Problem 6

Lin Yang

5/9/2022

```r
dat6 <- read.csv("data/hurricanoutcome2.csv") %>%
  janitor::clean_names() %>%
  mutate(damage = as.numeric(str_replace(damage, "\\$", "")),
         deaths = as.integer(gsub(",", "", deaths)),
         month = as.factor(month),
         nature = as.factor(nature)) %>%
  dplyr::select(-1)
```

```r
#damage
damage_x <- model.matrix(damage ~ ., dat6)[ ,-1]
damage_y <- dat6$damage

set.seed(1)
ctrl <- trainControl(method = "repeatedcv")
damage.fit <- train(damage_x, damage_y,
                    method = "glmnet",
                    preProcess = "scale",
                    tuneGrid = expand.grid(alpha = 1,
                                           lambda = exp(seq(5, -3, length = 100))),
                    trControl = ctrl)
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut =
## 10, : These variables have zero variances: monthNovember, natureNR
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut =
## 10, : These variables have zero variances: monthJuly
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :
## There were missing values in resampled performance measures.
```

```r
damage.fit$bestTune
```

```
##    alpha   lambda
## 54     1 3.606826
```

```r
coef(damage.fit$finalModel, damage.fit$bestTune$lambda) %>%
  as.matrix() %>%
  knitr::kable(col.names = gsub("[.]", " ", "Coefficients"))
```

	Coefficients
(Intercept)	-533.5099174
season	3.3837824
deaths	0.0000000
monthJuly	0.0000000
monthJune	0.0000000
monthNovember	0.0000000
monthOctober	0.0000000
monthSeptember	0.0000000
natureNR	0.0000000
natureTS	0.0000000
maxspeed	1.2117851
meanspeed	0.0000000
maxpressure	0.0000000
meanpressure	0.0000000
hours	0.0000000
total_pop	0.3187361
percent_poor	0.0000000
percent_usa	0.7073409

```r
#refit a linear regression model
damage.lm <- lm(damage ~ season + maxspeed + total_pop + percent_usa, data = dat6)
summary(damage.lm)
```

```
##
## Call:
## lm(formula = damage ~ season + maxspeed + total_pop + percent_usa,
##     data = dat6)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.607  -9.966  -3.352   2.112  92.129
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.317e+03  5.176e+02  -2.544   0.0152 *
## season       6.485e-01  2.580e-01   2.514   0.0163 *
## maxspeed     1.969e-01  1.227e-01   1.604   0.1169
## total_pop    3.284e-06  4.314e-06   0.761   0.4511
## percent_usa  1.356e-01  7.586e-02   1.788   0.0817 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.21 on 38 degrees of freedom
## Multiple R-squared:  0.2317, Adjusted R-squared:  0.1508
## F-statistic: 2.865 on 4 and 38 DF,  p-value: 0.03612
```

```r
fit <- lm(damage ~., data = dat6)
summary(fit)
```

```
##
```

```
## Call:
## lm(formula = damage ~ ., data = dat6)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -22.110  -9.505  -3.315   4.942  85.925
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -3.120e+03  1.918e+03  -1.626   0.1164
## season           9.922e-01  5.142e-01   1.930   0.0651 .
## deaths           7.425e-04  2.010e-03   0.369   0.7149
## monthJuly       -1.315e+01  2.332e+01  -0.564   0.5777
## monthJune        2.446e+00  2.114e+01   0.116   0.9088
## monthNovember    6.206e+00  2.768e+01   0.224   0.8245
## monthOctober     1.432e+01  1.481e+01   0.967   0.3429
## monthSeptember  -1.209e+00  1.155e+01  -0.105   0.9174
## natureNR        -1.555e+01  3.520e+01  -0.442   0.6625
## natureTS        -8.018e+00  1.578e+01  -0.508   0.6158
## maxspeed         2.804e-01  2.926e-01   0.958   0.3470
## meanspeed        2.275e-02  4.851e-01   0.047   0.9630
## maxpressure      1.147e+00  1.385e+00   0.828   0.4156
## meanpressure    -5.319e-02  3.862e-02  -1.377   0.1807
## hours            2.185e-03  4.815e-02   0.045   0.9642
## total_pop        6.179e-06  6.728e-06   0.918   0.3672
## percent_poor    -1.414e-01  2.959e-01  -0.478   0.6368
## percent_usa      1.545e-01  1.067e-01   1.448   0.1600
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.98 on 25 degrees of freedom
## Multiple R-squared:  0.3464, Adjusted R-squared:  -0.09797
## F-statistic: 0.7796 on 17 and 25 DF,  p-value: 0.6985
```

```r
#deaths
#quantile(dat6$deaths, 0.95)
#dat_death <- dat6 %>% filter(deaths <= 1000)
#boxplot(dat6$deaths)
#summary(dat_death)
death_x <- model.matrix(deaths ~ ., dat6)[ ,-1]
death_y <- dat6$deaths

set.seed(100)
death.fit <- train(deaths ~ season+damage+deaths+month+nature+maxspeed+meanspeed+maxpressure+meanpressu
                   data = dat6,
                   method = "glmnet",
                   family = "poisson",
                   preProcess = "scale",
                   tuneGrid = expand.grid(alpha = 1,
                                          lambda = exp(seq(-4, 5, length = 500))),
                   trControl = ctrl)
```

```
## Warning in model.matrix.default(Terms, m, contrasts): the response appeared on
## the right-hand side and was dropped
```

```
## Warning in model.matrix.default(Terms, m, contrasts): problem with term 3 in
## model.matrix: no columns are assigned
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut =
## 10, : These variables have zero variances: natureNR
```

```
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut =
## 10, : These variables have zero variances: monthNovember
```

```
plot(death.fit)
```
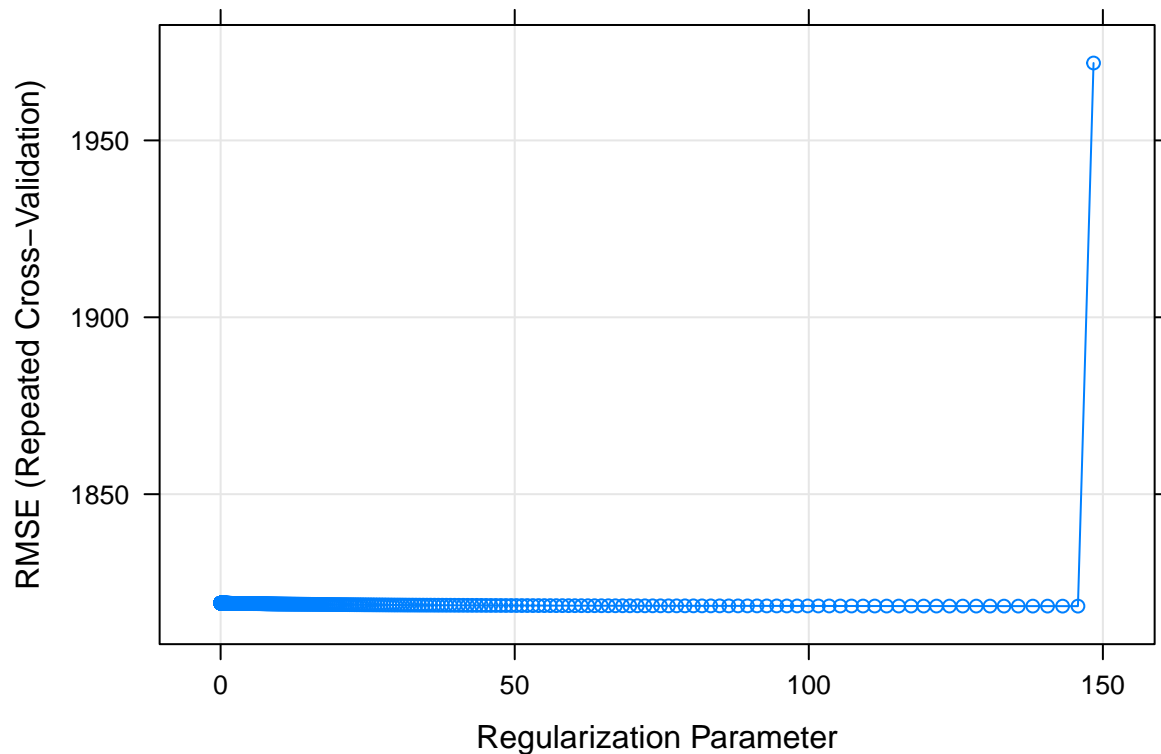


```
death.fit$bestTune
```

```
##     alpha   lambda
## 499     1 145.7604
```

```
coef(death.fit$finalModel, death.fit$bestTune$lambda) %>%
  as.matrix() %>%
  knitr::kable(col.names = gsub("[.]", " ", "Coefficients"))
```

	Coefficients
(Intercept)	5.6728507

4

	Coefficients
season	0.0000000
damage	0.0279429
monthJuly	-0.3610097
monthJune	0.0000000
monthNovember	0.0000000
monthOctober	0.2184573
monthSeptember	0.0000000
natureNR	0.0000000
natureTS	0.0000000
maxspeed	0.0000000
meanspeed	0.0000000
maxpressure	0.0000000
meanpressure	0.0000000
total_pop	0.0000000
percent_poor	0.8603086
percent_usa	0.0000000

```
poi_fit = glm(deaths ~ season+damage+deaths+month+nature+maxspeed+meanspeed+maxpressure+meanpressure+to
```

```
## Warning in model.matrix.default(mt, mf, contrasts): the response appeared on the
## right-hand side and was dropped
```

```
## Warning in model.matrix.default(mt, mf, contrasts): problem with term 3 in
## model.matrix: no columns are assigned
```

```
summary(poi_fit)
```

```
##
## Call:
## glm(formula = deaths ~ season + damage + deaths + month + nature +
##     maxspeed + meanspeed + maxpressure + meanpressure + total_pop +
##     percent_poor + percent_usa + offset(log(hours)), family = poisson(link = log),
##     data = dat6)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -38.151  -17.900   -3.638    4.481   64.167
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     2.043e+02  5.521e+00  37.000  < 2e-16 ***
## season          8.086e-03  1.370e-03   5.900 3.63e-09 ***
## damage          2.458e-02  3.918e-04  62.735  < 2e-16 ***
## monthJuly      -5.373e+00  1.139e-01 -47.162  < 2e-16 ***
## monthJune      -9.333e-02  8.791e-02  -1.062    0.288
## monthNovember   1.493e-01  1.538e-01   0.971    0.332
## monthOctober    8.907e-01  4.600e-02  19.361  < 2e-16 ***
## monthSeptember  1.813e+00  3.984e-02  45.509  < 2e-16 ***
## natureNR        6.051e+00  1.170e-01  51.719  < 2e-16 ***
## natureTS        2.576e+00  5.633e-02  45.729  < 2e-16 ***
```

```
## maxspeed      -1.715e-02  7.040e-04 -24.356  < 2e-16 ***
## meanspeed     -4.139e-02  1.411e-03 -29.331  < 2e-16 ***
## maxpressure   -2.209e-01  4.612e-03 -47.903  < 2e-16 ***
## meanpressure   2.421e-03  8.103e-05  29.883  < 2e-16 ***
## total_pop      9.244e-07  1.832e-08  50.458  < 2e-16 ***
## percent_poor   4.572e-02  4.451e-04 102.703  < 2e-16 ***
## percent_usa    1.337e-03  3.175e-04   4.210 2.55e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 148969  on 42  degrees of freedom
## Residual deviance:  21012  on 26  degrees of freedom
## AIC: 21321
##
## Number of Fisher Scoring iterations: 8
```

```r
#wave.nb=glm.nb(deaths ~season+damage+deaths+month+nature+maxspeed+meanspeed+maxpressure+mea#npressure+
#summary(wave.nb)
#sqrt(sum((predict(wave.nb, newdata = dat6) - dat6$deaths)^2))



#sqrt(sum((predict(poi_fit, newdata = dat6) - dat6$deaths)^2))
#summary(dat6$deaths)
```