Analysis-Brain-AllData

Setup

Enviroment Setup

```
#setwd('~/Desktop/andlab/code')
odf <- read.csv('./brain_cb.csv')
odf <- odf[complete.cases(odf), ]

library('caret')
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
#setwd('~/Desktop/andlab/code')
```

Function Define

```
process <- function(df, reg, upr, plot=FALSE){

  newdf <- data.frame('y' = df$aggressive_sumscore)
  newdf$fitted <- reg$fitted.values

  newdf$class <- ifelse(newdf$y <= (-2.02065), 'L', ifelse(newdf$y>(upr),'M','H'))
  #change == -2.020650971 to change <= -2.020650971 to solve prediction for 0 is 0
  newdf$pred_class <-ifelse(newdf$fitted <= (-2.02065), 'L', ifelse(newdf$fitted>(up
r),'M','H'))

  print(mean(newdf$pred_class == newdf$class))
  #cm <- confusionMatrix(factor(newdf$pred_class, levels = 0:2), factor(newdf$class,
levels = 0:2) )
  cm <- confusionMatrix(table(newdf$class,newdf$pred_class))
  if(plot)
  {
    par(mfrow = c(2, 2))
    plot(reg)
  }
  return(cm=cm)
}

process_2_level <- function(df, reg, plot=FALSE){

  newdf <- data.frame('y' = df$aggressive_sumscore)
  newdf$fitted <- reg$fitted.values

  newdf$class <- ifelse(newdf$y <= (-2.020650971), 'L', 'H')
  #change == -2.020650971 to change <= -2.020650971 to solve prediction for 0 is 0
  newdf$pred_class <-ifelse(newdf$fitted <= (-2.020650971), 'L', 'H')

  print(mean(newdf$pred_class == newdf$class))
  #cm <- confusionMatrix(factor(newdf$pred_class, levels = 0:2), factor(newdf$class,
levels = 0:2) )
  cm <- confusionMatrix(table(newdf$class,newdf$pred_class))
  if(plot)
  {
    par(mfrow = c(2, 2))
    plot(reg)
  }
  return(cm=cm)
}
```

Analysis W/ All Data, All Attributes

Attribute selection

```
df <- subset(odf, select=-c(prosocial_child, prosocial_parent, subjectkey))
```

OLS

Model Selection

```
reg <- lm(df$aggressive_sumscore ~ ., data=df)
```

Result

```
summary(reg)
```

```
##
## Call:
## lm(formula = df$aggressive_sumscore ~ ., data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.644 -1.914 -1.471  0.386 32.773
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.224087   0.089894   2.493  0.01269 *
## rsfmri_cor_ngd_cerc_scs_thp -0.584585   0.285513  -2.047  0.04063 *
## rsfmri_cor_ngd_cerc_scs_cde -0.053300   0.345053  -0.154  0.87724
## rsfmri_cor_ngd_cerc_scs_pt  -0.426288   0.301887  -1.412  0.15795
## rsfmri_cor_ngd_cerc_scs_hp  -0.483225   0.305502  -1.582  0.11374
## rsfmri_cor_ngd_cerc_scs_ag  -0.695918   0.216092  -3.220  0.00128 **
## rsfmri_cor_ngd_cerc_scs_aa   0.310640   0.238658   1.302  0.19308
## rsfmri_cor_ngd_cerc_scs_bs   0.008034   0.238067   0.034  0.97308
## rsfmri_cor_ngd_df_scs_thp   -0.124255   0.440536  -0.282  0.77791
## rsfmri_cor_ngd_df_scs_cde    0.309316   0.288583   1.072  0.28381
## rsfmri_cor_ngd_df_scs_pt     0.367106   0.346409   1.060  0.28928
## rsfmri_cor_ngd_df_scs_hp    -0.559522   0.281512  -1.988  0.04688 *
## rsfmri_cor_ngd_df_scs_ag     0.118085   0.310568   0.380  0.70379
## rsfmri_cor_ngd_df_scs_aa    -0.133659   0.199450  -0.670  0.50279
## rsfmri_cor_ngd_df_scs_bs    -0.085437   0.153257  -0.557  0.57722
## rsfmri_cor_ngd_dsa_scs_thp  -0.361705   0.266768  -1.356  0.17516
## rsfmri_cor_ngd_dsa_scs_cde  -0.175751   0.301879  -0.582  0.56045
## rsfmri_cor_ngd_dsa_scs_pt    0.153157   0.382552   0.400  0.68890
## rsfmri_cor_ngd_dsa_scs_hp   -0.132386   0.412192  -0.321  0.74808
## rsfmri_cor_ngd_dsa_scs_ag    0.912593   0.435551   2.095  0.03617 *
## rsfmri_cor_ngd_dsa_scs_aa   -0.293120   0.294888  -0.994  0.32024
## rsfmri_cor_ngd_dsa_scs_bs   -0.570311   0.275608  -2.069  0.03854 *
## rsfmri_cor_ngd_fopa_scs_thp  0.144710   0.254378   0.569  0.56945
## rsfmri_cor_ngd_fopa_scs_cde  0.072505   0.454776   0.159  0.87333
## rsfmri_cor_ngd_fopa_scs_pt  -0.309079   0.301522  -1.025  0.30536
## rsfmri_cor_ngd_fopa_scs_hp  -0.245102   0.294756  -0.832  0.40569
## rsfmri_cor_ngd_fopa_scs_ag  -0.396177   0.287707  -1.377  0.16854
## rsfmri_cor_ngd_fopa_scs_aa   0.053215   0.330627   0.161  0.87213
## rsfmri_cor_ngd_fopa_scs_bs   0.035860   0.177372   0.202  0.83979
## rsfmri_cor_ngd_sa_scs_thp    0.000434   0.279149   0.002  0.99876
## rsfmri_cor_ngd_sa_scs_cde   -0.048681   0.229523  -0.212  0.83204
## rsfmri_cor_ngd_sa_scs_pt    -0.700347   0.343962  -2.036  0.04176 *
## rsfmri_cor_ngd_sa_scs_hp     0.562652   0.262898   2.140  0.03236 *
## rsfmri_cor_ngd_sa_scs_ag     0.204990   0.284836   0.720  0.47174
## rsfmri_cor_ngd_sa_scs_aa     0.029209   0.454840   0.064  0.94880
## rsfmri_cor_ngd_sa_scs_bs     0.162003   0.238182   0.680  0.49641
## rsfmri_cor_ngd_vta_scs_thp  -0.113448   0.248408  -0.457  0.64790
## rsfmri_cor_ngd_vta_scs_cde  -0.285884   0.359014  -0.796  0.42587
## rsfmri_cor_ngd_vta_scs_pt   -0.049497   0.371129  -0.133  0.89391
## rsfmri_cor_ngd_vta_scs_hp   -0.502686   0.304682  -1.650  0.09900 .
## rsfmri_cor_ngd_vta_scs_ag   -0.621755   0.369771  -1.681  0.09270 .
```

```
## rsfmri_cor_ngd_vta_scs_aa    0.259894   0.280487   0.927  0.35417
## rsfmri_cor_ngd_vta_scs_bs    0.052903   0.410540   0.129  0.89747
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.479 on 11041 degrees of freedom
## Multiple R-squared:  0.008874,   Adjusted R-squared:  0.005104
## F-statistic: 2.354 on 42 and 11041 DF,  p-value: 1.894e-06
```

```
cm <- process(df, reg, -0.73254990)
```

```
## [1] 0.345904
```

```
print(cm)
```

```
## Confusion Matrix and Statistics
##
##
##        H    L    M
##   H   46    1 2852
##   L   57    2 4285
##   M   50    5 3786
##
## Overall Statistics
##
##                Accuracy : 0.3459
##                  95% CI : (0.337, 0.3548)
##     No Information Rate : 0.9855
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 8e-04
##
##  Mcnemar's Test P-Value : <2e-16
##
## Statistics by Class:
##
##                      Class: H  Class: L Class: M
## Sensitivity           0.30065 0.2500000  0.34661
## Specificity           0.73900 0.6079812  0.65839
## Pos Pred Value        0.01587 0.0004604  0.98568
## Neg Pred Value        0.98693 0.9991098  0.01463
## Prevalence            0.01380 0.0007218  0.98547
## Detection Rate        0.00415 0.0001804  0.34157
## Detection Prevalence  0.26155 0.3919163  0.34654
## Balanced Accuracy     0.51983 0.4289906  0.50250
```

```
print(cm$byClass)
```

```
##          Sensitivity Specificity Pos Pred Value Neg Pred Value    Precision
## Class: H   0.3006536   0.7389992   0.0158675405     0.98692731 0.0158675405
## Class: L   0.2500000   0.6079812   0.0004604052     0.99910979 0.0004604052
## Class: M   0.3466081   0.6583851   0.9856808123     0.01463482 0.9856808123
##             Recall           F1   Prevalence Detection Rate
## Class: H 0.3006536 0.0301441678 0.0138036810   0.0041501263
## Class: L 0.2500000 0.0009191176 0.0007217611   0.0001804403
## Class: M 0.3466081 0.5128691412 0.9854745579   0.3415734392
##          Detection Prevalence Balanced Accuracy
## Class: H            0.2615482         0.5198264
## Class: L            0.3919163         0.4289906
## Class: M            0.3465355         0.5024966
```

OLS W/ Interaction

Model Selection

```
reg <- lm(df$aggressive_sumscore ~ . + .^2, data=df)
```

Result

```
#summary(reg)
cm <- process(df, reg, -0.73254990)
```

```
## [1] 0.3612414
```

```
print(cm)
```

```
## Confusion Matrix and Statistics
##
##
##        H    L    M
##   H  393   92 2414
##   L  561  165 3618
##   M  357   38 3446
##
## Overall Statistics
##
##                Accuracy : 0.3612
##                  95% CI : (0.3523, 0.3703)
##     No Information Rate : 0.8551
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.0356
##
##  Mcnemar's Test P-Value : <2e-16
##
## Statistics by Class:
##
##                      Class: H Class: L Class: M
## Sensitivity           0.29977  0.55932   0.3636
## Specificity           0.74358  0.61266   0.7540
## Pos Pred Value        0.13556  0.03798   0.8972
## Neg Pred Value        0.88784  0.98071   0.1672
## Prevalence            0.11828  0.02661   0.8551
## Detection Rate        0.03546  0.01489   0.3109
## Detection Prevalence  0.26155  0.39192   0.3465
## Balanced Accuracy     0.52168  0.58599   0.5588
```

```
print(cm$byClass)
```

```
##          Sensitivity Specificity Pos Pred Value Neg Pred Value  Precision
## Class: H   0.2997712   0.7435792     0.13556399      0.8878436 0.13556399
## Class: L   0.5593220   0.6126610     0.03798343      0.9807122 0.03798343
## Class: M   0.3635788   0.7540473     0.89716220      0.1671959 0.89716220
##             Recall         F1 Prevalence Detection Rate Detection Prevalence
## Class: H 0.2997712 0.18669834 0.11827860     0.03545651            0.2615482
## Class: L 0.5593220 0.07113602 0.02661494     0.01488632            0.3919163
## Class: M 0.3635788 0.51745627 0.85510646     0.31089859            0.3465355
##          Balanced Accuracy
## Class: H         0.5216752
## Class: L         0.5859915
## Class: M         0.5588131
```

Stepwise

Model Selection

```
null <- lm(aggressive_sumscore ~ 1, data=df)
full <- lm(aggressive_sumscore ~ ., data=df)
reg <- step(null, scope=formula(full), direction="forward", k=log(nrow(df)), trace=0)
```

Result

```
summary(reg)
```

```
##
## Call:
## lm(formula = aggressive_sumscore ~ rsfmri_cor_ngd_cerc_scs_ag +
##     rsfmri_cor_ngd_vta_scs_hp, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.532 -1.943 -1.514  0.376 33.418
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.05564    0.04032   1.380 0.167681
## rsfmri_cor_ngd_cerc_scs_ag -0.85628    0.18915  -4.527 6.04e-06 ***
## rsfmri_cor_ngd_vta_scs_hp  -0.86647    0.26138  -3.315 0.000919 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.483 on 11081 degrees of freedom
## Multiple R-squared:  0.002821,   Adjusted R-squared:  0.002641
## F-statistic: 15.68 on 2 and 11081 DF,  p-value: 1.59e-07
```

```
cm <- process(df, reg, -0.73254990)
```

```
## [1] 0.3447311
```

```
print(cm)
```

```
## Confusion Matrix and Statistics
##
##
##        H    L    M
##   H    5    0 2894
##   L    9    0 4335
##   M   24    1 3816
##
## Overall Statistics
##
##                Accuracy : 0.3447
##                  95% CI : (0.3359, 0.3537)
##     No Information Rate : 0.9965
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : -0.0023
##
##  Mcnemar's Test P-Value : <2e-16
##
## Statistics by Class:
##
##                       Class: H  Class: L Class: M
## Sensitivity          0.1315789 0.000e+00 0.345496
## Specificity          0.7380047 6.080e-01 0.358974
## Pos Pred Value       0.0017247 0.000e+00 0.993491
## Neg Pred Value       0.9959682 9.999e-01 0.001933
## Prevalence           0.0034284 9.022e-05 0.996481
## Detection Rate       0.0004511 0.000e+00 0.344280
## Detection Prevalence 0.2615482 3.919e-01 0.346536
## Balanced Accuracy    0.4347918 3.040e-01 0.352235
```

```
print(cm$byClass)
```

```
##          Sensitivity Specificity Pos Pred Value Neg Pred Value   Precision
## Class: H   0.1315789   0.7380047    0.001724733    0.995968235 0.001724733
## Class: L   0.0000000   0.6080484    0.000000000    0.999851632 0.000000000
## Class: M   0.3454957   0.3589744    0.993491278    0.001932901 0.993491278
##             Recall          F1   Prevalence Detection Rate Detection Prevalence
## Class: H 0.1315789 0.003404835 3.428365e-03   0.0004511007            0.2615482
## Class: L 0.0000000         NaN 9.022014e-05   0.0000000000            0.3919163
## Class: M 0.3454957 0.512696493 9.964814e-01   0.3442800433            0.3465355
##          Balanced Accuracy
## Class: H         0.4347918
## Class: L         0.3040242
## Class: M         0.3522350
```

Stepwise W/ Interaction

Model Selection

```
null <- lm(aggressive_sumscore ~ 1, data=df)
full <- lm(aggressive_sumscore ~ . + .^2, data=df)
reg <- step(null, scope=formula(full), direction="forward", k=log(nrow(df)), trace=0)
```

Result

```
summary(reg)
```

```
##
## Call:
## lm(formula = aggressive_sumscore ~ rsfmri_cor_ngd_cerc_scs_ag +
##     rsfmri_cor_ngd_vta_scs_hp, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.532 -1.943 -1.514  0.376 33.418
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.05564    0.04032   1.380 0.167681
## rsfmri_cor_ngd_cerc_scs_ag -0.85628    0.18915  -4.527 6.04e-06 ***
## rsfmri_cor_ngd_vta_scs_hp  -0.86647    0.26138  -3.315 0.000919 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.483 on 11081 degrees of freedom
## Multiple R-squared:  0.002821,   Adjusted R-squared:  0.002641
## F-statistic: 15.68 on 2 and 11081 DF,  p-value: 1.59e-07
```

```
cm <- process(df, reg, -0.73254990)
```

```
## [1] 0.3447311
```

```
print(cm)
```

```
## Confusion Matrix and Statistics
##
##
##        H    L    M
##   H    5    0 2894
##   L    9    0 4335
##   M   24    1 3816
##
## Overall Statistics
##
##                Accuracy : 0.3447
##                  95% CI : (0.3359, 0.3537)
##     No Information Rate : 0.9965
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : -0.0023
##
##  Mcnemar's Test P-Value : <2e-16
##
## Statistics by Class:
##
##                       Class: H  Class: L Class: M
## Sensitivity          0.1315789 0.000e+00 0.345496
## Specificity          0.7380047 6.080e-01 0.358974
## Pos Pred Value       0.0017247 0.000e+00 0.993491
## Neg Pred Value       0.9959682 9.999e-01 0.001933
## Prevalence           0.0034284 9.022e-05 0.996481
## Detection Rate       0.0004511 0.000e+00 0.344280
## Detection Prevalence 0.2615482 3.919e-01 0.346536
## Balanced Accuracy    0.4347918 3.040e-01 0.352235
```

```
print(cm$byClass)
```

```
##          Sensitivity Specificity Pos Pred Value Neg Pred Value   Precision
## Class: H   0.1315789   0.7380047    0.001724733    0.995968235 0.001724733
## Class: L   0.0000000   0.6080484    0.000000000    0.999851632 0.000000000
## Class: M   0.3454957   0.3589744    0.993491278    0.001932901 0.993491278
##             Recall          F1   Prevalence Detection Rate Detection Prevalence
## Class: H 0.1315789 0.003404835 3.428365e-03   0.0004511007            0.2615482
## Class: L 0.0000000         NaN 9.022014e-05   0.0000000000            0.3919163
## Class: M 0.3454957 0.512696493 9.964814e-01   0.3442800433            0.3465355
##          Balanced Accuracy
## Class: H         0.4347918
## Class: L         0.3040242
## Class: M         0.3522350
```