LHCb MC group meeting: $B^0 \to K^{*0} \mu \mu$

Oliver Dahme

University of Zurich

o.dahme@cern.ch

October 9, 2017

$B^0 \to K^{*0} \, \mu \, \mu$

- Decay is a FCNC
- with four charged particles in final state:
- K^+ and π^- from the K^{*0} decay
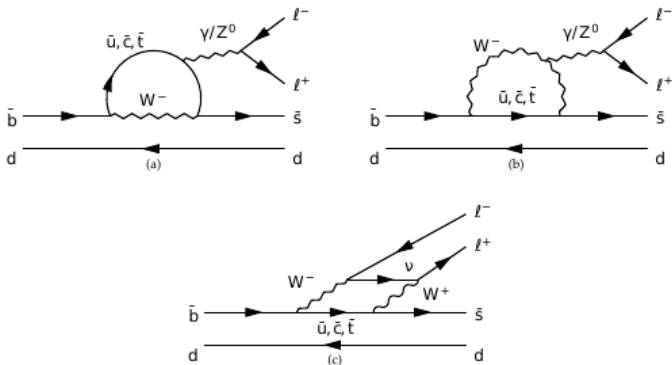- two leptons from loop or box diagrams:

Figure: Feynman diagrams for decay $B_d \to \mu^+ \mu^-$ at lowest order

$B^0 \rightarrow K^{*0} \mu \mu$

three angels define the kinematics of the decay:

- ThetaK or θ_K
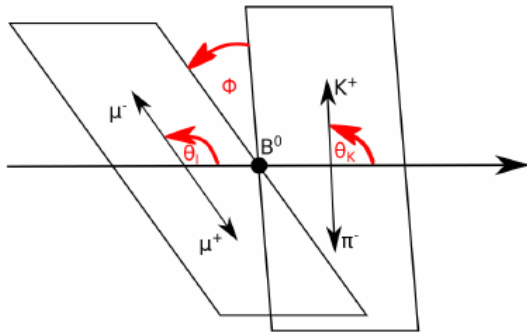- ThetaL or θ_L
- Phi or ϕ

Figure: kinematic variables of the decay $B^0 \rightarrow K^{*0} \mu \mu$

classification with Machine Learning (BDT)

- to eliminate combinatorial background, Boost Decision Trees are used as a black box to classify data into signal and background.
- They assign a 'probability' to each event.
- sk bdt from the scikit-learn package is used, and uBoost from the hep ml package.
- both responses are saved sepratly as one root file.
- training is performed on the Linux Cluster in Zurich

Region to train

- you want the bdt to train in the background region
- J/Psi has two resonance which will be cut off
- since a bdt needs real data for training the data is Kfolded. That means that data is splitted into parts to train and to classify.
- 10 runs are performed in each 90% are used for training to classify 10% in the end all the data is classified. One run is called a fold.
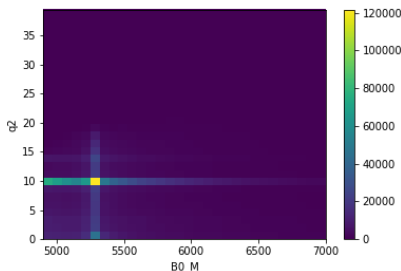
(a) Before Cut
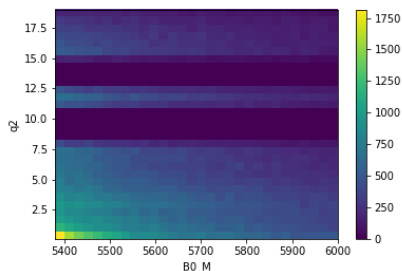
(b) After Cut

training of the BDTs

8 train parameters:
- B0 P
- B0 PT
- B0 ENDVERTEX CHI2
- B0 IP OWNPV
- B0 IPCHI2 OWNPV
- B0 FD OWNPV
- B0 FDCHI2 OWNPV
- B0 relinfo VTXISOBDTHARDFIRSTVALUE

4 uniform parameters:
- B0 M
- B0 ThetaK
- B0 ThetaL
- B0 Phi

uniform parameters have to be uncorrelated.
But BDT might find a correlation which is bad.

BDT correlation check

If there is no correlation this plot, between the bdt response and one of the uniform parameters, is flat. Therefore is no correlation between the sk bdt and the B0 mass. There is just a very small correlation between uBoost and the B0 mass.
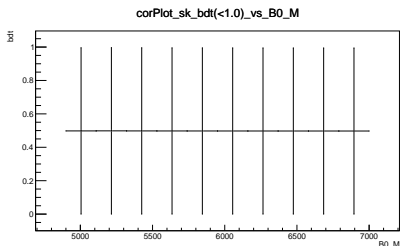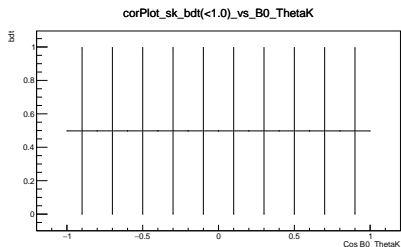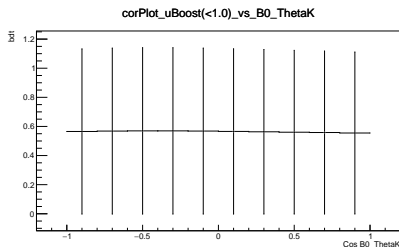

(a) Correlation sk bdt VS B0 M
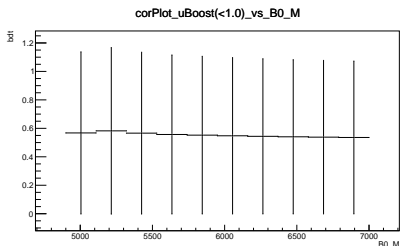
(b) Correlation uBoost VS B0 M

BDT correlation check

There is no correlation between the bdts and the cosine of the ThetaK angle.
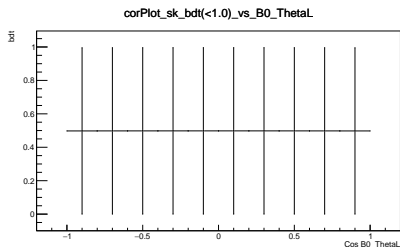

(a) Correlation sk bdt VS B0 ThetaK
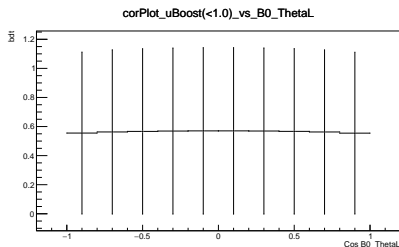

(b) Correlation uBoost VS B0 ThetaK

BDT correlation check

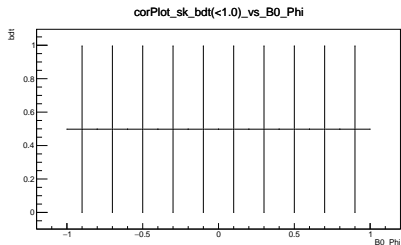There is no correlation between the bdts and the cosine of the ThetaL angle.
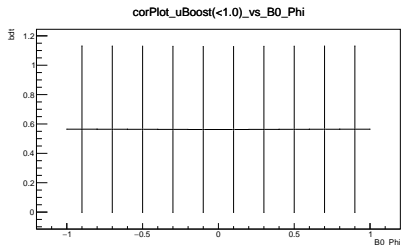

(a) Correlation sk bdt VS B0 ThetaL

(b) Correlation uBoost VS B0 ThetaL

BDT correlation check

There is no correlation between the bdts and Phi angle.

(a) Correlation sk bdt VS B0 Phi

(b) Correlation uBoost VS B0 Phi

Roc curves

The ROC curve gives an estimate how good a bdt performed. There is a roc curve for each of the 10 folds. This is just an example but all the others look similar. You can see that already at 0.6 true positive rate there is a 0.2 false positive rate. That means that if the bdt got 60% classified correctly there are already 20% classified wrongly. This is very bad I am looking into it.
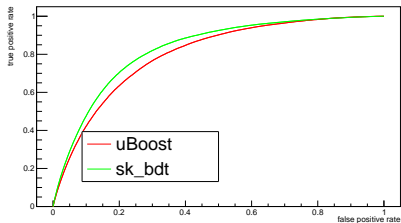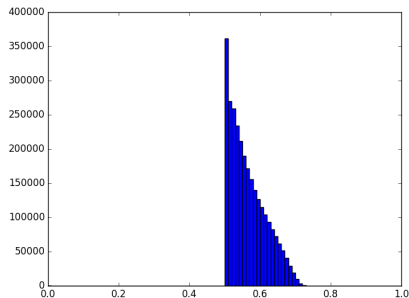

(a) ROC curve of fold 0
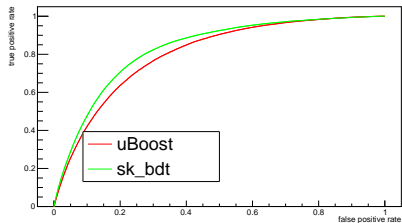
(b) ROC curve of fold 5

BDT response histogram

This are examples of histograms of the bdt responses for one fold. You can see clearly that most of the data is classified with a 0.5 probability to be signal. There is no information at all in that, this is connected to the bad roc curves above. At the moment I have no idea how to solve that problem. If you have one just write me.
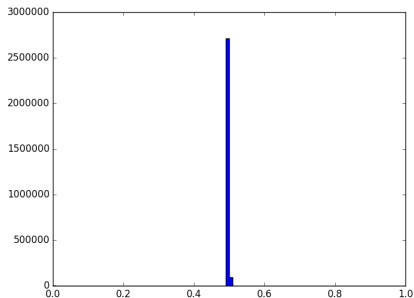
(a) sk bdt response of fold 0

(b) uBoost response of fold 0