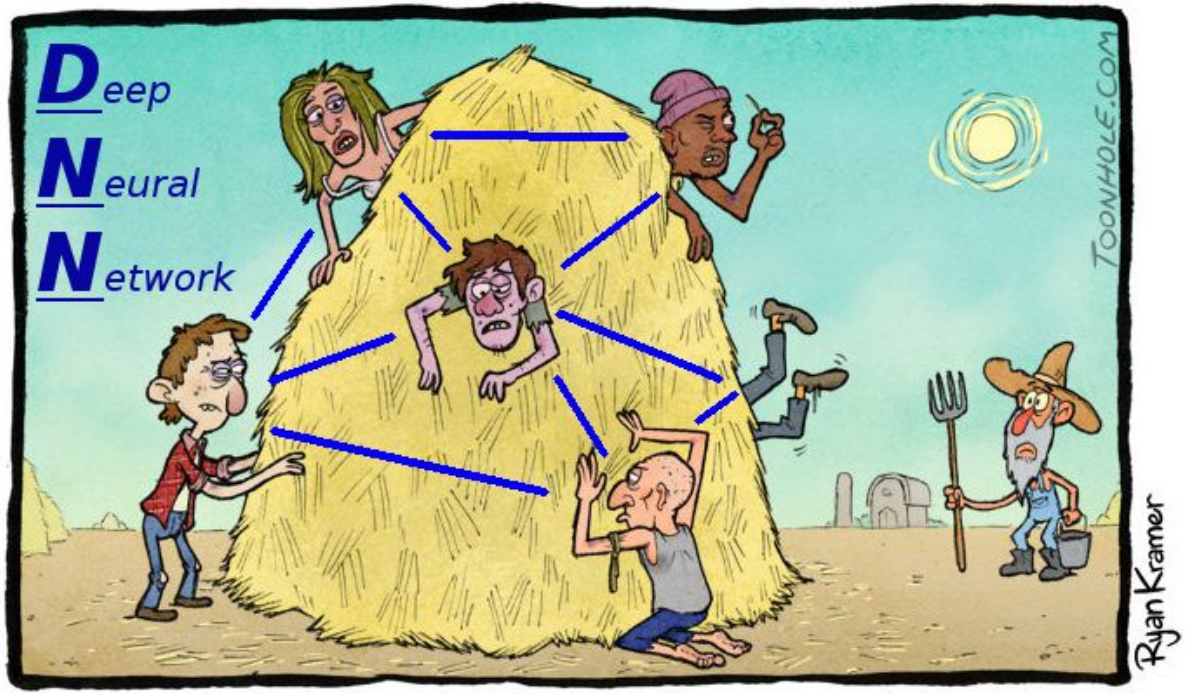
Direct optimization of discovery significance

Status report 07-08-2019

Contents

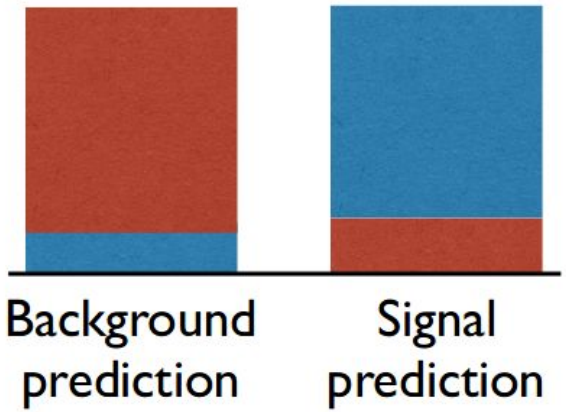
1. Reminder
2. Binary classification
3. Multiclass
4. Summary



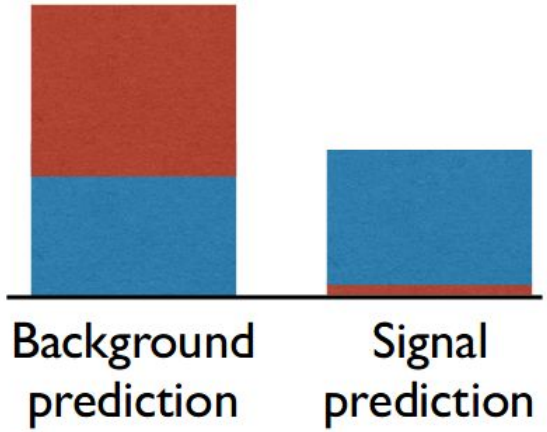
Reminder

Approaches for optimization

Accuarity



Statistical significance



Statistical significance

- *s* = N correctly classified signal events
- *b* = N background events classified as signal

Asimov significance:

$$Z_A = \left[2 \left((s+b) \ln \left[\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{1/2}$$

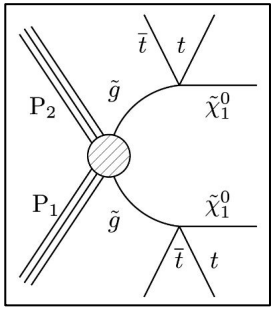
loss: $l_{Asimov} = 1 / Z_A^2$

Simplified statistical significance:

b >> s, σ_b + $\sigma_b \propto$ b ⇨ Z $s/\sqrt{s+b}$

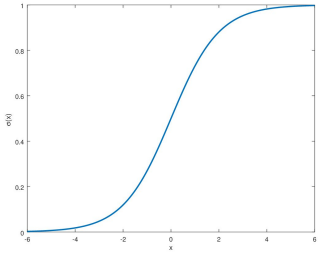
loss: $l_s = (s + b) / s^2$

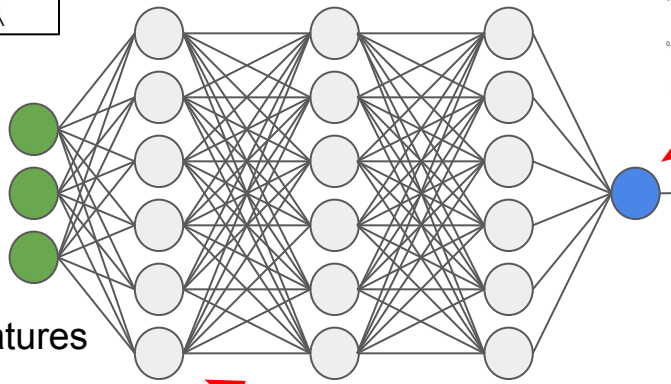
Setup



Two hidden layers x
256 nodes each

sigmoid

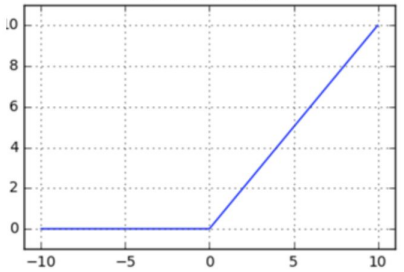




features

(0; 1)

ReLU



Binary classification:

- Primary optimization with I_s.
- Afterwards use I_{Asimov}.

Multiclass:

- Primary optimization with crossentropy.
- Afterwards ...some combination of crossentropy and significance.

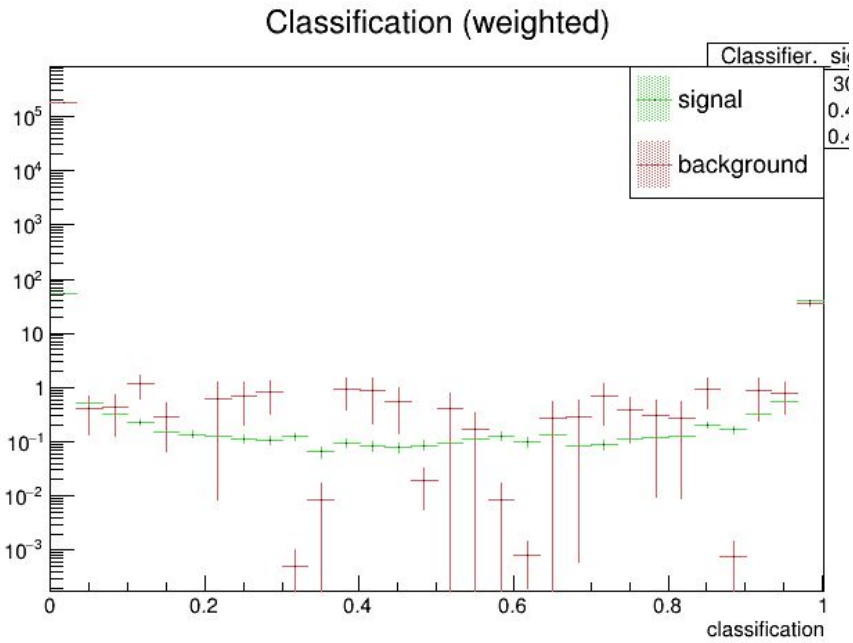
Binary classification

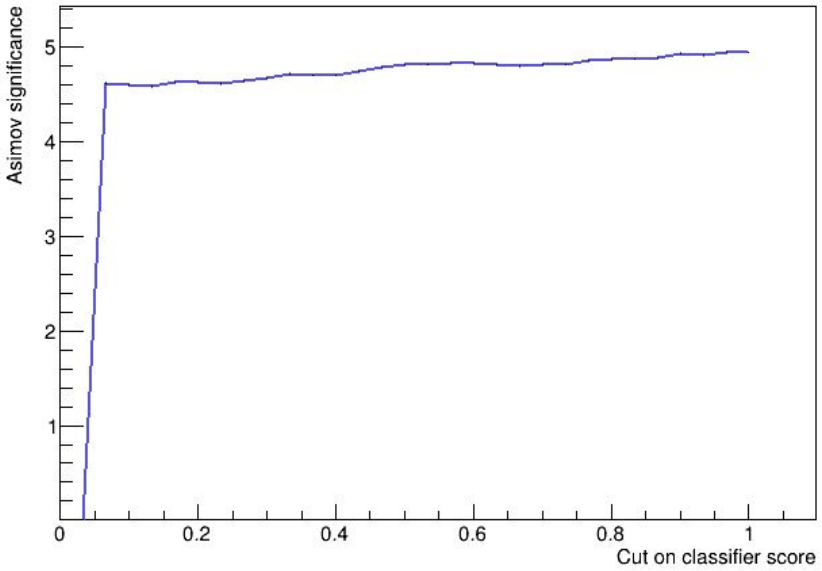
* Considered model: M_{gluino} = 1900 GeV amd M_{LSP} = 1000 GeV.

Original approach

* sigma (bg) = 10 %

loss = l_s -> l_{Asimov} → ~5σ

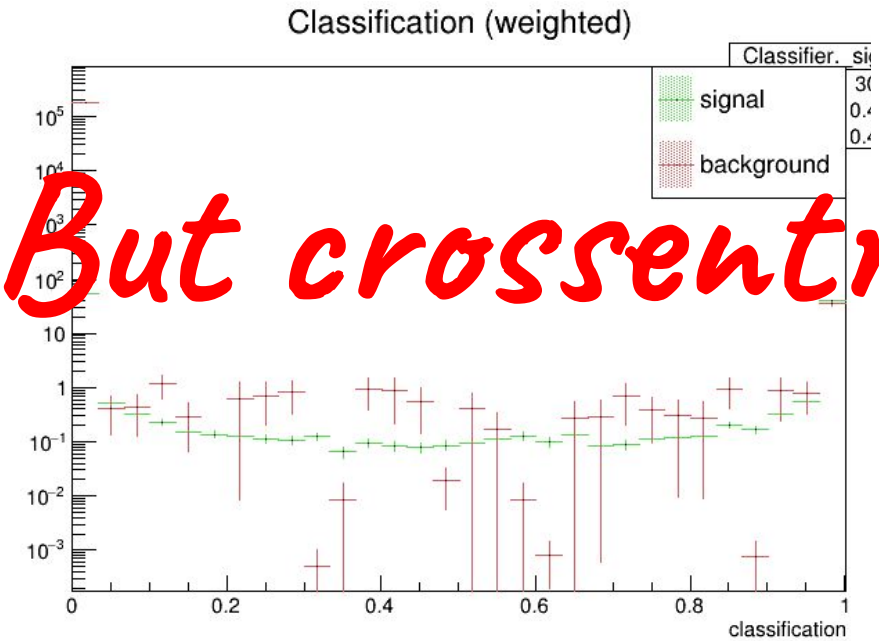




Original approach

* sigma (bg) = 10 %

loss = l_s -> l_{Asimov} → ~5σ



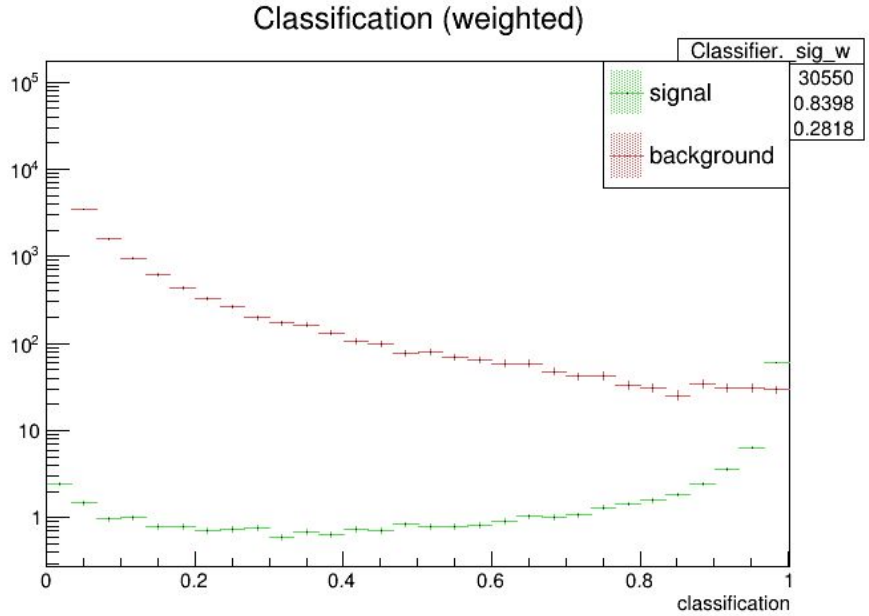
But crossentropy is better!

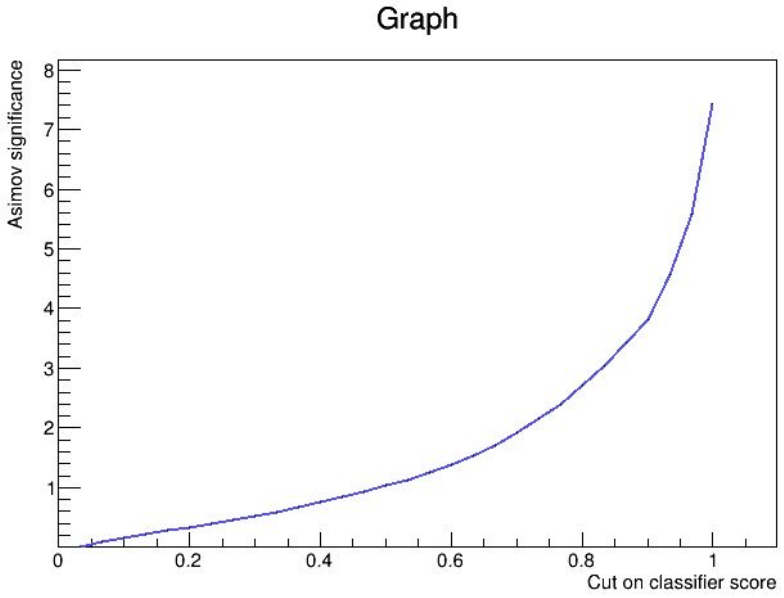
Bare crossentropy

* sigma (bg) = 10 %

loss = binary crossentropy (bCE)

~7σ

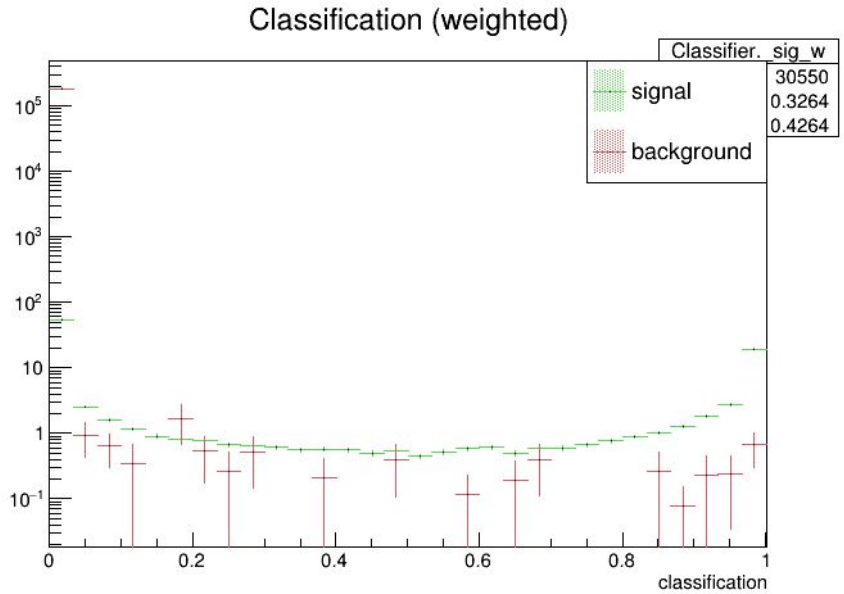


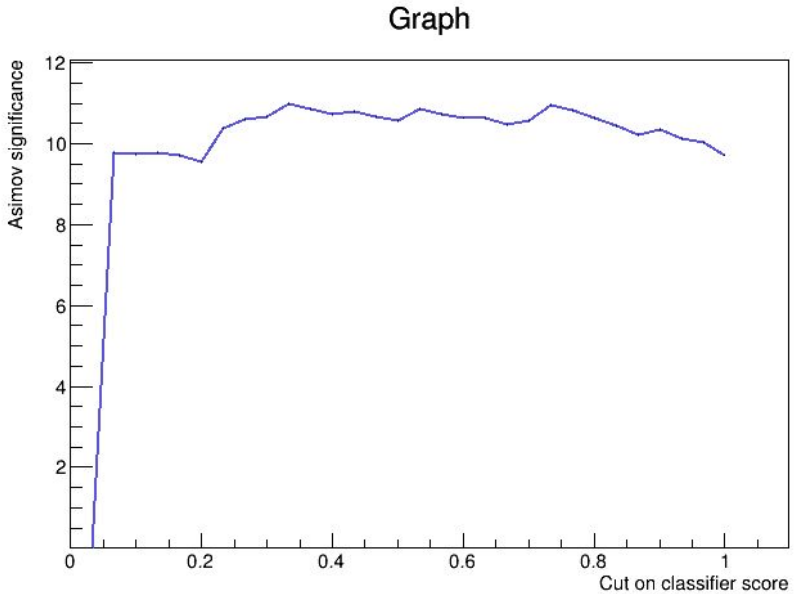


Pretrain with crossentropy

* sigma (bg) = 10 %

loss = crossentropy -> I_{Asimov} ⇒ ~10σ



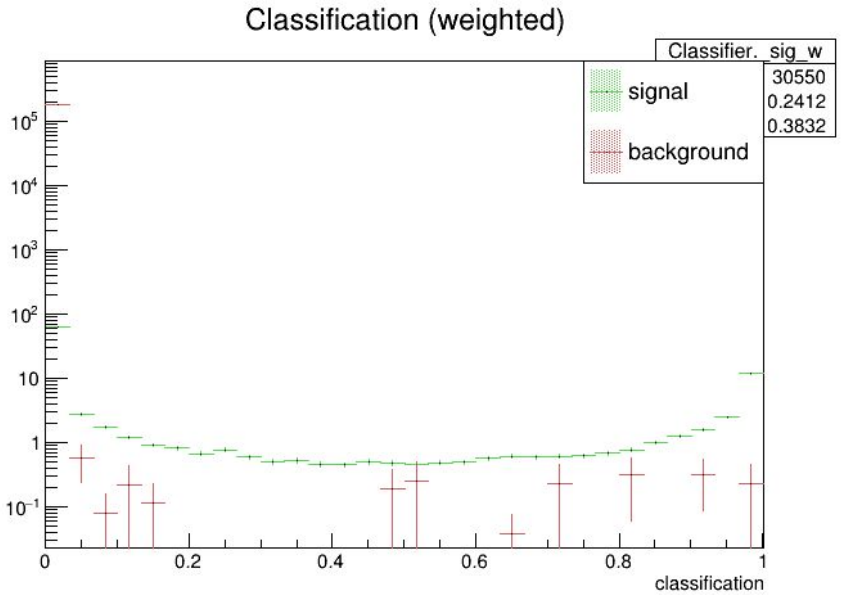


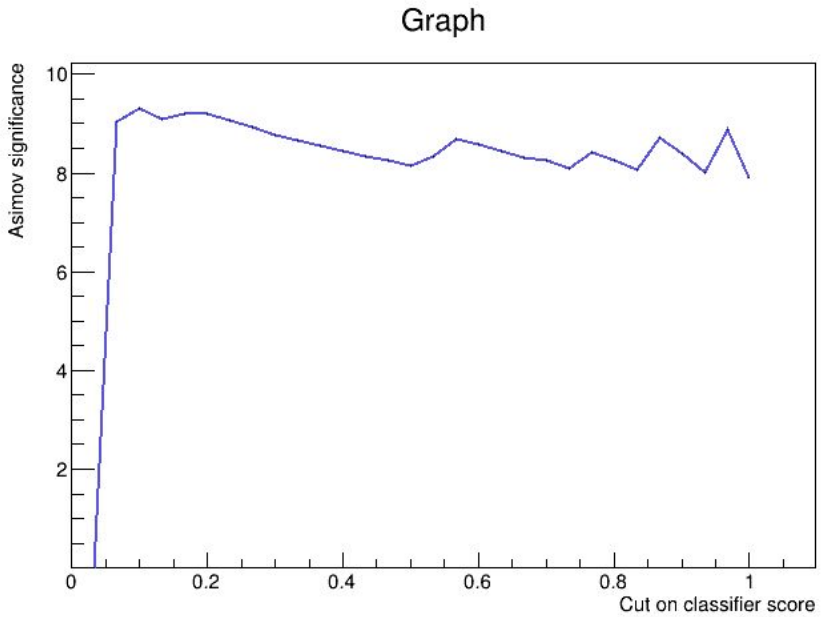
Pretrain with crossentropy

* sigma (bg) = 30 %

loss = crossentropy -> I_{Asimov}

~8.5σ



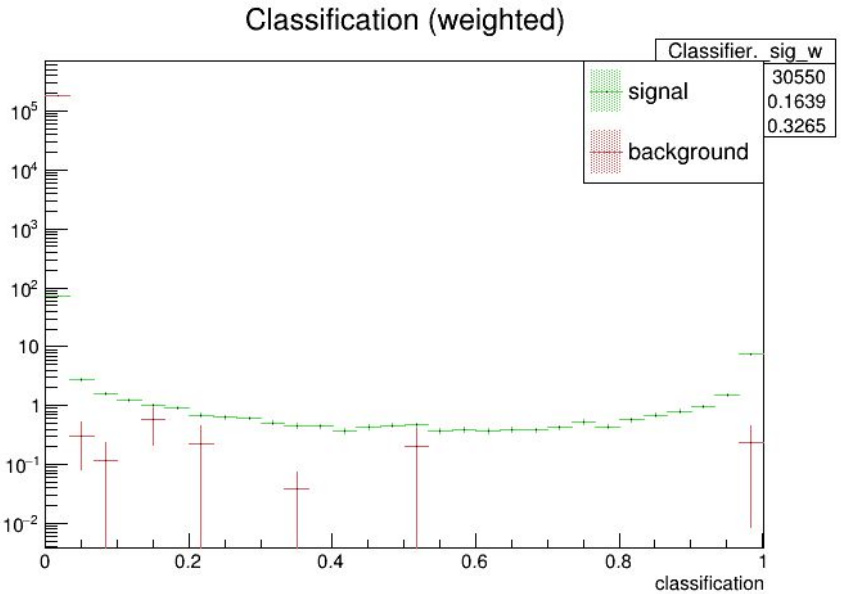


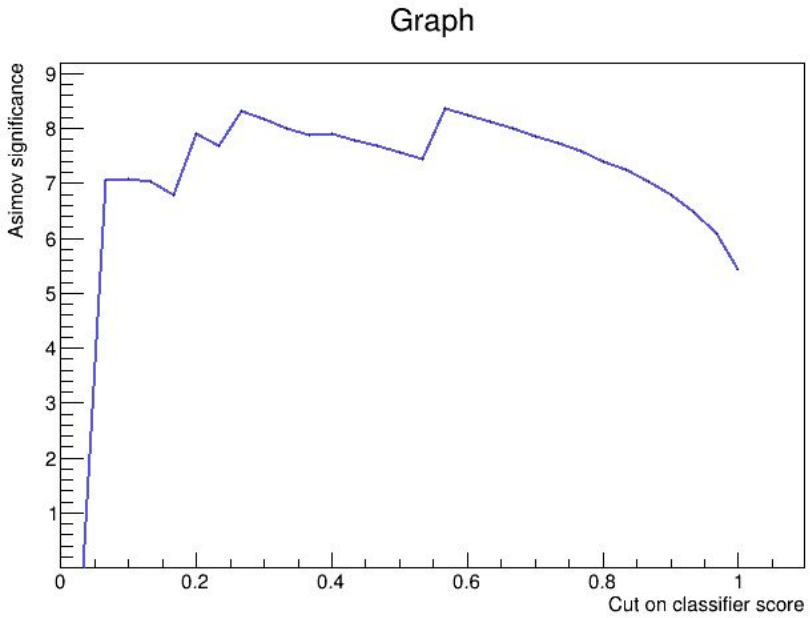
Pretrain with crossentropy

* sigma (bg) = 50 %

loss = crossentropy -> I_{Asimov}

~7σ



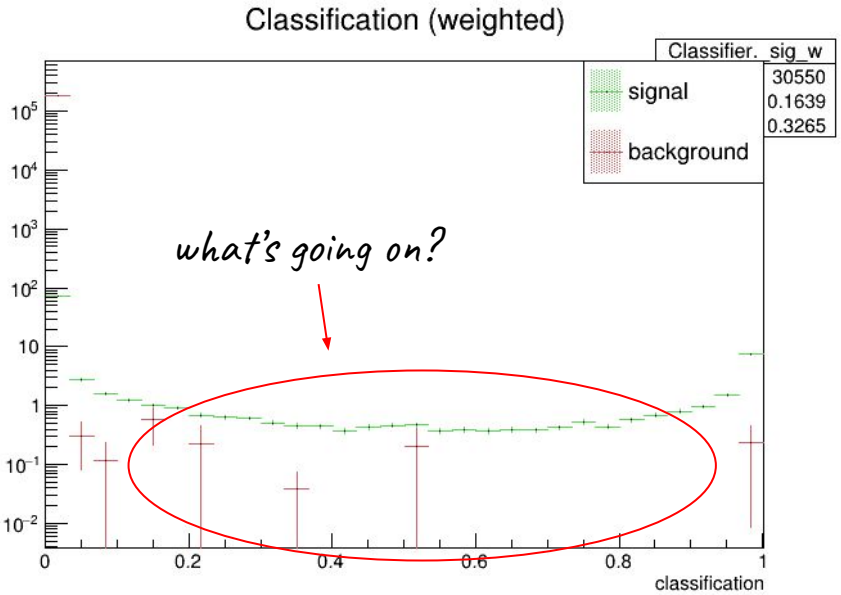


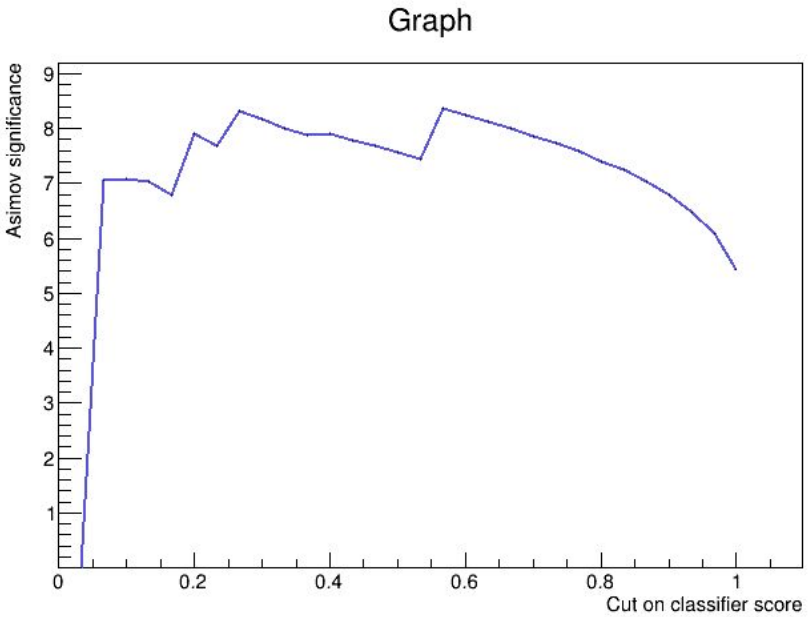
Pretrain with crossentropy

* sigma (bg) = 50 %

loss = crossentropy -> I_{Asimov}

~7σ





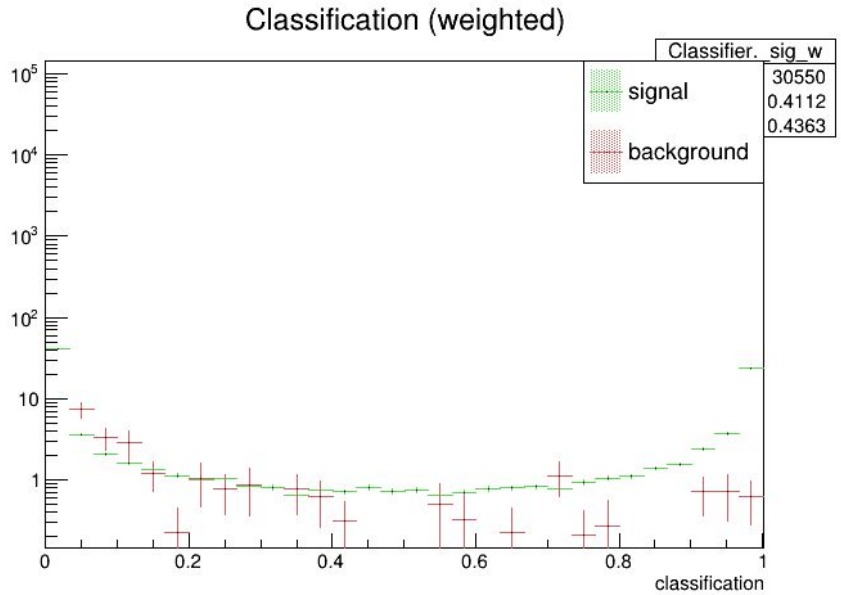
b = max(b, 2)

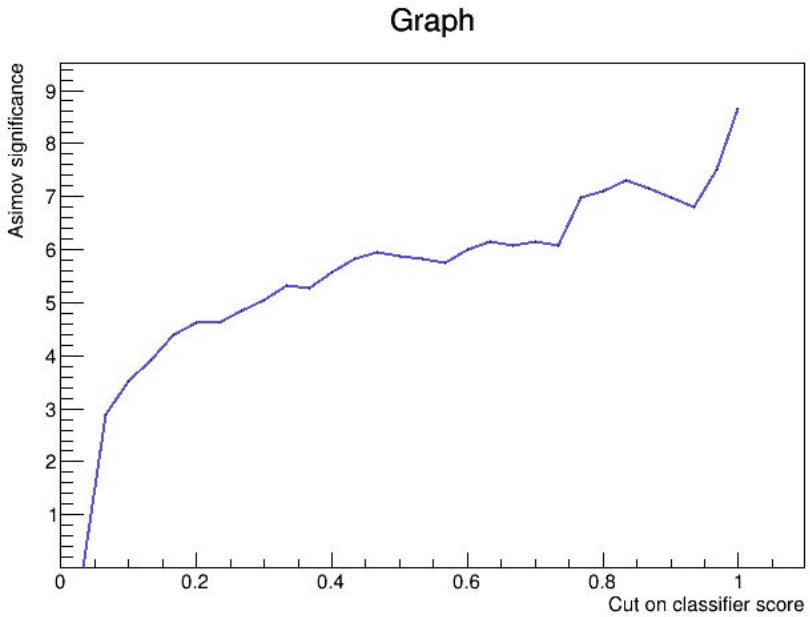
* sigma (bg) = 50 %

loss = crossentropy -> I_{Asimov}

~8σ

* Maybe just a statistical fluctuation





Multiclass

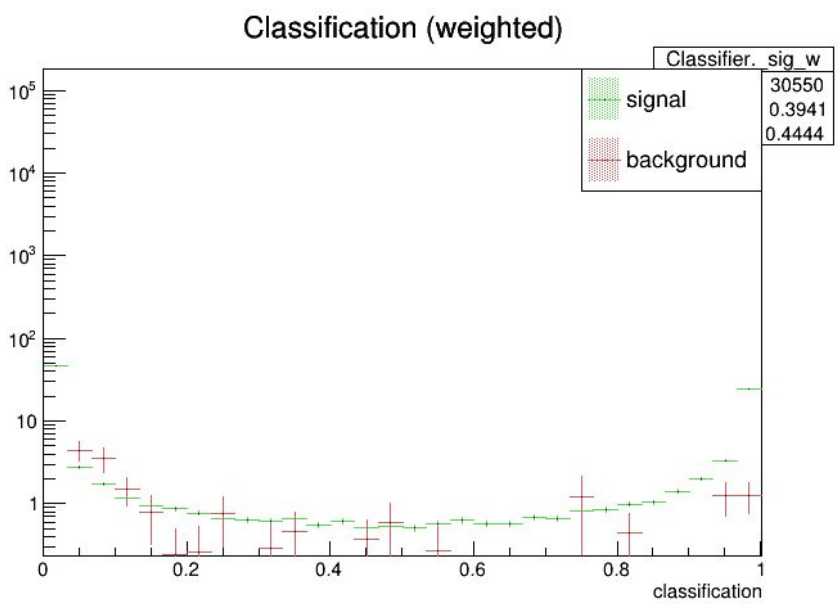
* Considered model: M_{gluino} = 1900 GeV amd M_{LSP} = 1000 GeV.
* loss = l_{Asimov} + scale . crossentropy

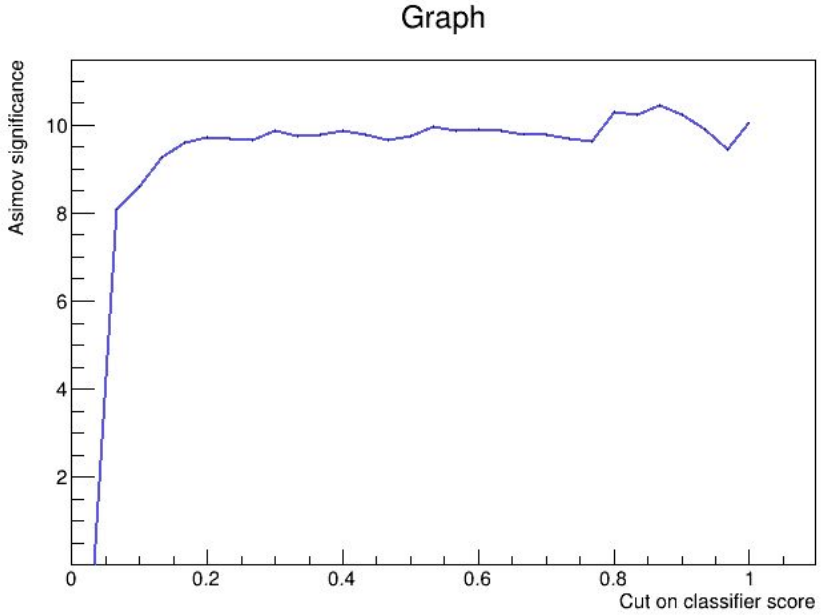
Multiclass

* sigma (bg) = 10 %

I_{Asimov}(... / sample_weights)
crossentropy(... / sample_weights)

~10σ



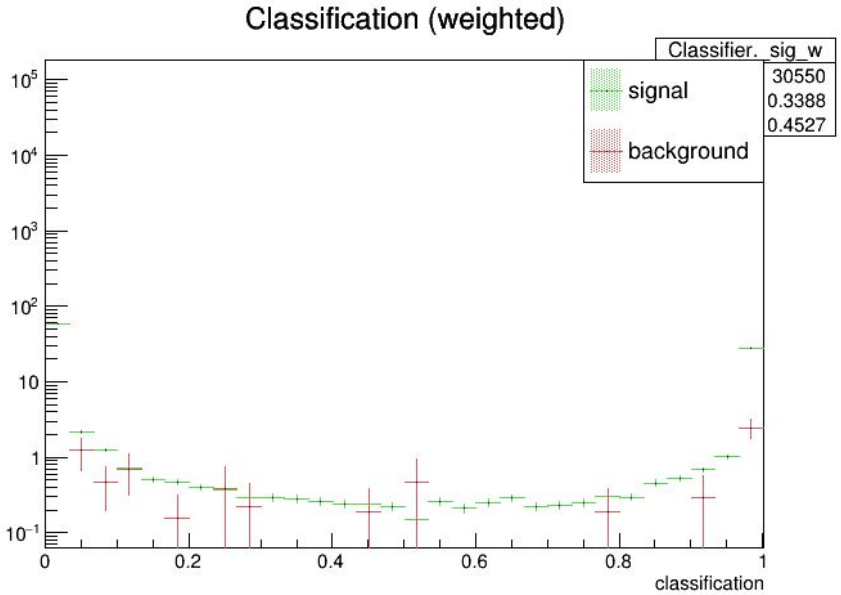


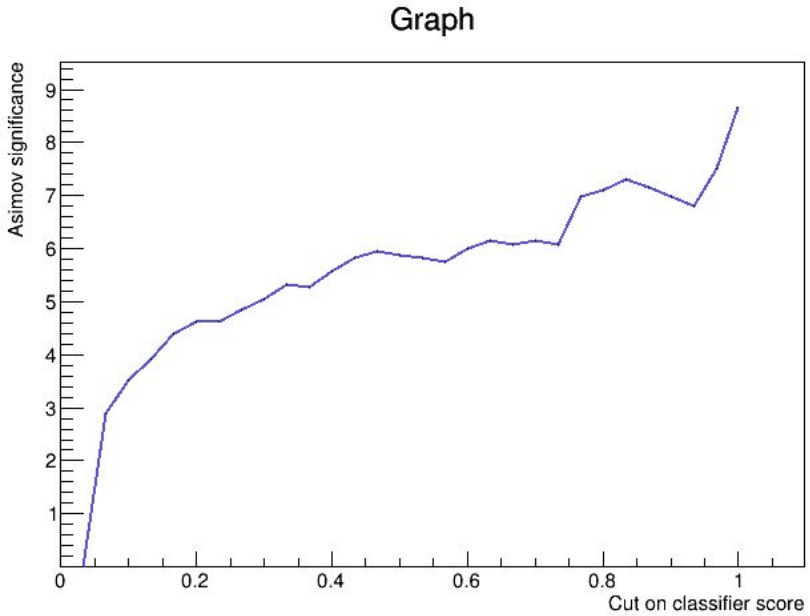
Multiclass

* sigma (bg) = 10 %

I_{Asimov}(... / sample_weights)
crossentropy(... / class_weights)

~8σ





To be continued...