On Machine Learning application in B physics spectroscopy

Oleg Filatov
oleg.filatov@cern.ch

Moscow International School of Physics ’19
Young Scientist Forum
Introduction

Motivation

- explore ML&PP joined boundaries
- complement one with the other
- probe this and that

Actual plan

- Features study
- Models comparison
- Boosting AMS
Data preprocessing

- $B_s^0 \rightarrow \psi(2S) \phi$
  - $\psi(2S) \rightarrow J/\psi\pi^+\pi^-$, $J/\psi \rightarrow \mu^+\mu^-$, $\phi \rightarrow K^+K^-$
- data collected with CMS at $\sqrt{s} = 8$ TeV
- some preselection cuts to reduce bkgr
- signal - MC, bkgr - data $B_s^0$ sidebands
- train/test/val split [0.6/0.2/0.2], drop nans, shuffle
- variables (features):
  - kinematics: $p_T$, $\eta$
  - vertices: detach, fit proba, cosine
  - masses: $m[KK]$, $m[\pi\pi]$, $m[J/\psi\pi\pi]$, $m[\mu\mu]$
  - number of B candidates in event: nB
Features
Numerous ways to get importances!

**Features**

- PIPI\_mass\_Cjp 0.25
- PHI\_mass\_Cjp 0.23
- K2\_pt 0.19
- K1\_pt 0.18
- nB 0.18
- phi\_pt 0.17
- BU\_pvdistsignif2\_Cjp 0.094
- BU\_pvdistsignif3\_Cjp 0.093
- mu\_1\_eta 0.088
- mu\_2\_pt 0.085
- JPSI\_mass\_Cmumu 0.085
- mu\_1\_pt 0.084
- mu\_2\_eta 0.083
- JPSI\_vtxprob\_Cmumu 0.082
- JPSI\_pvdistsignif2\_Cmumu 0.08
- BU\_vtxprob\_Cjp 0.07
- JP\_eta 0.05
- BU\_pvcos2\_Cjp 0.048
- PI1\_pt 0.048
- K2\_eta 0.045
The only consistent method! [1] [2]
Domain non-invariance!

nB  MC  $16.7\sigma$
    data  $16.5\sigma$

w/0  MC  $15.5\sigma$
    data  $16.8\sigma$

probit x iminuit [1], [2]
Models
mass features removed!

- **Logistic Regression**
  - penalty=L2, C=10.

- **Random Forest**
  - max_features=sqrt, n_estimators=200, max_depth=-1

- **BDT (LightGBM)**
  - num_leaves = 30, max_depth = -1, binary CE
  - LR = 0.01, ES = 30 epochs
  - subsample= .9, colsample_bytree = .9

- **NN (Keras)**
  - 2x Dense (200) + BatchNorm + Dropout(0.5), ReLU, He init
  - Adam (LR=0.003), ReduceLROnPlateau(0.1, 10), binary CE
  - batch_size=1000, ES = 30 epochs

- **NN (PyTorch)**
  - same
Approximate Median Significance

- In a one-bin experiment assuming \( n_{\text{obs}} = s + b \) (Asimov dataset) w/o bkgr uncertainty \([1]\):

\[
\text{AMS} = \sqrt{2 \left( (s + b) \ln \left( 1 + \frac{s}{b} \right) - s \right)}, \quad \text{large enough } s+b
\]

\[
\text{AMS} \approx \frac{s}{\sqrt{b}}, \quad s \ll b
\]

- Introduced in Higgs and \( \tau \to \mu\mu\mu \) Kaggle competitions \([2]\), \([3]\)

- Henceforth will use:
  - \( s=100, b=100000 \) reweighting
  - conventional training (e.g. with logloss)
  - bayesian optim. with GPyOpt \([4]\) for finding an AMS-optimal cut on a classifier output
logloss = −\sum_{i=0}^{N} [y_i \log(p_i) + (1 − y_i) \log(1 − p_i)]

<table>
<thead>
<tr>
<th>Model</th>
<th>Acc</th>
<th>AUC</th>
<th>logloss</th>
<th>AMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogReg</td>
<td>0.82</td>
<td>0.85</td>
<td>0.425</td>
<td>0.74</td>
</tr>
<tr>
<td>RF</td>
<td>0.87</td>
<td>0.94</td>
<td>0.330</td>
<td>1.19</td>
</tr>
<tr>
<td>BDT (LightGBM)</td>
<td>0.86</td>
<td>0.93</td>
<td>0.334</td>
<td>1.21</td>
</tr>
<tr>
<td>top20 BDT (LightGBM)</td>
<td>0.86</td>
<td>0.93</td>
<td>0.335</td>
<td>1.16</td>
</tr>
<tr>
<td>top10 BDT (LightGBM)</td>
<td>0.86</td>
<td>0.93</td>
<td>0.335</td>
<td>1.16</td>
</tr>
<tr>
<td>NN (PyTorch) w/o BN</td>
<td>0.87</td>
<td>0.94</td>
<td>0.325</td>
<td>1.02</td>
</tr>
<tr>
<td>NN (PyTorch) BN</td>
<td>0.90</td>
<td>0.96</td>
<td>0.256</td>
<td>1.57</td>
</tr>
<tr>
<td>NN (Keras) w/o BN</td>
<td>0.86</td>
<td>0.93</td>
<td>0.338</td>
<td>1.05</td>
</tr>
<tr>
<td>NN (Keras) BN &amp; DO</td>
<td>0.91</td>
<td>0.97</td>
<td>0.222</td>
<td>1.62</td>
</tr>
</tbody>
</table>
AMSBoost
\[ \text{logloss} = \sqrt{2 \left( (s + b) \ln \left( 1 + \frac{s}{b} \right) - s \right)} \]

- train initial booster as before
- boost it further replacing logloss with AMS as a loss function
- BDT output cut for AMS set to 0.95
- check quality on 100 samples bootstrapped out of val set

Requires fine tuning

Unstable gradients

+20% improvement
Learned

- drawing conclusions about features impact
- comparing different models using proper metrics
- improving BDT performance with extra AMSBoost

Actual plan

- Features study
- Models comparison
- Boosting AMS
Backup
CMS DETECTOR

- Total weight: 14,000 tonnes
- Overall diameter: 15.0 m
- Overall length: 28.7 m
- Magnetic field: 3.8 T
Selection cuts

$\mu^\pm$

- $p_T(\mu^\pm) > 4$ GeV; $|\eta(\mu^\pm)| < 2.2$
- $p_T(\mu^+\mu^-) > 7$ GeV
- opposite-sign, soft, high-purity tracks
- dimuon vtx prob $> 10\%$
- $\cos 2D(\mu^+\mu^-, PV) > 0.9$
- $3.04 < m(\mu^+\mu^-) < 3.15$ GeV
- $DS_{2D}(\mu^+\mu^-, PV) > 3$

$\pi^\pm$

- $p_T > 0.6$ GeV; $|\eta| < 2.5$
- opposite-sign, high-purity tracks

$K^\pm$

- $p_T > 0.7$ GeV; $|\eta| < 2.5$
- opposite-sign, high-purity tracks
- $m(K^+K^-) < 1.05$ GeV

$B^0_s$

- $p_T > 15$ GeV/c
- 6-track kinematic vertex fit with $J/\psi$ mass constraint
- vtx prob $> 10\%$
- $5.1 < m(B^0_s) < 5.6$ GeV
- $DS_{2D}(PV) > 5$
- $\cos 2D(PV) > 0.999$