Mushrooms

Ina Ding

Introduction

```
This section includes an introduction to the project motivation, data, and research question

The research question and motivation are clearly stated in the introduction, including citati
```

The primary goal of this project is to best predict whether or not a mushroom is poisonous depending on various physical characteristics, rarity, and habitat of the fungus. The data set consists of 8124 hypothetical samples, constructed from the Audobon Society Field Guide. The samples correspond to 23 species of mushrooms from the Agaricus and Lepiota Families. Each mushroom is categorized as either poisonous or edible, with mushrooms 'not reccomended for eating' or of unknown edibility are counted as poisonous. Though the observations are hypothetical mushrooms, analyzing them can still provide beneficial results that can be applied to help identify the edibility of the near 14,000 existing species of mushrooms.

```
 1. cap-shape:                bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
 2. cap-surface:              fibrous=f,grooves=g,scaly=y,smooth=s
 3. cap-color:                brown=n,buff=b,cinnamon=c,gray=g,green=r,
                              pink=p,purple=u,red=e,white=w,yellow=y

 4. bruises?:                 bruises=t,no=f
 5. odor:                     almond=a,anise=l,creosote=c,fishy=y,foul=f,
                              musty=m,none=n,pungent=p,spicy=s

 6. gill-attachment:          attached=a,descending=d,free=f,notched=n
 7. gill-spacing:             close=c,crowded=w,distant=d
 8. gill-size:                broad=b,narrow=n
 9. gill-color:               black=k,brown=n,buff=b,chocolate=h,gray=g,
                              green=r,orange=o,pink=p,purple=u,red=e, white=w,yellow=y

10. stalk-shape:              enlarging=e,tapering=t
```

```
11. stalk-root:               bulbous=b,club=c,cup=u,equal=e, rhizomorphs=z,rooted=r,missing=
12. stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s
13. stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s
14. stalk-color-above-ring:   brown=n,buff=b,cinnamon=c,gray=g,orange=o,
                              pink=p,red=e,white=w,yellow=y
15. stalk-color-below-ring:   brown=n,buff=b,cinnamon=c,gray=g,orange=o,
                              pink=p,red=e,white=w,yellow=y
16. veil-type:                partial=p,universal=u
17. veil-color:               brown=n,orange=o,white=w,yellow=y
18. ring-number:              none=n,one=o,two=t
19. ring-type:                cobwebby=c,evanescent=e,flaring=f,large=l,
                              none=n,pendant=p,sheathing=s,zone=z
20. spore-print-color:        black=k,brown=n,buff=b,chocolate=h,green=r,
                              orange=o,purple=u,white=w,yellow=y
21. population:               abundant=a,clustered=c,numerous=n, scattered=s,several=v,solita
22. habitat:                  grasses=g,leaves=l,meadows=m,paths=p, urban=u,waste=w,woods=d
```

Citations Mushroom. (1987). UCI Machine Learning Repository. https://doi.org/10.24432/C5959T.

```
install.packages("leaps")
```

```
mushrooms <- read.csv("Mushrooms - Sheet1.csv")
library(tidymodels)
library(tidyverse)
library(leaps)
library(glmnet)
```

```
no_veil <- subset(mushrooms, select = -c(veil))
sapply(lapply(no_veil, unique), length)
```

```
    poisonous       cap.shape     cap.surface       cap.color         bruises
            2               6               4              10               2
         odor gill.attachment    gill.spacing       gill.size      gill.color
            9               2               2               2              12
  stalk.shape      stalk.root        ss.above        ss.below        sc.above
            2               5               4               4               9
    sc.below      veil.color        ring.num      ring.type     spore.color
            9               4               3               5               9
   population         habitat
            6               7
```

```
no_veil$poisonous[no_veil$poisonous == 'p'] <- 1
no_veil$poisonous[no_veil$poisonous == 'e'] <- 0

no_veil$poisonous <- as.numeric(as.character(no_veil$poisonous))

##m1 <- glm(poisonous ~ .,
          ## data = no_veil,
          ## family = "binomial")
##summary(m1)

m_all <- regsubsets(poisonous ~ .,
                    data = no_veil,
                    nbest = 1, nvmax = 5, really.big=T)
```

Reordering variables and trying again:

```
m_all
```

```
Subset selection object
Call: regsubsets.formula(poisonous ~ ., data = no_veil, nbest = 1,
    nvmax = 5, really.big = T)
95 Variables  (and intercept)
                Forced in Forced out
cap.shapec          FALSE      FALSE
cap.shapef          FALSE      FALSE
cap.shapek          FALSE      FALSE
cap.shapes          FALSE      FALSE
cap.shapex          FALSE      FALSE
cap.surfaceg        FALSE      FALSE
cap.surfaces        FALSE      FALSE
cap.surfacey        FALSE      FALSE
cap.colorc          FALSE      FALSE
cap.colore          FALSE      FALSE
cap.colorg          FALSE      FALSE
cap.colorn          FALSE      FALSE
cap.colorp          FALSE      FALSE
cap.colorr          FALSE      FALSE
cap.coloru          FALSE      FALSE
cap.colorw          FALSE      FALSE
cap.colory          FALSE      FALSE
```

```
bruisest             FALSE      FALSE
odorc                FALSE      FALSE
odorf                FALSE      FALSE
odorl                FALSE      FALSE
odorm                FALSE      FALSE
odorn                FALSE      FALSE
odorp                FALSE      FALSE
odors                FALSE      FALSE
odory                FALSE      FALSE
gill.attachmentf     FALSE      FALSE
gill.spacingw        FALSE      FALSE
gill.sizen           FALSE      FALSE
gill.colore          FALSE      FALSE
gill.colorg          FALSE      FALSE
gill.colorh          FALSE      FALSE
gill.colork          FALSE      FALSE
gill.colorn          FALSE      FALSE
gill.coloro          FALSE      FALSE
gill.colorp          FALSE      FALSE
gill.colorr          FALSE      FALSE
gill.coloru          FALSE      FALSE
gill.colorw          FALSE      FALSE
gill.colory          FALSE      FALSE
stalk.shapet         FALSE      FALSE
stalk.rootb          FALSE      FALSE
stalk.rootc          FALSE      FALSE
stalk.roote          FALSE      FALSE
stalk.rootr          FALSE      FALSE
ss.abovek            FALSE      FALSE
ss.aboves            FALSE      FALSE
ss.abovey            FALSE      FALSE
ss.belowk            FALSE      FALSE
ss.belows            FALSE      FALSE
ss.belowy            FALSE      FALSE
sc.abovee            FALSE      FALSE
sc.aboveg            FALSE      FALSE
sc.aboven            FALSE      FALSE
sc.aboveo            FALSE      FALSE
sc.abovep            FALSE      FALSE
sc.abovew            FALSE      FALSE
sc.abovey            FALSE      FALSE
sc.belowe            FALSE      FALSE
sc.belowg            FALSE      FALSE
```

```
sc.belown           FALSE      FALSE
sc.belowp           FALSE      FALSE
sc.beloww           FALSE      FALSE
sc.belowy           FALSE      FALSE
veil.coloro         FALSE      FALSE
ring.numo           FALSE      FALSE
ring.typef          FALSE      FALSE
ring.typep          FALSE      FALSE
spore.colork        FALSE      FALSE
spore.colorn        FALSE      FALSE
spore.coloro        FALSE      FALSE
spore.colorr        FALSE      FALSE
spore.coloru        FALSE      FALSE
spore.colorw        FALSE      FALSE
spore.colory        FALSE      FALSE
populationc         FALSE      FALSE
populationn         FALSE      FALSE
populations         FALSE      FALSE
populationv         FALSE      FALSE
populationy         FALSE      FALSE
habitatg            FALSE      FALSE
habitatl            FALSE      FALSE
habitatm            FALSE      FALSE
habitatp            FALSE      FALSE
habitatu            FALSE      FALSE
sc.abovec           FALSE      FALSE
sc.belowc           FALSE      FALSE
sc.belowo           FALSE      FALSE
veil.colorw         FALSE      FALSE
veil.colory         FALSE      FALSE
ring.numt           FALSE      FALSE
ring.typel          FALSE      FALSE
ring.typen          FALSE      FALSE
spore.colorh        FALSE      FALSE
habitatw            FALSE      FALSE
1 subsets of each size up to 6
Selection Algorithm: exhaustive
```

```
summary(m_all)$which
```

```
  (Intercept) cap.shapec cap.shapef cap.shapek cap.shapes cap.shapex
```

```
1        TRUE      FALSE      FALSE      FALSE      FALSE      FALSE
2        TRUE      FALSE      FALSE      FALSE      FALSE      FALSE
3        TRUE      FALSE      FALSE      FALSE      FALSE      FALSE
4        TRUE      FALSE      FALSE      FALSE      FALSE      FALSE
5        TRUE      FALSE      FALSE      FALSE      FALSE      FALSE
6        TRUE      FALSE      FALSE      FALSE      FALSE      FALSE
  cap.surfaceg cap.surfaces cap.surfacey cap.colorc cap.colore cap.colorg
1        FALSE        FALSE        FALSE      FALSE      FALSE      FALSE
2        FALSE        FALSE        FALSE      FALSE      FALSE      FALSE
3        FALSE        FALSE        FALSE      FALSE      FALSE      FALSE
4        FALSE        FALSE        FALSE      FALSE      FALSE      FALSE
5        FALSE        FALSE        FALSE      FALSE      FALSE      FALSE
6        FALSE        FALSE        FALSE      FALSE      FALSE      FALSE
  cap.colorn cap.colorp cap.colorr cap.coloru cap.colorw cap.colory bruisest
1      FALSE      FALSE      FALSE      FALSE      FALSE      FALSE    FALSE
2      FALSE      FALSE      FALSE      FALSE      FALSE      FALSE    FALSE
3      FALSE      FALSE      FALSE      FALSE      FALSE      FALSE    FALSE
4      FALSE      FALSE      FALSE      FALSE      FALSE      FALSE    FALSE
5      FALSE      FALSE      FALSE      FALSE      FALSE      FALSE    FALSE
6      FALSE      FALSE      FALSE      FALSE      FALSE      FALSE    FALSE
  odorc odorf odorl odorm odorn odorp odors odory gill.attachmentf
1 FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE            FALSE
2 FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE            FALSE
3 FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE            FALSE
4 FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE            FALSE
5 FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE            FALSE
6 FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE            FALSE
  gill.spacingw gill.sizen gill.colore gill.colorg gill.colorh gill.colork
1         FALSE      FALSE       FALSE       FALSE       FALSE       FALSE
2         FALSE      FALSE       FALSE       FALSE       FALSE       FALSE
3         FALSE      FALSE       FALSE       FALSE       FALSE       FALSE
4         FALSE      FALSE       FALSE       FALSE       FALSE       FALSE
5         FALSE      FALSE       FALSE       FALSE       FALSE       FALSE
6         FALSE      FALSE       FALSE       FALSE       FALSE       FALSE
  gill.colorn gill.coloro gill.colorp gill.colorr gill.coloru gill.colorw
1       FALSE       FALSE       FALSE       FALSE       FALSE       FALSE
2       FALSE       FALSE       FALSE       FALSE       FALSE       FALSE
3       FALSE       FALSE       FALSE       FALSE       FALSE       FALSE
4       FALSE       FALSE       FALSE       FALSE       FALSE       FALSE
5       FALSE       FALSE       FALSE       FALSE       FALSE       FALSE
6       FALSE       FALSE       FALSE       FALSE       FALSE       FALSE
  gill.colory stalk.shapet stalk.rootb stalk.rootc stalk.roote stalk.rootr
1       FALSE        FALSE       FALSE       FALSE       FALSE       FALSE
```

```
2       FALSE        FALSE       FALSE        TRUE       FALSE       FALSE
3       FALSE        FALSE       FALSE        TRUE       FALSE        TRUE
4       FALSE        FALSE       FALSE        TRUE       FALSE        TRUE
5       FALSE        FALSE       FALSE        TRUE       FALSE        TRUE
6       FALSE        FALSE       FALSE        TRUE       FALSE        TRUE
  ss.abovek ss.aboves ss.abovey ss.belowk ss.belows ss.belowy sc.abovec
1     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
2     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
3     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
4     FALSE     FALSE     FALSE     FALSE     FALSE      TRUE     FALSE
5     FALSE     FALSE     FALSE     FALSE     FALSE      TRUE     FALSE
6     FALSE     FALSE     FALSE     FALSE     FALSE      TRUE     FALSE
  sc.abovee sc.aboveg sc.aboven sc.aboveo sc.abovep sc.abovew sc.abovey
1     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
2     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
3     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
4     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
5     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
6     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
  sc.belowc sc.belowe sc.belowg sc.belown sc.belowo sc.belowp sc.beloww
1     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
2     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
3     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
4     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
5     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
6     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
  sc.belowy veil.coloro veil.colorw veil.colory ring.numo ring.numt ring.typef
1     FALSE       FALSE       FALSE       FALSE     FALSE     FALSE      FALSE
2     FALSE       FALSE       FALSE       FALSE     FALSE     FALSE      FALSE
3     FALSE       FALSE       FALSE       FALSE     FALSE     FALSE      FALSE
4     FALSE       FALSE       FALSE       FALSE     FALSE     FALSE      FALSE
5     FALSE       FALSE       FALSE       FALSE     FALSE     FALSE      FALSE
6     FALSE       FALSE       FALSE       FALSE     FALSE     FALSE      FALSE
  ring.typel ring.typen ring.typep spore.colorh spore.colork spore.colorn
1      FALSE      FALSE      FALSE        FALSE        FALSE        FALSE
2      FALSE      FALSE      FALSE        FALSE        FALSE        FALSE
3      FALSE      FALSE      FALSE        FALSE        FALSE        FALSE
4      FALSE      FALSE      FALSE        FALSE        FALSE        FALSE
5      FALSE      FALSE      FALSE        FALSE        FALSE        FALSE
6      FALSE      FALSE      FALSE        FALSE        FALSE        FALSE
  spore.coloro spore.colorr spore.coloru spore.colorw spore.colory populationc
1        FALSE        FALSE        FALSE        FALSE        FALSE       FALSE
2        FALSE        FALSE        FALSE        FALSE        FALSE       FALSE
```

```
3        FALSE        FALSE        FALSE        FALSE        FALSE       FALSE
4        FALSE        FALSE        FALSE        FALSE        FALSE       FALSE
5        FALSE         TRUE        FALSE        FALSE        FALSE       FALSE
6        FALSE         TRUE         TRUE        FALSE        FALSE       FALSE
  populationn populations populationv populationy habitatg habitatl habitatm
1       FALSE       FALSE       FALSE       FALSE    FALSE    FALSE    FALSE
2       FALSE       FALSE       FALSE       FALSE    FALSE    FALSE    FALSE
3       FALSE       FALSE       FALSE       FALSE    FALSE    FALSE    FALSE
4       FALSE       FALSE       FALSE       FALSE    FALSE    FALSE    FALSE
5       FALSE       FALSE       FALSE       FALSE    FALSE    FALSE    FALSE
6       FALSE       FALSE       FALSE       FALSE    FALSE    FALSE    FALSE
  habitatp habitatu habitatw
1    FALSE    FALSE    FALSE
2    FALSE    FALSE    FALSE
3    FALSE    FALSE    FALSE
4    FALSE    FALSE    FALSE
5    FALSE    FALSE    FALSE
6    FALSE    FALSE    FALSE
```

```r
summary(m_all)$rsq
```

```
[1] 0.6170992 0.7850914 0.8701623 0.9041082 0.9387674 0.9614494
```

```r
install.packages("glmnet")
```

```r
library(glmnet)

y <- no_veil$poisonous
x <- model.matrix(poisonous ~ .,
                  data = no_veil)

m_lasso_cv <- cv.glmnet(x, y, alpha = 1)
best_lambda <- m_lasso_cv$lambda.min
best_lambda
```

```
[1] 0.0002523176
```

```r
m_best <- glmnet(x, y, alpha = 1, lambda = best_lambda)
m_best$beta
```

```
96 x 1 sparse Matrix of class "dgCMatrix"
                            s0
(Intercept)       .
cap.shapec         2.469270e-01
cap.shapef         .
cap.shapek         .
cap.shapes        -4.780328e-03
cap.shapex         .
cap.surfaceg       7.481027e-01
cap.surfaces       .
cap.surfacey       7.228459e-04
cap.colorc        -6.613558e-02
cap.colore         .
cap.colorg         .
cap.colorn        -9.789276e-04
cap.colorp         .
cap.colorr         .
cap.coloru         .
cap.colorw         2.720506e-03
cap.colory         .
bruisest           6.459105e-04
odorc              9.665179e-01
odorf              7.966668e-01
odorl             -4.046078e-03
odorm              5.659035e-02
odorn             -4.517302e-02
odorp              9.423022e-01
odors              7.963789e-01
odory              7.964045e-01
gill.attachmentf   .
gill.spacingw     -1.719982e-02
gill.sizen         .
gill.colore        .
gill.colorg        .
gill.colorh        .
gill.colork        .
gill.colorn        .
gill.coloro        .
gill.colorp        .
gill.colorr        .
gill.coloru        .
gill.colorw       -1.959178e-04
gill.colory        .
```

```
stalk.shapet      6.117153e-03
stalk.rootb      -7.871970e-03
stalk.rootc      -4.573025e-02
stalk.roote       7.148989e-03
stalk.rootr      -8.618155e-01
ss.abovek         2.321158e-03
ss.aboves         .
ss.abovey        -8.035657e-01
ss.belowk         6.740357e-04
ss.belows         .
ss.belowy         8.175232e-01
sc.abovec         3.588411e-02
sc.abovee         .
sc.aboveg         .
sc.aboven         .
sc.aboveo        -7.569320e-03
sc.abovep         1.198385e-04
sc.abovew         .
sc.abovey         4.736608e-01
sc.belowc         2.978803e-06
sc.belowe         .
sc.belowg         .
sc.belown        -9.342302e-03
sc.belowo        -1.576968e-02
sc.belowp         .
sc.beloww         1.350767e-03
sc.belowy         2.835038e-02
veil.coloro       .
veil.colorw       .
veil.colory       2.771655e-01
ring.numo         9.290282e-02
ring.numt        -4.116992e-02
ring.typef       -1.562118e-01
ring.typel        8.587265e-03
ring.typen        .
ring.typep        .
spore.colorh      1.555109e-01
spore.colork      .
spore.colorn     -6.014691e-04
spore.coloro      .
spore.colorr      1.135084e+00
spore.coloru     -2.520651e-02
spore.colorw      1.481708e-01
```

```
spore.colory      .
populationc       5.019038e-02
populationn      -7.732554e-04
populations       .
populationv       .
populationy       .
habitatg          .
habitatl          .
habitatm          .
habitatp          8.385491e-05
habitatu          .
habitatw         -6.500746e-02
```