Classification of Left and right brainers

Code

Venkat Koushik Muthyapu

October 21, 2020

0.1 R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

1 Loading the data

Hide

```
library(tidyverse)

datcc <- read_csv(
    "/Users/venkatkoushikmuthyapu/desktop/rakin/Phase3.csv"
  )
summary(datcc)
```

```
 Participant #     LB or Rb         HR Change Math test HR Change Memory Test
 Min.   : 1.00   Length:50          Min.   :-8.00       Min.   :-2.00
 1st Qu.:13.25   Class :character   1st Qu.: 1.00       1st Qu.: 1.00
 Median :25.50   Mode  :character   Median : 3.00       Median : 2.00
 Mean   :25.50                      Mean   : 2.72       Mean   : 2.68
 3rd Qu.:37.75                      3rd Qu.: 4.75       3rd Qu.: 4.00
 Max.   :50.00                      Max.   :11.00       Max.   :10.00
 PDC Math Test  PDC Memory Test Systolic Pressure Change Math Test
 Min.   :0.00   Min.   :-1.0    Min.   :-23.00
 1st Qu.:1.25   1st Qu.: 1.0    1st Qu.:  0.25
 Median :2.00   Median : 1.5    Median :  2.00
 Mean   :2.60   Mean   : 1.8    Mean   :  2.90
 3rd Qu.:3.00   3rd Qu.: 3.0    3rd Qu.:  5.00
 Max.   :7.00   Max.   : 9.0    Max.   : 17.00
 Systolic Pressure Change Memory Test Diastolic Pressure Change Math Test
 Min.   :-23.00                       Min.   :-30.00
 1st Qu.:  0.00                       1st Qu.: -2.75
 Median :  1.00                       Median :  1.50
 Mean   :  2.38                       Mean   :  3.16
 3rd Qu.:  5.75                       3rd Qu.:  7.75
 Max.   : 34.00                       Max.   : 31.00
 Diastolic Pressure Change Memory Test
 Min.   :-26.00
 1st Qu.: -5.25
 Median :  2.00
 Mean   :  0.90
 3rd Qu.:  6.75
 Max.   : 38.00
```

2 Renaming the variables and cleaning the data

Hide

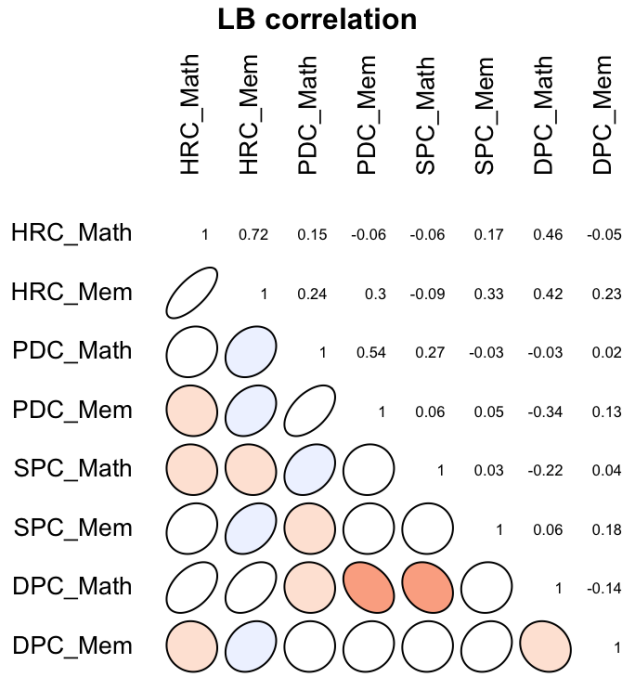
```
datcc <-
  datcc %>%
  dplyr::rename(
    LB_RB = `LB or Rb`
,   HRC_Math = `HR Change Math test`
,   HRC_Mem = `HR Change Memory Test`
,   PDC_Math = `PDC Math Test`
,   PDC_Mem = `PDC Memory Test`
,   SPC_Math = `Systolic Pressure Change Math Test`
,   SPC_Mem = `Systolic Pressure Change Memory Test`
,   DPC_Math = `Diastolic Pressure Change Math Test`
,   DPC_Mem = `Diastolic Pressure Change Memory Test`
)

datcc <-
datcc %>%
  select(-`Participant #`)%>%
  filter(  PDC_Mem < 7
         , SPC_Math > -20
         , SPC_Mem < 21) %>%
  mutate(
    LB_RB = factor(LB_RB)
  )
summary(datcc)
```

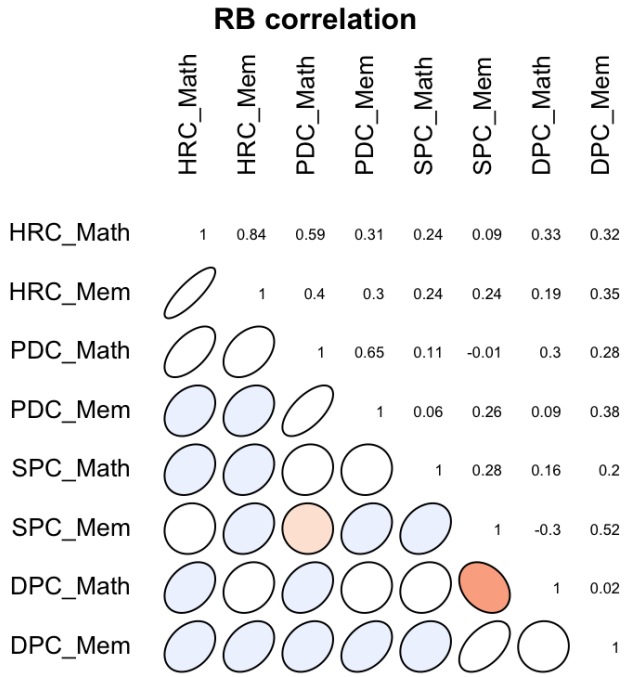
```
 LB_RB      HRC_Math         HRC_Mem          PDC_Math        PDC_Mem
 LB:25   Min.   :-8.000   Min.   :-2.000   Min.   :0.000   Min.   :-1.00
 RB:22   1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.:1.500   1st Qu.: 0.50
         Median : 3.000   Median : 2.000   Median :2.000   Median : 1.00
         Mean   : 2.936   Mean   : 2.766   Mean   :2.553   Mean   : 1.66
         3rd Qu.: 5.000   3rd Qu.: 4.000   3rd Qu.:3.000   3rd Qu.: 3.00
         Max.   :11.000   Max.   :10.000   Max.   :6.000   Max.   : 6.00
    SPC_Math          SPC_Mem          DPC_Math          DPC_Mem
 Min.   :-13.000   Min.   :-23.00   Min.   :-30.000   Min.   :-26.0000
 1st Qu.:  1.000   1st Qu.:  0.00   1st Qu.: -2.500   1st Qu.: -3.0000
 Median :  2.000   Median :  1.00   Median :  1.000   Median :  2.0000
 Mean   :  3.574   Mean   :  2.17   Mean   :  3.532   Mean   :  0.8723
 3rd Qu.:  5.000   3rd Qu.:  5.50   3rd Qu.:  8.500   3rd Qu.:  6.5000
 Max.   : 17.000   Max.   : 18.00   Max.   : 31.000   Max.   : 23.0000
```

3 Data exploration

Hide

```
# Scatterplot matrix
library(ggplot2)
library(GGally)
p <- ggpairs(datcc
, mapping = ggplot2::aes(colour = LB_RB, alpha = 0.5)
, progress=FALSE
)
print(p)
```





4 Model Selection

Hide

```
dat_datcc_d <- datcc %>% select(HRC_Math:DPC_Mem) # the data
dat_datcc_c <- datcc %>% pull(LB_RB)       # the classes

# start random number generator in same place for everyone
# and so that random partitions are the same each time code is run
set.seed(7)

#library(klaR)  # don't run this since it does library(MASS) and breaks select() f
rom dplyr
# Backward
step_datcc_b <-
  klaR::stepclass(
    dat_datcc_d
  , dat_datcc_c
  , method = "qda"
  , improvement = 0.001 # stop criterion: improvement less than
  , direction = "backward"
  , start.vars = colnames(dat_datcc_d)
  )
```

```
correctness rate: 0.5;  starting variables (8): HRC_Math, HRC_Mem, PDC_Math, PDC_M
em, SPC_Math, SPC_Mem, DPC_Math, DPC_Mem
correctness rate: 0.555;  out: "SPC_Mem";  variables (7): HRC_Math, HRC_Mem, PDC_M
ath, PDC_Mem, SPC_Math, DPC_Math, DPC_Mem
correctness rate: 0.615;  out: "DPC_Mem";  variables (6): HRC_Math, HRC_Mem, PDC_M
ath, PDC_Mem, SPC_Math, DPC_Math
correctness rate: 0.645;  out: "DPC_Math";  variables (5): HRC_Math, HRC_Mem, PDC_
Math, PDC_Mem, SPC_Math
correctness rate: 0.66;  out: "PDC_Math";  variables (4): HRC_Math, HRC_Mem, PDC_M
em, SPC_Math
correctness rate: 0.67;  out: "SPC_Math";  variables (3): HRC_Math, HRC_Mem, PDC_M
em

 hr.elapsed min.elapsed sec.elapsed
      0.000       0.000       3.158
```

Hide

```
## NOTE HERE
step_datcc_b$formula
```

```
dat_datcc_c ~ HRC_Math + HRC_Mem + PDC_Mem
<environment: 0x7f9ff5782630>
```

Hide

```
# estimated correct/error rates
step_datcc_b$result.pm
```

```
crossval.rate      apparent
    0.6700000     0.3191489
```

Hide

```
# Forward
step_datcc_f <-
  klaR::stepclass(
    dat_datcc_d
  , dat_datcc_c
  , method =  "qda"
  , improvement = 0.001 # stop criterion: improvement less than
  , direction = "forward"
  , start.vars = ""
  )
```

```
correctness rate: 0;  starting variables (0): ,
correctness rate: 0.575;  in: "DPC_Math";  variables (1): DPC_Math
correctness rate: 0.6;  in: "SPC_Math";  variables (2): DPC_Math, SPC_Math

 hr.elapsed min.elapsed sec.elapsed
      0.000       0.000       1.569
```

Hide

```
## NOTE HERE
step_datcc_f$formula
```

```
dat_datcc_c ~ SPC_Math + DPC_Math
<environment: 0x7f9fed8bfd70>
```

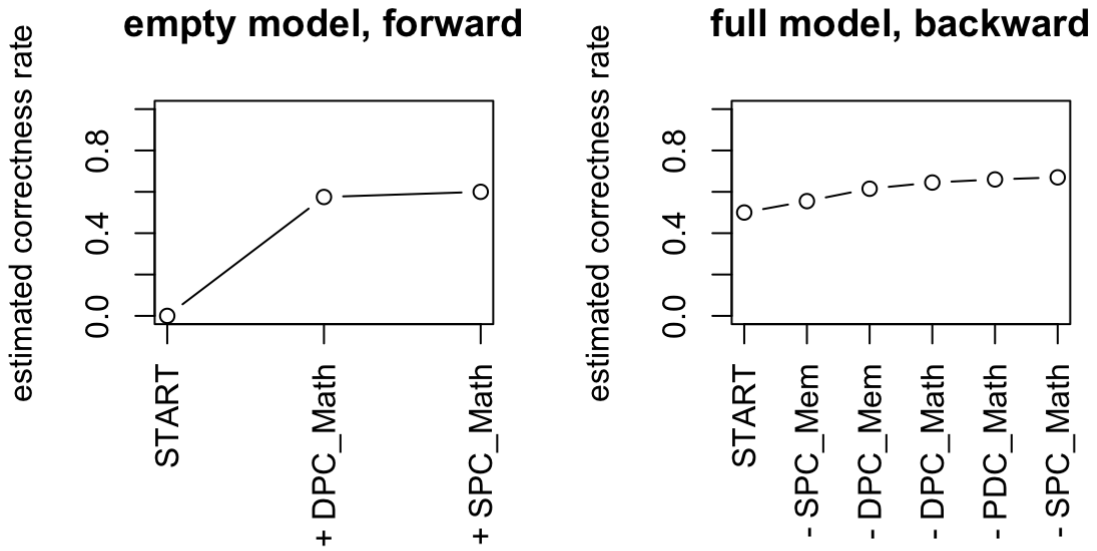
Hide

```
# estimated correct/error rates
step_datcc_f$result.pm
```

```
crossval.rate      apparent
    0.6000000     0.3829787
```

Hide

```
op <- par(no.readonly = TRUE) # the whole list of settable par's.
  # make wider left margin to fit contrast labels
  par(mfrow = c(1,2), mar = 0*rep(1, 4)) # order is c(bottom, left, top, right)
  plot(step_datcc_f, ylim = c(0, 1), main = "empty model, forward")
  plot(step_datcc_b, ylim = c(0, 1), main = "full model, backward")
```



Hide

```
par(op) # reset plotting options
```

5 Using The final model to do out qda

Hide

```
## NOTE HERE
# set the formula you're using here, then it will be used throughout the rest
datcc_formula <- step_datcc_b

# Select and print the final model
#library(MASS)  # don't run library(MASS) because it breaks select() from dplyr
qda_datcc_final <-
  MASS::qda(LB_RB ~ HRC_Math + HRC_Mem + PDC_Mem
            , data = datcc
  )
qda_datcc_final
```

```
Call:
qda(LB_RB ~ HRC_Math + HRC_Mem + PDC_Mem, data = datcc)

Prior probabilities of groups:
       LB        RB
0.5319149 0.4680851

Group means:
   HRC_Math  HRC_Mem  PDC_Mem
LB 1.840000 2.240000 1.640000
RB 4.181818 3.363636 1.681818
```

Hide

```
qda_datcc_final$result.pm
```

```
NULL
```

6 traing using CV and making Confusion matrix

Hide

```
# CV = TRUE does jackknife (leave-one-out) crossvalidation
#library(MASS)  # don't run library(MASS) because it breaks select() from dplyr
qda_datcc_cv <-
  MASS::qda(LB_RB ~ HRC_Math + HRC_Mem + PDC_Mem
            , data = datcc
  , CV = TRUE
  )
#qda_datcc_cv

# Create a table of classification and posterior probabilities for each observatio
n
classify_datcc <-
  data.frame(
    Source = datcc$LB_RB
  , class = qda_datcc_cv$class
  , error = ""
  , round(qda_datcc_cv$posterior, 3)
  )
colnames(classify_datcc) <-
  c(
    "Source"
  , "class"
  , "error"
  , paste("post", colnames(qda_datcc_cv$posterior), sep="_")
  )

# error column
classify_datcc$error <-
  as.character(classify_datcc$error)
classify_agree <-
  as.character(as.numeric(datcc$LB_RB) - as.numeric(qda_datcc_cv$class))
# print table
  as.character(as.numeric(datcc$LB_RB) - as.numeric(qda_datcc_cv$class))
```

```
 [1] "0"  "-1" "1"  "-1" "0"  "0"  "0"  "-1" "-1" "-1" "0"  "1"  "1"  "-1" "0"
[16] "0"  "0"  "1"  "1"  "0"  "0"  "0"  "-1" "1"  "0"  "-1" "0"  "1"  "0"  "0"
[31] "0"  "0"  "0"  "0"  "-1" "-1" "-1" "0"  "0"  "0"  "0"  "-1" "0"  "0"  "0"
[46] "0"  "1"
```

Hide

```
classify_datcc$error[!(classify_agree == 0)] <-
  classify_agree[!(classify_agree == 0)]
# print table
#classify_sjrs

# A list of classification statistics
library(caret)
confusionMatrix(
    data      = qda_datcc_cv$class # predictions
  , reference = datcc$LB_RB   # true labels
  , mode      = "sens_spec"       # restrict output to relevant summaries
)
```

```
Confusion Matrix and Statistics

          Reference
Prediction LB RB
        LB 13  8
        RB 12 14

               Accuracy : 0.5745
                 95% CI : (0.4218, 0.7174)
    No Information Rate : 0.5319
    P-Value [Acc > NIR] : 0.3317

                  Kappa : 0.1547

 Mcnemar's Test P-Value : 0.5023

            Sensitivity : 0.5200
            Specificity : 0.6364
         Pos Pred Value : 0.6190
         Neg Pred Value : 0.5385
             Prevalence : 0.5319
         Detection Rate : 0.2766
   Detection Prevalence : 0.4468
      Balanced Accuracy : 0.5782

       'Positive' Class : LB
```

Hide

```
classify_datcc
```

	Source	class	error	post_LB	post_RB
1	RB	RB		0.332	0.668
2	LB	RB	-1	0.381	0.619
3	RB	LB	1	0.997	0.003
4	LB	RB	-1	0.205	0.795
5	RB	RB		0.411	0.589
6	RB	RB		0.265	0.735
7	RB	RB		0.383	0.617
8	LB	RB	-1	0.225	0.775
9	LB	RB	-1	0.167	0.833
10	LB	RB	-1	0.477	0.523
11	LB	LB		0.963	0.037
12	RB	LB	1	0.726	0.274
13	RB	LB	1	0.531	0.469
14	LB	RB	-1	0.458	0.542
15	RB	RB		0.317	0.683
16	LB	LB		0.708	0.292
17	RB	RB		0.391	0.609
18	RB	LB	1	0.604	0.396
19	RB	LB	1	0.676	0.324
20	RB	RB		0.320	0.680
21	LB	LB		0.828	0.172
22	RB	RB		0.455	0.545
23	LB	RB	-1	0.469	0.531
24	RB	LB	1	0.533	0.467
25	LB	LB		0.563	0.437
26	LB	RB	-1	0.461	0.539
27	RB	RB		0.388	0.612
28	RB	LB	1	0.536	0.464
29	LB	LB		0.960	0.040
30	LB	LB		0.677	0.323
31	RB	RB		0.357	0.643
32	LB	LB		0.608	0.392
33	LB	LB		0.686	0.314
34	LB	LB		0.997	0.003
35	LB	RB	-1	0.249	0.751
36	LB	RB	-1	0.249	0.751
37	LB	RB	-1	0.407	0.593
38	LB	LB		0.537	0.463
39	RB	RB		0.294	0.706
40	LB	LB		0.988	0.012
41	LB	LB		0.964	0.036
42	LB	RB	-1	0.390	0.610
43	LB	LB		1.000	0.000
44	RB	RB		0.285	0.715
45	RB	RB		0.315	0.685
46	RB	RB		0.302	0.698
47	RB	LB	1	0.993	0.007

7 Testing

Hide

```
dat_test <- read_csv(
    "/Users/venkatkoushikmuthyapu/desktop/rakin/Phase4.csv"
  )

dat_test <-
  dat_test %>%
  dplyr::rename(
    LB_RB = `LB or Rb`
,   HRC_Math = `HR Change Math test`
,   HRC_Mem = `HR Change Memory Test`
,   PDC_Math = `PDC Math Test`
,   PDC_Mem = `PDC Memory Test`
,   SPC_Math = `Systolic Pressure Change Math Test`
,   SPC_Mem = `Systolic Pressure Change Memory Test`
,   DPC_Math = `Diastolic Pressure Change Math Test`
,   DPC_Mem = `Diastolic Pressure Change Memory Test`
)

dat_test <-
dat_test %>%
  select(-`Participant #`)%>%
  mutate(
    LB_RB = factor(LB_RB)
  )


final.pred <- predict(
   qda_datcc_final
 , newdata = dat_test
)


classify_dat_test <-
  data.frame(
    Source = dat_test$LB_RB
  , class = final.pred$class
  , round(final.pred$posterior, 3)
  )
colnames(classify_dat_test) <-
  c(
    "Source"
  , "class"
  , paste("post", colnames(final.pred$posterior), sep="_")
  )

library(caret)
confusionMatrix(
    data      = final.pred$class # predictions
```

```
  , reference = dat_test$LB_RB   # true labels
  , mode      = "sens_spec"       # restrict output to relevant summaries
)
```

```
Confusion Matrix and Statistics

          Reference
Prediction LB RB
        LB  2  1
        RB  0  1

               Accuracy : 0.75
                 95% CI : (0.1941, 0.9937)
    No Information Rate : 0.5
    P-Value [Acc > NIR] : 0.3125

                  Kappa : 0.5

 Mcnemar's Test P-Value : 1.0000

            Sensitivity : 1.0000
            Specificity : 0.5000
         Pos Pred Value : 0.6667
         Neg Pred Value : 1.0000
             Prevalence : 0.5000
         Detection Rate : 0.5000
   Detection Prevalence : 0.7500
      Balanced Accuracy : 0.7500

       'Positive' Class : LB
```

Hide

```
classify_dat_test
```

```
  Source class post_LB post_RB
1     RB    RB   0.340   0.660
2     LB    LB   0.845   0.155
3     LB    LB   0.708   0.292
4     RB    LB   0.923   0.077
```