Flavor Tagging - NN

- replace BDT in LCFIPlus with simple NN
- four categories (BDTs / NNs) depending on number of reconstructed vertices & pseudovertices
- multi-class NN with three output classes (b jets, c jets, other)
- simple feed-forward NN, four hidden layers, 100 nodes each
- activation function: relu
- loss: Cross entropy, optimizer: Adam
- 19 - 29 input features
- batch size: 50, 50 epochs trained
- standardization applied
- all classes weighted equally in loss
- 50% training, 25% validation, 25% test

News

- choose training with minimum loss for interference
- added confusion matrices, misidentification rates
- preselection done as in LCFIPlus algorithm (trk1d0sig != 0)

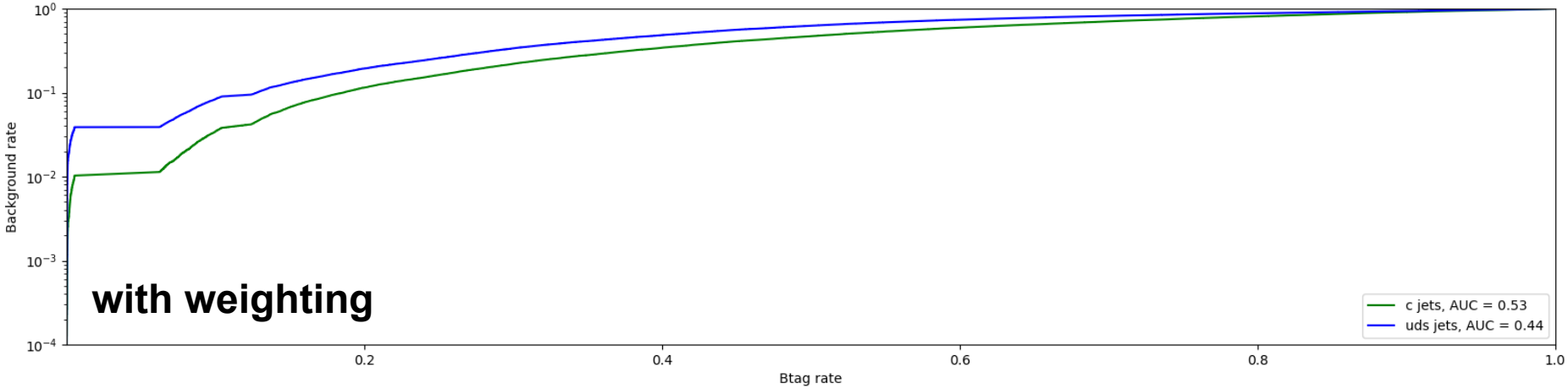
- optimized training in different categories (layers, nodes, learning rate, with/without standardization..)
- tried one training without categorization

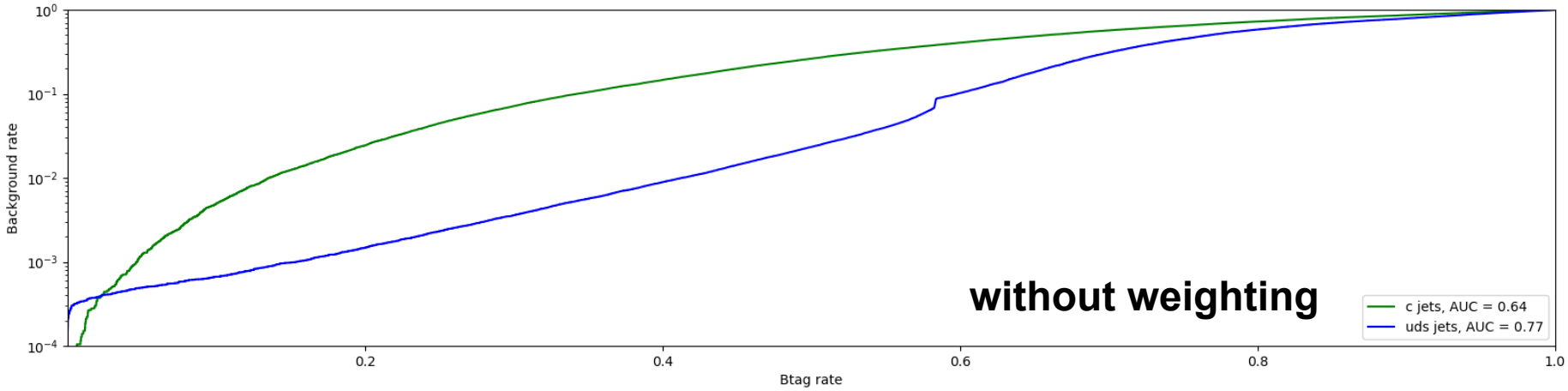
Category 1

- no secondary vertex found inside the jet, 0-2 pseudovertices found
  - category dominated by light jets
  - total number of jets: 2090245
  - number of b jets: 118190
  - number of c jets: 343759
  - number of uds jets: 1628296

- training with/without weighting of classes in loss

Category 1

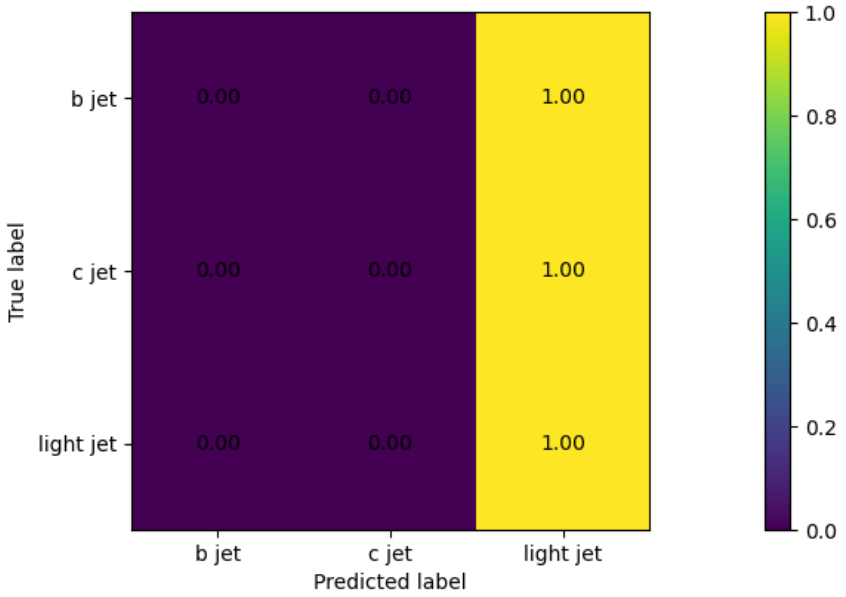


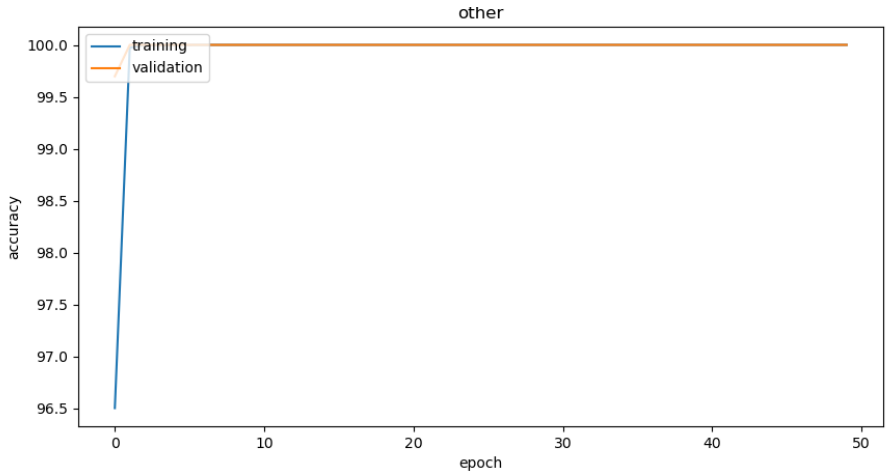


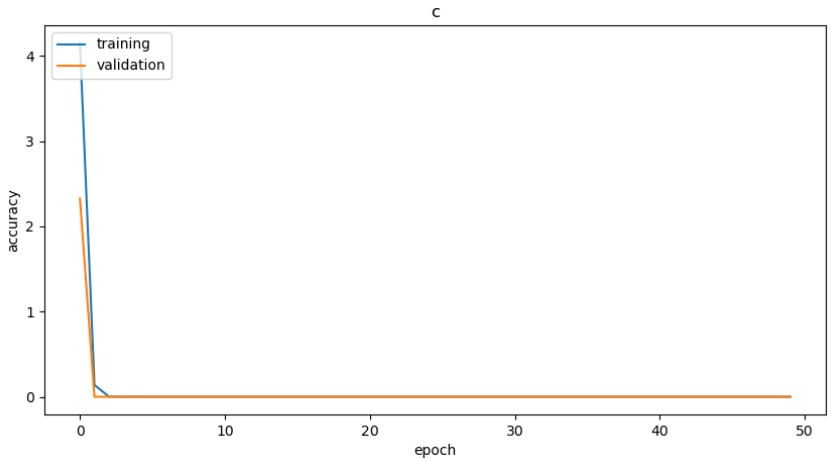
➡ without weighting much better separation between light jets and b jets

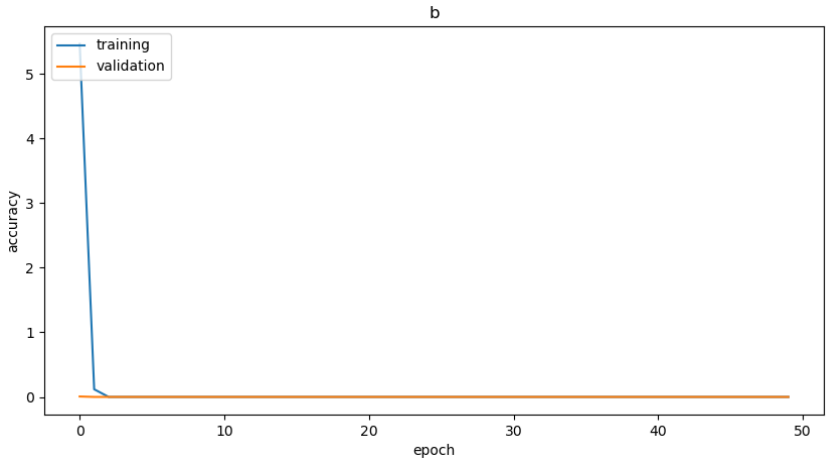
Further optimization

More layers, smaller learning rates, more nodes









➡ since category is dominated by light jets, small loss can be reached by sorting all jets into one class

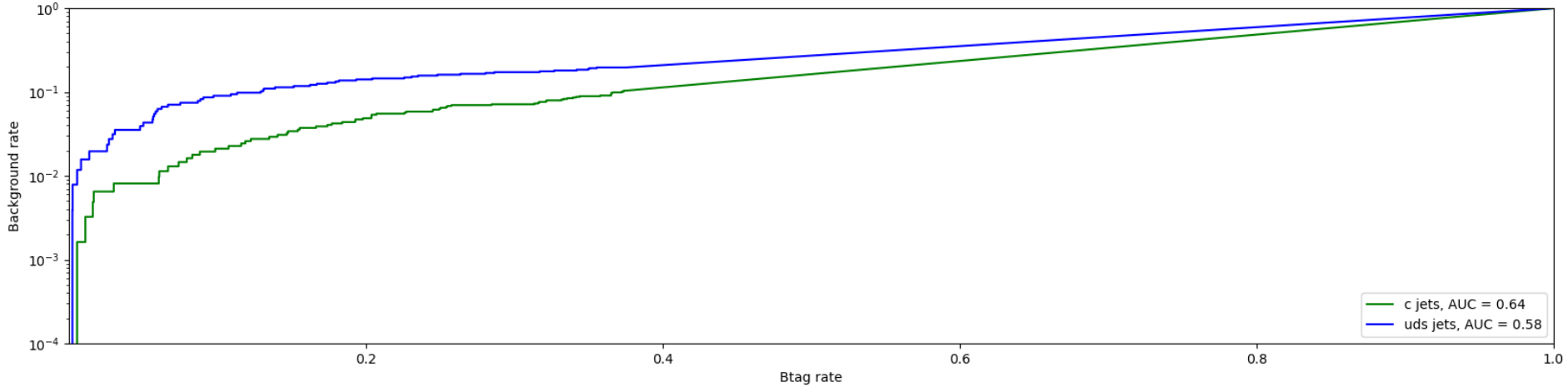
Category 4

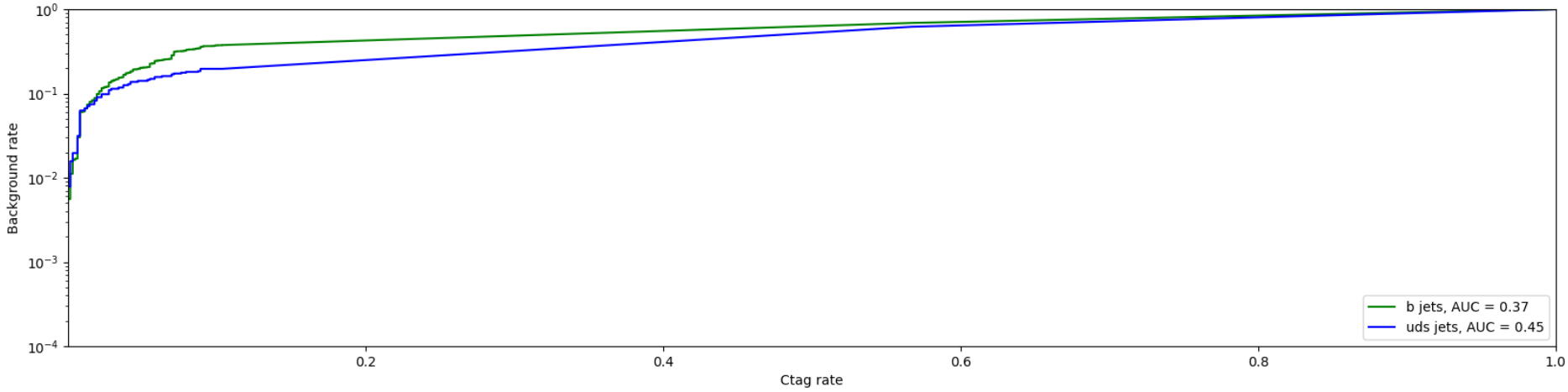
- two secondary vertices found inside the jet
  - ➡ category dominated by b jets
  - total number of jets: 130179
  - number of b jets: 126773
  - number of c jets: 2488
  - number of uds jets: 918

- problem even worse, could not get a good result for this category so far

Category 4

Four layers, 150 nodes, 200 epochs, lr = 0.0001

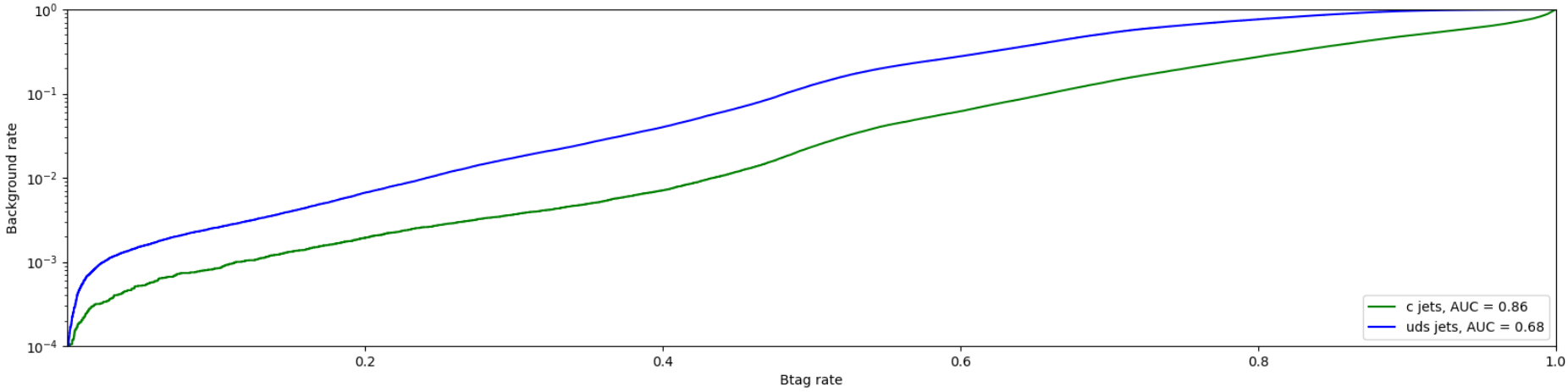


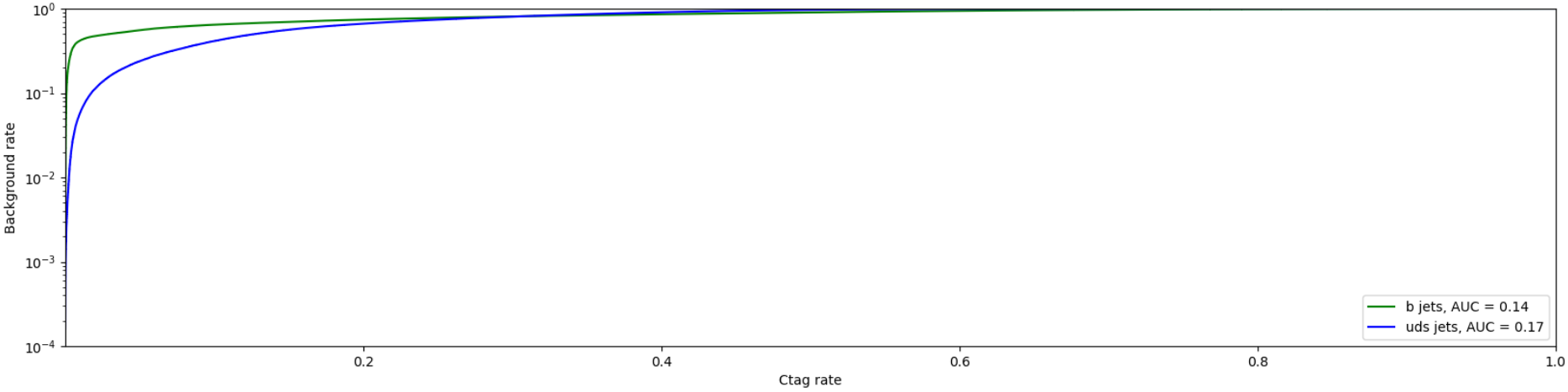


New idea: no categorization

- do not use any categorization
  - ➡ train one NN with all features from 4 categories + number of secondary vertices, number of pseudovertices
- four layers, 150 nodes each, trained 50 epochs, lr = 0.0001

New idea: no categorization

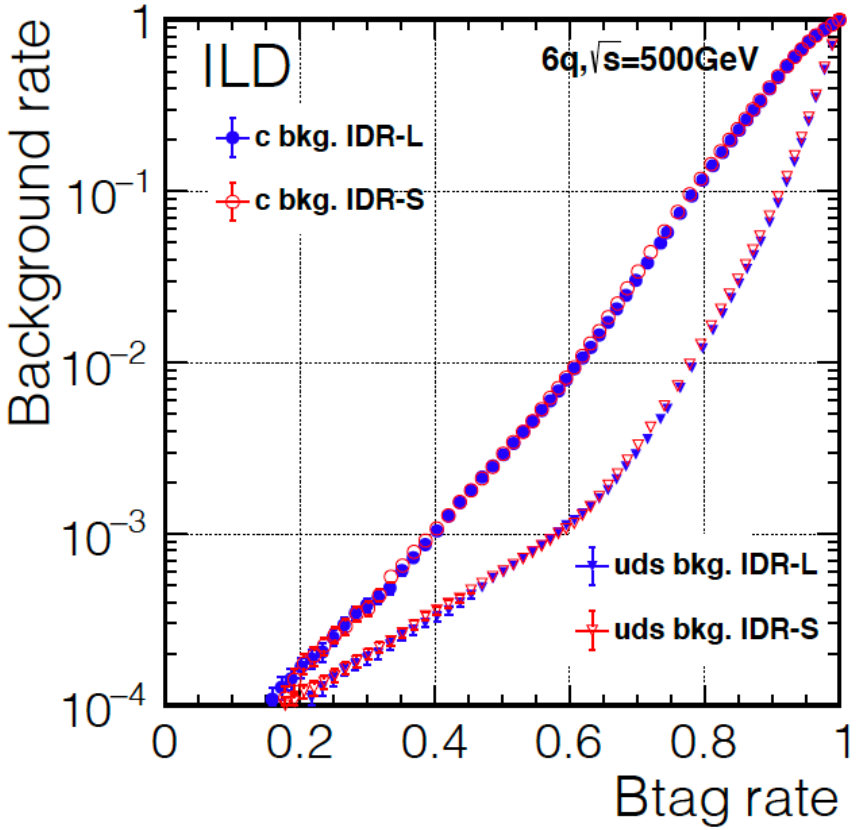


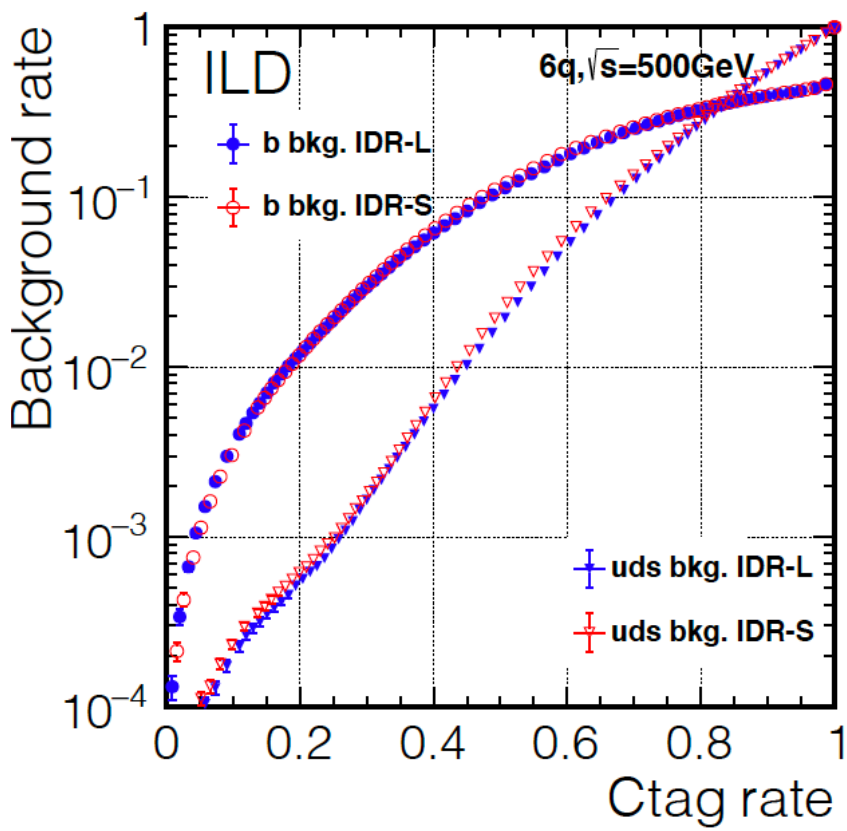


Next steps / options

- use the catgeorization: idea for solving the problem that all events are sorted into one class?
- no categorization: can the NN be optimized?
- try something more complicated like e.g. CMS

Flavor Tagging - Performance LCFIPlus





Features

"features" : ["trk1d0sig", "trk2d0sig", "trk1z0sig", "trk2z0sig", "trk1pt_jete", "trk2pt_jete", "jprobr25sigma", "jprobz25sigma", "d0bprob2", "d0cprob2", "d0qprob2", "z0bprob2", "z0cprob2", "z0qprob2", "nmuon", "nelectron", "trkmass"],

"features" : ["trk1d0sig", "trk2d0sig", "trk1z0sig", "trk2z0sig", "trk1pt_jete", "trk2pt_jete", "jprobr2", "jprobz2", "vtxlen1_jete", "vtxsig1_jete", "vtxdirang1_jete", "vtxmom1_jete", "vtxmass1", "vtxmult1", "vtxmasspc", "vtxprob", "d0bprob2", "d0cprob2", "d0qprob2", "z0bprob2", "z0cprob2", "z0qprob2", "trkmass", "nelectron", "nmuon"],

"features" : ["trk1d0sig", "trk2d0sig", "trk1z0sig", "trk2z0sig", "trk1pt_jete", "trk2pt_jete", "jprobr2", "jprobz2", "vtxlen1_jete", "vtxsig1_jete", "vtxdirang1_jete", "vtxmom1_jete", "vtxmass1", "vtxmult1", "vtxmasspc", "vtxprob","1vtxprob", "vtxlen12all_jete", "vtxmassall"],

"features" : ["trk1d0sig", "trk2d0sig", "trk1z0sig", "trk2z0sig", "trk1pt_jete", "trk2pt_jete", "jprobr2", "jprobz2","vtxlen1_jete", "vtxsig1_jete", "vtxdirang1_jete", "vtxmom1_jete", "vtxmass1", "vtxmult1", "vtxmasspc", "vtxprob", "vtxlen2_jete", "vtxsig2_jete", "vtxdirang2_jete", "vtxmom2_jete", "vtxmass2", "vtxmult2", "vtxlen12_jete", "vtxsig12_jete", "vtxdirang12_jete", "vtxmom_jete", "vtxmass", "vtxmult", "1vtxprob"],

Features

Name	Description	Normalization factor	Used by category
trk1d0sig	d0 significance of track with highest d0 significance	1	A, B, C, D
trk2d0sig	d0 significance of track with second highest d0 significance	1	A, B, C, D
trk1z0sig	z0 significance of track with highest d0 significance	1	A, B, C, D
trk2z0sig	z0 significance of track with second highest d0 significance	1	A, B, C, D
trk1pt	transverse momentum of track with highest d0 significance	$1/E_{\text{jet}}$	A, B, C, D
trk2pt	transverse momentum of track with second highest d0 significance	$1/E_{\text{jet}}$	A, B, C, D
jprobr	joint probability in the r-phi plane using all tracks	1	A, B, C, D
jprobr5sigma	joint probability in the r-phi plane using all tracks having impact parameter significance exceeding 5 sigma	1	A, B, C, D
jprobz	joint probability in the z projection using all tracks	1	A, B, C, D
jprobz5sigma	joint probability in the z projection using all tracks having impact parameter significance exceeding 5 sigma	1	A, B, C, D
d0bprob	product of b-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
d0cprob	product of c-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
d0qprob	product of q-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
z0bprob	product of b-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
z0cprob	product of c-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
z0qprob	product of q-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
nmuon	number of identified muons	1	A, B, C, D
nelectron	number of identified electrons	1	A, B, C, D
trkmass	mass of all tracks exceeding 5 sigma significance in d0/z0 values	1	A, B, C, D

Features

Name	Description	Normalization factor	Used by category
1vtxprob	vertex probability with all tracks associated in vertices combined	1	B, C, D
vtxlen1	decay length of the first vertex in the jet (zero if no vertex is found)	$1/E_{\text{jet}}$	B, C, D
vtxlen2	decay length of the second vertex in the jet (zero if number of vertex is less than two)	$1/E_{\text{jet}}$	D
vtxlen12	distance between the first and second vertex (zero if number of vertex is less than two)	$1/E_{\text{jet}}$	D
vtxsig1	decay length significance of the first vertex in the jet (zero if no vertex is found)	$1/E_{\text{jet}}$	B, C, D
vtxsig2	decay length significance of the second vertex in the jet (zero if number of vertex is less than two)	$1/E_{\text{jet}}$	D
vtxsig12	vtxlen12 divided by its error as computed from the sum of the covariance matrix of the first and second vertices, projected along the line connecting the two vertices	$1/E_{\text{jet}}$	D
vtxdirang1	the angle between the momentum (computed as a vector sum of track momenta) and the displacement of the first vertex	E_{jet}	B, C, D
vtxdirang2	the angle between the momentum (computed as a vector sum of track momenta) and the displacement of the second vertex	E_{jet}	D
vtxmult1	number of tracks included in the first vertex (zero if no vertex is found)	1	B, C, D
vtxmult2	number of tracks included in the second vertex (zero if number of vertex is less than two)	1	D
vtxmult	number of tracks which are used to form secondary vertices (summed for all vertices)	1	D
vtxmom1	magnitude of the vector sum of the momenta of all tracks combined into the first vertex	$1/E_{\text{jet}}$	B, C, D
vtxmom2	magnitude of the vector sum of the momenta of all tracks combined into the second vertex	$1/E_{\text{jet}}$	D
vtxmass1	mass of the first vertex computed from the sum of track four-momenta	1	B, C, D
vtxmass2	mass of the second vertex computed from the sum of track four-momenta	1	D
vtxmass	vertex mass as computed from the sum of four momenta of all tracks forming secondary vertices	1	B, C, D
vtxmasspc	mass of the vertex with minimum pt correction allowed by the error matrices of the primary and secondary vertices	1	B, C, D
vtxprob	vertex probability; for multiple vertices, the probability P is computed as 1-P = (1-P1)(1-P2)...(1-PN)	1	B, C, D