

# Speeding up TMVA with Multiprocessing Techniques

Kim Albertsson (CERN/LTU)

2018-04-19

# Introduction

- Based on work by Mammad Hajili (Summer Student 2017)
- In CV, data is split into independent folds
- Uses `ROOT::TProcessExecutor` for each fold

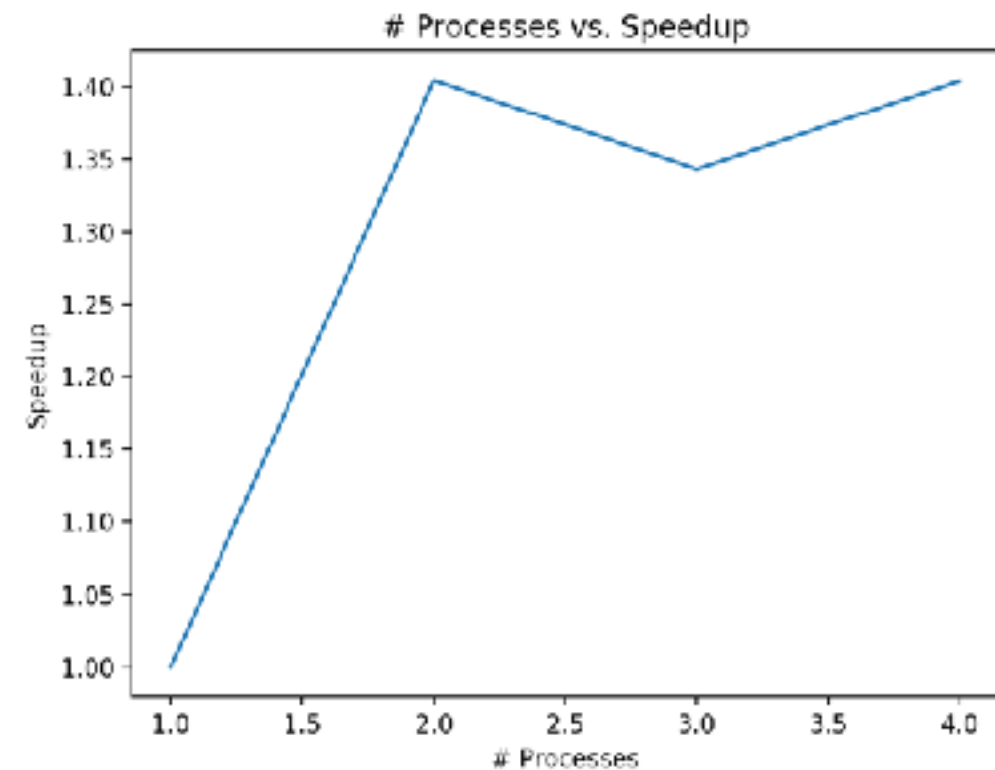


```
for each method:  
    for each fold:  
        method.evaluate(fold)
```

# Benchmark

- MacBook Intel Core i7 @ 3GHz, 8Gb RAM
- 6000 samples from 2d gaussian, 10-fold cv
- Method parameters chosen to take “some” time

```
cv.BookMethod(TMVA::Types::kBDT, "BDTG",  
"!H:!V:NTrees=300:MinNodeSize=2.5%"  
":BoostType=Grad:NegWeightTreatment=Pray"  
":Shrinkage=0.10:nCuts=20"  
":MaxDepth=6");
```



# Issues

- Text output is mangled.

```
Creating standalone class: dataset/w...
<HEADER> Factory : Test all methods
<HEADER> Factory : Test method: BDTG_fold2 for
:
<HEADER> BDTG_fold2 : [dataset] : Evaluation of BD
<HEADER> Factory : Test all methods
<HEADER> Factory : Test method: BDTG_fold1 for
:
<HEADER> BDTG_fold1 : [dataset] : Evaluation of BD
: Elapsed time for evaluation of 2999 e
<HEADER> Factory : Evaluate all methods
<HEADER> Factory : Evaluate classifier: BDTG_fo
:
<HEADER> BDTG_fold2 : [dataset] : Loop over test e
:
```