Problem 6

Lin Yang

5/9/2022

Data Preparation

```
load("generated_data/B_final.RData")
source("shared_code/data_cleaning.R")
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

## Rows: 22038 Columns: 8

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (4): ID, Month, Nature, time
## dbl (4): Season, Latitude, Longitude, Wind.kt

##
## 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.

## New names:
## * id -> id...1
## * season -> season...2
## * nature -> nature...3
## * time -> time...4
## * latitude -> latitude...5
## * ...

## New names:
## * id -> id...1
## * wkt_cur -> wkt_cur...2
## * id -> id...10
## * wkt_cur -> wkt_cur...11

## Warning: `add_rownames()` was deprecated in dplyr 1.0.0.
## Please use `tibble::rownames_to_column()` instead.
```

```
## Joining, by = "id"
```

```
id <- train$id %>% unique()

B_reg <- tibble(
  id = id,
  beta0 = B_final[,1],
  beta1 = B_final[,2],
  beta2 = B_final[,3],
  beta3 = B_final[,4],
  beta4 = B_final[,5]
) %>% mutate(id = tolower(id))

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)) %>%
  rename(id = hurrican_id)

dat6new <- left_join(dat6, B_reg, by = "id") %>% dplyr::select(-1)

dat6new$nature
```

```
##  [1] TS TS TS TS TS TS NR TS TS TS TS TS TS TS TS TS TS TS TS TS TS TS TS TS
## [26] TS TS TS TS TS TS TS TS TS TS TS TS TS DS DS TS DS DS
## Levels: DS NR TS
```

Model for Damage

```
#damage
damage_x <- model.matrix(damage ~ ., dat6new)[ ,-1]
damage_y <- dat6new$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)

damage.fit$bestTune
```

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

```
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
beta0	0.0000000
beta1	0.0000000
beta2	0.0000000
beta3	0.0000000
beta4	0.0000000

```
#refit a linear regression model
damage.lm <- lm(damage ~ season + maxspeed + total_pop + percent_usa, data = dat6new)
summary(damage.lm)$coefficient[, 1] %>%
  knitr::kable(col.names = gsub("[.]", " ", "Coefficients"))
```

	Coefficients
(Intercept)	-1316.7386136
season	0.6485139
maxspeed	0.1968674
total_pop	0.0000033
percent_usa	0.1356486

Model for Deaths

```
death_x <- model.matrix(deaths ~ ., dat6new)[ ,-1]
death_y <- dat6new$deaths

death.fit <- glm(deaths ~ season + damage + month + nature + maxspeed + meanspeed + maxpressure + meanpr
                 family = poisson(link = log), data = dat6new)
```

```
summary(death.fit)$coefficients %>% knitr::kable()
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-199.5331639	11.8792784	-16.7967411	0.0000000
season	-0.0404185	0.0028048	-14.4104991	0.0000000
damage	0.0220163	0.0005679	38.7649762	0.0000000
monthJuly	-10.2286750	0.1604645	-63.7441688	0.0000000
monthJune	0.3928062	0.0989170	3.9710698	0.0000716
monthNovember	1.8733767	0.1664682	11.2536625	0.0000000
monthOctober	-1.6041896	0.0787720	-20.3649754	0.0000000
monthSeptember	1.2490015	0.0575033	21.7205350	0.0000000
natureNR	2.0903864	0.1371766	15.2386495	0.0000000
natureTS	-1.1903051	0.1118619	-10.6408484	0.0000000
maxspeed	0.0035207	0.0013988	2.5168778	0.0118400
meanspeed	-0.1978651	0.0039977	-49.4953412	0.0000000
maxpressure	0.0048106	0.0075485	0.6372945	0.5239331
meanpressure	0.0021204	0.0001759	12.0515409	0.0000000
total_pop	0.0000009	0.0000000	31.4237737	0.0000000
percent_poor	0.0873434	0.0010058	86.8433730	0.0000000
percent_usa	-0.0080185	0.0004884	-16.4173000	0.0000000
beta0	41.3531048	0.5634443	73.3934206	0.0000000
beta1	132.8784572	1.9305164	68.8305252	0.0000000
beta2	-10.7339527	0.5001340	-21.4621524	0.0000000
beta3	-0.4736994	0.5091748	-0.9303277	0.3522014
beta4	4.4919244	0.1971025	22.7897893	0.0000000

```
#set.seed(100)
#death.fit <- train(deaths ~ season + damage + month + nature + maxspeed + maxpressure + #meanpressure
#                   data = dat6new,
#                   method = "glmnet",
#                   family = "poisson",
#                   preProcess = "scale",
#                   tuneGrid = expand.grid(alpha = seq(0, 1, length = 21),
#                                          lambda = exp(seq(-4, 4, length = 200))),
#                   trControl = ctrl)
#plot(death.fit)
#death.fit$bestTune
#coef(death.fit$finalModel, death.fit$bestTune$lambda) %>%
#  as.matrix() %>%
#  knitr::kable(col.names = gsub("[.]", " ", "Coefficients"))
```