Higgs_Final_Report

Particle Physics

I was in Zurich working for UBS during the period of 2018-2020. When my family visited me in Zurich during the summer of July/Aug 2019, I took them for 1 day trip to Geneva to visit the CERN laboratory, the birth place of "God particle". The trip was amazing, it takes one through the journey of atomic particle to sub-atomic. There were lots of short video clips, practical demonstrations and exhibitions etc.

What caught my eyes was a poster with title "Machine Learning for High Energy Physics". The aim of the poster was to explore the new ways to improve the cross-fertilization of the two fields: astrophysics community (dark matter and galaxy zoo challenges) and neurobiology (connectomics and decoding the human brain).

Therefore, I decided to explore what is particle physics and how can the machine learning and deep learning can help solve the particle physics problems. The summary of this short article was the result of my visit to CERN laboratory.

Problem Statement

After the visit to the CERN laboratory, I started exploring the fundamentals of particle physics, its fundamental equation of particles etc, even though I didn't understand much as it has been a decade since I looked into physics closely but it was very interesting research. As part of this exploration I came across a small simulated dataset at Kaggle website (2014). The goal of the Higgs Boson Machine Learning Challenge is to explore the potential of advanced machine learning methods to improve the discovery significance of the experiment. The good thing about this challenge is no knowledge of particle physics is required. Therefore, I decided to use this dataset to explore how the Machine learning algorithms can be used for Higgs Boson research.

Data

The data was sourced from the CERN public domain website. For the link please refer to the references.

Data Pre-processing

- For experimentation, a small portion of data was extracted and written to the local disk.

```
##   EventId DER_mass_MMC DER_mass_transverse_met_lep DER_mass_vis DER_pt_h
## 1  100000      138.470                     51.655       97.827   27.980
## 2  100001      160.937                     68.768      103.235   48.146
## 3  100002     -999.000                    162.172      125.953   35.635
## 4  100003      143.905                     81.417       80.943    0.414
## 5  100004      175.864                     16.915      134.805   16.405
## 6  100005       89.744                     13.550       59.149  116.344
##   DER_deltaeta_jet_jet DER_mass_jet_jet DER_prodeta_jet_jet DER_deltar_tau_lep
## 1                0.910          124.711               2.666              3.064
## 2             -999.000         -999.000            -999.000              3.473
## 3             -999.000         -999.000            -999.000              3.148
## 4             -999.000         -999.000            -999.000              3.310
## 5             -999.000         -999.000            -999.000              3.891
```

```
## 6                2.636          284.584              -0.540              1.362
##   DER_pt_tot DER_sum_pt DER_pt_ratio_lep_tau DER_met_phi_centrality
## 1     41.928    197.760                1.582                  1.396
## 2      2.078    125.157                0.879                  1.414
## 3      9.336    197.814                3.776                  1.414
## 4      0.414     75.968                2.354                 -1.285
## 5     16.405     57.983                1.056                 -1.385
## 6     61.619    278.876                0.588                  0.479
##   DER_lep_eta_centrality PRI_tau_pt PRI_tau_eta PRI_tau_phi PRI_lep_pt
## 1                  0.200     32.638       1.017       0.381     51.626
## 2               -999.000     42.014       2.039      -3.011     36.918
## 3               -999.000     32.154      -0.705      -2.093    121.409
## 4               -999.000     22.647      -1.655       0.010     53.321
## 5               -999.000     28.209      -2.197      -2.231     29.774
## 6                  0.975     53.651       0.371       1.329     31.565
##   PRI_lep_eta PRI_lep_phi PRI_met PRI_met_phi PRI_met_sumet PRI_jet_num
## 1       2.273      -2.414  16.824      -0.277       258.733           2
## 2       0.501       0.103  44.704      -1.916       164.546           1
## 3      -0.953       1.052  54.283      -2.186       260.414           1
## 4      -0.522      -3.100  31.082       0.060        86.062           0
## 5       0.798       1.569   2.723      -0.871        53.131           0
## 6      -0.884       1.857  40.735       2.237       282.849           3
##   PRI_jet_leading_pt PRI_jet_leading_eta PRI_jet_leading_phi
## 1             67.435               2.150                0.444
## 2             46.226               0.725                1.158
## 3             44.251               2.053               -2.028
## 4           -999.000            -999.000            -999.000
## 5           -999.000            -999.000            -999.000
## 6             90.547              -2.412               -0.653
##   PRI_jet_subleading_pt PRI_jet_subleading_eta PRI_jet_subleading_phi
## 1                46.062                  1.240                 -2.475
## 2              -999.000               -999.000               -999.000
## 3              -999.000               -999.000               -999.000
## 4              -999.000               -999.000               -999.000
## 5              -999.000               -999.000               -999.000
## 6                56.165                  0.224                  3.106
##   PRI_jet_all_pt      Weight Label
## 1        113.497 0.002653311     s
## 2         46.226 2.233584487     b
## 3         44.251 2.347388944     b
## 4          0.000 5.446378212     b
## 5          0.000 6.245332687     b
## 6        193.660 0.083414031     b
```

```
##  [1] "EventId"                     "DER_mass_MMC"
##  [3] "DER_mass_transverse_met_lep" "DER_mass_vis"
##  [5] "DER_pt_h"                    "DER_deltaeta_jet_jet"
##  [7] "DER_mass_jet_jet"            "DER_prodeta_jet_jet"
##  [9] "DER_deltar_tau_lep"          "DER_pt_tot"
## [11] "DER_sum_pt"                  "DER_pt_ratio_lep_tau"
## [13] "DER_met_phi_centrality"      "DER_lep_eta_centrality"
## [15] "PRI_tau_pt"                  "PRI_tau_eta"
## [17] "PRI_tau_phi"                 "PRI_lep_pt"
## [19] "PRI_lep_eta"                 "PRI_lep_phi"
```

```
## [21] "PRI_met"                    "PRI_met_phi"
## [23] "PRI_met_sumet"              "PRI_jet_num"
## [25] "PRI_jet_leading_pt"         "PRI_jet_leading_eta"
## [27] "PRI_jet_leading_phi"        "PRI_jet_subleading_pt"
## [29] "PRI_jet_subleading_eta"     "PRI_jet_subleading_phi"
## [31] "PRI_jet_all_pt"             "Weight"
## [33] "Label"
```

```
## [1] 10000    32
```

Data Exploration

- The correlation matrix structure is presented below.

```
correlationMatrix <- cor(df_small[,1:31])
print(correlationMatrix)
```

```
##                              DER_mass_MMC DER_mass_transverse_met_lep
## DER_mass_MMC                 1.0000000000                -0.448497005
## DER_mass_transverse_met_lep -0.4484970054                 1.000000000
## DER_mass_vis                 0.1812117691                 0.181802981
## DER_pt_h                     0.1812179616                -0.189210054
## DER_deltaeta_jet_jet         0.1638699919                -0.185861051
## DER_mass_jet_jet             0.1595273791                -0.192890815
## DER_prodeta_jet_jet          0.1637461042                -0.185530480
## DER_deltar_tau_lep           0.2373392462                 0.043643126
## DER_pt_tot                   0.0412832070                 0.084530649
## DER_sum_pt                   0.2010938186                -0.153463346
## DER_pt_ratio_lep_tau        -0.0129550071                 0.339506535
## DER_met_phi_centrality       0.3472909545                -0.421903101
## DER_lep_eta_centrality       0.1638363322                -0.185774925
## PRI_tau_pt                   0.1350814413                -0.151837061
## PRI_tau_eta                 -0.0030825163                 0.005939972
## PRI_tau_phi                  0.0083560783                 0.002594532
## PRI_lep_pt                   0.1018718018                 0.294813516
## PRI_lep_eta                  0.0056060563                -0.009712292
## PRI_lep_phi                 -0.0007556288                 0.014085201
## PRI_met                     -0.1645583448                 0.196071072
## PRI_met_phi                 -0.0061185499                -0.013359124
## PRI_met_sumet                0.2250319887                -0.168250101
## PRI_jet_num                  0.2204164417                -0.221169414
## PRI_jet_leading_pt           0.2464260253                -0.234458320
## PRI_jet_leading_eta          0.2436133167                -0.225780908
## PRI_jet_leading_phi          0.2436453889                -0.225796756
## PRI_jet_subleading_pt        0.1641114401                -0.185989154
## PRI_jet_subleading_eta       0.1638355966                -0.185731289
## PRI_jet_subleading_phi       0.1638150273                -0.185726809
## PRI_jet_all_pt               0.1837734349                -0.211206232
## Weight                      -0.3318222502                 0.424403474
##                              DER_mass_vis     DER_pt_h DER_deltaeta_jet_jet
## DER_mass_MMC                 0.181211769  0.181217962          0.163869992
## DER_mass_transverse_met_lep  0.181802981 -0.189210054         -0.185861051
## DER_mass_vis                 1.000000000 -0.065783885         -0.041133148
## DER_pt_h                    -0.065783885  1.000000000          0.484909288
## DER_deltaeta_jet_jet        -0.041133148  0.484909288          1.000000000
```

```
## DER_mass_jet_jet            -0.046303244  0.488430319          0.943832289
## DER_prodeta_jet_jet         -0.041024863  0.484950970          0.999981158
## DER_deltar_tau_lep           0.595092623 -0.492006910         -0.305334798
## DER_pt_tot                  -0.004243658  0.507396522          0.171616265
## DER_sum_pt                   0.084029481  0.762867297          0.674478631
## DER_pt_ratio_lep_tau         0.100627520  0.075721928          0.044656038
## DER_met_phi_centrality      -0.101314838  0.493014895          0.378826018
## DER_lep_eta_centrality      -0.041110246  0.484982136          0.999998363
## PRI_tau_pt                   0.280575657  0.379240308          0.184817256
## PRI_tau_eta                  0.001414333  0.013108298          0.010313774
## PRI_tau_phi                  0.005254652 -0.011701121         -0.004981598
## PRI_lep_pt                   0.408990071  0.320349084          0.164713512
## PRI_lep_eta                 -0.002119010  0.017627993          0.018119001
## PRI_lep_phi                 -0.007487205 -0.011578806         -0.006049651
## PRI_met                     -0.082592299  0.741755489          0.232851720
## PRI_met_phi                  0.002798646 -0.005889562         -0.003584592
## PRI_met_sumet                0.054415148  0.717358866          0.626152466
## PRI_jet_num                 -0.036573637  0.576766192          0.870130172
## PRI_jet_leading_pt          -0.024288196  0.569306396          0.551476651
## PRI_jet_leading_eta         -0.019557893  0.518266150          0.528501173
## PRI_jet_leading_phi         -0.019563843  0.518297072          0.528496385
## PRI_jet_subleading_pt       -0.041685214  0.491993095          0.999367476
## PRI_jet_subleading_eta      -0.041099485  0.485006222          0.999995038
## PRI_jet_subleading_phi      -0.041097051  0.485000401          0.999995455
## PRI_jet_all_pt              -0.054255096  0.741689440          0.715699589
## Weight                       0.095900648 -0.391103714         -0.401466347
##                            DER_mass_jet_jet DER_prodeta_jet_jet
## DER_mass_MMC                  0.1595273791         0.163746104
## DER_mass_transverse_met_lep  -0.1928908153        -0.185530480
## DER_mass_vis                 -0.0463032437        -0.041024863
## DER_pt_h                      0.4884303192         0.484950970
## DER_deltaeta_jet_jet          0.9438322895         0.999981158
## DER_mass_jet_jet              1.0000000000         0.942189170
## DER_prodeta_jet_jet           0.9421891704         1.000000000
## DER_deltar_tau_lep           -0.3034763687        -0.305411353
## DER_pt_tot                    0.1570441700         0.171970122
## DER_sum_pt                    0.6768307778         0.674699177
## DER_pt_ratio_lep_tau          0.0275502552         0.045012188
## DER_met_phi_centrality        0.3748294462         0.378565060
## DER_lep_eta_centrality        0.9433605653         0.999989542
## PRI_tau_pt                    0.1995833664         0.184693076
## PRI_tau_eta                   0.0066547001         0.010356643
## PRI_tau_phi                   0.0009440849        -0.005102918
## PRI_lep_pt                    0.1596102175         0.164975366
## PRI_lep_eta                   0.0135384362         0.018143049
## PRI_lep_phi                  -0.0054544359        -0.006075834
## PRI_met                       0.2437447466         0.232898784
## PRI_met_phi                  -0.0009602820        -0.003614500
## PRI_met_sumet                 0.6184186484         0.626497032
## PRI_jet_num                   0.8149474942         0.870327261
## PRI_jet_leading_pt            0.5258389510         0.551479401
## PRI_jet_leading_eta           0.4984942216         0.528497666
## PRI_jet_leading_phi           0.4985241649         0.528492729
## PRI_jet_subleading_pt         0.9454074761         0.999368238
```

```
## PRI_jet_subleading_eta          0.9433341236         0.999987499
## PRI_jet_subleading_phi          0.9432707137         0.999988535
## PRI_jet_all_pt                  0.7162648929         0.715929433
## Weight                         -0.3989355549        -0.401091136
##                            DER_deltar_tau_lep   DER_pt_tot   DER_sum_pt
## DER_mass_MMC                       0.2373392462  0.041283207  0.201093819
## DER_mass_transverse_met_lep        0.0436431261  0.084530649 -0.153463346
## DER_mass_vis                       0.5950926234 -0.004243658  0.084029481
## DER_pt_h                          -0.4920069098  0.507396522  0.762867297
## DER_deltaeta_jet_jet              -0.3053347984  0.171616265  0.674478631
## DER_mass_jet_jet                  -0.3034763687  0.157044170  0.676830778
## DER_prodeta_jet_jet               -0.3054113526  0.171970122  0.674699177
## DER_deltar_tau_lep                 1.0000000000 -0.096600747 -0.426436281
## DER_pt_tot                        -0.0966007468  1.000000000  0.246517306
## DER_sum_pt                        -0.4264362812  0.246517306  1.000000000
## DER_pt_ratio_lep_tau               0.0480275544  0.020577327  0.108657323
## DER_met_phi_centrality            -0.2114081970  0.112054261  0.421228998
## DER_lep_eta_centrality            -0.3053911584  0.171730827  0.674616839
## PRI_tau_pt                        -0.2057272861  0.075152712  0.484916354
## PRI_tau_eta                        0.0003330023  0.010509458  0.006692191
## PRI_tau_phi                        0.0091190606 -0.008814493 -0.004391554
## PRI_lep_pt                        -0.0605473771  0.068474906  0.455233962
## PRI_lep_eta                       -0.0054640638  0.013488847  0.010658199
## PRI_lep_phi                       -0.0003013820 -0.002236615 -0.015350077
## PRI_met                           -0.3056794398  0.639600587  0.398842064
## PRI_met_phi                        0.0142913663  0.013742360 -0.012884406
## PRI_met_sumet                     -0.3996954012  0.286913358  0.905601998
## PRI_jet_num                       -0.3555891716  0.225257870  0.765206650
## PRI_jet_leading_pt                -0.3418976103  0.124077976  0.641876700
## PRI_jet_leading_eta               -0.3121247244  0.113243750  0.583015882
## PRI_jet_leading_phi               -0.3121279283  0.113210296  0.583062027
## PRI_jet_subleading_pt             -0.3091933966  0.176738663  0.689811103
## PRI_jet_subleading_eta            -0.3053960192  0.171745097  0.674651224
## PRI_jet_subleading_phi            -0.3053899051  0.171772333  0.674633814
## PRI_jet_all_pt                    -0.4420756432  0.258052026  0.967222767
## Weight                             0.2014868257 -0.143833147 -0.423322283
##                            DER_pt_ratio_lep_tau DER_met_phi_centrality
## DER_mass_MMC                       -0.012955007            0.347290954
## DER_mass_transverse_met_lep         0.339506535           -0.421903101
## DER_mass_vis                        0.100627520           -0.101314838
## DER_pt_h                            0.075721928            0.493014895
## DER_deltaeta_jet_jet                0.044656038            0.378826018
## DER_mass_jet_jet                    0.027550255            0.374829446
## DER_prodeta_jet_jet                 0.045012188            0.378565060
## DER_deltar_tau_lep                  0.048027554           -0.211408197
## DER_pt_tot                          0.020577327            0.112054261
## DER_sum_pt                          0.108657323            0.421228998
## DER_pt_ratio_lep_tau                1.000000000           -0.040874289
## DER_met_phi_centrality             -0.040874289            1.000000000
## DER_lep_eta_centrality              0.044758714            0.378768326
## PRI_tau_pt                         -0.483272204            0.146045117
## PRI_tau_eta                         0.008736513           -0.003189370
## PRI_tau_phi                         0.008326872           -0.004569967
## PRI_lep_pt                          0.702619596            0.046058296
```

```
## PRI_lep_eta                          0.007495023           0.014888348
## PRI_lep_phi                          0.001575012          -0.005511828
## PRI_met                              0.036655745           0.140243146
## PRI_met_phi                          0.005605954          -0.005624688
## PRI_met_sumet                        0.051159391           0.423668472
## PRI_jet_num                          0.056750334           0.497074129
## PRI_jet_leading_pt                   0.042930889           0.554255468
## PRI_jet_leading_eta                  0.036838980           0.540812837
## PRI_jet_leading_phi                  0.036796319           0.540809746
## PRI_jet_subleading_pt                0.045719794           0.379746491
## PRI_jet_subleading_eta               0.044796237           0.378764976
## PRI_jet_subleading_phi               0.044836021           0.378746589
## PRI_jet_all_pt                       0.082111845           0.452572350
## Weight                               0.193688638          -0.466890980
##                            DER_lep_eta_centrality   PRI_tau_pt   PRI_tau_eta
## DER_mass_MMC                           0.163836332  0.135081441 -0.0030825163
## DER_mass_transverse_met_lep           -0.185774925 -0.151837061  0.0059399715
## DER_mass_vis                          -0.041110246  0.280575657  0.0014143326
## DER_pt_h                               0.484982136  0.379240308  0.0131082981
## DER_deltaeta_jet_jet                   0.999998363  0.184817256  0.0103137736
## DER_mass_jet_jet                       0.943360565  0.199583366  0.0066547001
## DER_prodeta_jet_jet                    0.999989542  0.184693076  0.0103566433
## DER_deltar_tau_lep                    -0.305391158 -0.205727286  0.0003330023
## DER_pt_tot                             0.171730827  0.075152712  0.0105094583
## DER_sum_pt                             0.674616839  0.484916354  0.0066921909
## DER_pt_ratio_lep_tau                   0.044758714 -0.483272204  0.0087365131
## DER_met_phi_centrality                 0.378768326  0.146045117 -0.0031893700
## DER_lep_eta_centrality                 1.000000000  0.184813941  0.0103198100
## PRI_tau_pt                             0.184813941  1.000000000 -0.0078603888
## PRI_tau_eta                            0.010319810 -0.007860389  1.0000000000
## PRI_tau_phi                           -0.005012132 -0.011080433  0.0001104502
## PRI_lep_pt                             0.164810854  0.093181185  0.0012947723
## PRI_lep_eta                            0.018127579 -0.012991412  0.5636073291
## PRI_lep_phi                           -0.006057094 -0.001931584  0.0047705001
## PRI_met                                0.232909313  0.124201175  0.0139251640
## PRI_met_phi                           -0.003594708 -0.024162820 -0.0076492492
## PRI_met_sumet                          0.626324594  0.448130448  0.0053208243
## PRI_jet_num                            0.870206211  0.212621084  0.0138426453
## PRI_jet_leading_pt                     0.551486179  0.207787374  0.0157812472
## PRI_jet_leading_eta                    0.528502317  0.180928815  0.0161194014
## PRI_jet_leading_phi                    0.528497475  0.181014334  0.0156666928
## PRI_jet_subleading_pt                  0.999375714  0.189060607  0.0102114814
## PRI_jet_subleading_eta                 0.999996954  0.184778707  0.0105109979
## PRI_jet_subleading_phi                 0.999997559  0.184804606  0.0103321255
## PRI_jet_all_pt                         0.715841532  0.324881525  0.0093584174
## Weight                                -0.401366201 -0.296258450 -0.0029753519
##                              PRI_tau_phi   PRI_lep_pt   PRI_lep_eta
## DER_mass_MMC                0.0083560783  0.101871802  0.0056060563
## DER_mass_transverse_met_lep 0.0025945322  0.294813516 -0.0097122915
## DER_mass_vis                0.0052546519  0.408990071 -0.0021190098
## DER_pt_h                   -0.0117011210  0.320349084  0.0176279935
## DER_deltaeta_jet_jet       -0.0049815978  0.164713512  0.0181190010
## DER_mass_jet_jet            0.0009440849  0.159610217  0.0135384362
## DER_prodeta_jet_jet        -0.0051029181  0.164975366  0.0181430491
```

```
## DER_deltar_tau_lep           0.0091190606 -0.060547377 -0.0054640638
## DER_pt_tot                  -0.0088144932  0.068474906  0.0134888466
## DER_sum_pt                  -0.0043915544  0.455233962  0.0106581994
## DER_pt_ratio_lep_tau         0.0083268718  0.702619596  0.0074950233
## DER_met_phi_centrality      -0.0045699674  0.046058296  0.0148883478
## DER_lep_eta_centrality      -0.0050121316  0.164810854  0.0181275794
## PRI_tau_pt                  -0.0110804326  0.093181185 -0.0129914119
## PRI_tau_eta                  0.0001104502  0.001294772  0.5636073291
## PRI_tau_phi                  1.0000000000 -0.001151654 -0.0101381089
## PRI_lep_pt                  -0.0011516541  1.000000000  0.0018188127
## PRI_lep_eta                 -0.0101381089  0.001818813  1.0000000000
## PRI_lep_phi                 -0.2099886318 -0.001453657  0.0023766868
## PRI_met                     -0.0102634387  0.122518006  0.0098681397
## PRI_met_phi                  0.0487265946 -0.012602868 -0.0006731478
## PRI_met_sumet               -0.0016658899  0.362493824  0.0092214080
## PRI_jet_num                 -0.0009857714  0.195002216  0.0246693893
## PRI_jet_leading_pt           0.0108413341  0.181309116  0.0295357277
## PRI_jet_leading_eta          0.0115934007  0.157352816  0.0304451339
## PRI_jet_leading_phi          0.0111459917  0.157367362  0.0299681367
## PRI_jet_subleading_pt       -0.0049315925  0.168495645  0.0179827433
## PRI_jet_subleading_eta      -0.0049756781  0.164859880  0.0183387216
## PRI_jet_subleading_phi      -0.0051406683  0.164892244  0.0181361329
## PRI_jet_all_pt              -0.0024189246  0.294680848  0.0150643093
## Weight                      -0.0045378087 -0.023863980 -0.0182722205
##                              PRI_lep_phi      PRI_met   PRI_met_phi
## DER_mass_MMC                -0.0007556288 -0.164558345 -0.0061185499
## DER_mass_transverse_met_lep  0.0140852011  0.196071072 -0.0133591237
## DER_mass_vis                -0.0074872045 -0.082592299  0.0027986465
## DER_pt_h                    -0.0115788062  0.741755489 -0.0058895621
## DER_deltaeta_jet_jet        -0.0060496513  0.232851720 -0.0035845920
## DER_mass_jet_jet            -0.0054544359  0.243744747 -0.0009602820
## DER_prodeta_jet_jet         -0.0060758340  0.232898784 -0.0036144995
## DER_deltar_tau_lep          -0.0003013820 -0.305679440  0.0142913663
## DER_pt_tot                  -0.0022366146  0.639600587  0.0137423600
## DER_sum_pt                  -0.0153500771  0.398842064 -0.0128844063
## DER_pt_ratio_lep_tau         0.0015750118  0.036655745  0.0056059537
## DER_met_phi_centrality      -0.0055118279  0.140243146 -0.0056246881
## DER_lep_eta_centrality      -0.0060570941  0.232909313 -0.0035947080
## PRI_tau_pt                  -0.0019315843  0.124201175 -0.0241628199
## PRI_tau_eta                  0.0047705001  0.013925164 -0.0076492492
## PRI_tau_phi                 -0.2099886318 -0.010263439  0.0487265946
## PRI_lep_pt                  -0.0014536573  0.122518006 -0.0126028678
## PRI_lep_eta                  0.0023766868  0.009868140 -0.0006731478
## PRI_lep_phi                  1.0000000000 -0.003202782  0.0239994057
## PRI_met                     -0.0032027816  1.000000000 -0.0010680575
## PRI_met_phi                  0.0239994057 -0.001068058  1.0000000000
## PRI_met_sumet               -0.0160409125  0.371482445 -0.0143398016
## PRI_jet_num                 -0.0100033943  0.256003195  0.0007389220
## PRI_jet_leading_pt          -0.0058461831  0.227911669  0.0028492541
## PRI_jet_leading_eta         -0.0043004739  0.191513237  0.0040541787
## PRI_jet_leading_phi         -0.0049318668  0.191507633  0.0033933550
## PRI_jet_subleading_pt       -0.0068033199  0.238326972 -0.0039487046
## PRI_jet_subleading_eta      -0.0060364228  0.232945866 -0.0036074257
## PRI_jet_subleading_phi      -0.0062541409  0.232916138 -0.0037530327
```

```
## PRI_jet_all_pt              -0.0173050467  0.414336006 -0.0069470067
## Weight                       0.0090014644 -0.095872796 -0.0013223525
##                            PRI_met_sumet   PRI_jet_num PRI_jet_leading_pt
## DER_mass_MMC                0.225031989  0.2204164417        0.246426025
## DER_mass_transverse_met_lep -0.168250101 -0.2211694141       -0.234458320
## DER_mass_vis                0.054415148 -0.0365736365       -0.024288196
## DER_pt_h                    0.717358866  0.5767661921        0.569306396
## DER_deltaeta_jet_jet        0.626152466  0.8701301720        0.551476651
## DER_mass_jet_jet            0.618418648  0.8149474942        0.525838951
## DER_prodeta_jet_jet         0.626497032  0.8703272611        0.551479401
## DER_deltar_tau_lep         -0.399695401 -0.3555891716       -0.341897610
## DER_pt_tot                  0.286913358  0.2252578704        0.124077976
## DER_sum_pt                  0.905601998  0.7652066498        0.641876700
## DER_pt_ratio_lep_tau        0.051159391  0.0567503345        0.042930889
## DER_met_phi_centrality      0.423668472  0.4970741291        0.554255468
## DER_lep_eta_centrality      0.626324594  0.8702062111        0.551486179
## PRI_tau_pt                  0.448130448  0.2126210840        0.207787374
## PRI_tau_eta                 0.005320824  0.0138426453        0.015781247
## PRI_tau_phi                -0.001665890 -0.0009857714        0.010841334
## PRI_lep_pt                  0.362493824  0.1950022157        0.181309116
## PRI_lep_eta                 0.009221408  0.0246693893        0.029535728
## PRI_lep_phi                -0.016040912 -0.0100033943       -0.005846183
## PRI_met                     0.371482445  0.2560031945        0.227911669
## PRI_met_phi                -0.014339802  0.0007389220        0.002849254
## PRI_met_sumet               1.000000000  0.7147499761        0.598622564
## PRI_jet_num                 0.714749976  1.0000000000        0.832828780
## PRI_jet_leading_pt          0.598622564  0.8328287799        1.000000000
## PRI_jet_leading_eta         0.546037315  0.8173917756        0.996197093
## PRI_jet_leading_phi         0.546085185  0.8174184819        0.996199574
## PRI_jet_subleading_pt       0.639676661  0.8722899579        0.552499994
## PRI_jet_subleading_eta      0.626351137  0.8702331538        0.551488963
## PRI_jet_subleading_phi      0.626343542  0.8702178602        0.551483391
## PRI_jet_all_pt              0.884840328  0.8095213359        0.668521565
## Weight                     -0.432813431 -0.4999881223       -0.516693334
##                            PRI_jet_leading_eta PRI_jet_leading_phi
## DER_mass_MMC                       0.243613317         0.243645389
## DER_mass_transverse_met_lep       -0.225780908        -0.225796756
## DER_mass_vis                      -0.019557893        -0.019563843
## DER_pt_h                           0.518266150         0.518297072
## DER_deltaeta_jet_jet               0.528501173         0.528496385
## DER_mass_jet_jet                   0.498494222         0.498524165
## DER_prodeta_jet_jet                0.528497666         0.528492729
## DER_deltar_tau_lep                -0.312124724        -0.312127928
## DER_pt_tot                         0.113243750         0.113210296
## DER_sum_pt                         0.583015882         0.583062027
## DER_pt_ratio_lep_tau               0.036838980         0.036796319
## DER_met_phi_centrality             0.540812837         0.540809746
## DER_lep_eta_centrality             0.528502317         0.528497475
## PRI_tau_pt                         0.180928815         0.181014334
## PRI_tau_eta                        0.016119401         0.015666693
## PRI_tau_phi                        0.011593401         0.011145992
## PRI_lep_pt                         0.157352816         0.157367362
## PRI_lep_eta                        0.030445134         0.029968137
## PRI_lep_phi                       -0.004300474        -0.004931867
```

```
## PRI_met                              0.191513237         0.191507633
## PRI_met_phi                          0.004054179         0.003393355
## PRI_met_sumet                        0.546037315         0.546085185
## PRI_jet_num                          0.817391776         0.817418482
## PRI_jet_leading_pt                   0.996197093         0.996199574
## PRI_jet_leading_eta                  1.000000000         0.999991932
## PRI_jet_leading_phi                  0.999991932         1.000000000
## PRI_jet_subleading_pt                0.528173229         0.528170702
## PRI_jet_subleading_eta               0.528499740         0.528496050
## PRI_jet_subleading_phi               0.528501166         0.528495994
## PRI_jet_all_pt                       0.610586496         0.610618333
## Weight                              -0.507384165        -0.507397794
##                              PRI_jet_subleading_pt PRI_jet_subleading_eta
## DER_mass_MMC                          0.164111440           0.163835597
## DER_mass_transverse_met_lep          -0.185989154          -0.185731289
## DER_mass_vis                         -0.041685214          -0.041099485
## DER_pt_h                              0.491993095           0.485006222
## DER_deltaeta_jet_jet                  0.999367476           0.999995038
## DER_mass_jet_jet                      0.945407476           0.943334124
## DER_prodeta_jet_jet                   0.999368238           0.999987499
## DER_deltar_tau_lep                   -0.309193397          -0.305396019
## DER_pt_tot                            0.176738663           0.171745097
## DER_sum_pt                            0.689811103           0.674651224
## DER_pt_ratio_lep_tau                  0.045719794           0.044796237
## DER_met_phi_centrality                0.379746491           0.378764976
## DER_lep_eta_centrality                0.999375714           0.999996954
## PRI_tau_pt                            0.189060607           0.184778707
## PRI_tau_eta                           0.010211481           0.010510998
## PRI_tau_phi                          -0.004931593          -0.004975678
## PRI_lep_pt                            0.168495645           0.164859880
## PRI_lep_eta                           0.017982743           0.018338722
## PRI_lep_phi                          -0.006803320          -0.006036423
## PRI_met                               0.238326972           0.232945866
## PRI_met_phi                          -0.003948705          -0.003607426
## PRI_met_sumet                         0.639676661           0.626351137
## PRI_jet_num                           0.872289958           0.870233154
## PRI_jet_leading_pt                    0.552499994           0.551488963
## PRI_jet_leading_eta                   0.528173229           0.528499740
## PRI_jet_leading_phi                   0.528170702           0.528496050
## PRI_jet_subleading_pt                 1.000000000           0.999373561
## PRI_jet_subleading_eta                0.999373561           1.000000000
## PRI_jet_subleading_phi                0.999373564           0.999994739
## PRI_jet_all_pt                        0.731951330           0.715879137
## Weight                               -0.401951770          -0.401356704
##                              PRI_jet_subleading_phi PRI_jet_all_pt       Weight
## DER_mass_MMC                          0.163815027    0.183773435 -0.331822250
## DER_mass_transverse_met_lep          -0.185726809   -0.211206232  0.424403474
## DER_mass_vis                         -0.041097051   -0.054255096  0.095900648
## DER_pt_h                              0.485000401    0.741689440 -0.391103714
## DER_deltaeta_jet_jet                  0.999995455    0.715699589 -0.401466347
## DER_mass_jet_jet                      0.943270714    0.716264893 -0.398935555
## DER_prodeta_jet_jet                   0.999988535    0.715929433 -0.401091136
## DER_deltar_tau_lep                   -0.305389905   -0.442075643  0.201486826
## DER_pt_tot                            0.171772333    0.258052026 -0.143833147
```

```
## DER_sum_pt                           0.674633814    0.967222767 -0.423322283
## DER_pt_ratio_lep_tau                 0.044836021    0.082111845  0.193688638
## DER_met_phi_centrality               0.378746589    0.452572350 -0.466890980
## DER_lep_eta_centrality               0.999997559    0.715841532 -0.401366201
## PRI_tau_pt                           0.184804606    0.324881525 -0.296258450
## PRI_tau_eta                          0.010332125    0.009358417 -0.002975352
## PRI_tau_phi                         -0.005140668   -0.002418925 -0.004537809
## PRI_lep_pt                           0.164892244    0.294680848 -0.023863980
## PRI_lep_eta                          0.018136133    0.015064309 -0.018272221
## PRI_lep_phi                         -0.006254141   -0.017305047  0.009001464
## PRI_met                              0.232916138    0.414336006 -0.095872796
## PRI_met_phi                         -0.003753033   -0.006947007 -0.001322352
## PRI_met_sumet                        0.626343542    0.884840328 -0.432813431
## PRI_jet_num                          0.870217860    0.809521336 -0.499988122
## PRI_jet_leading_pt                   0.551483391    0.668521565 -0.516693334
## PRI_jet_leading_eta                  0.528501166    0.610586496 -0.507384165
## PRI_jet_leading_phi                  0.528495994    0.610618333 -0.507397794
## PRI_jet_subleading_pt                0.999373564    0.731951330 -0.401951770
## PRI_jet_subleading_eta               0.999994739    0.715879137 -0.401356704
## PRI_jet_subleading_phi               1.000000000    0.715845704 -0.401328623
## PRI_jet_all_pt                       0.715845704    1.000000000 -0.426097952
## Weight                              -0.401328623   -0.426097952  1.000000000
```

```
highlyCorrelated <- findCorrelation(correlationMatrix, cutoff=0.5)
print(highlyCorrelated)
```

```
##  [1] 23 30 10 27 29 28 13  5  7  6 22 24  4 25 26  8 20 17 15
```

Variable Importance

- The predictors important plot is shown below.

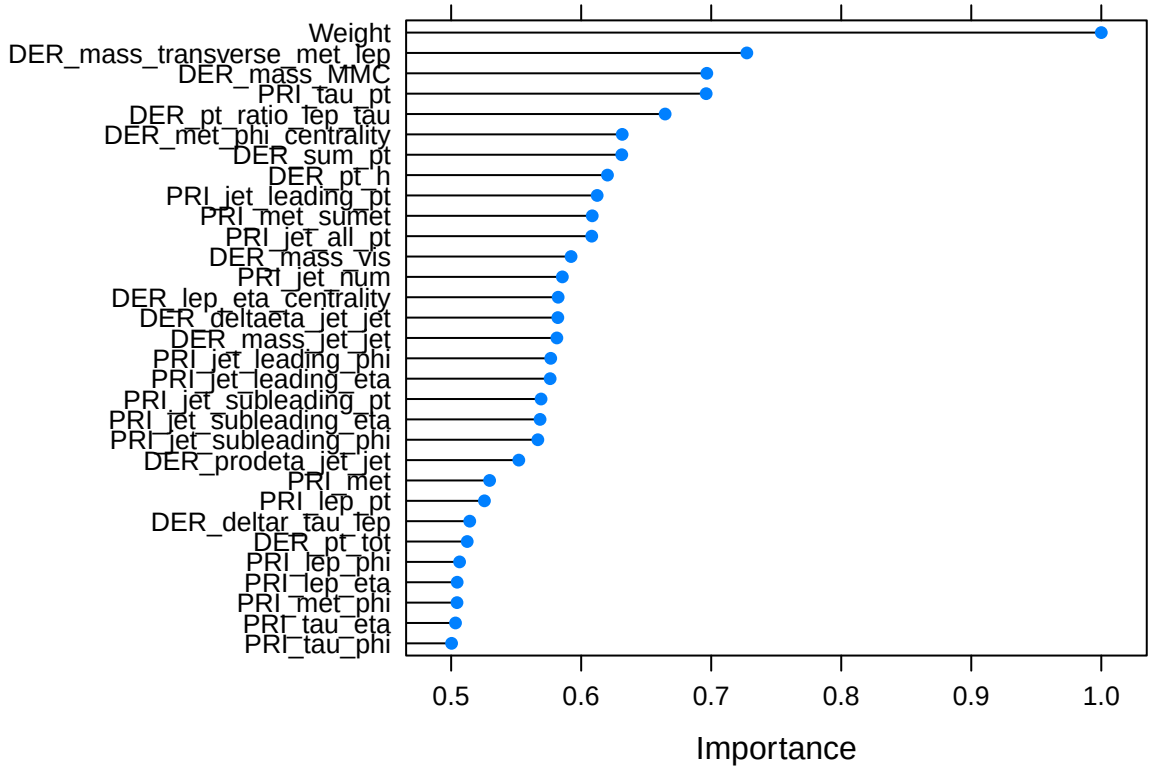
```
control <- trainControl(method = "repeatedcv",number=3, repeats=2)

model1<- train(Label~., data=df_small, method="lvq",  trControl=control)
importance<-varImp(model1,scale=FALSE)
print(importance)
```

```
## ROC curve variable importance
##
##   only 20 most important variables shown (out of 31)
##
##                              Importance
## Weight                           1.0000
## DER_mass_transverse_met_lep      0.7274
## DER_mass_MMC                     0.6966
## PRI_tau_pt                       0.6961
## DER_pt_ratio_lep_tau             0.6646
## DER_met_phi_centrality           0.6316
## DER_sum_pt                       0.6312
## DER_pt_h                         0.6202
## PRI_jet_leading_pt               0.6123
## PRI_met_sumet                    0.6085
## PRI_jet_all_pt                   0.6081
## DER_mass_vis                     0.5922
## PRI_jet_num                      0.5855
```

```
## DER_lep_eta_centrality          0.5823
## DER_deltaeta_jet_jet            0.5821
## DER_mass_jet_jet                0.5813
## PRI_jet_leading_phi             0.5766
## PRI_jet_leading_eta             0.5761
## PRI_jet_subleading_pt           0.5692
## PRI_jet_subleading_eta          0.5684
```

```
# plot importance
plot(importance)
```



Lasso Regression

- Modelled Lasso regression for feature selection.

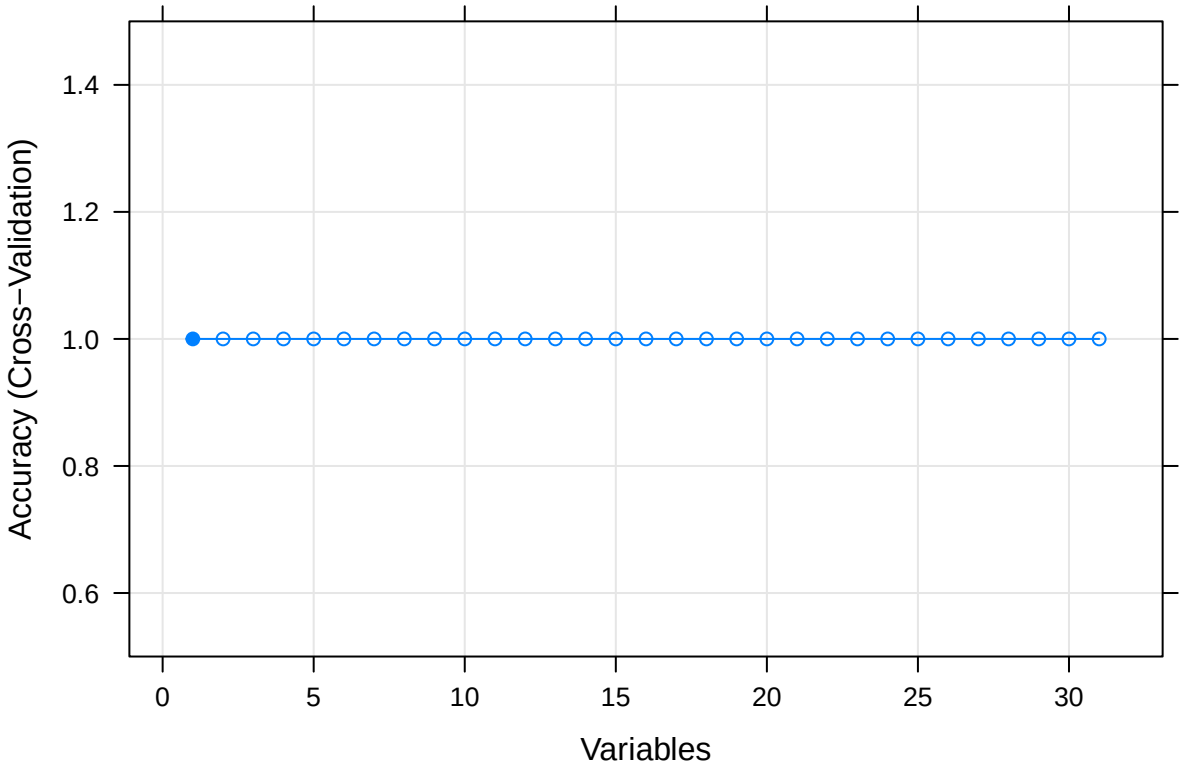
```
set.seed(7)
# load the library
#library(mlbench)
#library(caret)
# load the data
#data(PimaIndiansDiabetes)
# define the control using a random forest selection function
control <- rfeControl(functions=rfFuncs, method="cv", number=3)
# run the RFE algorithm
results <- rfe(df_small[,1:31], df_small[,32], sizes=c(1:31), rfeControl=control)
# summarize the results
print(results)
```

```
##
## Recursive feature selection
##
## Outer resampling method: Cross-Validated (3 fold)
##
## Resampling performance over subset size:
##
##  Variables Accuracy Kappa AccuracySD KappaSD Selected
##          1        1     1          0       0        *
##          2        1     1          0       0
##          3        1     1          0       0
##          4        1     1          0       0
##          5        1     1          0       0
##          6        1     1          0       0
##          7        1     1          0       0
##          8        1     1          0       0
##          9        1     1          0       0
##         10        1     1          0       0
##         11        1     1          0       0
##         12        1     1          0       0
##         13        1     1          0       0
##         14        1     1          0       0
##         15        1     1          0       0
##         16        1     1          0       0
##         17        1     1          0       0
##         18        1     1          0       0
##         19        1     1          0       0
##         20        1     1          0       0
##         21        1     1          0       0
##         22        1     1          0       0
##         23        1     1          0       0
##         24        1     1          0       0
##         25        1     1          0       0
##         26        1     1          0       0
##         27        1     1          0       0
##         28        1     1          0       0
##         29        1     1          0       0
##         30        1     1          0       0
##         31        1     1          0       0
##
## The top 1 variables (out of 1):
##    Weight
```

```
# list the chosen features
predictors(results)
```

```
## [1] "Weight"
```

```
# plot the results
plot(results, type=c("g", "o"))
```



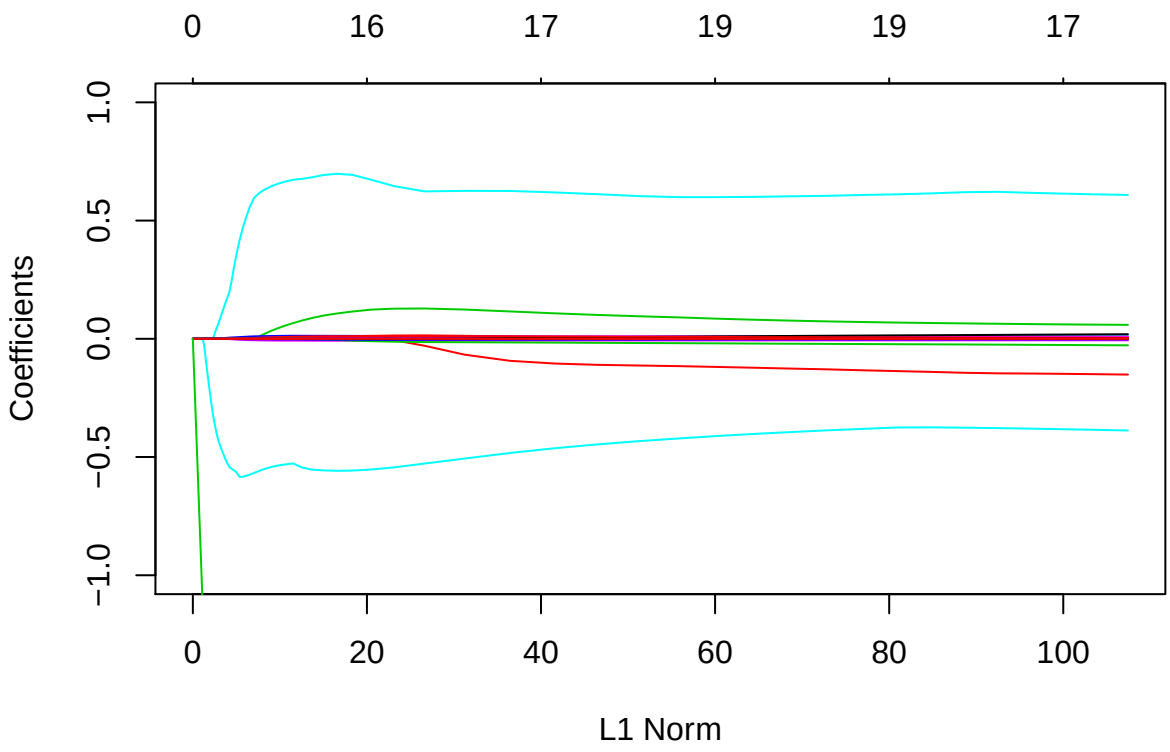

```
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 3.0-2
```

```
y=df_small[,32]
x=df_small[,1:31]
lassoRes <- glmnet(as.matrix(x),as.matrix(df_small$Label),alpha=1, lambda=10^((-80:80)/20), family = 'b
plot(lassoRes, ylim=c(-1,1))
```

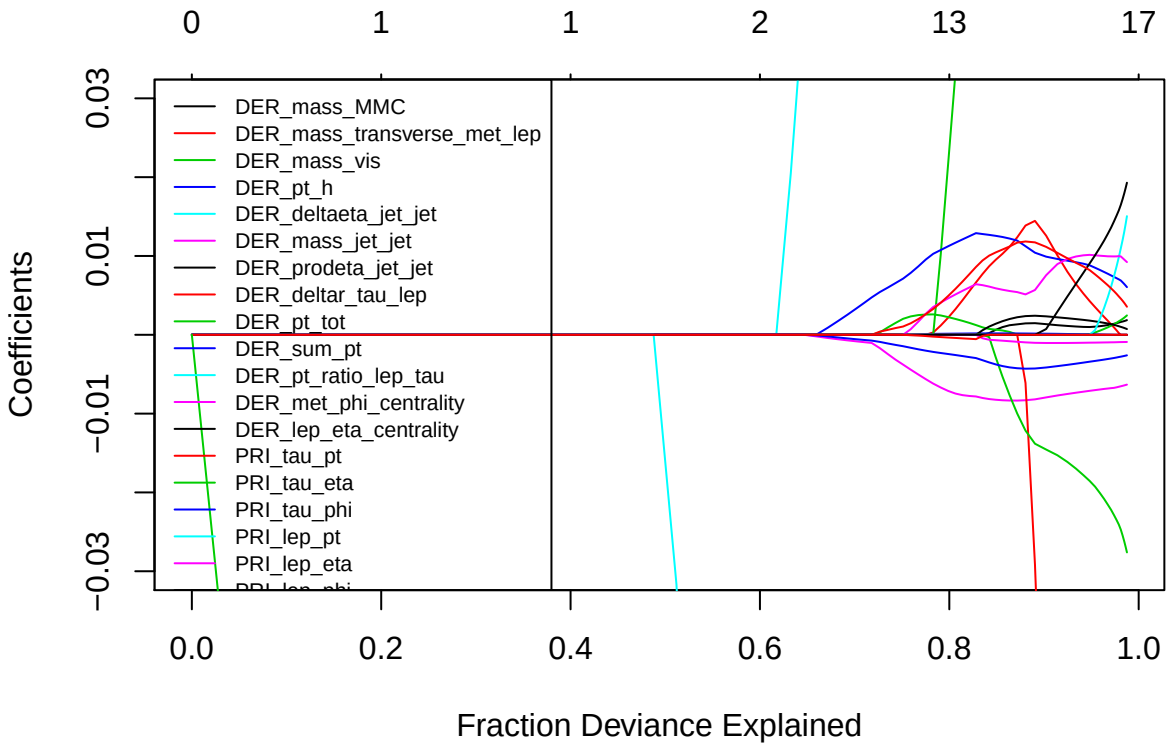
```
## Warning in regularize.values(x, y, ties, missing(ties)): collapsing to unique
## 'x' values
```




```
plot(lassoRes, xvar='dev',ylim=c(-0.03, 0.03))
```

```
## Warning in regularize.values(x, y, ties, missing(ties)): collapsing to unique
## 'x' values
```

```
legend("topleft", lwd = 1, col = 1:6, legend = colnames(x), cex = 0.7)
```



Modelling

Conclusions