Machine Learning Methods

Jieqi Tu (jt3098)

2/19/2020

Import dataset

```
# Import dataset
CL = readxl::read_excel("./data_new/ABC_Cord Blood_Metabolomics_CL data_15Jan2020.xlsx")
BA = readxl::read_excel("./data_new/ABC_Cord Blood_Metabolomics_BA data_15Jan2020.xlsx")
PM = readxl::read_excel("./data_new/ABC_Cord Blood_Metabolomics_PM data_15Jan2020.xlsx")
```

```
## New names:
## * lactamide -> lactamide...94
## * lactamide -> lactamide...95
```

```
OL =  readxl::read_excel("./data_new/ABC_Cord Blood_Metabolomics_OL data_15Jan2020.xlsx")
```

Convert 0 to half of the minimum values

```
CL_data = CL[11:491]
BA_data = BA[11:266]
PM_data = PM[11:193]
OL_data = OL[11:81]
CL_data[CL_data == 0] = NA
BA_data[BA_data == 0] = NA
PM_data[PM_data == 0] = NA
OL_data[OL_data == 0] = NA
CL_min = sapply(CL_data[1:481], function(x) min(x, na.rm = T))
BA_min = sapply(BA_data[1:256], function(x) min(x, na.rm = T))
PM_min = sapply(PM_data[1:183], function(x) min(x, na.rm = T))
OL_min = sapply(OL_data[1:71], function(x) min(x, na.rm = T))

CL_data = CL[11:491]
BA_data = BA[11:266]
PM_data = PM[11:193]
OL_data = OL[11:81]
# Convert 0 to half of the minimum value
for(i in 1:481) {
  CL_data[i][CL_data[i]==0] = 0.5*CL_min[i]
}
```

```r
for(i in 1:256) {
  BA_data[i][BA_data[i]==0] = 0.5*BA_min[i]
}

for(i in 1:183) {
  PM_data[i][PM_data[i]==0] = 0.5*PM_min[i]
}

for(i in 1:71) {
  OL_data[i][OL_data[i]==0] = 0.5*OL_min[i]
}

CL_info = CL[1:10]
CL_new = cbind.data.frame(CL_info, CL_data)

BA_info = BA[1:10]
BA_new = cbind.data.frame(BA_info, BA_data)

PM_info = PM[1:10]
PM_new = cbind.data.frame(PM_info, PM_data)

OL_info = OL[1:10]
OL_new = cbind.data.frame(OL_info, OL_data)
```

Scaling and Transformation

```r
# Divide variables by the sd of control groups
BA_c = BA_new %>%
  group_by(Strata) %>%
  filter(PROT_ASD_2015 == 0)
CL_c = CL_new %>%
  group_by(Strata) %>%
  filter(PROT_ASD_2015 == 0)
PM_c = PM_new %>%
  group_by(Strata) %>%
  filter(PROT_ASD_2015 == 0)
OL_c = OL_new %>%
  group_by(Strata) %>%
  filter(PROT_ASD_2015 == 0)
sd_BA = sapply(BA_c[11:266], function(x) sd(x))
sd_CL = sapply(CL_c[11:491], function(x) sd(x))
sd_PM = sapply(PM_c[11:193], function(x) sd(x))
sd_OL = sapply(OL_c[11:81], function(x) sd(x))

# Divide the standard deviation
for(i in 1:256) {
  BA_new[i+10] = BA_new[i+10]/sd_BA[i]
}
for(i in 1:481) {
  CL_new[i+10] = CL_new[i+10]/sd_CL[i]
}
```

```r
for(i in 1:183) {
  PM_new[i+10] = PM_new[i+10]/sd_PM[i]
}
for(i in 1:71) {
  OL_new[i+10] = OL_new[i+10]/sd_OL[i]
}

# Log Transformation
BA_log = BA_new
CL_log = CL_new
PM_log = PM_new
OL_log = OL_new
for(i in 1:256) {
  BA_log[i+10] = log10(BA_new[i+10])
}

for(i in 1:481) {
  CL_log[i+10] = log10(CL_new[i+10])
}

for(i in 1:183) {
  PM_log[i+10] = log10(PM_new[i+10])
}

for(i in 1:71) {
  OL_log[i+10] = log10(OL_new[i+10])
}
```

Classification Methods

Initial Setting

```r
# Initial setting for machine learning methods
ctrl = trainControl(method = "repeatedcv", number = 5,
                    summaryFunction = twoClassSummary,
                    classProbs = TRUE)
```

Linear Methods for Classification

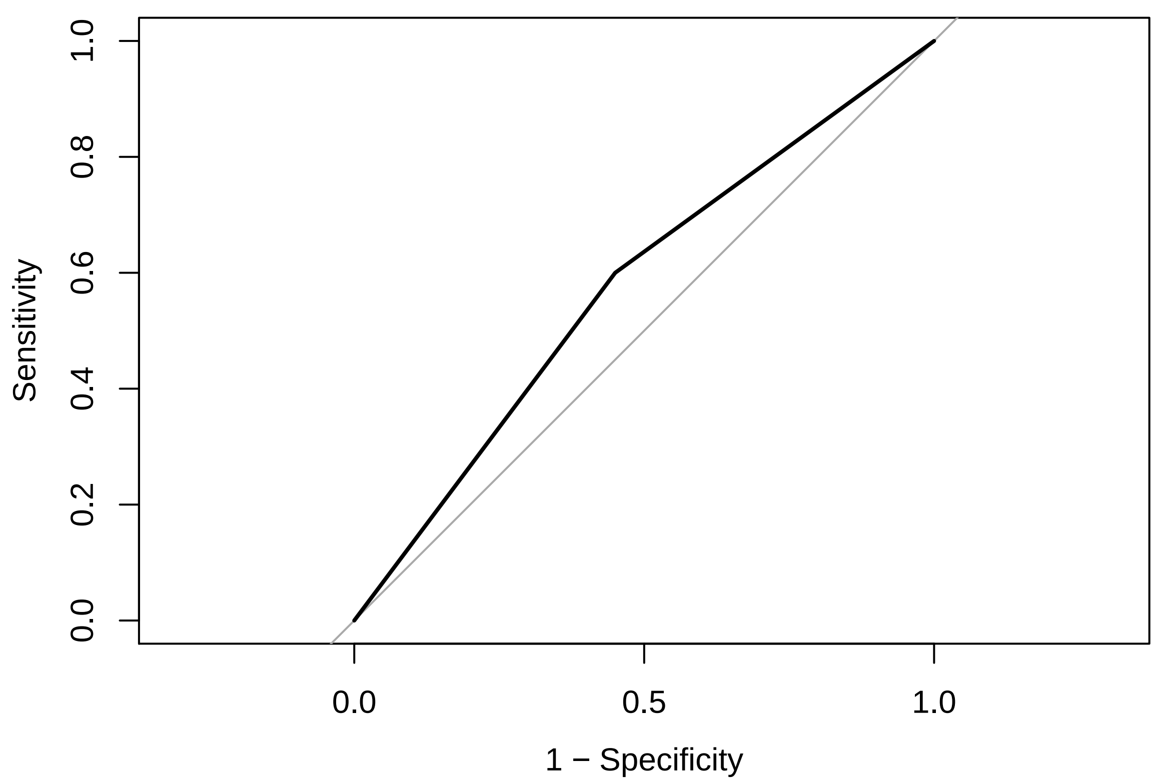
```r
# Delete matching variables
CL_analysis = CL_log[,-c(1, 2, 3, 4, 5, 6, 7, 8, 10)]

# Divide data into two parts
set.seed(1029)
rowTrain = createDataPartition(y = CL_analysis$PROT_ASD_2015,
                               p = 0.8,
                               list = FALSE)
```

```r
# Rename the outcome variable
CL_analysis$PROT_ASD_2015 = ifelse(CL_analysis$PROT_ASD_2015 == "1", "Positive", "Negative")

# Rename the outcome variable
model.glm = train(x = CL_analysis[rowTrain, 2:482],
                  y = CL_analysis$PROT_ASD_2015[rowTrain],
                  method = "glm",
                  metric = "ROC",
                  trControl = ctrl)

# Test performance
glm.pred = predict(model.glm, newdata = CL_analysis[-rowTrain,], type = "prob")[,2]
roc.glm = roc(CL_analysis$PROT_ASD_2015[-rowTrain], glm.pred)
```

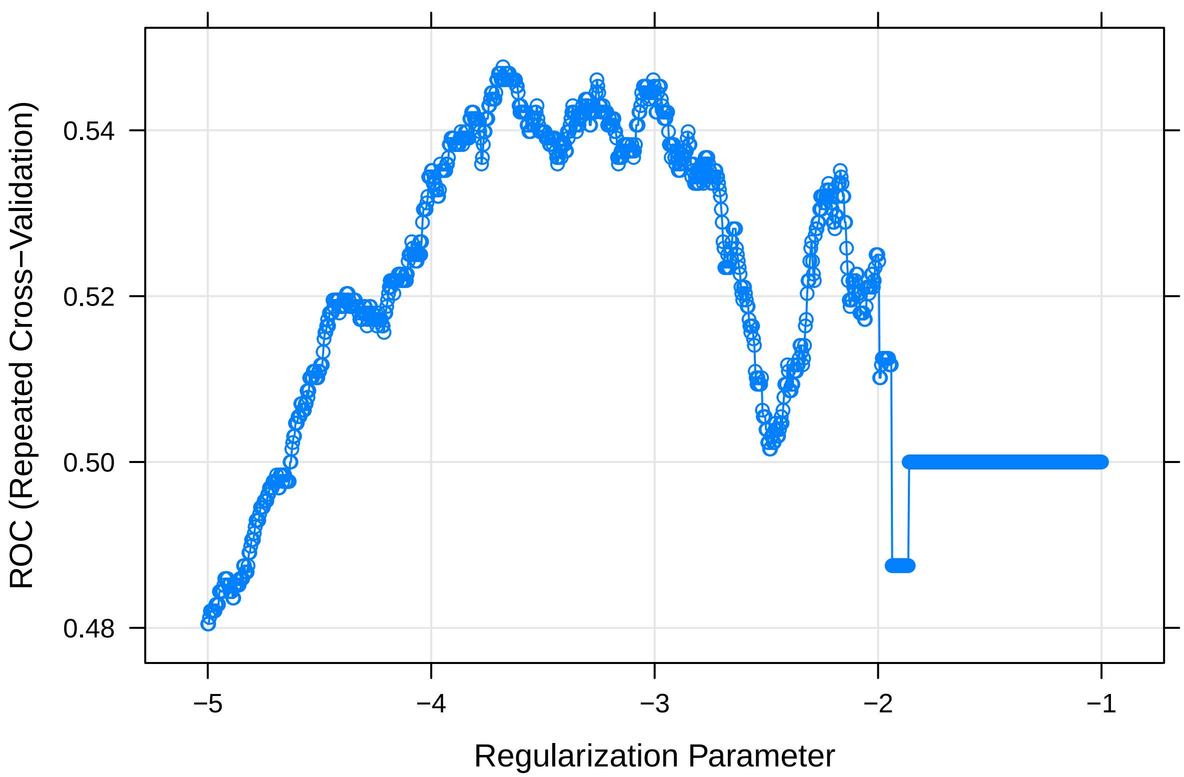
Logistic Regression

```
## Setting levels: control = Negative, case = Positive

## Setting direction: controls < cases
```

```r
plot(roc.glm, legacy.axes = T)
```



```r
# Lasso regression
set.seed(1029)

# Set outcome and features
x_CL = model.matrix(CL_analysis$PROT_ASD_2015~., CL_analysis)[rowTrain, -1]
y_CL = CL_analysis$PROT_ASD_2015[rowTrain]
# Fit LASSO Regression
model.lasso = train(x = CL_analysis[rowTrain, 2:482],
                    y = CL_analysis$PROT_ASD_2015[rowTrain],
                    method = "glmnet",
                    tuneGrid = expand.grid(alpha = 1,
                                           lambda = exp(seq(-5, -1, length = 1000))),
                    metric = "ROC",
                    family = "binomial",
                    trControl = ctrl)
plot(model.lasso, xTrans = function(x) log(x))
```



LASSO

```r
model.lasso$bestTune
```

```
##     alpha     lambda
## 331     1 0.02525632
```

```r
coef(model.lasso$finalModel, model.lasso$bestTune$lambda)
```

```
## 482 x 1 sparse Matrix of class "dgCMatrix"
##                                                    1
## (Intercept)                             -4.068637662
## Ceramide (d32:1)                          .
## Ceramide (d33:1)                          .
## Ceramide (d34:0)                          .
## Ceramide (d34:1)                          .
## Ceramide (d34:2)                          .
## Ceramide (d36:1)                          .
## Ceramide (d38:1)                          .
## Ceramide (d39:1)                          .
## Ceramide (d40:0)                          .
## Ceramide (d40:1)                          .
## Ceramide (d40:2)                          .
## Ceramide (d41:1)                          .
## Ceramide (d42:0)                          .
## Ceramide (d42:1)                          .
## Ceramide (d42:2) A                        .
## Ceramide (d42:2) B                      -0.226830868
## Ceramide (d43:1)                          .
## Ceramide (d44:1)                          .
## FA (10:0) (capric acid)                   .
## FA (11:0) (undecylic acid)                .
## FA (12:0) (lauric acid)                  0.549348692
## FA (13:0) (tridecylic acid)               .
## FA (14:0) (myristic acid)                 .
## FA (14:1) (physeteric acid)              0.628460087
## FA (15:0) (pentadecylic acid)             .
## FA (15:1)                               -0.864577021
## FA (16:1) (palmitoleic acid)              .
## FA (18:1) (oleic acid)                    .
## FA (18:2) (linoleic acid)                 .
## FA (18:3) (linolenic acid)                .
## FA (20:1) (eicosenoic acid)               .
## FA (20:2) (eicosadienoic acid)            .
## FA (20:3) (eicosatrienoic acid)           .
## FA (20:3) (homo-gamma-linolenic acid)     .
## FA (20:4) (arachidonic acid)              .
## FA (20:5) (eicosapentaenoic acid)         .
## FA (22:2) (docosadienoic acid)            .
## FA (22:6) (docosahexaenoic acid)          .
## FA (24:0) (lignoceric acid)               .
## FA (24:1) (nervonic acid)                 .
## FA (8:0) (caprylic acid)                  .
## FA 20:6;                                  .
## GlcCer (d38:1)                            .
## GlcCer (d40:1)                            .
## GlcCer (d41:1)                            .
## GlcCer (d42:1)                            .
## GlcCer (d42:2)                            .
## GlcCer(d14:1(4E)/20:0(2OH))               .
```

```
## LPC (14:0)                              .
## LPC (16:0)                              .
## LPC (16:1)                              .
## LPC (18:0) A                            .
## LPC (18:0) B                            .
## LPC (18:1)                              .
## LPC (18:2)                              .
## LPC (20:1)                              .
## LPC (20:2)                             -0.060926508
## LPC (20:3)                              .
## LPC (22:5)                              .
## LPC 20:4;                               .
## LPE (16:0)                              .
## LPE (18:2)                              .
## LPE (20:4)                              .
## LPE (22:6)                              .
## PC (32:0)                               .
## PC (32:1)                               .
## PC (32:2)                               .
## PC (33:1)                               .
## PC (33:2)                               .
## PC (34:0)                               .
## PC (34:1)                               .
## PC (34:2)                               .
## PC (34:3)                               .
## PC (34:4)                               .
## PC (35:1)                               .
## PC (35:2)                               .
## PC (35:4)                               .
## PC (36:1)                               .
## PC (36:2)                               .
## PC (36:3) A                             .
## PC (36:4) A                             .
## PC (36:4) B                             .
## PC (36:5) A                             .
## PC (36:5) B                             .
## PC (37:2)                               .
## PC (37:4)                               .
## PC (38:2)                               .
## PC (38:3)                               .
## PC (38:4) A                             .
## PC (38:5) A                             .
## PC (38:5) B                             .
## PC (38:6)                               .
## PC (40:4)                               .
## PC (40:5) A                             .
## PC (40:5) B                             .
## PC (40:6) A                             .
## PC (40:6) B                             .
## PC (40:7)                               .
## PC (40:8)                               .
## PC (o-32:0)                             .
## PC (p-32:0) or PC (o-32:1)              .
## PC (p-34:0) or PC (o-34:1)              .
```

```
## PC (p-34:1) or PC (o-34:2) A           .
## PC (p-34:1) or PC (o-34:2) B           .
## PC (p-34:2) or PC (o-34:3)             .
## PC (p-36:1) or PC (o-36:2)             .
## PC (p-36:3) or PC (o-36:4)             .
## PC (p-36:4) or PC (o-36:5)             .
## PC (p-38:3) or PC (o-38:4)             .
## PC (p-38:4) or PC (o-38:5) A           .
## PC (p-38:4) or PC (o-38:5) B           .
## PC (p-38:5) or PC (o-38:6)             .
## PC (p-40:1) or PC (o-40:2)             .
## PC (p-40:3) or PC (o-40:4)             .
## PC (p-40:4) or PC (o-40:5)             .
## PC (p-42:4) or PC (o-42:5)             .
## PC (p-42:5) or PC (o-42:6)             .
## PC (p-44:4) or PC (o-44:5)             .
## PC 38:4e; PC 16:0e/22:4;               .
## PE (34:1)                              .
## PE (34:2)                              .
## PE (36:1)                              .
## PE (36:2)                              .
## PE (36:3)                              .
## PE (36:4)                              .
## PE (38:2)                              .
## PE (38:4) A                            .
## PE (38:4) B                            .
## PE (38:6)                              0.028649303
## PE (40:6)                              .
## PE (p-34:1) or PE (o-34:2)             .
## PE (p-34:2) or PE (o-34:3)             .
## PE (p-36:1) or PE (o-36:2)             .
## PE (p-36:2) or PE (o-36:3)            -0.291009313
## PE (p-36:4) or PE (o-36:5)             .
## PE (p-36:5) or PE (o-36:6)             .
## PE (p-38:3) or PE (o-38:4)             .
## PE (p-38:4) or PE (o-38:5)             .
## PE (p-38:5) or PE (o-38:6)             .
## PE (p-38:6) or PE (o-38:7)             .
## PE (p-40:4) or PE (o-40:5)             .
## PE (p-40:5) or PE (o-40:6)             .
## PE (p-40:6) or PE (o-40:7)             .
## PE (p-40:7) or PE (o-40:8)             3.407886533
## PE 38:5; PE 16:0-22:5;                 .
## PE 38:5; PE 18:1-20:4;                 .
## PE 38:5e; PE 16:1e/22:4;               .
## PE 40:5e; PE 18:1e/22:4;               .
## PI 34:2; PI 16:0-18:2;                 .
## PI 36:4; PI 16:0-20:4;                 .
## PI 38:4; PI 18:0-20:4;                 .
## PI 38:5; PI 18:1-20:4;                 .
## PI 38:6; PI 16:0-22:6;                 .
## PS 38:4; PS 18:0-20:4;                 .
## PS 40:6; PS 18:0-22:6;                 .
## SM (d30:1)                             .
```

```
## SM (d32:0)                          .
## SM (d32:1)                          .
## SM (d32:2)                          .
## SM (d33:1)                          .
## SM (d34:0)                          .
## SM (d34:1)                          .
## SM (d34:2)                          .
## SM (d36:0)                          .
## SM (d36:1)                          .
## SM (d36:2)                          .
## SM (d36:3)                          .
## SM (d37:1)                          .
## SM (d38:0)                          .
## SM (d38:1)                          .
## SM (d38:2)                          .
## SM (d39:1)                          .
## SM (d39:2)                          .
## SM (d40:0)                          .
## SM (d40:1)                          .
## SM (d40:2) A                        0.002351198
## SM (d40:2) B                        .
## SM (d40:3)                          .
## SM (d41:1)                          .
## SM (d41:2)                          .
## SM (d42:0)                          .
## SM (d42:1)                          .
## SM (d42:2) A                        2.658694958
## SM (d42:2) B                        .
## SM (d42:3)                          .
## SM (d43:1)                          .
## SM (d43:2)                          .
## SM (d44:2)                          .
## SM d42:1; SM d23:1/19:0;            .
## AC(10:0)                            0.129886029
## AC(10:1)                            .
## AC(12:0)                            .
## AC(12:1)                            .
## AC(14:1)                            .
## AC(14:2)                            .
## AC(16:0)                            .
## AC(18:0)                            .
## AC(18:1)                            .
## AC(18:2)                            .
## CE(16:0)                            .
## CE(16:1)                            .
## CE(18:0)                            .
## CE(18:1)                            1.307962597
## CE(18:2)                            .
## CE(18:3)                            .
## CE(20:3)                            .
## CE(20:4)                            .
## CE(22:2)                            .
## CE(22:6)                            .
## Cer(d34:1)                          .
```

```
## Cer(d36:1)                              .
## Cer(d38:1)                              .
## Cer(d40:1)                              .
## Cer(d41:1)                              .
## Cer(d42:1)                              .
## Cer(d42:2) A                            .
## Cholesterol                             .
## DG(34:0)                                .
## DG(34:1)                                .
## DG(34:2)                                .
## DG(34:3)                                1.345809213
## DG(36:1)                                .
## DG(36:2)                                .
## DG(36:3)                                .
## DG(36:4) A                              0.225688859
## DG(38:5)                                0.132419350
## GlcCer(d34:1)                           .
## GlcCer(d40:1)                           .
## GlcCer(d42:1)                           .
## GlcCer(d42:2)                           .
## Lactosylceramide (d18:1/24:1(15Z))     -0.323753209
## LPC(14:0)                               .
## LPC(15:0)                               .
## LPC(16:0)                               .
## LPC(16:1)                               .
## LPC(17:1)                               .
## LPC(18:0)                               .
## LPC(18:1)                               .
## LPC(18:2)                               .
## LPC(18:3)                               .
## LPC(20:0)                               .
## LPC(20:2)                               .
## LPC(20:3)                               .
## LPC(20:4)                               .
## LPC(20:5)                               .
## LPC(22:4)                               .
## LPC(22:5)                               .
## LPC(22:6)                               .
## LPE(18:0)                               .
## LPE(18:2)                               .
## PC 18:0e; PC 16:0e/2:0;                 .
## PC 34:1e;                               .
## PC 34:2e;                               .
## PC(28:0)                                .
## PC(30:0)                                .
## PC(31:0)                                .
## PC(31:1)                                .
## PC(32:0)                                .
## PC(32:1)                                .
## PC(32:2)                                .
## PC(33:0)                                .
## PC(33:1)                                .
## PC(33:2)                                .
## PC(34:0)                                .
```

```
## PC(34:1)                              .
## PC(34:2)                              .
## PC(34:3) A                            .
## PC(34:3) B                            .
## PC(34:3) C                          -1.002641307
## PC(34:4)                            -0.258211349
## PC(35:1)                              .
## PC(35:2) A                            .
## PC(35:2) B                            .
## PC(35:3)                              .
## PC(35:4)                              .
## PC(36:1)                              .
## PC(36:2)                              .
## PC(36:3) A                            .
## PC(36:3) B                            .
## PC(36:4) A                            .
## PC(36:4) B                            .
## PC(36:4) C                            .
## PC(36:5) C                            .
## PC(36:5) D                           0.627422753
## PC(36:5)A                            1.876124952
## PC(36:6)                              .
## PC(37:2)                              .
## PC(37:3)                            -0.719912878
## PC(37:4)                              .
## PC(37:5)                              .
## PC(37:6)                              .
## PC(38:2)                              .
## PC(38:3)                              .
## PC(38:4) A                            .
## PC(38:4) B                          -0.410027677
## PC(38:4) C                            .
## PC(38:5) A                           2.451053450
## PC(38:5) B                            .
## PC(38:6) A                            .
## PC(38:6) B                            .
## PC(38:7)                            -0.069373321
## PC(39:4)                              .
## PC(39:6)                              .
## PC(40:4)                              .
## PC(40:5) A                            .
## PC(40:5) B                            .
## PC(40:6) A                            .
## PC(40:6)B                             .
## PC(40:7) A                          -0.473267562
## PC(40:7) B                            .
## PC(40:8)                              .
## PC(42:10)                           -1.848689778
## PC(42:5)                              .
## PC(42:6)                              .
## PC(o-32:0)                            .
## PC(o-34:0)                            .
## PC(p-32:0) or PC (o-32:1)             .
## PC(p-32:1)/PC(o-32:2)                0.743691974
```

```
## PC(p-34:1)/PC(o-34:2)                    .
## PC(p-34:2)/PC(o-34:3)                    .
## PC(p-36:1)/PC(o-36:2) A                  .
## PC(p-36:1)/PC(o-36:2) B                  .
## PC(p-36:2)/PC(o-36:3)                    .
## PC(p-36:3) or PC (o-36:4)                .
## PC(p-36:4)/PC(o-36:5)                    .
## PC(p-36:5)/PC(o-36:6)                    .
## PC(p-38:2) or PC (o-38:3)                .
## PC(p-38:3)/PC(o-38:4) A                  .
## PC(p-38:3)/PC(o-38:4) B                  .
## PC(p-38:4)/PC(o-38:5) A                  .
## PC(p-38:5)/PC(o-38:6) A                  .
## PC(p-38:5)/PC(o-38:6) B                  .
## PC(p-38:6)/PC(o-38:7)                    .
## PC(p-40:3)/PC(o-40:4)                    .
## PC(p-40:4)/PC(o-40:5)                    .
## PC(p-40:5)/PC(o-40:6)                    .
## PC(p-40:6)/PC(o-40:7) A                  .
## PC(p-40:6)/PC(o-40:7) B                  .
## PC(p-42:3) or PC (o-42:4)                .
## PC(p-42:4)/PC(o-42:5)                    .
## PC(p-42:5)/PC(o-42:6) A                 -1.162487429
## PC(p-44:4)/PC(o-44:5)                    .
## PE(36:1)                                 .
## PE(36:4)                                 .
## PE(38:4)                                 .
## PE(38:6)                                 .
## PE(p-34:1)/PE(o-34:2)                    .
## PE(p-36:1)/PE(o-36:2)                    .
## PE(p-36:2)/PE(o-36:3)                    .
## PE(p-36:4)/PE(o-36:5)                    .
## PE(p-38:4)/PE(o-38:5)                   -0.133058398
## PE(p-38:5)/PE(o-38:6)                    .
## PE(p-40:4)/PE(o-40:5)A                  -0.115894692
## PE(p-40:4)/PE(o-40:5)B                   .
## SM d41:1;                                .
## SM d42:1;                                .
## SM d42:2;                                .
## SM d44:3;                                .
## SM(d30:1)                                .
## SM(d32:0)                                .
## SM(d32:1)                                .
## SM(d32:2)                                .
## SM(d33:1)                                .
## SM(d34:0)                                .
## SM(d34:1)                                .
## SM(d34:2)                                .
## SM(d36:0)                                .
## SM(d36:1)                                .
## SM(d36:2)                                .
## SM(d36:3)                                .
## SM(d38:0)                                .
## SM(d38:1)                                .
```

```
## SM(d38:2)                           .
## SM(d39:1)                           .
## SM(d39:2)                           .
## SM(d40:0)                           .
## SM(d40:1)                           .
## SM(d40:2) A                         .
## SM(d40:2) B                        -1.498790238
## SM(d41:1)                           .
## SM(d41:2) A                         .
## SM(d41:2) B                         .
## SM(d42:0)                           .
## SM(d42:1)                           .
## SM(d42:2)                           .
## SM(d42:3)                           .
## SM(d43:1)                           .
## SM(d43:2)                           .
## SM(d44:2)                           .
## TAG 46:2; TAG 12:0-16:1-18:1;       .
## TAG 50:4; TAG 16:1-16:1-18:2;       .
## TAG 52:5; TAG 16:1-18:2-18:2;       .
## TAG 54:6; TAG 16:0-16:0-22:6;       .
## TAG(42:0)                           .
## TAG(42:1)                           .
## TAG(42:2)                           .
## TAG(42:3)                          -0.011234882
## TAG(44:0)                           .
## TAG(44:1)                           .
## TAG(46:0)                           .
## TAG(46:1)                           .
## TAG(46:4) B                         .
## TAG(48:0)                           .
## TAG(48:1)                           .
## TAG(48:2)                           .
## TAG(48:3)                           .
## TAG(48:4) A                         0.646404204
## TAG(48:4) B                         .
## TAG(49:0)                           .
## TAG(49:1)                           .
## TAG(49:2)                           .
## TAG(49:3)                           .
## TAG(50:1)                           .
## TAG(50:2)                           .
## TAG(50:3) A                         .
## TAG(50:4)                           .
## TAG(50:5)                           .
## TAG(50:6)                           .
## TAG(51:1)                           .
## TAG(51:2)                           .
## TAG(51:3)                           .
## TAG(51:4)                           .
## TAG(51:5)                          -0.006533713
## TAG(52:1)                           .
## TAG(52:2)                           .
## TAG(52:3)                           .
```

```
## TAG(52:4)                        .
## TAG(52:5)                        .
## TAG(52:6)                        .
## TAG(53:0)                        .
## TAG(53:1)                        .
## TAG(53:2)                        .
## TAG(53:3)                        .
## TAG(53:4)                        .
## TAG(54:1)                        .
## TAG(54:2)                        .
## TAG(54:3)                        0.440895955
## TAG(54:4)                        .
## TAG(54:5) A                      .
## TAG(54:5) B                      .
## TAG(54:6) B                      .
## TAG(54:7) A                      .
## TAG(54:7) B                      .
## TAG(54:8)                        .
## TAG(54:9)                        .
## TAG(55:1)                        .
## TAG(55:2)                        .
## TAG(55:3)                        .
## TAG(56:0)                        .
## TAG(56:1)                        .
## TAG(56:2)                        .
## TAG(56:3)                        .
## TAG(56:4)                        .
## TAG(56:5) A                      .
## TAG(56:5) B                      .
## TAG(56:5) C                      .
## TAG(56:6)                        .
## TAG(56:7) A                      .
## TAG(56:7) B                      .
## TAG(56:8) A                      .
## TAG(56:9)                        .
## TAG(57:1)                        .
## TAG(57:2)                        .
## TAG(58:1)                        .
## TAG(58:2)                        .
## TAG(58:3)                        .
## TAG(58:4)                        .
## TAG(58:5)                        .
## TAG(58:6)                        .
## TAG(58:8)                        .
## TAG(58:9)                        .
## TAG(59:2)                        1.280980109
## TAG(59:3)                        .
## TAG(60:1)                        .
## TAG(60:2)                        .
## TAG(60:3)                        .
## TAG(60:4)                       -0.177570873
## TAG(60:6)                        .
## TAG(62:1)                        .
## TAG(62:2)                        0.030193919
```

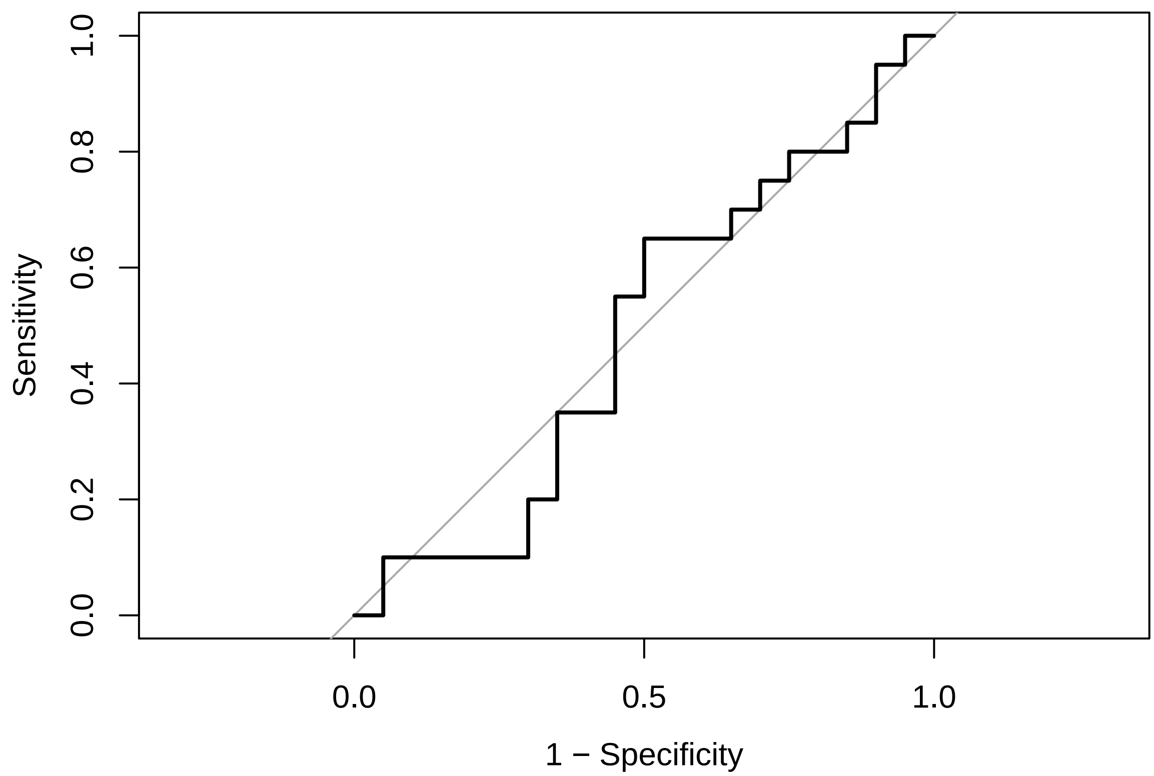
```
## TAG(62:3)                              .
```

```
lasso.pred = predict(model.lasso, newdata = CL_analysis[-rowTrain,], type = "prob")[,2]
roc.lasso = roc(CL_analysis$PROT_ASD_2015[-rowTrain], lasso.pred)
```

```
## Setting levels: control = Negative, case = Positive
```

```
## Setting direction: controls < cases
```

```
plot(roc.lasso, legacy.axes = T)
```



Classification Trees

```
# Random Forest
rf.grid = expand.grid(mtry = 1:160,
                      splitrule = "gini",
                      min.node.size = 1:6)

set.seed(1029)

model.rf = train(PROT_ASD_2015~., CL_analysis,
```

```
                 subset = rowTrain,
                 method = "ranger",
                 tuneGrid = rf.grid,
                 metric = "ROC",
                 trControl = ctrl)
```

Random Forest

```
# Adaboost loss function
ada.grid = expand.grid(n.trees = c(1000, 2000, 3000),
                       interaction.depth = 1:6,
                       shrinkage = c(0.001, 0.003, 0.005),
                       n.minobsinnode = 1)

set.seed(1029)

model.ada = train(PROT_ASD_2015~., CL_analysis,
                  subset = rowTrain,
                  tuneGrid = ada.grid,
                  trControl = ctrl,
                  method = "gbm",
                  distribution = "adaboost",
                  metric = "ROC",
                  verbose = FALSE)
```

AdaBoost

```
# XGBoost
xgboost.grid = expand.grid(eta = 0.1,
                           colsample_bytree=c(0.5,0.7),
                           max_depth=c(3,6),
                           nrounds=100,
                           gamma=1,
                           min_child_weight=2,
                           subsample = 1)

set.seed(1029)

modelxgboost = train(PROT_ASD_2015~.,
                     CL_analysis,
                     subset = rowTrain,
                     method = "xgbTree",
                     trControl = ctrl,
                     tuneGrid = xgboost.grid,
                     metric = "ROC",
                     verbose = F)
```

XGBoost

Model Comparison

```
# Calculate ROC for each model
glm.pred = predict(model.glm, newdata = CL_analysis[-rowTrain,], type = "prob")[,2]
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from a rank-deficient fit may be misleading
```

```
roc.glm = roc(CL_analysis$PROT_ASD_2015[-rowTrain], glm.pred)
```

```
## Setting levels: control = Negative, case = Positive

## Setting direction: controls < cases
```

```
lasso.pred = predict(model.lasso, newdata = CL_analysis[-rowTrain,], type = "prob")[,2]
roc.lasso = roc(CL_analysis$PROT_ASD_2015[-rowTrain], lasso.pred)
```

```
## Setting levels: control = Negative, case = Positive
## Setting direction: controls < cases
```

```
ada.pred = predict(model.ada, newdata = CL_analysis[-rowTrain,], type = "prob")[,2]
roc.ada = roc(CL_analysis$PROT_ASD_2015[-rowTrain], ada.pred)
```

```
## Setting levels: control = Negative, case = Positive

## Setting direction: controls > cases
```

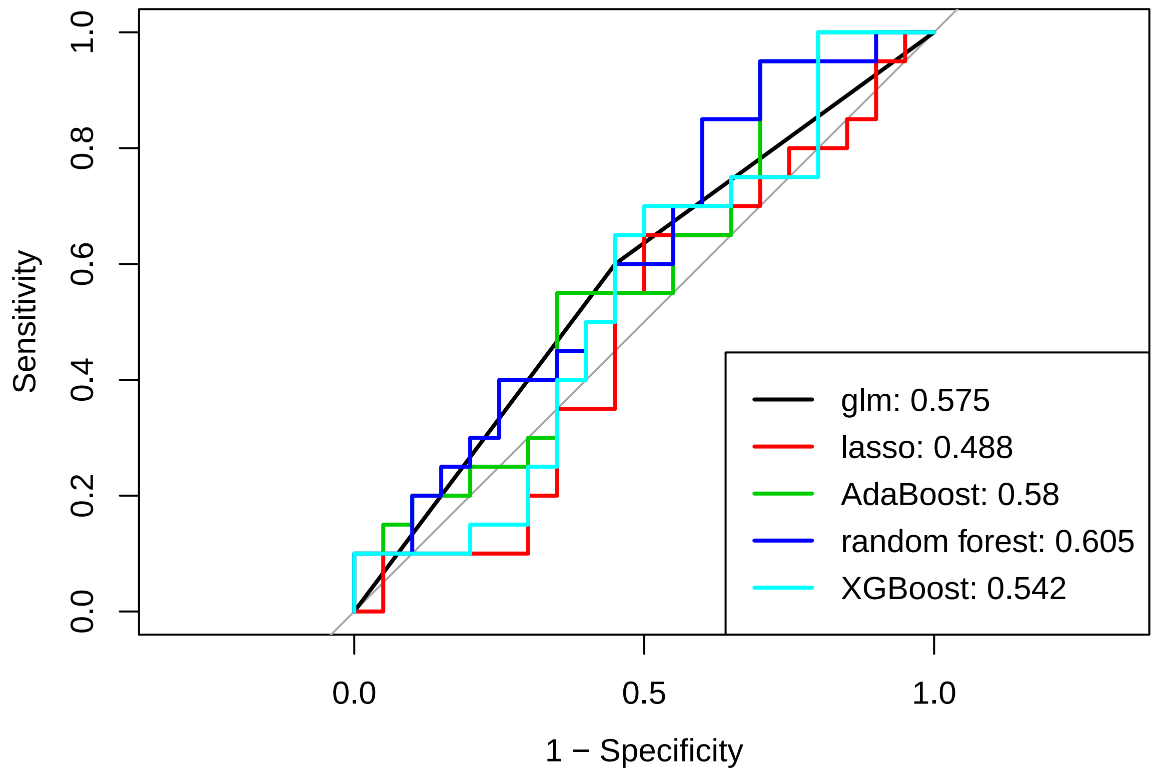
```
rf.pred = predict(model.rf, newdata = CL_analysis[-rowTrain,], type = "prob")[,2]
roc.rf = roc(CL_analysis$PROT_ASD_2015[-rowTrain], rf.pred)
```

```
## Setting levels: control = Negative, case = Positive
## Setting direction: controls > cases
```

```
xgboost.pred = predict(modelxgboost, newdata = CL_analysis[-rowTrain,], type = "prob")[,2]
roc.xgboost = roc(CL_analysis$PROT_ASD_2015[-rowTrain], xgboost.pred)
```

```
## Setting levels: control = Negative, case = Positive
## Setting direction: controls > cases
```

```
auc = c(roc.glm$auc[1],roc.lasso$auc[1],roc.ada$auc[1],roc.rf$auc[1],roc.xgboost$auc[1])

plot(roc.glm, legacy.axes = T)
plot(roc.lasso, col = 2, add = T)
plot(roc.ada, col= 3, add = T)
plot(roc.rf, col= 4, add = T)
plot(roc.xgboost, col= 5, add = T)
modelNames = c("glm", "lasso", "AdaBoost", "random forest", "XGBoost")
legend("bottomright", legend = paste0(modelNames, ": ", round(auc, 3)), col = 1:5, lwd = 2)
```




```
res = resamples(list(GLM = model.glm,
                     LASSO = model.lasso,
                     AdaBoost = model.ada,
                     RandomForest = model.rf,
                     XGBoost =modelxgboost))

summary(res)
```

```
##
## Call:
## summary.resamples(object = res)
##
## Models: GLM, LASSO, AdaBoost, RandomForest, XGBoost
## Number of resamples: 5
##
## ROC
##                   Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## GLM          0.3847656 0.4062500 0.5039062 0.4820313 0.5214844 0.5937500    0
## LASSO        0.5117188 0.5156250 0.5625000 0.5476562 0.5703125 0.5781250    0
## AdaBoost     0.4335938 0.5078125 0.5351562 0.5328125 0.5585938 0.6289062    0
## RandomForest 0.4531250 0.4726562 0.4960938 0.5406250 0.6015625 0.6796875    0
## XGBoost      0.4101562 0.5078125 0.5273438 0.5304688 0.5859375 0.6210938    0
##
## Sens
##                 Min. 1st Qu. Median    Mean 3rd Qu.    Max. NA's
```

```
## GLM         0.2500  0.5000 0.5000 0.4875  0.5625 0.6250    0
## LASSO       0.2500  0.4375 0.5000 0.5000  0.6250 0.6875    0
## AdaBoost    0.5000  0.5625 0.5625 0.5625  0.5625 0.6250    0
## RandomForest 0.3125  0.5000 0.5625 0.5125  0.5625 0.6250    0
## XGBoost     0.4375  0.5000 0.5000 0.5125  0.5000 0.6250    0
##
## Spec
##               Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's
## GLM         0.3125  0.3750 0.5000 0.4750  0.5625 0.6250    0
## LASSO       0.2500  0.5000 0.5625 0.5375  0.6250 0.7500    0
## AdaBoost    0.3125  0.3750 0.5000 0.4625  0.5000 0.6250    0
## RandomForest 0.3750  0.4375 0.5000 0.5125  0.5625 0.6875    0
## XGBoost     0.3750  0.3750 0.5000 0.4625  0.5000 0.5625    0
```

```r
bwplot(res, metric = "ROC")
```