Combined Predictors(BIO+CHEM)

1. Non linear Discrimination Analysis

```
mda(formula = trainY ~ ., data = train)

  Dimension: 8
  Pre Process: Center and Scale
  Percent Between-Group Variance Explained:
      v1     v2     v3     v4     v5     v6     v7     v8
   45.62  70.54  79.58  87.30  92.60  96.98  99.22 100.00


  Deviance: 0

  Confusion Matrix and Statistics

            Reference
  Prediction Mild None Severe
      Mild     11   11      3
      None     11    4      1
      Severe    7    6      2

  Overall Statistics for Testing set

                 Accuracy : 0.3036
                   95% CI : (0.1878, 0.441)
      No Information Rate : 0.5179
      P-Value [Acc > NIR] : 0.9996

                    Kappa : -0.1003
   Mcnemar's Test P-Value : 0.1597

  Statistics by Class:

                       Class: Mild Class: None Class: Severe
  Sensitivity               0.3793     0.19048       0.33333
  Specificity               0.4815     0.65714       0.74000
```

2. Neural Network

```
  225 samples
  202 predictors
    3 classes: 'Mild', 'None', 'Severe'

  Pre-processing: spatial sign transformation, scaled, centered
  Resampling: Bootstrapped (25 reps)

  Summary of sample sizes: 225, 225, 225, 225, 225, 225, ...

  Resampling results across tuning parameters:
```

```
  size  decay  Accuracy   Kappa       Accuracy SD  Kappa SD
  1     0.0    0.4537836  0.07471607  0.08057553   0.06684354
  1     0.1    0.5133845  0.12088479  0.06281646   0.10210564
  2     0.0    0.4428505  0.08085378  0.06288195   0.07227809
  2     0.1    0.5023176  0.11916586  0.05634693   0.09123446
  3     0.0    0.4676187  0.08745088  0.04241189   0.07151448
  3     0.1    0.4956529  0.10897236  0.05547206   0.08962320
  4     0.0    0.4686464  0.09977135  0.04895218   0.07394145


Kappa was used to select the optimal model using  the largest value.
The final values used for the model were size = 1 and decay = 0.1.

Confusion Matrix and Statistics for Testing set

          Reference
Prediction Mild None Severe
    Mild     17   16      5
    None     12    5      1
    Severe    0    0      0

Overall Statistics

               Accuracy : 0.3929
                 95% CI : (0.265, 0.5325)
    No Information Rate : 0.5179
    P-Value [Acc > NIR] : 0.97778

                  Kappa : -0.1498
 Mcnemar's Test P-Value : 0.08689

Statistics by Class:

                     Class: Mild Class: None Class: Severe
Sensitivity               0.5862     0.23810        0.0000
Specificity               0.2222     0.62857        1.0000
```

3.Flexible Discriminant Analysis

```
225 samples
202 predictors
  3 classes: 'Mild', 'None', 'Severe'

Pre-processing : Center and Scale
Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 225, 225, 225, 225, 225, 225, ...
```

```
Resampling results across tuning parameters:

  nprune  Accuracy   Kappa       Accuracy SD  Kappa SD
   2      0.4850801  0.01561390  0.04515211   0.06310917
  38      0.4773120  0.11601226  0.05975337   0.09750591
  74      0.4413228  0.08431081  0.07197247   0.10911220

Tuning parameter 'degree' was held constant at a value of 1
Kappa was used to select the optimal model using  the largest value.
The final values used for the model were degree = 1 and nprune = 38.
```

Confusion Matrix and Statistics **Testing set**

```
          Reference
Prediction Mild None Severe
    Mild     22   14      2
    None      6    6      2
    Severe    1    1      2

Overall Statistics

               Accuracy : 0.5357
                 95% CI : (0.3974, 0.6701)
    No Information Rate : 0.5179
    P-Value [Acc > NIR] : 0.4475

                  Kappa : 0.1515
 Mcnemar's Test P-Value : 0.2762

Statistics by Class:

                     Class: Mild Class: None Class: Severe
Sensitivity               0.7586      0.2857       0.33333
Specificity               0.4074      0.7714       0.96000
```

4.Support Vector Machines with Radial Basis Function Kernel

```
225 samples
202 predictors
  3 classes: 'Mild', 'None', 'Severe'

Pre-processing : Center and Scale
Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 225, 225, 225, 225, 225, 225, ...

Resampling results across tuning parameters:

  C       Accuracy   Kappa        Accuracy SD  Kappa SD
   0.0625  0.5122456  0.000000000  0.03794856   0.00000000
   0.1250  0.5122456  0.000000000  0.03794856   0.00000000
   0.2500  0.5119341  0.005702462  0.04175516   0.02721642
```

```
    0.5000  0.5147571  0.040719794  0.03682783   0.06075267
    1.0000  0.5228223  0.078864123  0.03592325   0.06688309
    2.0000  0.5374330  0.128709971  0.03744246   0.06212469
    4.0000  0.5374355  0.147896508  0.04652034   0.08207168
    8.0000  0.5232891  0.138126080  0.04664973   0.07857214
   16.0000  0.5107618  0.123936561  0.05396726   0.09679151

Tuning parameter 'sigma' was held constant at a value of 0.001278265
Kappa was used to select the optimal model using  the largest value.
The final values used for the model were sigma = 0.001278265 and C = 4.

Confusion Matrix and Statistics for testing set

          Reference
Prediction Mild None Severe
    Mild     20   17      6
    None      8    4      0
    Severe    1    0      0

Overall Statistics

               Accuracy : 0.4286
                 95% CI : (0.2971, 0.5678)
    No Information Rate : 0.5179
    P-Value [Acc > NIR] : 0.9294

                  Kappa : -0.0987
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: Mild Class: None Class: Severe
Sensitivity               0.6897     0.19048       0.00000
Specificity               0.1481     0.77143       0.98000
```

5. k-Nearest Neighbors

225 samples

202 predictors

3 classes: 'Mild', 'None', 'Severe'

pre-processing : Center and Scale

Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 225, 225, 225, 225, 225, 225, ...

Resampling results across tuning parameters:

k	Accuracy	Kappa	Accuracy SD	Kappa SD
1	0.4604128	0.1026969919	0.04571765	0.077625979
5	0.4770297	0.1049465821	0.06632193	0.077520150
9	0.4769611	0.0710456864	0.07567141	0.095280731
13	0.4775598	0.0496226638	0.07097258	0.077295030
17	0.4853737	0.0475059905	0.06182266	0.062205010
21	0.4850623	0.0382689613	0.06228243	0.079176958
41	0.4952677	0.0311927608	0.05836922	0.081458116

Kappa was used to select the optimal model using the largest value.

The final value used for the model was k = 5.

```
Confusion Matrix and Statistics for Testing set
          Reference
Prediction Mild None Severe
    Mild     18   12      4
    None      9    8      2
    Severe    2    1      0

Overall Statistics

               Accuracy : 0.4643
                 95% CI : (0.3299, 0.6026)
    No Information Rate : 0.5179
    P-Value [Acc > NIR] : 0.8254

                  Kappa : 0.0306
 Mcnemar's Test P-Value : 0.6989
```

Statistics by Class:

	Class: Mild	Class: None	Class: Severe
Sensitivity	0.6207	0.3810	0.00000
Specificity	0.4074	0.6857	0.94000

6.Naive Bayes

225 samples

202 predictors

3 classes: 'Mild', 'None', 'Severe'

Pre-processing : Center and Scale

Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 225, 225, 225, 225, 225, 225, ...

Resampling results across tuning parameters:

usekernel	Accuracy	Kappa	Accuracy SD	Kappa SD
FALSE	NaN	NaN	NA	NA
TRUE	0.4643774	0.08191371	0.08246112	0.07702208

Tuning parameter 'fL' was held constant at a value of 0

Kappa was used to select the optimal model using the largest value.

The final values used for the model were fL = 0 and usekernel = TRUE.

Confusion Matrix and Statistics **Testing set**

Prediction	Reference Mild	None	Severe
Mild	22	18	5
None	4	2	0
Severe	3	1	1

Overall Statistics

Accuracy : 0.4464
95% CI : (0.3134, 0.5853)
No Information Rate : 0.5179
P-Value [Acc > NIR] : 0.88561

Kappa : -0.0364

```
Statistics by Class:

                     Class: Mild Class: None Class: Severe
Sensitivity               0.7586     0.09524       0.16667
Specificity               0.1481     0.88571       0.92000
```

Testing set:

Linear	Accuracy	Kappa	Sensitivity	Specificity
LDA	0.3571	0.03	0.412	0.695
PLSDA	0.5357	0.09	0.4562	0.6298
NSC	0.4643	0.09	0	0.905

Non Linear Models	Accuracy	Kappa
MDA	0.3069	0.032
NNet	0.3929	0.0132
FDA	0.5357	0.1515
SVM	0.4268	0.031
KNN	0.4643	0.036
Naïve Bayes	0.4464	0.03

Both Linear and Non-linear models discussed in here do NOT do better on the combined data set. Only F DA shows somewhat better performance.

b)

20 most important variables shown (out of 96)
for **Biological Data**

	Mild	None	Severe
Z15	0.6013	0.6355	0.6355
Z100	0.6015	0.6108	0.6108
Z116	0.5990	0.5515	0.5990
Z59	0.5803	0.5434	0.5803
Z44	0.5801	0.5578	0.5801
Z56	0.5700	0.5782	0.5782
Z167	0.5756	0.5441	0.5756
Z64	0.5695	0.5695	0.4603
Z34	0.5658	0.5658	0.5241
Z121	0.5368	0.5581	0.5581
Z18	0.5580	0.5580	0.5090
Z101	0.5571	0.5461	0.5571
Z7	0.5548	0.5424	0.5548
Z46	0.5523	0.5523	0.4343
Z11	0.5514	0.5272	0.5514
Z71	0.5498	0.5498	0.4652
Z50	0.5487	0.5487	0.5208
Z42	0.5477	0.5477	0.4682
Z53	0.5016	0.5453	0.5453
Z48	0.5450	0.5450	0.4447

20 most important variables shown (out of 96)
for **Chemical Predictors**

	Mild	None	Severe
X139	0.6694	0.6870	0.6870
X145	0.6566	0.6804	0.6804
X1	0.6386	0.6711	0.6711
X133	0.5903	0.6701	0.6701
X132	0.6307	0.6672	0.6672
X144	0.6471	0.6600	0.6600
X101	0.6228	0.6576	0.6576
X35	0.5867	0.6544	0.6544
X138	0.6480	0.6346	0.6480
X81	0.6221	0.6471	0.6471
X150	0.6386	0.5990	0.6386
X120	0.5744	0.6348	0.6348
X171	0.6060	0.6297	0.6297
X103	0.5997	0.6260	0.6260
X127	0.6058	0.6240	0.6240
X24	0.5961	0.6208	0.6208
X28	0.5894	0.6196	0.6196
X62	0.5824	0.6137	0.6137
X142	0.6128	0.5953	0.6128
X23	0.6124	0.5738	0.6124

c) Combined Predictors

only 20 most important variables shown (out of 202) **BIO+CHEM Combined**

	Mild	None	Severe
X1	0.6640	0.6757	0.6757
X172	0.6652	0.6373	0.6652
X139	0.6426	0.6608	0.6608
X150	0.6573	0.5880	0.6573
X142	0.6518	0.6162	0.6518
X132	0.6403	0.6458	0.6458
X138	0.6406	0.6272	0.6406
X141	0.6404	0.6071	0.6404
X28	0.6160	0.6370	0.6370
X24	0.5907	0.6331	0.6331
X120	0.5916	0.6331	0.6331
X144	0.6306	0.6284	0.6306
X151	0.6304	0.5716	0.6304
Z15	0.5675	0.6265	0.6265
X171	0.5729	0.6243	0.6243
X133	0.5445	0.6225	0.6225
X145	0.6036	0.6196	0.6196
X123	0.6144	0.6154	0.6154
Z40	0.5671	0.6125	0.6125
X85	0.6058	0.6110	0.61

d) **Biological data with Non-Linear model(SVM)** performs best out of other cases consider in this exercise.