$cc                                                     -1.353747e-01
## df$ec:df$dc                                                      4.584725e+01
## df$ec:df$`Predicted Gender`male                                  2.392189e+00
## df$ec:df$`Predicted Ethnicity`black                              4.392340e+01
## df$ec:df$`Predicted Ethnicity`hispanic                          -4.033550e+01
## df$ec:df$`Predicted Ethnicity`white                              4.196871e-01
## df$cc:df$dc                                                      2.115944e+00
## df$cc:df$`Predicted Gender`male                                 -1.231721e+00
## df$cc:df$`Predicted Ethnicity`black                             -2.429899e+00
## df$cc:df$`Predicted Ethnicity`hispanic                          -9.938116e-01
## df$cc:df$`Predicted Ethnicity`white                             -1.672829e+00
## df$dc:df$`Predicted Gender`male                                 -1.118700e+00
## df$dc:df$`Predicted Ethnicity`black                             -2.158242e+01
## df$dc:df$`Predicted Ethnicity`hispanic                          -4.427393e-01
## df$dc:df$`Predicted Ethnicity`white                              1.293783e-02
## df$`Predicted Gender`male:df$`Predicted Ethnicity`black          5.450884e-01
## df$`Predicted Gender`male:df$`Predicted Ethnicity`hispanic      -1.401436e-01
## df$`Predicted Gender`male:df$`Predicted Ethnicity`white          4.131083e-02
```