

p_T Classification Study in EMTF

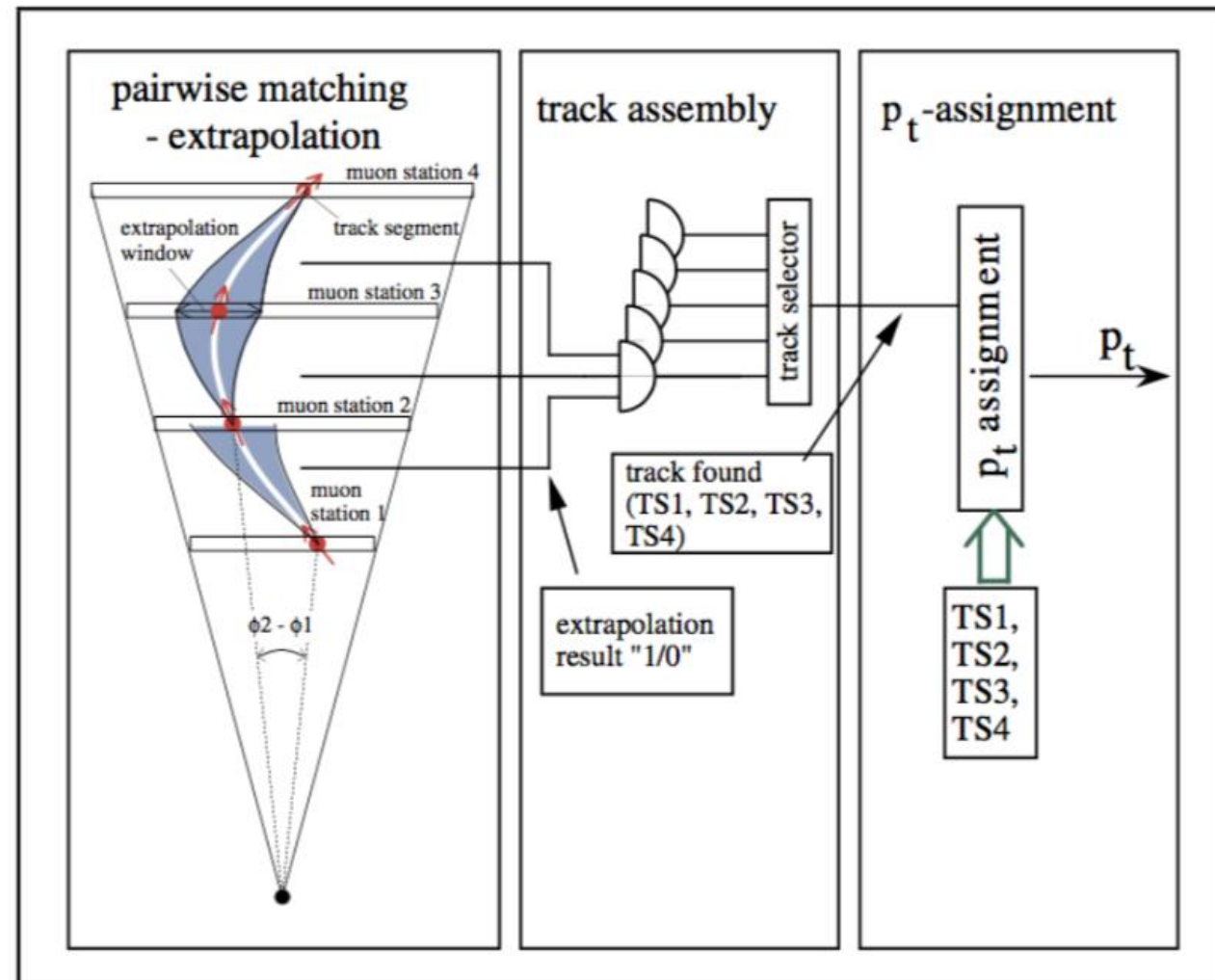
Wei Shi

CMS@Rice Meeting

Feb 2, 2018

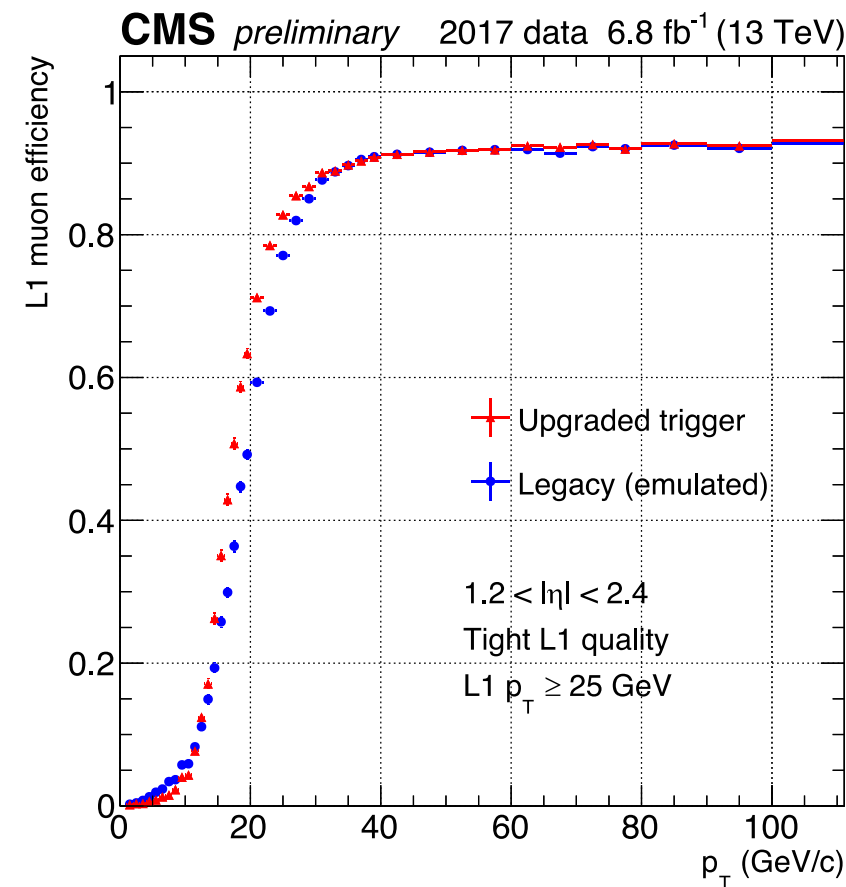
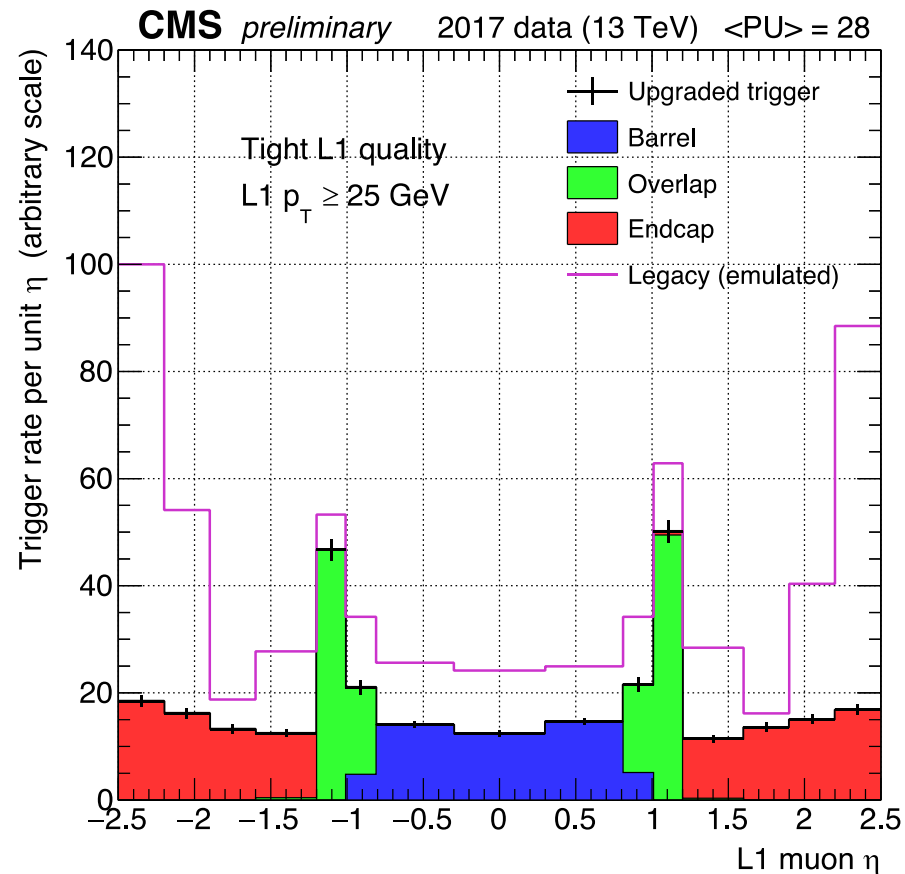
Introduction

- EMTF
 - Endcap Muon Track Finder
- Offline p_T training
 - Machine learning
 - Regression ^[1]
 - Classification



[1] CMS CR -2017/361: http://cds.cern.ch/record/2289251/files/CR2017_361.pdf

Introduction



*Plots are from https://cds.cern.ch/record/2286327/files/DP2017_041.pdf

2/2/2018

weishi@rice.edu

Motivation & Strategies

- Motivation

- Compare BDT (Boosted-Decision-Trees) classifiers with 2017 BDT regression

- Strategies

- 1) Use classifiers as reference (set signal efficiency as 95%, 93%)

- Require 2017 BDT regression to achieve 90% efficiency at same GEN pT as classifiers
 - Compare rate & efficiency

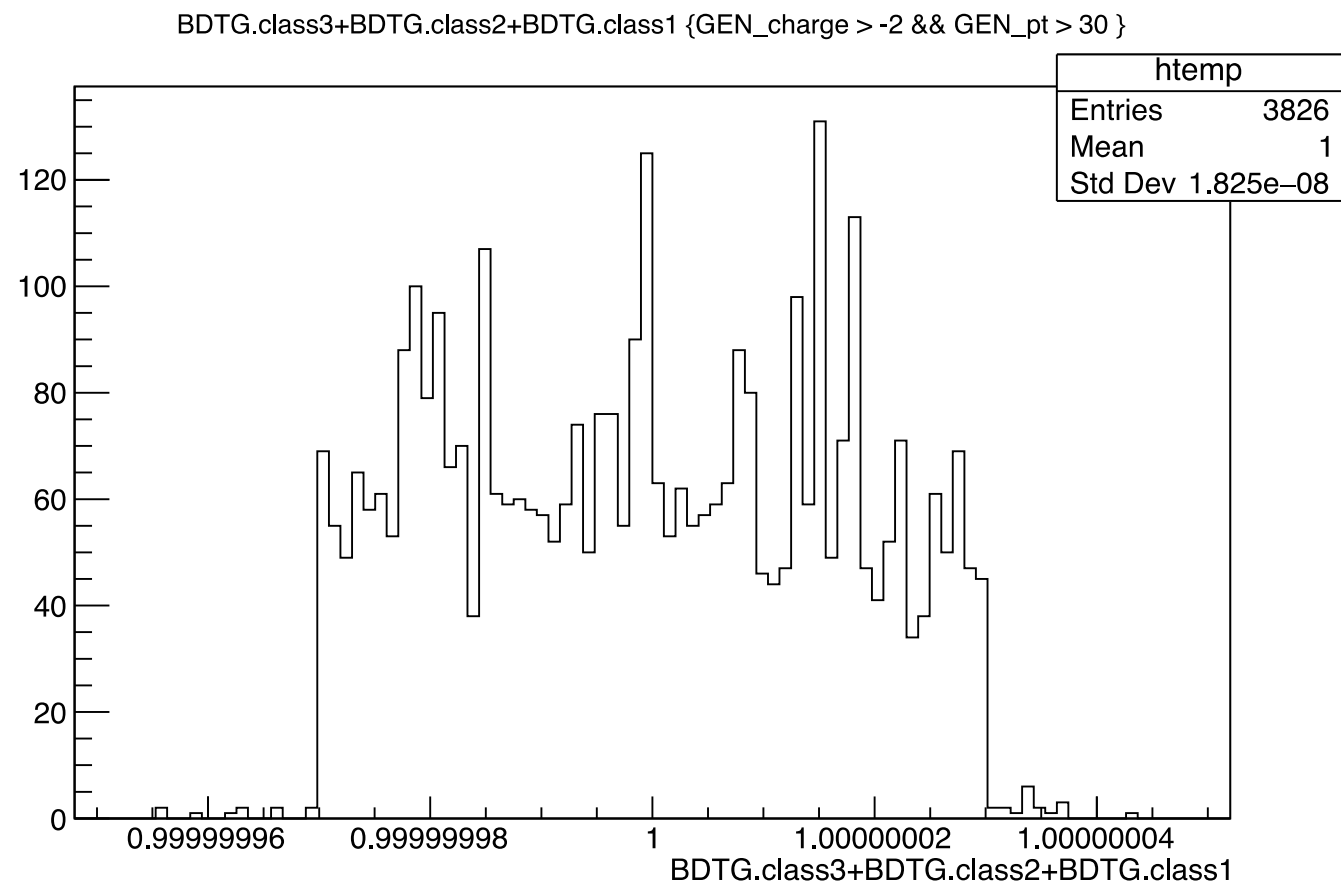
- 2) Use 2017 BDT regression $p_T > 16$ GeV as reference

- 16 GeV is scaled to 25.1 GeV in EMTF firmware
 - Require classifiers achieve 90% efficiency at same GEN pT as regression

Classifiers

- Binary classifiers on $p_T = X$ GeV
 - Only presents $X=25$
 - Typical L1 muon trigger threshold
 - Binary Classifier
 - Train 2 classes: background is GEN p_T [1, 8) GeV, signal is GEN p_T [X, 256] GeV
 - Binary ZB Classifier
 - Train 2 classes: Zerobias data as background, GEN p_T [X, 256] GeV as signal
- Both classifiers use same settings as BDT regression
 - Binary classifier has best performance under this setting
 - No weight applied in both classifiers

Total probability $\sum_{i=0}^n \text{class } i = 1$



Motivation & Strategies

- Motivation

- Compare BDT (Boosted-Decision-Trees) classifiers with 2017 BDT regression

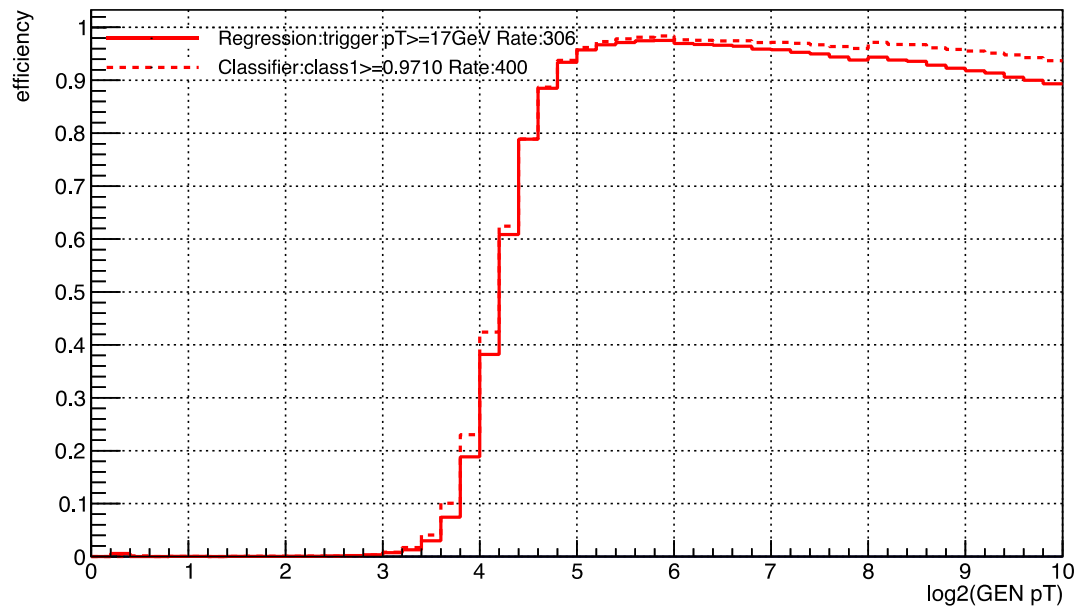
- Strategies

- 1) Use classifiers as reference (set signal efficiency as 95%, 93%)

- Require 2017 BDT regression to achieve 90% efficiency at same GEN pT as classifiers
 - Compare rate & efficiency

Binary Classifier

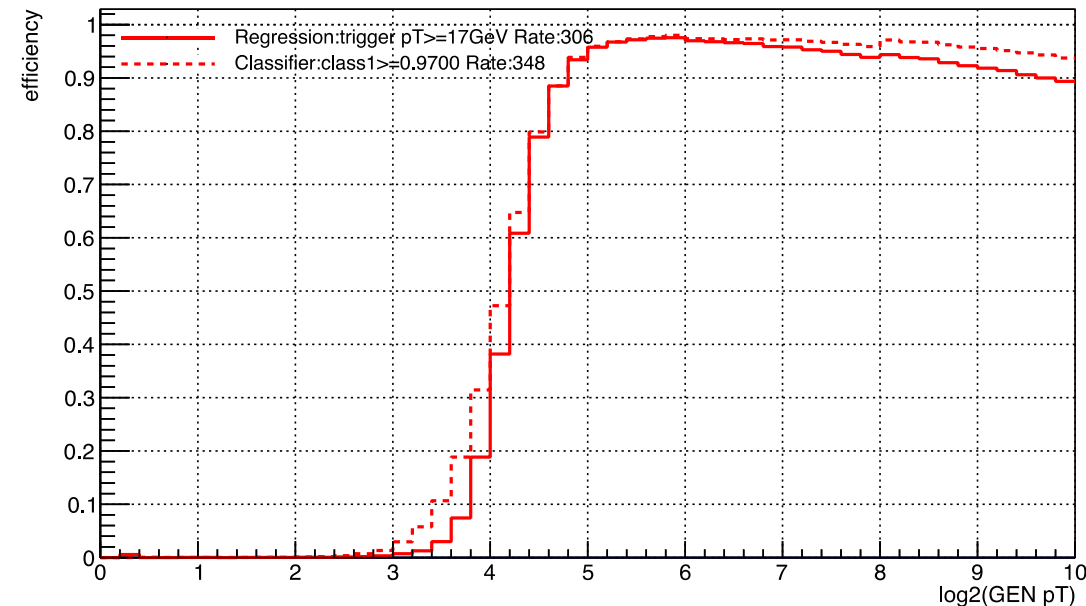
Regression CSC-only 90% Efficiency as Classifier



Binary ZB Classifier



Regression CSC-only 90% Efficiency as Classifier

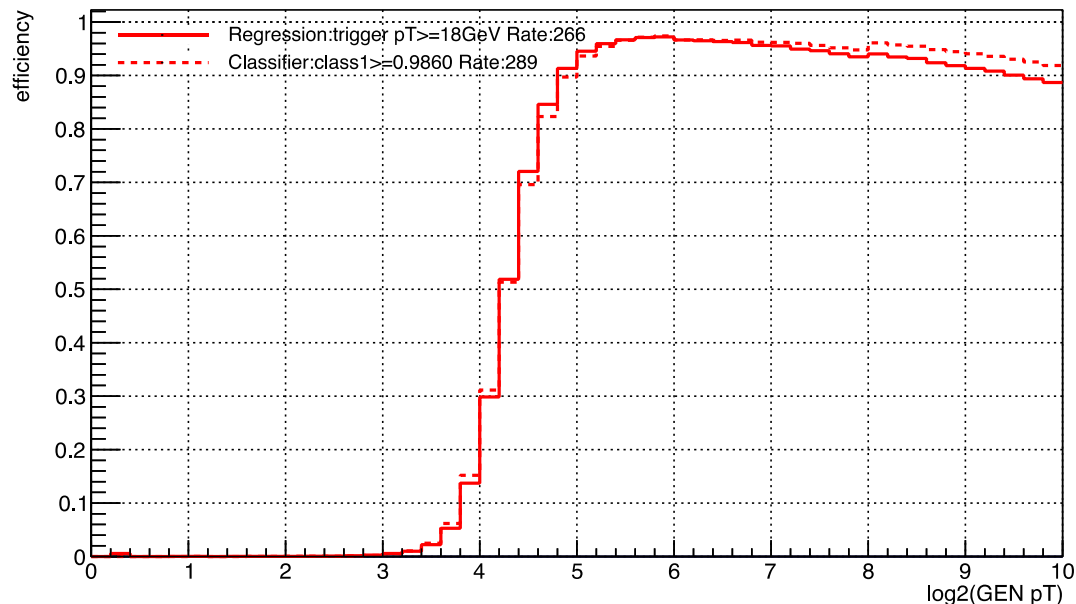


Regression 90% efficiency at same GEN pT as Classifiers with signal eff **95%**

	2 classes (<8, >=25)	2 classes (ZB, >=25)
Classifier cuts	class1>0.971	class1>0.970
Ratio in all ZB events (classifiers rate)	0.202% (400/197713)	0.354% (348/98384)
Regression pT cuts	17 GeV	17 GeV
Ratio in all ZB events (regression rate)	0.155% (306/197713)	0.155% (306/197713)

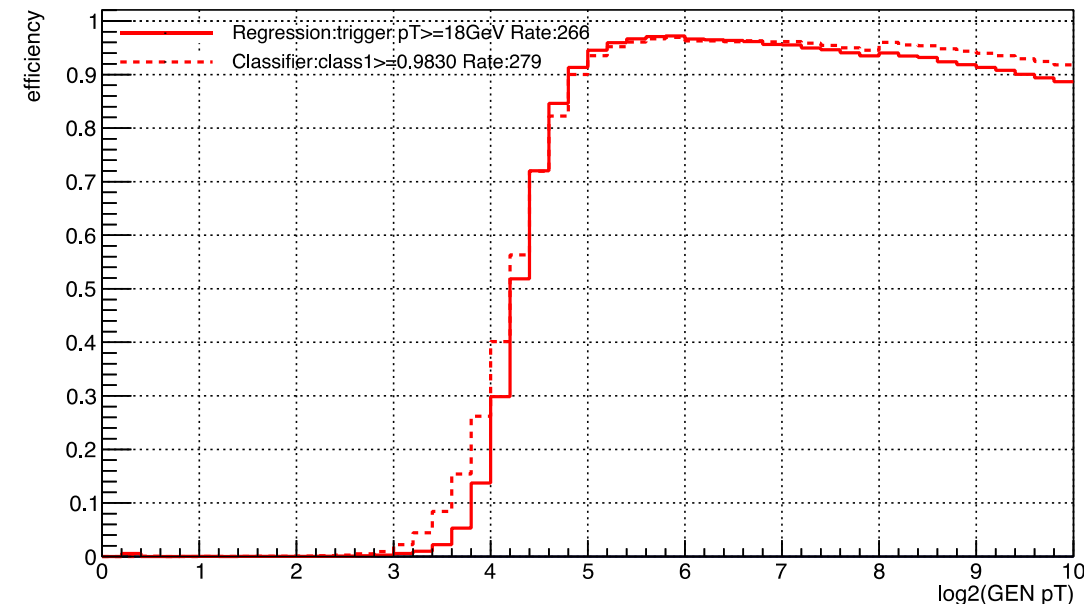
Binary Classifier

Regression CSC-only 90% Efficiency as Classifier



Binary ZB Classifier

Regression CSC-only 90% Efficiency as Classifier



Regression 90% efficiency at same GEN pT as Classifiers with signal eff **93%**

	2 classes (<8, >=25)	2 classes (ZB, >=25)
Classifier cuts	class1>0.986	class1>0.983
Ratio in all ZB events (classifiers rate)	0.146% (289/197713)	0.284% (279/98384)
Regression pT cuts	18 GeV	18 GeV
Ratio in all ZB events (regression rate)	0.135% (266/197713)	0.135% (266/197713)

Motivation & Strategies

- Motivation

- Compare BDT (Boosted-Decision-Trees) classifiers with 2017 BDT regression

- Strategies

- 1) Use classifiers as reference (set signal efficiency as 95%, 93%)

- Require 2017 BDT regression to achieve 90% efficiency at same GEN pT as classifiers
 - Compare rate & efficiency

- 2) Use 2017 BDT regression $p_T > 16$ GeV as reference

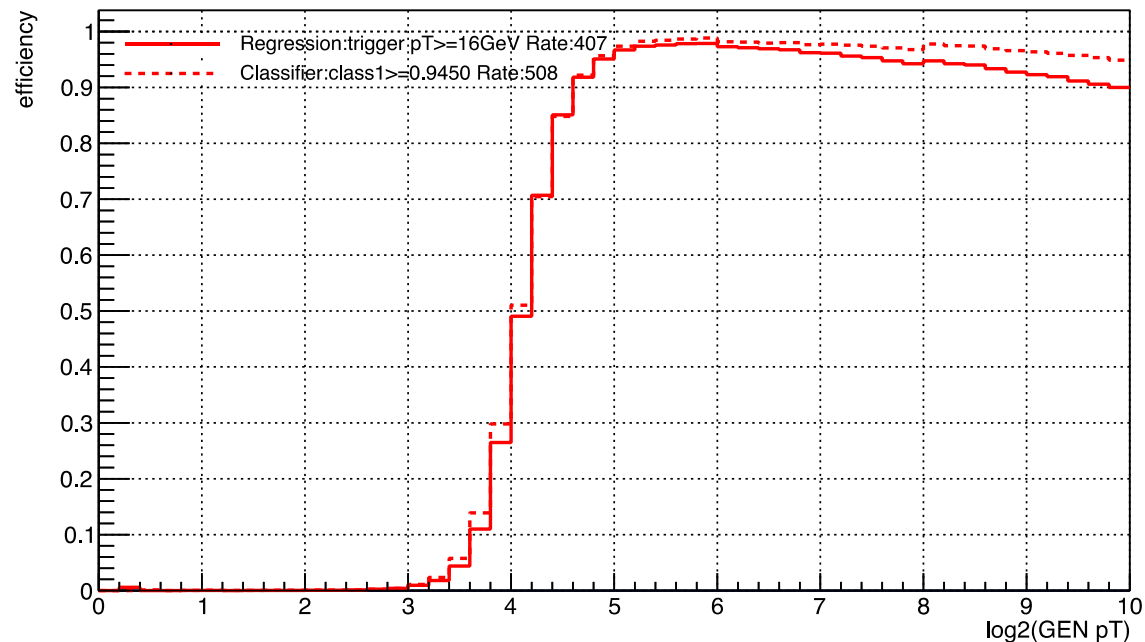
- 16 GeV is scaled to 25.1 GeV in EMTF firmware
 - Require classifiers achieve 90% efficiency at same GEN pT as regression

Binary Classifier

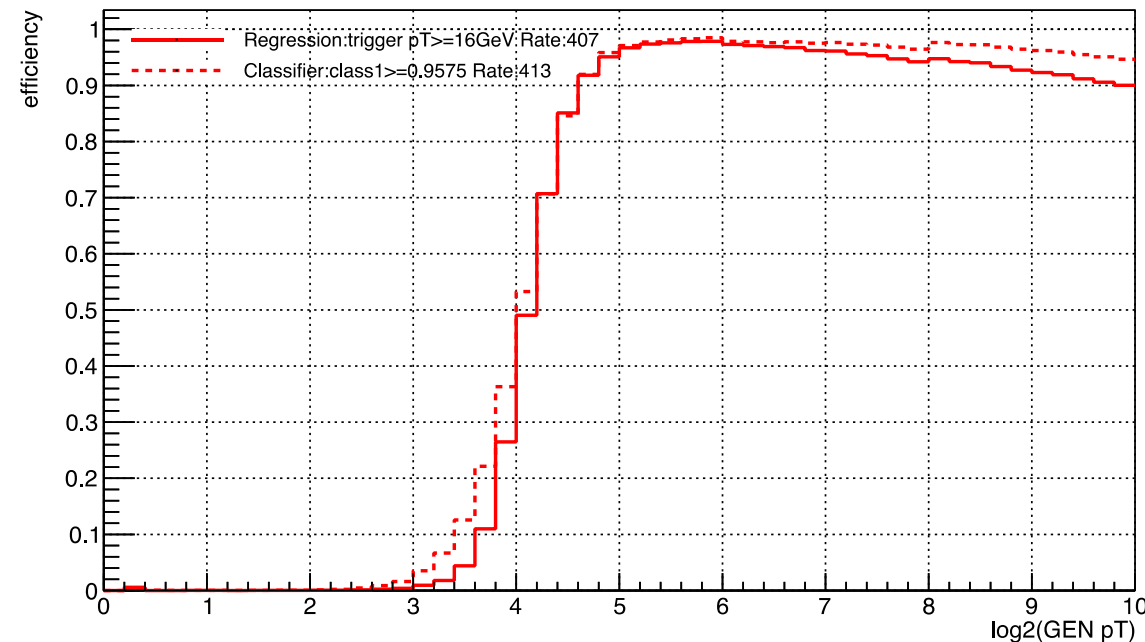


Binary ZB Classifier

Classifier CSC-only 90% Efficiency as Regression 16 GeV



Classifier CSC-only 90% Efficiency as Regression 16 GeV



Classifiers 90% efficiency at same GEN pT as Regression

Regression	2 classes (<8, >=25)	2 classes (ZB, >=25)
pT > 16 GeV	class1>0.945	class1>0.9575
0.206% (407/197713)	0.257% (508/197713)	0.420% (413/98384)

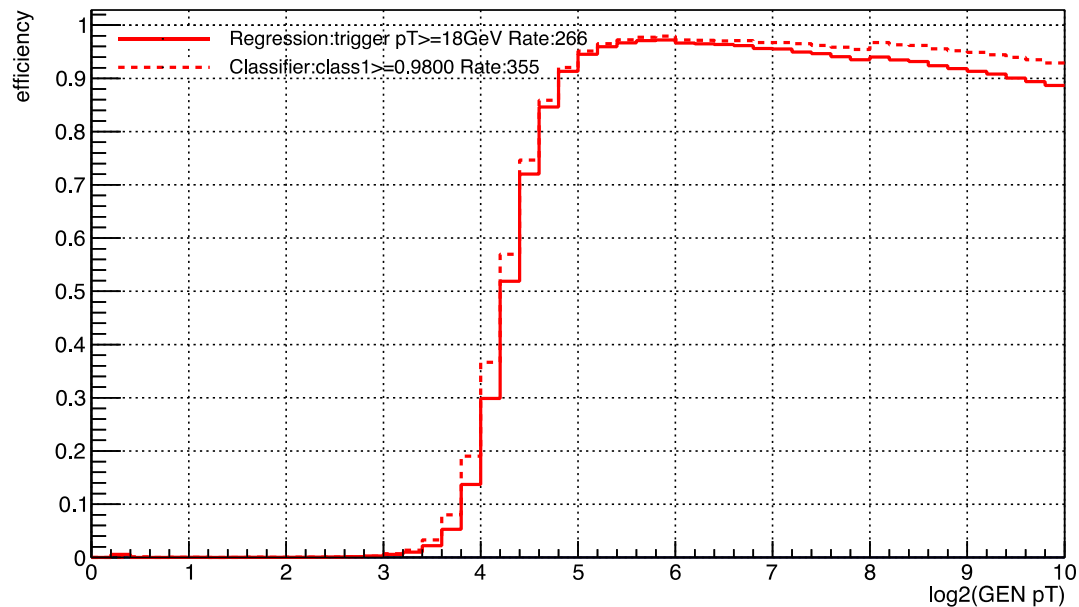
Summary & Plans

- Binary classifier has better performance
 - Comparable to the 2017 BDT regression (similar rate, higher efficiency at high p_T)
 - Rate ratio 50% lower than classifier trained with Zerobias events
 - Steeper turn-on curve than binary ZB classifier (especially below threshold 25 GeV)
- Plans
 - Train on real data (for example, ZeroBias+Muonia+SingleMu+MC)
 - Introduce pileup events from real data
 - New EMTF Ntuple adding RECO p_T branch
 - Improve regression
 - Combine classifier and regression

BACK UP

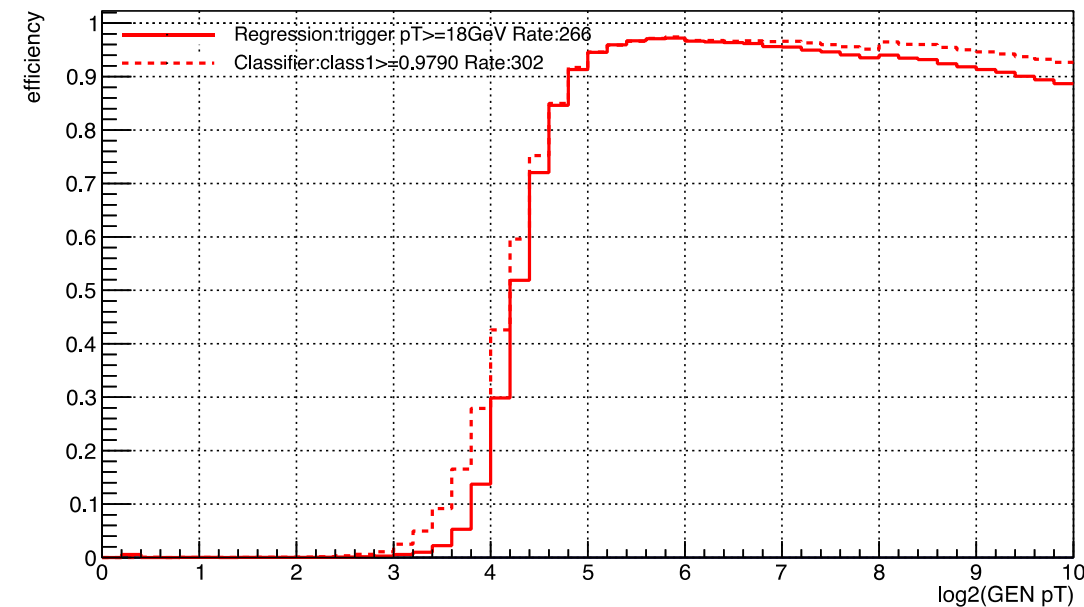
Binary Classifier

Regression CSC-only 90% Efficiency as Classifier



Binary ZB Classifier

Regression CSC-only 90% Efficiency as Classifier

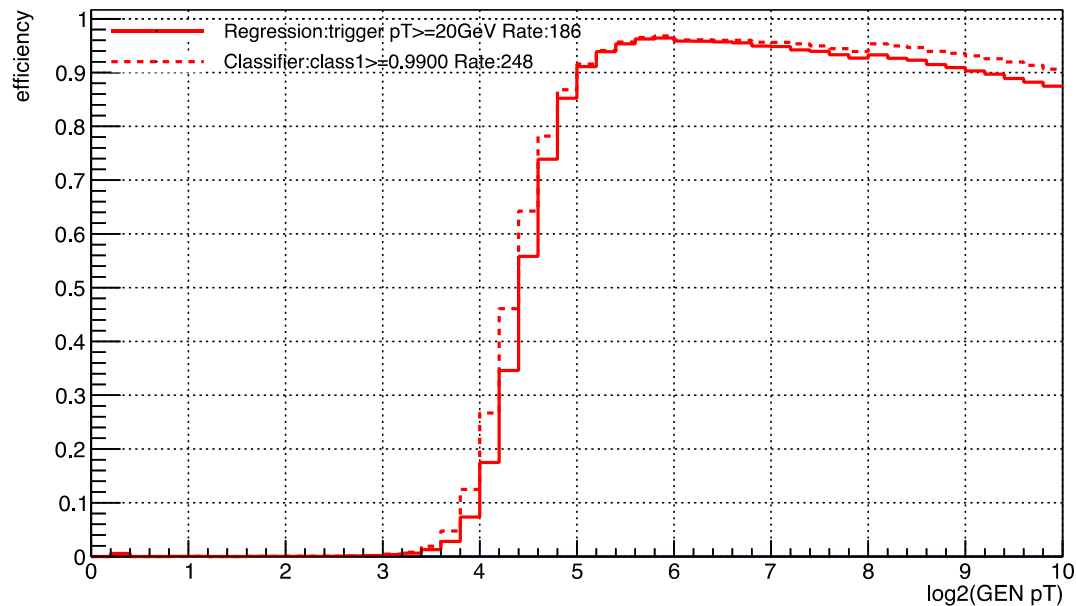


Regression 90% efficiency at same GEN pT as Classifiers with signal eff **94%**

	2 classes (<8, >=25)	2 classes (ZB, >=25)
Classifier cuts	class1>0.980	class1>0.979
Ratio in all ZB events (classifiers rate)	0.180% (355/197713)	0.307% (302/98384)
Regression pT cuts	18 GeV	18 GeV
Ratio in all ZB events (regression rate)	0.135% (266/197713)	0.135% (266/197713)

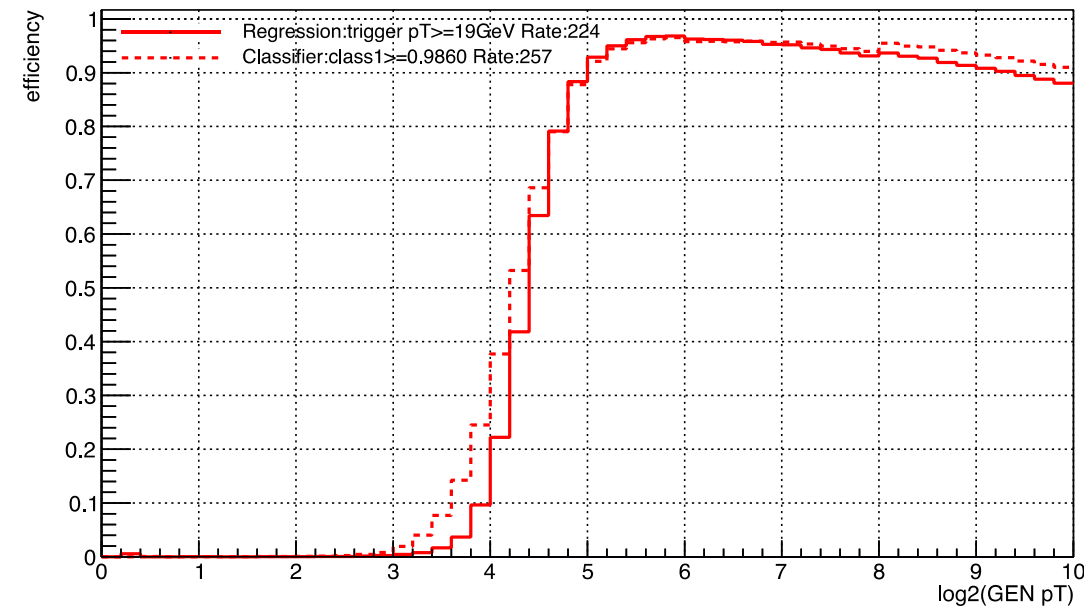
Binary Classifier

Regression CSC-only 90% Efficiency as Classifier



Binary ZB Classifier

Regression CSC-only 90% Efficiency as Classifier



Regression 90% efficiency at same GEN pT as Classifiers with signal eff **92%**

	2 classes ($<8, \geq 25$)	2 classes (ZB, ≥ 25)
Classifier cuts	$\text{class1} > 0.990$	$\text{class1} > 0.986$
Ratio in all ZB events (classifiers rate)	0.125% (248/197713)	0.261% (257/98384)
Regression pT cuts	20 GeV	19 GeV
Ratio in all ZB events (regression rate)	0.094% (186/197713)	0.113% (224/197713)

Tools

- Multi-class(binary class inclusive) training
 - <https://github.com/weishi10141993/EMTFPtAssign2017/blob/master/pTMultiClass.C>
- Compare classifier and regression:
 - Binary classifier
 - <https://github.com/weishi10141993/EMTFPtAssign2017/blob/master/macros/ClassifierROC.C>
 - Multi-class classifier
 - <https://github.com/weishi10141993/EMTFPtAssign2017/blob/master/macros/MultiClassifierROC.C>

2017 BDT Regression Rate

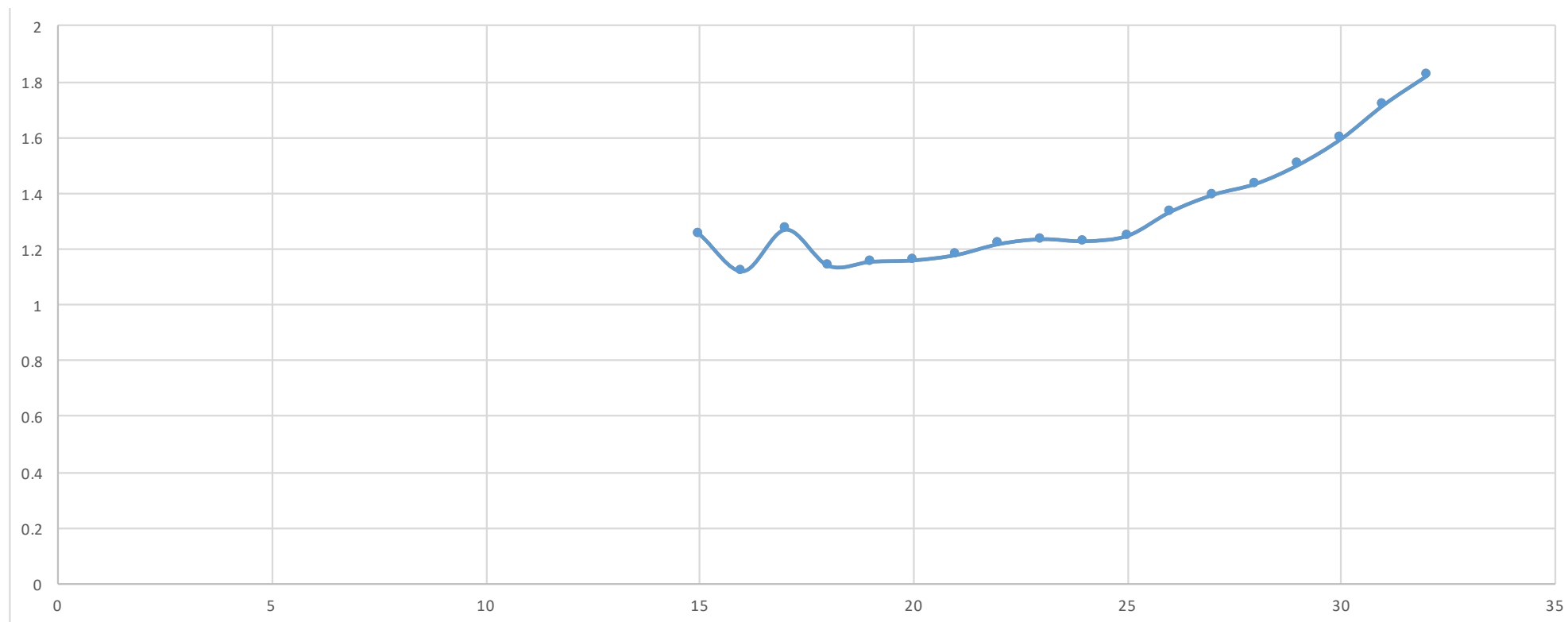
Total ZeroBias events: 197713 Regression pT (scaled pT @P5 [GeV])	2017 BDT Regression Rate
pT > 20 GeV (34.3)	186
pT > 19 GeV (31.8)	224
pT > 18 GeV (29.5)	266
pT > 17 GeV (27.3)	306
pT > 16 GeV (25.2)	407
pT > 15 GeV (23.2)	507
pT > 14 GeV (21.26)	635

2017 pT scale:

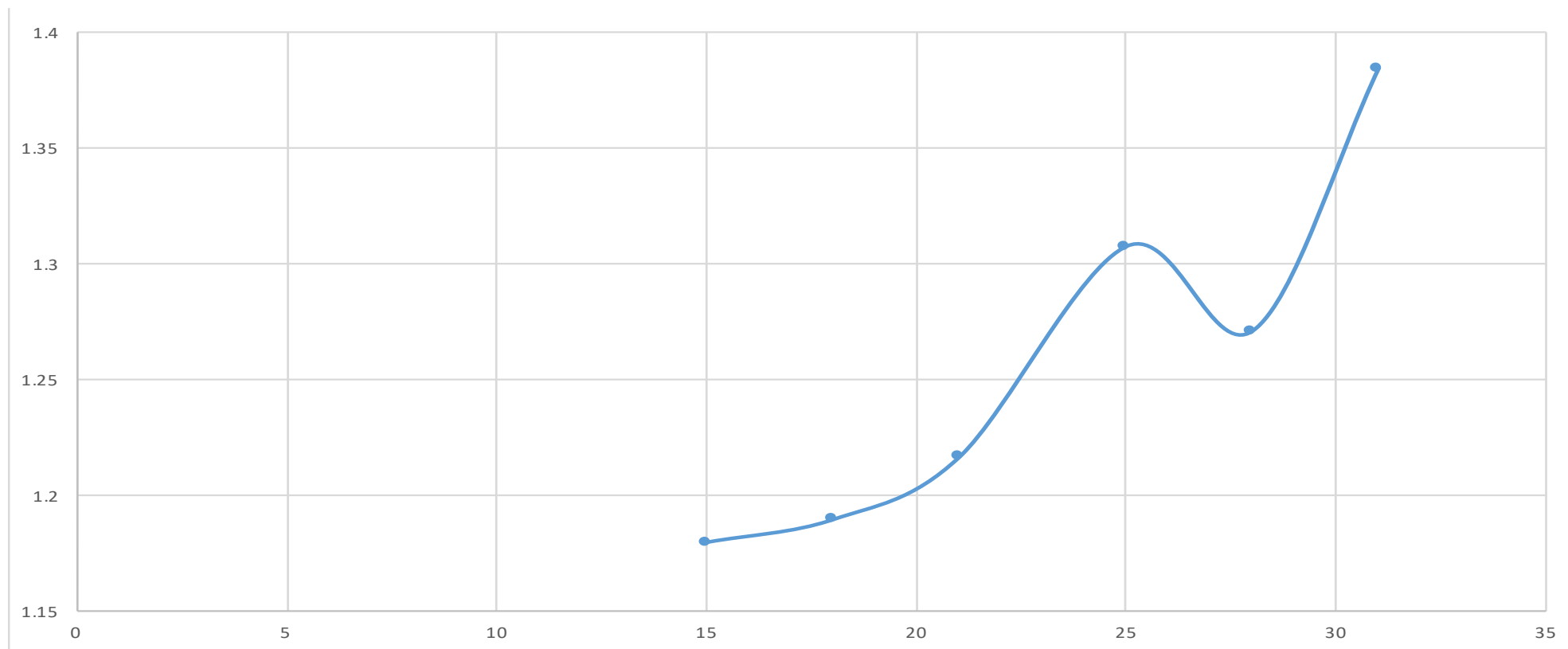
$\text{pt_xml} = \text{fmin}(20., \text{pt});$

$\text{pt_scale} = 1.2 / (1 - 0.015 * \text{pt_xml});$

Rate Ratio (Classifier/Regression) vs Classifier trained pT cut
 Require Classifier achieve 90% at same GEN pT as Regression pT > 16 GeV



Rate Ratio (Classifier/Regression) vs Classifier signal efficiency is 95% vs Classifier trained pT cut
Require regression achieve 90% at same GEN pT as classifier



Training & evaluation files

- Binary classifier
 - `root://eosuser.cern.ch/eos/user/w/wshi/Binary_Classifier_NonContinuous_Cut`
- Binary ZB classifier (train on zerobias events)
 - `root://eosuser.cern.ch/eos/user/w/wshi/Binary_Classifier_TrainZBbkg`
- Other classifiers
 - Binary lassifiers with continuous cut
 - `root://eosuser.cern.ch/eos/user/w/wshi/Binary_Classifier_Continuous_Cut`
 - 5 classes classifier
 - `root://eosuser.cern.ch/eos/user/w/wshi/MultiClass_Classifier`
 - 2017 BDT Regression: different targets & weights
 - `root://eosuser.cern.ch/eos/user/w/wshi/Targets_Weights_Regression`

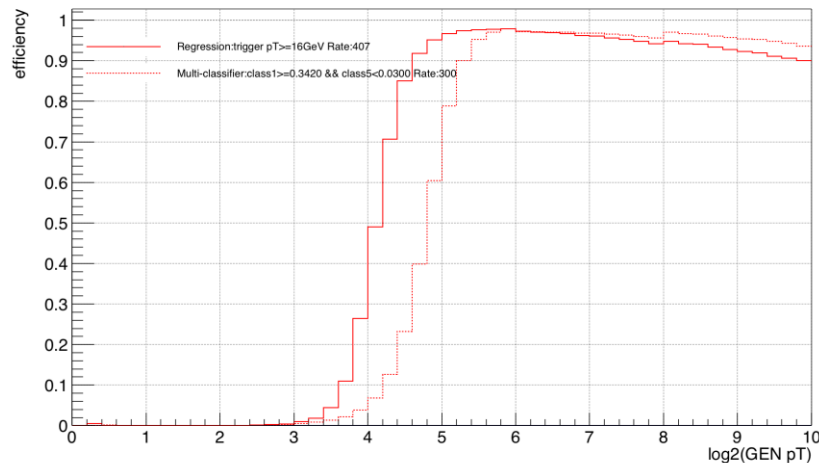
Classifiers

- Trained three classifiers
 - 5 classes: GEN pt [32, 256], [24, 32), [16, 24), [8, 16), [1, 8) GeV
 - 2 classes: GEN pt [32, 256], [1, 8) GeV
 - 2 classes: GEN pt [32, 256], [1, 32) GeV
- All classifiers have same BDT settings as 2017 BDT regression
 - No weight applied in all classifiers

Comparison Strategy

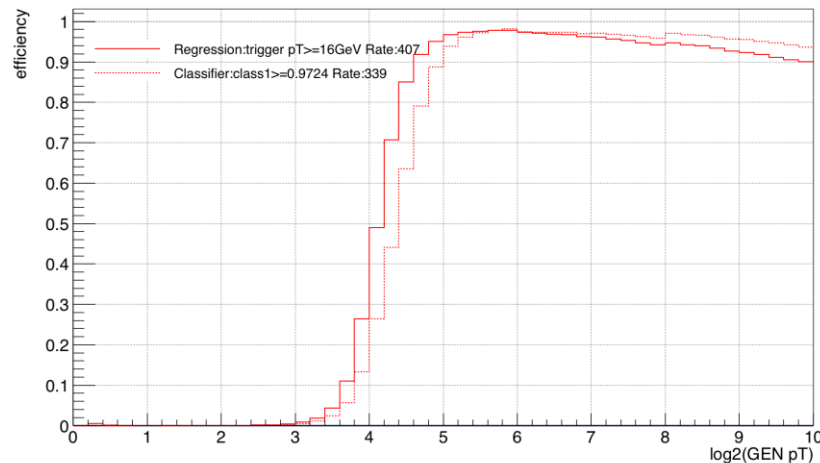
- Compare classifiers best rate & efficiency with regression $p_T > 16$ GeV(not scaled)
 - Require classifier signal efficiency to be 95%, 94%, 93% and 92%
 - Signal means [32, 256] GeV
- Compare classifiers best rate & efficiency with 2017 BDT regression
 - Require classifier signal efficiency to be 95%, 94%, 93% and 92%
 - Require regression achieve 90% efficiency at same GEN p_T as classifiers
- Compare classifiers rate & efficiency with 2017 BDT regression $p_T > 16$ GeV
 - Require three classifiers achieve 90% efficiency at same GEN p_T as regression $p_T > 16$ GeV

CSC only Efficiency: Regression vs Multi-classifier



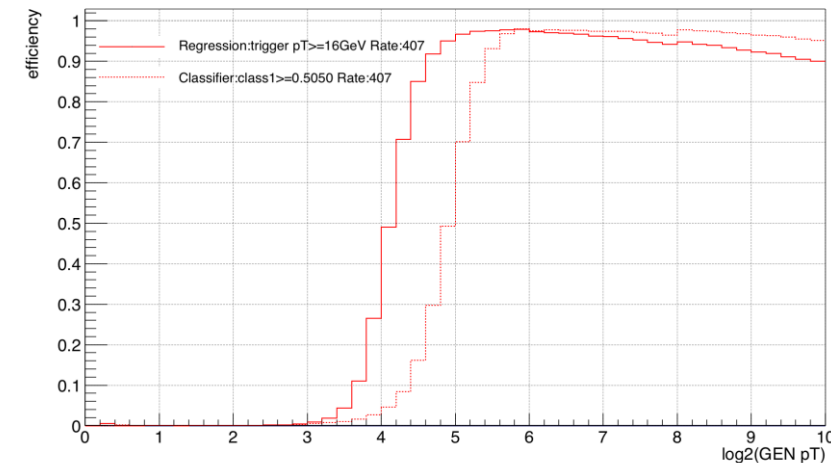
5 classes

CSC only Efficiency: Regression vs Classifier



2 classes non-continuous pT

CSC only Efficiency: Regression vs Classifier

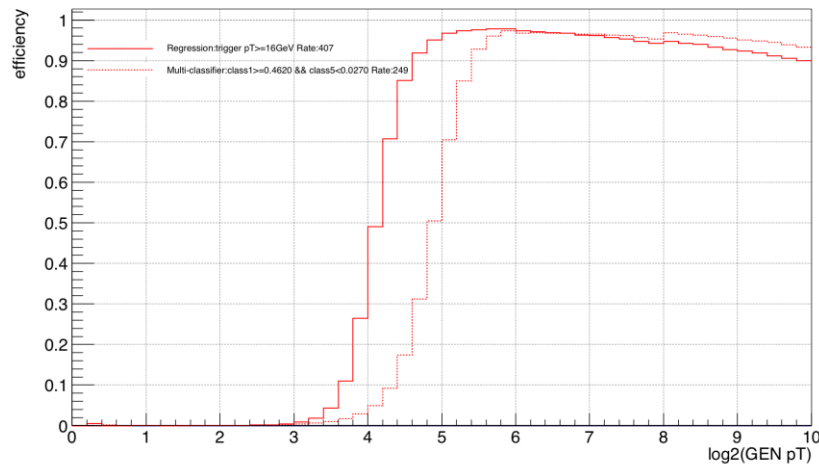


2 classes continuous pT

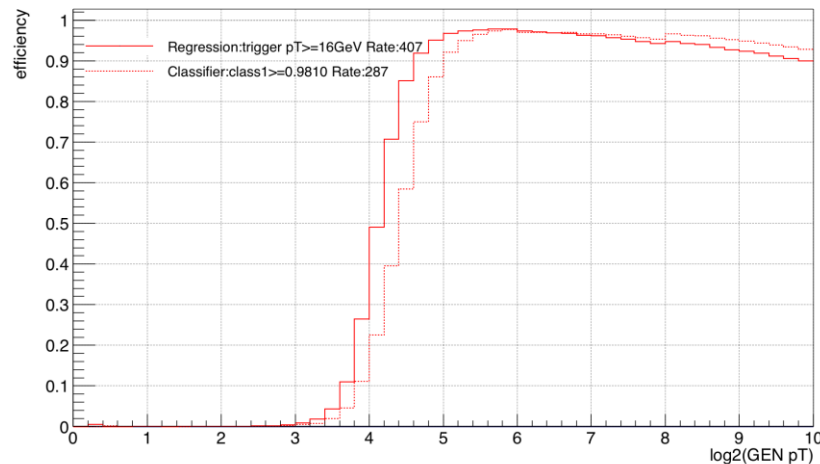
Rate: Classifiers signal efficiency **95%**

Regression	5 classes [1,8][8,16][16,24][24,32][32,1000]	2 classes (<8 , ≥ 32)	2 classes (<32 , ≥ 32)
pT > 16GeV	class1 > 0.342 && class5 < 0.03	class1 > 0.9724	class1 > 0.505
407 (0.206%)	300 (0.152%)	339 (0.172%)	407 (0.206%)

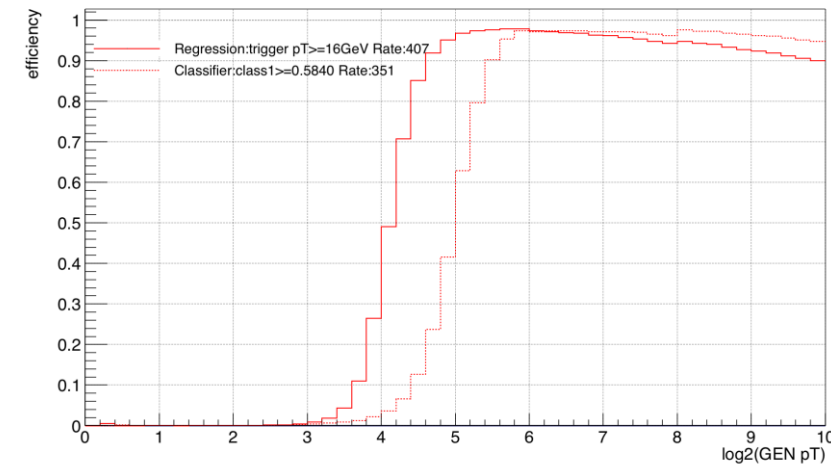
CSC only Efficiency: Regression vs Multi-classifier



CSC only Efficiency: Regression vs Classifier



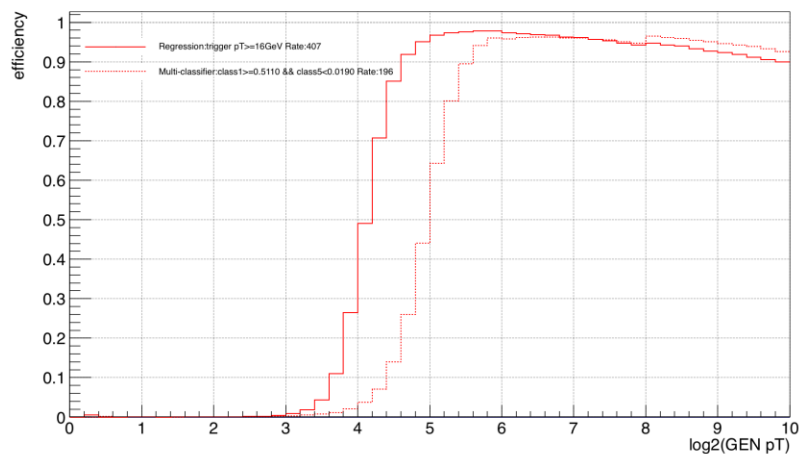
CSC only Efficiency: Regression vs Classifier



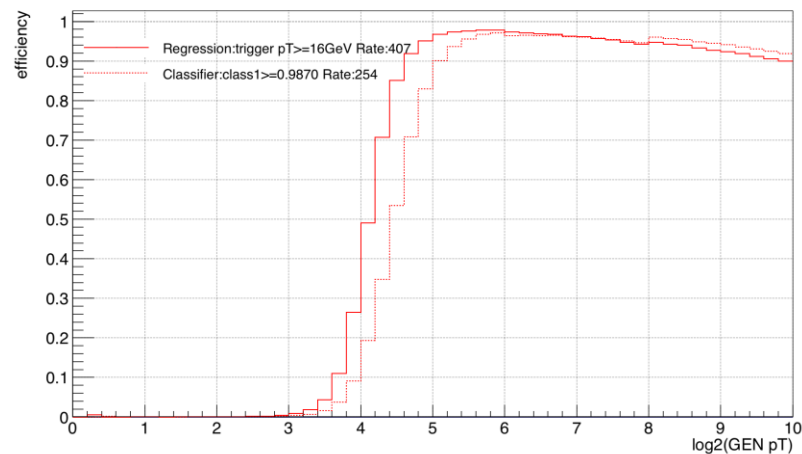
Rate: Classifiers signal efficiency **94%**

Regression	5 classes [1,8][8,16][16,24][24,32][32,1000]	2 classes (<8, >=32)	2 classes (<32, >=32)
$pT > 16\text{GeV}$	$\text{class1} > 0.462 \ \&\& \ \text{class5} < 0.027$	$\text{class1} > 0.981$	$\text{class1} > 0.584$
407 (0.206%)	249 (0.126%)	287 (0.145%)	351 (0.178%)

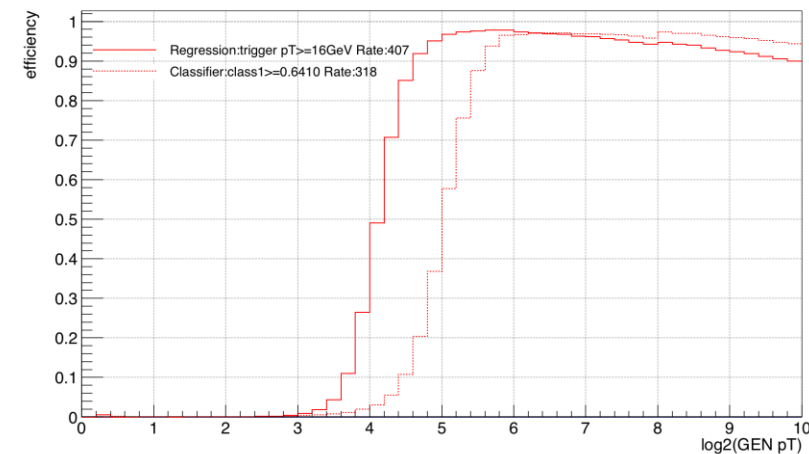
CSC only Efficiency: Regression vs Multi-classifier



CSC only Efficiency: Regression vs Classifier



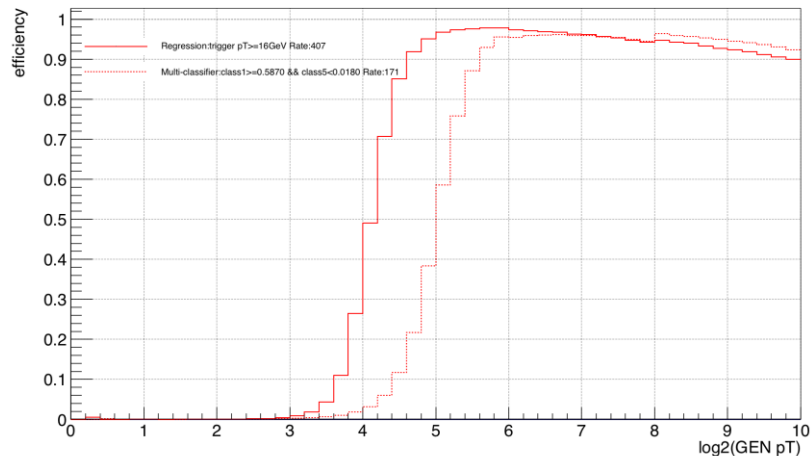
CSC only Efficiency: Regression vs Classifier



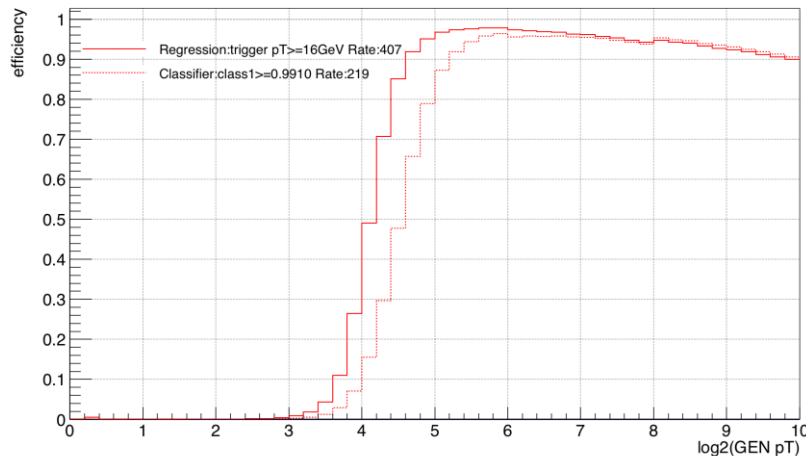
Rate: Classifiers signal efficiency **93%**

Regression	5 classes [1,8][8,16][16,24][24,32][32,1000]	2 classes (<8 , ≥ 32)	2 classes (<32 , ≥ 32)
$p_T > 16\text{GeV}$	$\text{class1} > 0.511 \ \&\& \ \text{class5} < 0.019$	$\text{class1} > 0.987$	$\text{class1} > 0.641$
407 (0.206%)	196 (0.099%)	254 (0.128%)	318 (0.161%)

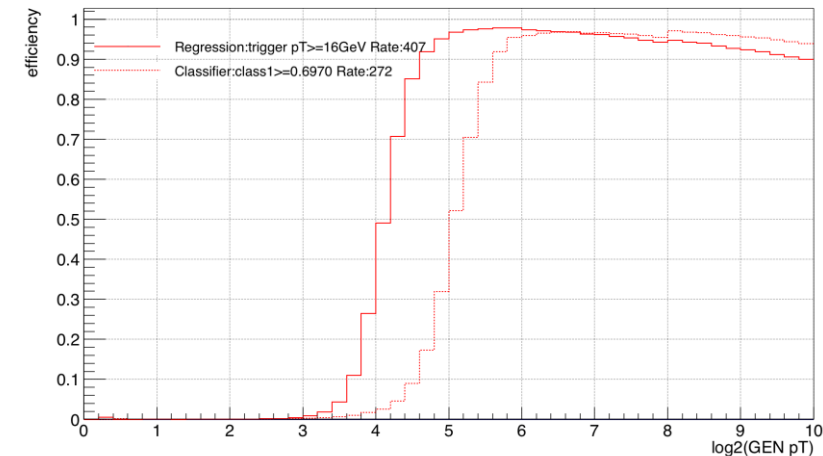
CSC only Efficiency: Regression vs Multi-classifier



CSC only Efficiency: Regression vs Classifier



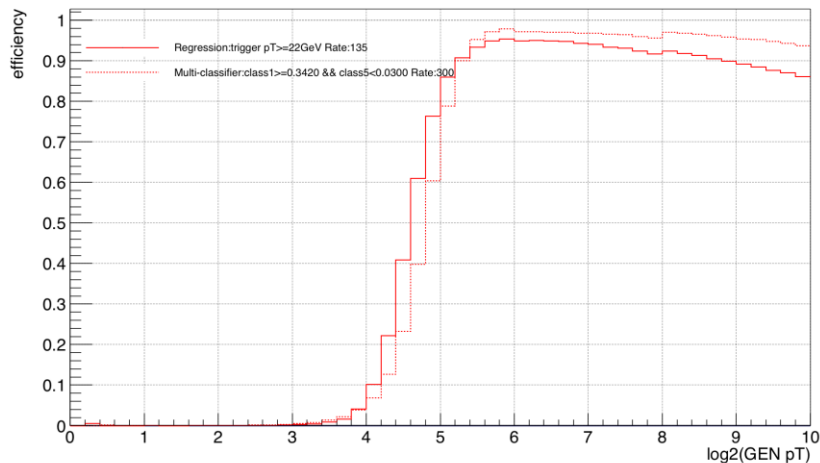
CSC only Efficiency: Regression vs Classifier



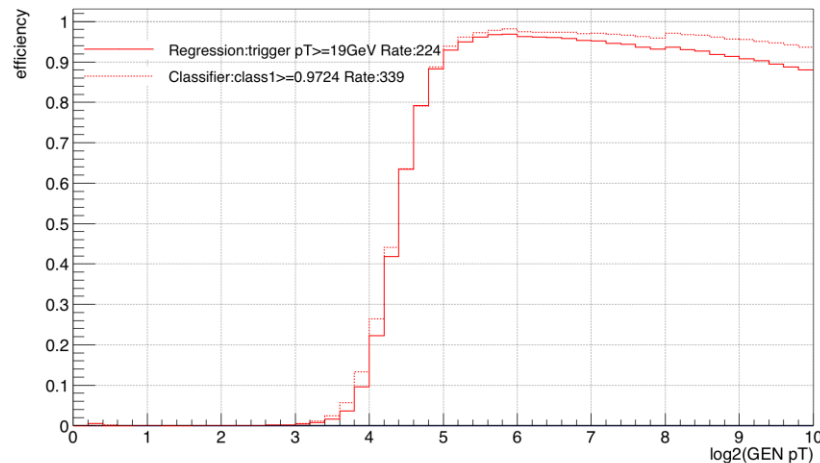
Rate: Classifiers signal efficiency **92%**

Regression	5 classes [1,8][8,16][16,24][24,32][32,1000]	2 classes (<8, >=32)	2 classes (<32, >=32)
$pT > 16\text{GeV}$	$\text{class1} > 0.587 \ \&\& \ \text{class5} < 0.018$	$\text{class1} > 0.991$	$\text{class1} > 0.697$
407 (0.206%)	171 (0.0865%)	219 (0.1108%)	272 (0.138%)

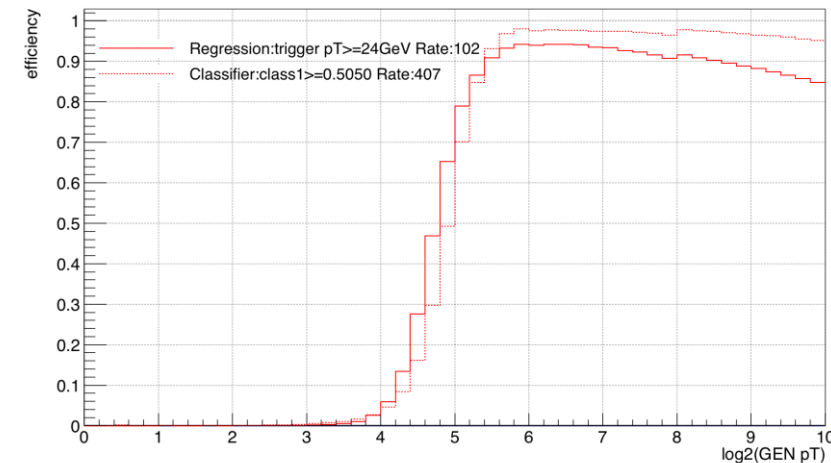
Regression CSC-only 90% Efficiency as Multi-classifier



Regression CSC-only 90% Efficiency as Classifier



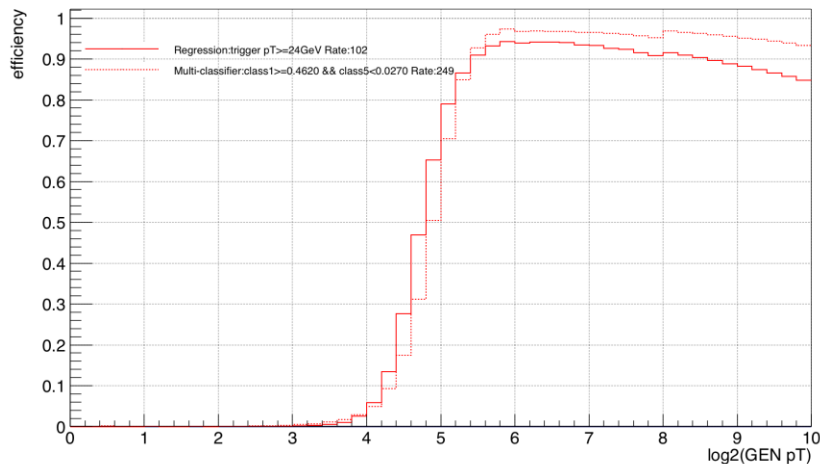
Regression CSC-only 90% Efficiency as Classifier



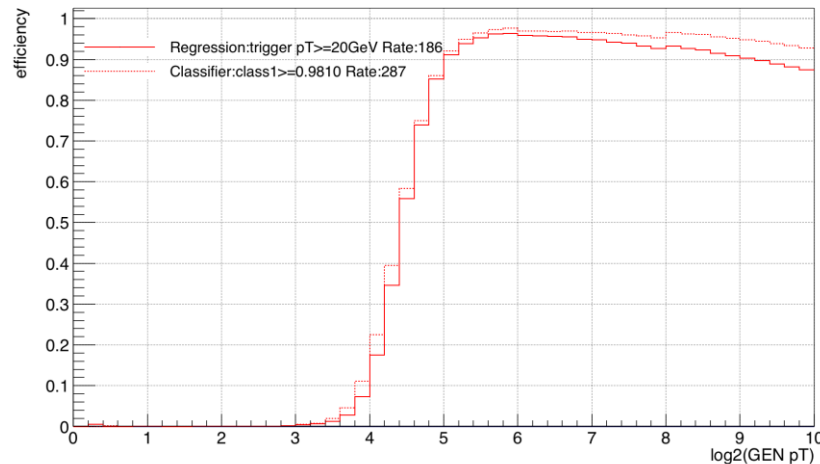
Regression 90% efficiency at same GEN pT as Classifiers with signal eff **95%**

	5 classes	2 classes (<8 , ≥ 32)	2 classes (<32 , ≥ 32)
Classifier cuts	$\text{class1} > 0.342 \ \&\& \ \text{class5} < 0.03$	$\text{class1} > 0.9724$	$\text{class1} > 0.505$
Classifiers Rate (ratio in all ZB events)	300 (0.152%)	339 (0.172%)	407 (0.206%)
Regression pT cuts	22 GeV	19 GeV	24 GeV
Regression Rate	135 (0.068%)	224 (0.113%)	102 (0.052%)

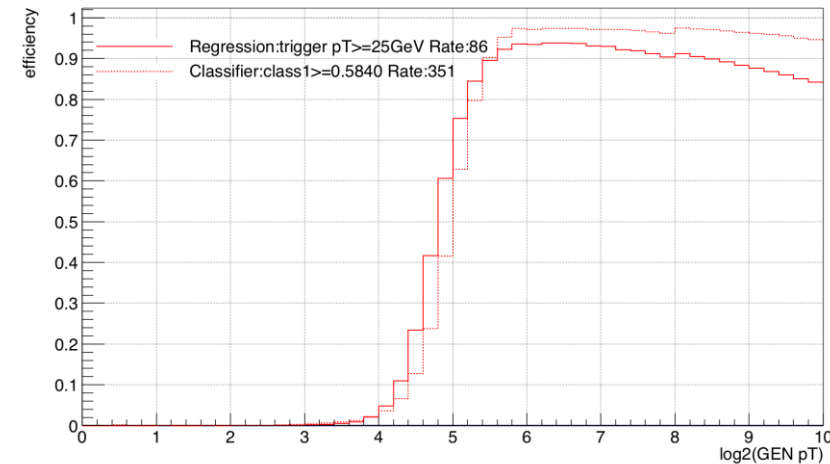
Regression CSC-only 90% Efficiency as Multi-classifier



Regression CSC-only 90% Efficiency as Classifier



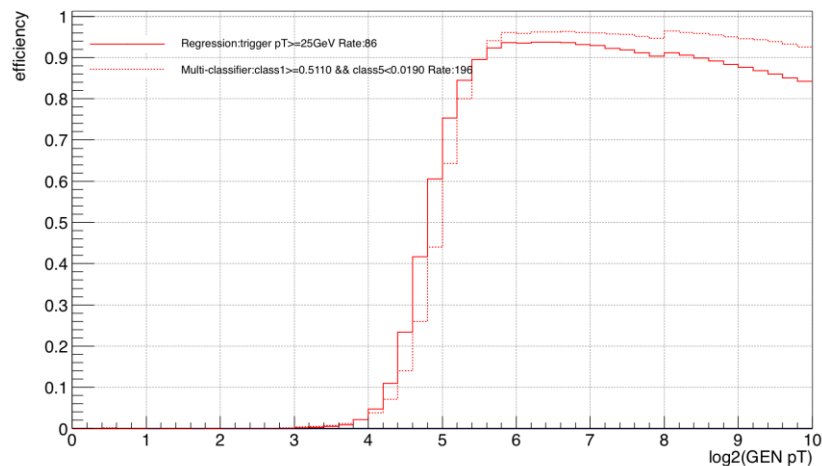
Regression CSC-only 90% Efficiency as Classifier



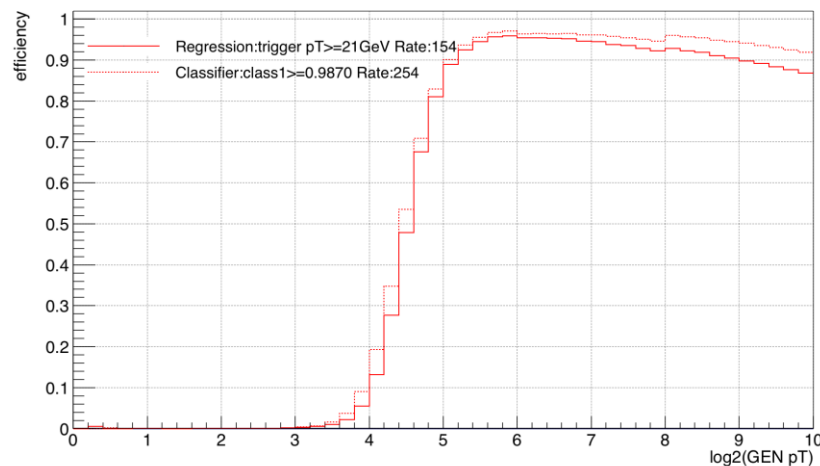
Regression 90% efficiency at same GEN pT as Classifiers with signal eff **94%**

	5 classes	2 classes (<8 , ≥ 32)	2 classes (<32 , ≥ 32)
Classifier cuts	$\text{class1} > 0.462 \ \&\& \ \text{class5} < 0.027$	$\text{class1} > 0.981$	$\text{class1} > 0.584$
Classifiers Rate (ratio in all ZB events)	249 (0.126%)	287 (0.145%)	351 (0.177%)
Regression pT cuts	24 GeV	20 GeV	25 GeV
Regression Rate	102 (0.051%)	186 (0.094%)	86 (0.043%)

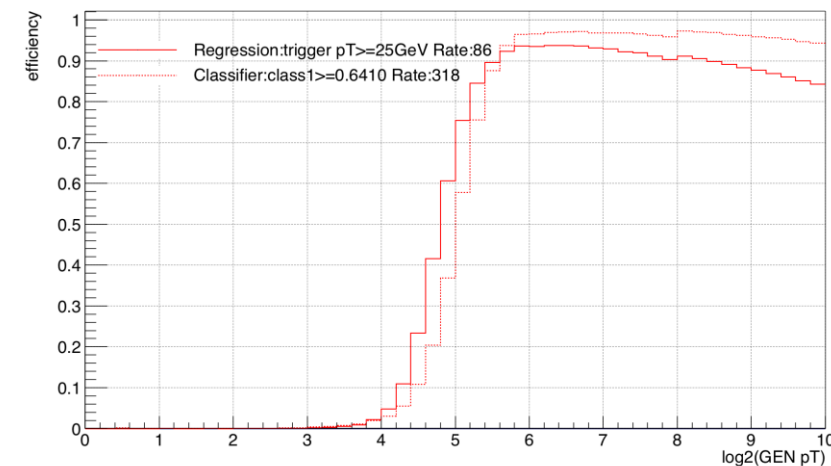
Regression CSC-only 90% Efficiency as Multi-classifier



Regression CSC-only 90% Efficiency as Classifier



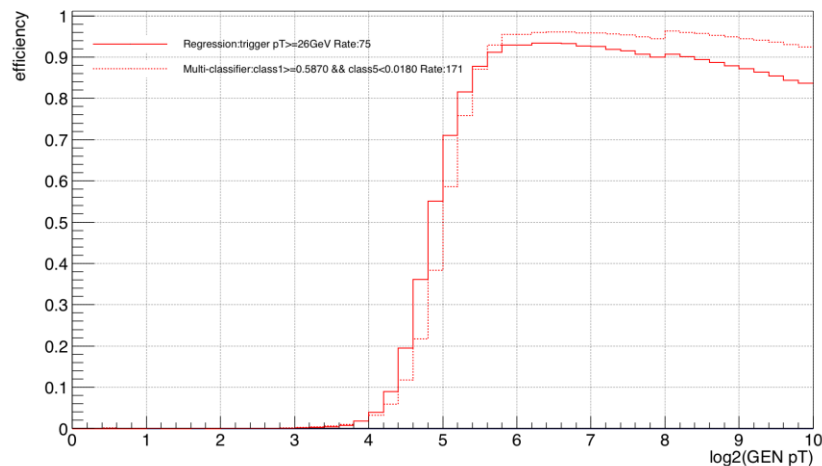
Regression CSC-only 90% Efficiency as Classifier



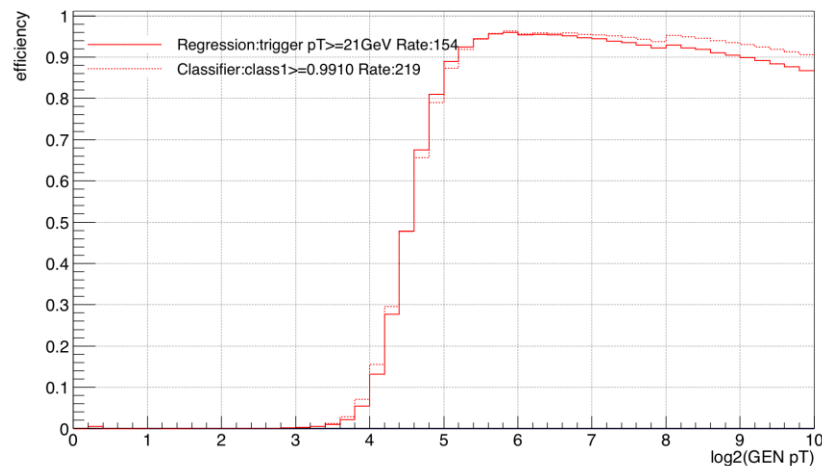
Regression 90% efficiency at same GEN pT as Classifiers with signal eff **93%**

	5 classes	2 classes ($<8, \geq 32$)	2 classes ($<32, \geq 32$)
Classifier cuts	$\text{class1} > 0.511 \ \&\& \ \text{class5} < 0.019$	$\text{class1} > 0.987$	$\text{class1} > 0.641$
Classifiers Rate (ratio in all ZB events)	196 (0.099%)	254 (0.128%)	318 (0.161%)
Regression pT cuts	25 GeV	21 GeV	25 GeV
Regression Rate	86 (0.043%)	154 (0.078%)	86 (0.043%)

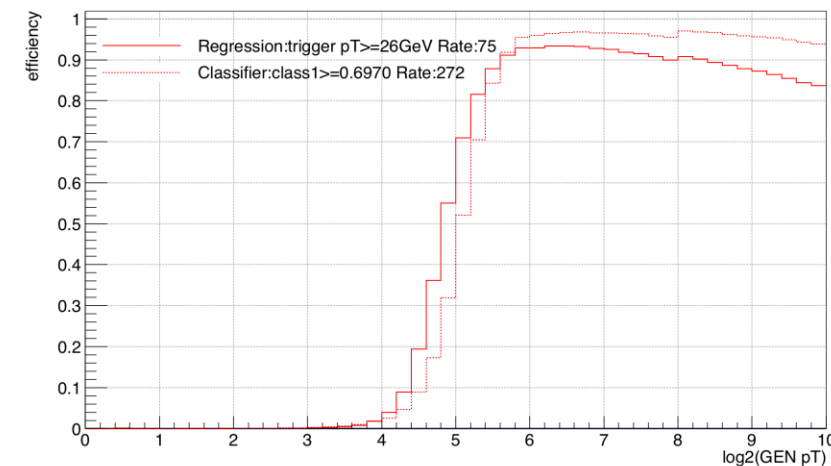
Regression CSC-only 90% Efficiency as Multi-classifier



Regression CSC-only 90% Efficiency as Classifier



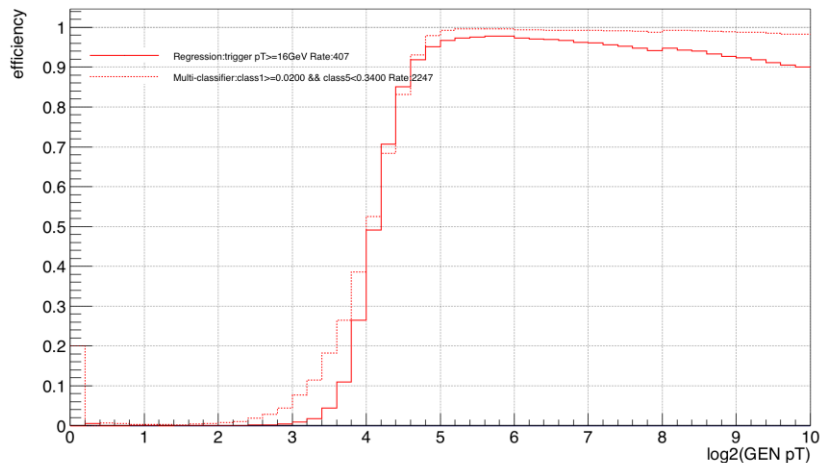
Regression CSC-only 90% Efficiency as Classifier



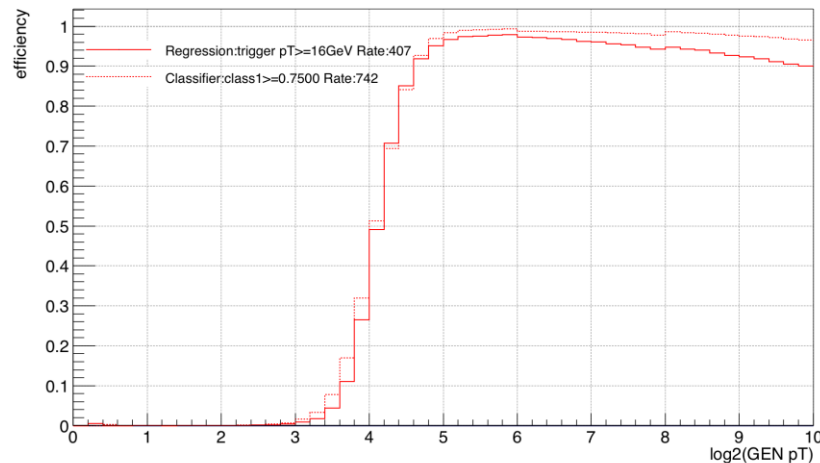
Regression 90% efficiency at same GEN pT as Classifiers with signal eff **92%**

	5 classes	2 classes (<8 , ≥ 32)	2 classes (<32 , ≥ 32)
Classifier cuts	class1 >0.587 && class5 <0.018	class1 >0.991	class1 >0.697
Classifiers Rate (ratio in all ZB events)	171 (0.08%)	219 (0.11%)	272 (0.137%)
Regression pT cuts	26 GeV	21 GeV	26 GeV
Regression Rate	75 (0.037%)	154 (0.077%)	75 (0.037%)

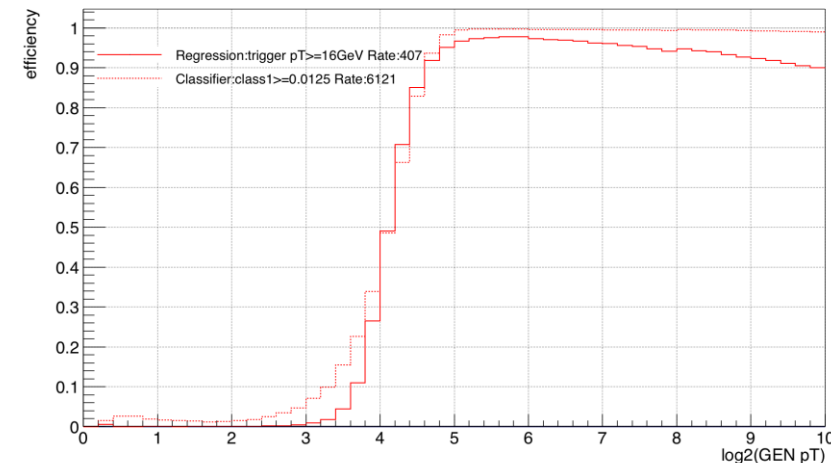
Multi-classifier CSC-only 90% Efficiency as Regression 16 GeV



Classifier CSC-only 90% Efficiency as Regression 16 GeV



Classifier CSC-only 90% Efficiency as Regression 16 GeV



Classifiers 90% efficiency at same GEN pT as Regression

Regression	5 classes [1,8][8,16][16,24][24,32][32,1000]	2 classes (<8, >=32)	2 classes (<32, >=32)
pT > 16GeV	class1>0.02 && class5<0.34	class1>0.75	class1>0.0125
407 (0.206%)	2247 (1.14%)	742 (0.375%)	6121 (3.10%)

Conclusion

- Best classifier trained with 2 classes setting: GEN pT [32, 256], [1, 8) GeV
 - “Best” under same BDT settings as regression (didn’t optimize for all classifiers)
 - Rate ratio close to 2017 BDT regression
 - Always achieve 90% eff at 32 GeV while other classifiers fail (slide 22-25)
 - Steeper turn on curve (slide 26)
- Depending on trained classes settings, the efficiency turn on curve for same pT cut (32 GeV) is different

MVA settings

BDT

```
factX->BookMethod( loadX, TMVA::Types::kBDT, "BDTG",  
"!H:!V:NTrees=400::BoostType=Grad:Shrinkage=0.1:nCuts=1000:MaxDepth=5:Min  
NodeSize=0.000001:RegressionLossFunctionBDTG=LeastSquares");
```