Bao Tai Le

Masterthesis update

SoSe 2024
8.10.2024

So how did my last week look like?

Dataset	**Create the batches**	**Train the Embedding Layer**	**Train the Classification Neural Network**
Variables are found and looks promising	The functions are created and work as intended	The functions is created and works as intended with a smaller sample	Needs to be set up
Just need a bigger dataset	Bigger batch sizes	PCA is set up	
		Model Loading and saving is setup	
	May need to do some unittesting to be 1000% sure	**Needs optimization** - regularization - different lossfunctions - deeper layers Etc.	
		Bigger batchsizes	

```
                                                             Sum momentum and inv. mass:         5700000
Number of Jets in Event 1199000 : 5
Event Nummer =1199100        Anzahl Eintraege fuer dieses Event =  864
Event Nummer =1199200        Anzahl Eintraege fuer dieses Event =  792
Event Nummer =1199300        Anzahl Eintraege fuer dieses Event =  698
Event Nummer =1199400        Anzahl Eintraege fuer dieses Event =  993
Event Nummer =1199500        Anzahl Eintraege fuer dieses Event =  743
Event Nummer =1199600        Anzahl Eintraege fuer dieses Event =  898
Event Nummer =1199700        Anzahl Eintraege fuer dieses Event =  646
Event Nummer =1199800        Anzahl Eintraege fuer dieses Event = 1005
Event Nummer =1199900        Anzahl Eintraege fuer dieses Event =  859
Event Nummer =1200000        Anzahl Eintraege fuer dieses Event = 1186
Event # 1200000
  end of file reached, close it and continue
End of File reached -1
1********* PYSTAT:  Statistics on Number of Events and Cross-sections *********

 ==============================================================================
 I                                I                               I          I
 I          Subprocess            I     Number of points          I   Sigma  I
 I                                I                               I          I
 I-------------------------------I-----------------------------I   (mb)   I
 I                                I                               I          I
 I N:o Type                       I     Generated      Tried  I          I
 I                                I                               I          I
 ==============================================================================
 I                                I                               I          I
 I   0 All included subprocesses  I         0          0 I  0.000D+00 I
 I  81 q + qbar -> Q + Qbar, mass  I         0          0 I  0.000D+00 I
 I  82 g + g -> Q + Qbar, massive  I         0          0 I  0.000D+00 I
 I                                I                               I          I
 ==============================================================================

 ********* Total number of errors, excluding junctions =    0 *************
 ********* Total number of errors, including junctions =    0 *************
 ********* Total number of warnings =                       0 *************
 ********* Fraction of events that fail fragmentation cuts =  1.00000 *********
```

```cpp
for (Int_t pp = 0; pp <= nParticles; pp++) {

if( !(
    (abs(pythia-> GetK(pp,2))== pdgcode)                                  //my particle is a b-quark/bbar-quark
    &&(abs(pythia->GetK(pythia->GetK(pp,3),2)) != abs(pythia->GetK(pp,2)))  //myparticle is not the same as its motherparticle
    &&(abs(pythia->GetK(pythia->GetK(pp,4),2)) != abs(pythia->GetK(pp,2)))  //myparticle is not the same as its daughterparticle
    &&(abs(pythia->GetK(pythia->GetK(pp,5),2)) != abs(pythia->GetK(pp,2)))  //myparticle is not the same as its daughterparticle
    //&&(abs(pythia->GetK(pythia->GetK(pp,3),2)) != 21)                      //the mother particle is not a gluon
    &&(abs(pythia->GetK(pythia->GetK(pp,4),2)) != 21)                       //the daughter particle is not a gluon
    &&(abs(pythia->GetK(pythia->GetK(pp,5),2)) != 21)
))
{
```

File Edit View Projects Bookmarks Sessions Tools Settings Help

```
1628    76.4987,75.6864,-11.1182,167.614,-149.139,247.624,-96.8152,227.913,-75.2104,258.793,4.31322,0.100562,289.904,1.53955,0
1629    83.8558,-63.5769,54.6789,109.983,71.1656,500.613,-287.204,410.034,299.493,583.361,0.321569,0.64262,65.8281,0.260621,0
1630    83.8558,-63.5769,54.6789,109.983,71.1656,500.613,-287.204,410.034,299.493,583.361,0.321569,0.64262,65.8281,0.260621,0
1631    83.8558,-63.5769,54.6789,109.983,71.1656,500.613,-287.204,410.034,299.493,583.361,0.321569,0.64262,65.8281,0.260621,0
1632    106.84,104.031,-24.3387,136.873,85.5528,136.805,-131.867,-36.4216,108.894,174.852,2.64229,0.163189,234.3,1.71823,0
1633    106.84,104.031,-24.3387,136.873,85.5528,167.641,167.627,-2.14807,-11.8507,168.059,0.832665,0.143844,114.244,0.77231,0
1634    24.6626,2.07709,-24.575,61.2181,56.0304,25.682,-5.05064,25.1805,-1.50948,25.7263,3.43302,0.541632,67.6579,2.04124,0
1635    51.5747,-46.0139,-23.2954,62.762,-35.7647,76.8189,76.7828,2.35514,50.6631,92.0212,3.79702,0.20776,149.501,2.77899,0
1636    51.5747,-46.0139,-23.2954,62.762,-35.7647,76.8189,76.7828,2.35514,50.6631,92.0212,3.79702,0.20776,149.501,2.77899,0
1637    43.2886,12.6858,-41.388,668.231,666.827,276.31,253.244,110.519,-35.3521,278.562,5.81315,0.544615,649.737,1.70516,0
1638    24.2949,23.893,-4.40099,25.164,-6.55619,44.4,-4.33879,44.1875,56.187,71.6124,4.62586,0.730196,70.2647,1.94965,0
1639    24.2949,23.893,-4.40099,25.164,-6.55619,124.314,24.5652,-121.862,-37.5204,129.852,1.19012,0.841623,61.6176,1.13841,0
1640    27.0681,15.628,22.1009,271.441,270.088,52.9334,-35.7482,39.0386,-47.9117,71.3966,4.04321,0.685917,253.046,2.28105,0
1641    25.5504,-5.94212,-24.8498,156.717,-154.62,55.3888,49.1522,-25.5339,50.7447,75.1195,3.57613,0.493447,196.347,2.26191,0
1642    76.4073,15.7546,-74.7654,91.8883,-51.0429,103.88,-57.0181,-86.8331,55.4568,117.756,1.38424,0.0417818,126.943,1.31257,0
1643    23.5685,23.4011,2.80383,24.5507,-6.87501,242.685,17.732,242.036,7.39722,242.798,1.41467,0.917092,99.1781,1.39496,0
1644    35.9922,-21.6115,-28.7815,36.1243,3.08711,114.455,-106.458,-42.031,340.424,359.15,1.81042,0.795361,129.713,1.21154,0
1645    54.9533,42.9899,34.2306,107.825,-92.7702,131.146,-122.032,-48.0373,-182.012,224.339,2.84888,0.232015,168.49,1.14484,0
1646    22.8963,22.8161,-1.91367,60.1572,55.6296,47.5083,-43.3015,-19.5453,280.356,284.353,2.7691,0.566374,70.1477,0.542987,0
1647    55.2657,45.4251,-31.4779,102.9,-86.7997,71.5301,51.0145,-50.1407,-125.919,144.818,0.196763,0.0517139,12.3612,0.101304,0
1648    55.2657,45.4251,-31.4779,102.9,-86.7997,71.5301,51.0145,-50.1407,-125.919,144.818,0.196763,0.0517139,12.3612,0.101304,0
1649    21.7214,21.4405,-3.48172,59.9539,55.8807,204.853,-200.925,39.9188,-22.8291,206.121,3.64391,0.444934,190.16,2.05226,0
1650    35.6984,6.75791,35.0529,41.3822,-20.9311,74.6058,-70.5837,24.1655,-134.798,154.067,1.63816,0.454802,79.8007,1.04652,0
1651    53.8094,-35.0334,-40.8424,55.9939,-15.4877,166.35,-125.549,109.133,612.528,634.715,2.78776,0.822615,300.285,1.84272,0
1652    87.9579,87.9535,0.878189,209.936,190.621,93.6821,85.9973,-37.1592,118.472,151.036,5.88329,0.281837,56.4537,0.318379,0
1653    29.5974,5.85735,-29.0121,31.3722,10.4024,298.629,245.948,-169.378,75.2499,307.964,0.774373,0.752031,71.0513,0.739594,0
1654    29.5974,5.85735,-29.0121,31.3722,10.4024,298.629,245.948,-169.378,75.2499,307.964,0.774373,0.752031,71.0513,0.739594,0
1655    29.5974,5.85735,-29.0121,31.3722,10.4024,298.629,245.948,-169.378,75.2499,307.964,0.774373,0.752031,71.0513,0.739594,0
1656    75.5101,-35.1559,-66.8269,629.45,624.904,111.766,-80.4088,-77.6272,879.213,886.289,0.322658,0.00140732,29.5192,0.0395244,
1657    49.6518,-19.4119,-45.6999,121.631,-111.036,243.005,-229.448,-80.0311,-437.102,500.11,0.855966,0.493048,91.4737,0.373046,0
1658    29.6579,-19.8746,22.0134,54.2986,45.4835,118.179,88.6436,-78.157,195.77,228.674,3.25605,0.523808,118.277,1.1189,0
1659    55.285,-29.4151,46.8101,60.3551,24.2139,124.686,-29.4865,-121.149,-174.716,214.644,2.81593,0.484523,209.71,2.3425,0
1660    35.2813,-30.2779,-18.1112,372.057,370.381,85.1889,-56.5524,-63.7101,39.8714,94.0579,2.6121,0.695842,186.339,1.04269,0
1661    34.8618,4.11137,-34.6185,62.2693,-51.5957,222.525,207.244,-81.039,-870.033,898.039,1.39869,0.821748,121.431,0.519322,0
1662    64.1921,-31.2182,-56.0897,76.623,41.8386,170.179,-4.43443,-170.122,345.963,385.553,0.973199,0.591508,103.8,0.613488,0
1663    87.0991,-27.5283,82.6344,89.8685,22.138,304.056,298.748,-56.5658,1506.9,1537.27,4.6779,0.854083,485.161,1.42214,0
1664    87.0991,-27.5283,82.6344,89.8685,22.138,304.056,298.748,-56.5658,1506.9,1537.27,4.6779,0.854083,485.161,1.42214,0
1665    89.9852,-6.54508,-89.7469,93.8281,26.5779,279.34,-12.7557,-279.049,347.338,445.73,0.752743,0.542052,122.174,0.606676,0
1666    46.9077,43.7838,-16.832,360.837,357.775,62.1374,-48.0316,-39.421,74.3631,96.9068,2.70166,0.678817,140.013,0.767441,0
```

1.1 Remove duplicates from my csv data

```
In [3]: with open(DATASET_DIR + BKG_DATA, 'r') as f, open(DATASET_DIR + BKG_nodup, 'w') as out_file :
            out_file.writelines(unique_everseen(f))

        with open(DATASET_DIR + SIG_DATA, 'r') as f, open(DATASET_DIR + SIG_nodup, 'w') as out_file :
            out_file.writelines(unique_everseen(f))
```

1.1.2 FullyConnected Dataset

1.1.2.1 Determine how many unique jets are in my bkg and my signal

```
In [4]: def sum1forline(filename):
            with open(filename) as f:
                return sum(1 for line in f)

        NUMBER_OF_SIGNAL = sum1forline(DATASET_DIR + BKG_nodup)
        NUMBER_OF_BKG = sum1forline(DATASET_DIR + SIG_nodup)

        print(f"num_events_BKG : {NUMBER_OF_BKG}")
        print(f"num_events_SIG : {NUMBER_OF_SIGNAL}")

        num_events_BKG : 1344
        num_events_SIG : 1344
```

```cpp
  int _listing_indx;
  int _pdg_id;


Symbolischer Link zu Datensaetzen:
ln -sf /project/etp1/biebel/programme/powheg/ggHH/powheg-ggHZ-events-1.2Mio-pythia8-fixed.dat MCatNLO-events.dat
ln -sf /project/etp1/biebel/programme/powheg/ggHH/powheg-ggHH-events-1.2Mio-pythia8.dat MCatNLO-events.dat
ln -sf /project/etp1/biebel/MC/MCatNLO-4.01/Linux13TeV-mt172.5-12.5Mio/MCatNLO-12.5Mio-events-orig.dat

efine FILENAME    "0106Pythia_ttb_Wb_All.root"
efine TREENAME    "tree"
efine BRANCHNAME  "particles"
efine PDGNUMBER   211

Definieren der Histogrammnamen

Alle b-quark-histogramme
efine HISTNAME1 "Number of b-quarks in an event"
efine HISTNAME2 "All b-quark phi"
efine HISTNAME3 "All b-quark pseudo_rap"
efine HISTNAME4 "Relative angular difference between b-qu

Alle Jet-histogramme
efine HISTNAME5  "Anzahl_der_Jets"
efine HISTNAME6  "Masse_der_Jets"
efine HISTNAME7  "Phi der Jets"
efine HISTNAME8  "Jets_Pseudorapiditaet"
efine HISTNAME9 "Relative angular difference between jets

Define Sanity-Variables for b-quarks

efine HISTNAME10 "Phi of jet#1"
efine HISTNAME11 "Pseudorap of jet#1"

efine HISTNAME12 "Phi of jet#2"
efine HISTNAME13 "Pseudorap of jet#2"

efine HISTNAME14 "DeltaR BKG"
efine HISTNAME20 "DeltaR BKG (leaving out same jet index)"

Define Histogramms for Variables coming from H->bb
```

Terminal window:

```
tai.le@gar-ws-etp02:/project/etp3/BaoTai.Le/MeineAnalyse/basic-event-analysis
tai.le@gar-ws-etp02:/project/etp3/BaoTai.Le/MeineAnalyse/basic-event-analysis 95x24
-rw-r--r-- 1 tai.le ls-schaile     2510 Oct  7 12:16 CERN.cxx
-rw-r--r-- 1 tai.le ls-schaile      323 Jun 17 15:13 CERNfunctions.h
-rw-r--r-- 1 tai.le ls-schaile     1674 Jun 17 15:13 CERN.h
-rw-r--r-- 1 tai.le ls-schaile      191 Jun 17 15:13 CERNLinkDef.h
-rw-r--r-- 1 tai.le ls-schaile      821 Oct  7 12:16 CERN_rdict.pcm
-rwxr-xr-x 1 tai.le ls-schaile     1217 Jun 17 15:13 compilertest.sh
drwxr-sr-x 2 tai.le ls-schaile       11 Oct  7 14:23 Datasets
drwxr-sr-x 2 tai.le ls-schaile       38 Jul  5 14:24 default
-rw-r--r-- 1 tai.le ls-schaile     3963 Jul  3 15:49 FH_setdefaults.C
-rw-r--r-- 1 tai.le ls-schaile    47292 Jun 18 12:19 Hepevt_FastJet.C
-rw-r--r-- 1 tai.le ls-schaile        0 Jun 17 15:17 Jetmassen_im_ersten_4-Jet-Ereignis.dat
-rwxr-xr-x 1 tai.le ls-schaile     3210 Jul 23 21:17 makemyroot_FastJet
lrwxrwxrwx 1 tai.le ls-schaile       80 Jun 17 15:16 MCatNLO-events.dat -> /project/etp1/biebel/
programme/powheg/ggHH/powheg-ggHH-events-1.2Mio-pythia8.dat
-rw-r--r-- 1 tai.le ls-schaile   195091 Jun 17 15:13 PythiaParticleListe.pdf
-rw-r--r-- 1 tai.le ls-schaile     6433 Jun 17 15:13 readhepevt.f
-rw-r--r-- 1 tai.le ls-schaile    14176 Oct  7 12:16 readhepevt.o
-rwxr-xr-x 1 tai.le ls-schaile    21328 Oct  7 12:16 readhepevt.so
-rw-r--r-- 1 tai.le ls-schaile     2456 Jun 17 15:13 README.md
-rwxr-xr-x 1 tai.le ls-schaile  1595872 Oct  7 12:16 rootCERN
-rw-r--r-- 1 tai.le ls-schaile  1245308 Jul 24 13:30 rootCERN.log
-rw-r--r-- 1 tai.le ls-schaile  1245308 Jul 24 13:30 rootCERN.loh
-rw-r--r-- 1 tai.le ls-schaile     8106 Jul  3 15:58 workfile.root
tai.le@gar-ws-etp02:/project/etp3/BaoTai.Le/MeineAnalyse/basic-event-analysis$
```

Small question:

How do I change
the file directory
for the dataset