My BDT worked. Until I looked at data.

My ML failure story

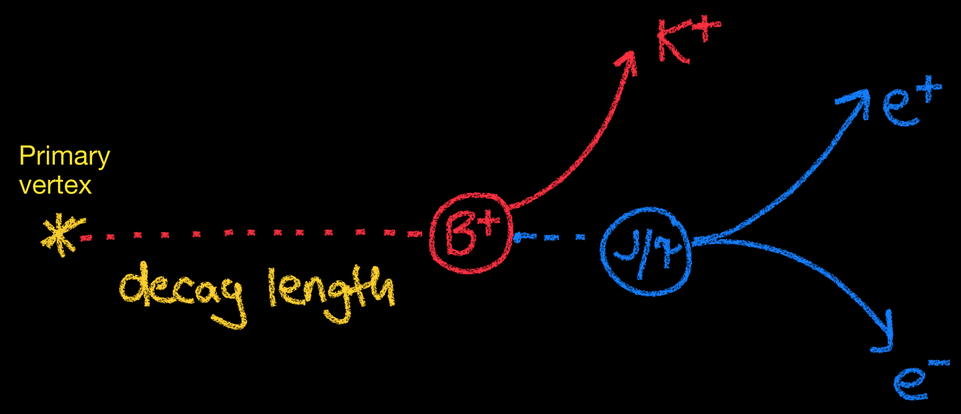
Ida Storehaug
University of Oslo
HAMLET-PHYSICS 2026, Copenhagen

I had a problem perfectly suited for ML…

- Rare signal, plenty of background
  - 500.000 proton-proton collisions per second
  - ~0.7% produce a beauty quark pair
  - A few tens of % of these hadronize to a $B^{\pm}$ meson
  - ~0.006% of these decay in «my» channel

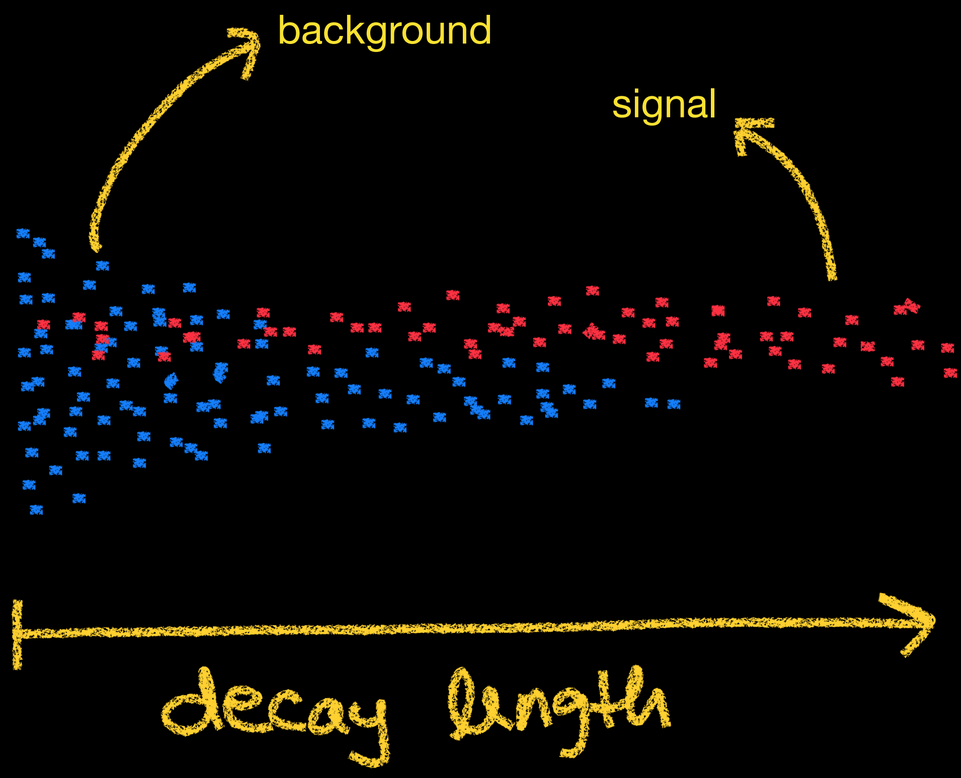
I had a problem perfectly suited for ML…

- Rare signal, plenty of background
  - 500.000 proton-proton collisions per second
  - ~0.7% produce a beauty quark pair
  - A few tens of % of these hadronize to a $B^{\pm}$ meson
  - ~0.006% of these decay in «my» channel



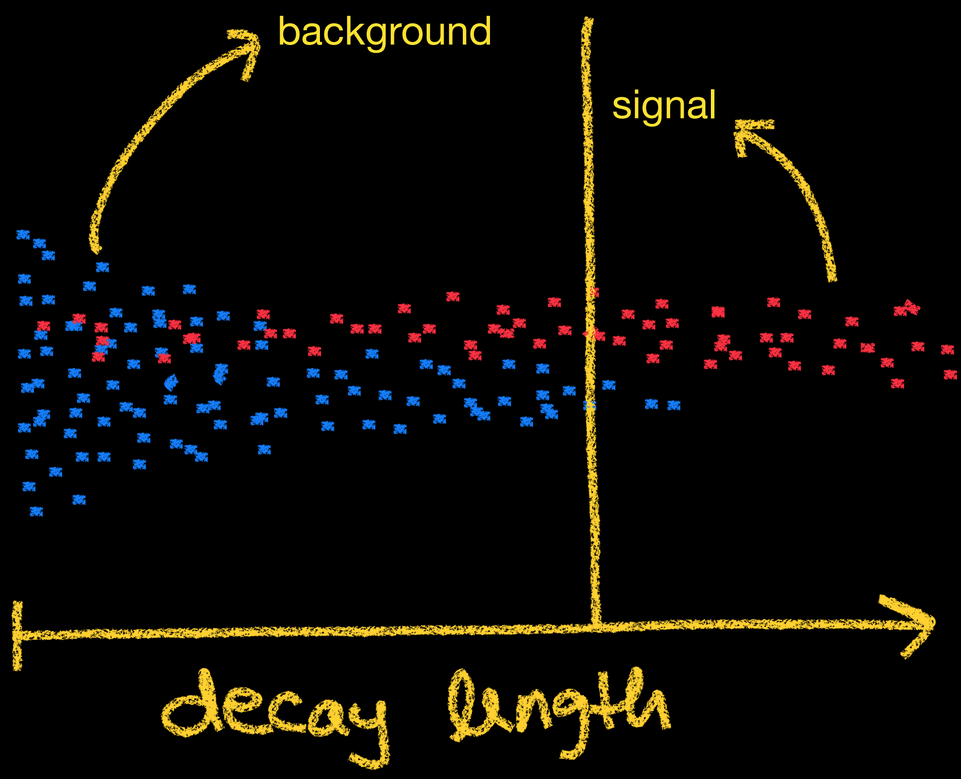
I had a problem perfectly suited for ML…

- Rare signal, plenty of background
  - 500.000 proton-proton collisions per second
  - ~0.7% produce a beauty quark pair
  - A few tens of % of these hadronize to a $B^{\pm}$ meson
  - ~0.006% of these decay in «my» channel
- Several discriminating variables



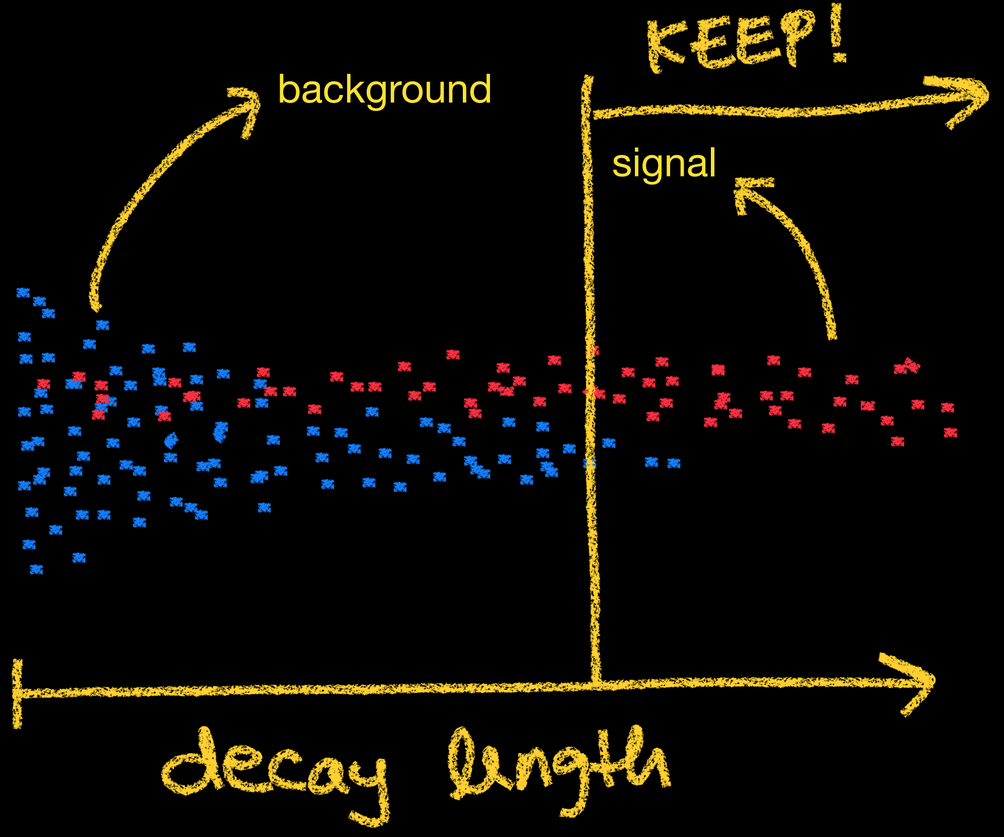
I had a problem perfectly suited for ML…

- Rare signal, plenty of background
  - 500.000 proton-proton collisions per second
  - ~0.7% produce a beauty quark pair
  - A few tens of % of these hadronize to a $B^{\pm}$ meson
  - ~0.006% of these decay in «my» channel
- Several discriminating variables



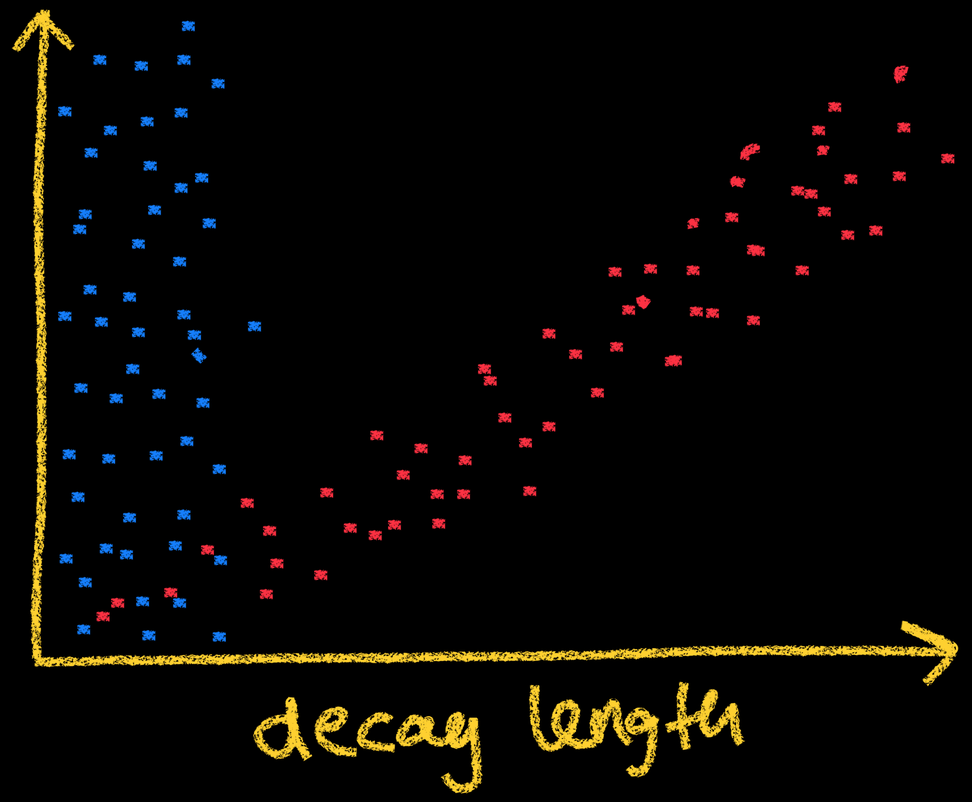
I had a problem perfectly suited for ML…

- Rare signal, plenty of background
  - 500.000 proton-proton collisions per second
  - ~0.7% produce a beauty quark pair
  - A few tens of % of these hadronize to a $B^{\pm}$ meson
  - ~0.006% of these decay in «my» channel
- Several discriminating variables



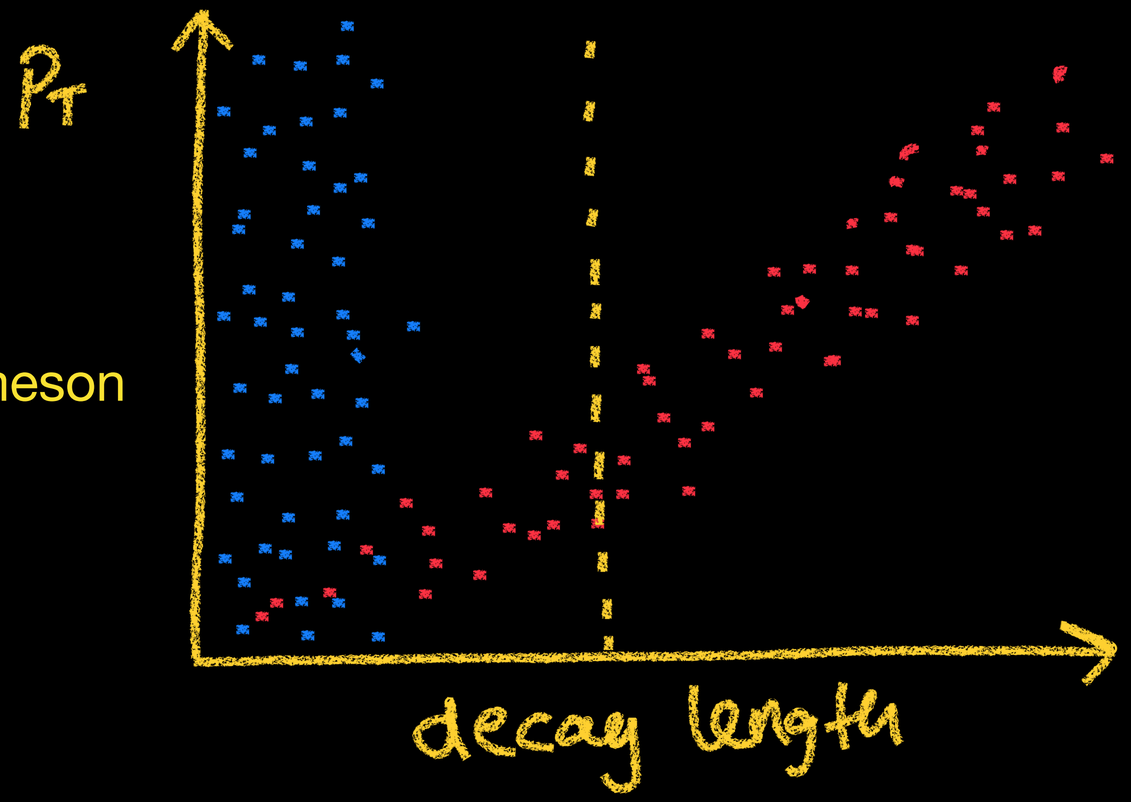
I had a problem perfectly suited for ML…

- Rare signal, plenty of background
  - 500.000 proton-proton collisions per second
  - ~0.7% produce a beauty quark pair
  - A few tens of % of these hadronize to a $B^{\pm}$ meson
  - ~0.006% of these decay in «my» channel
- Several discriminating variables



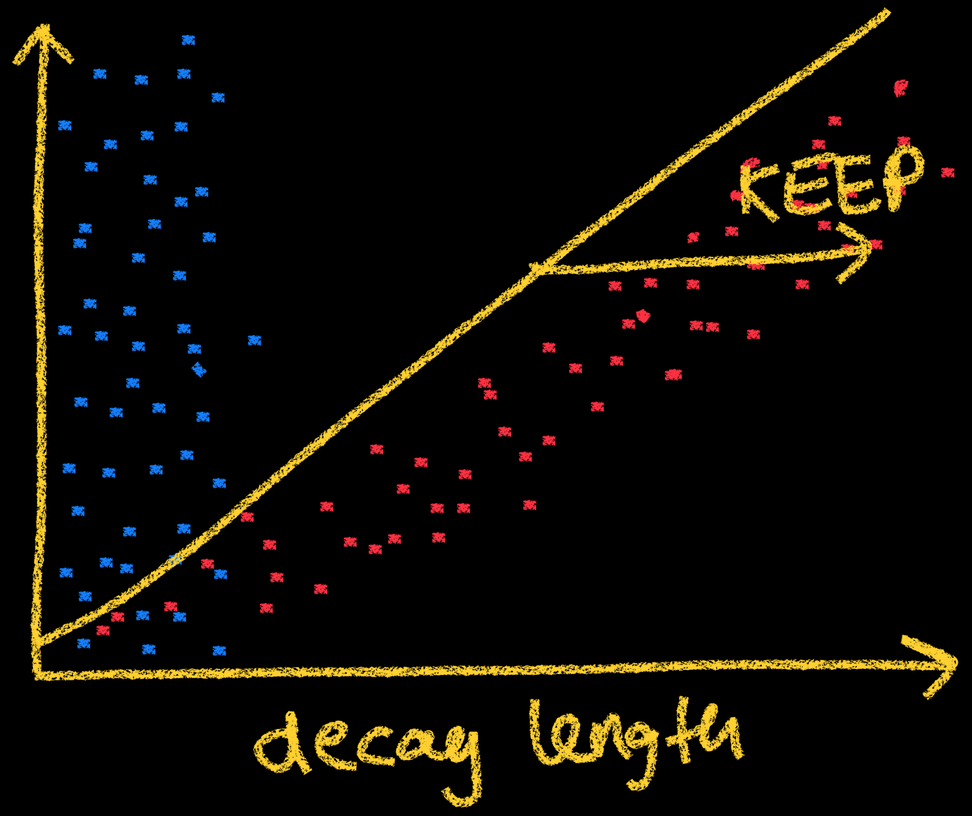
I had a problem perfectly suited for ML…

- Rare signal, plenty of background
  - 500.000 proton-proton collisions per second
  - ~0.7% produce a beauty quark pair
  - A few tens of % of these hadronize to a $B^{\pm}$ meson
  - ~0.006% of these decay in «my» channel
- Several discriminating variables
- Strong correlation between them

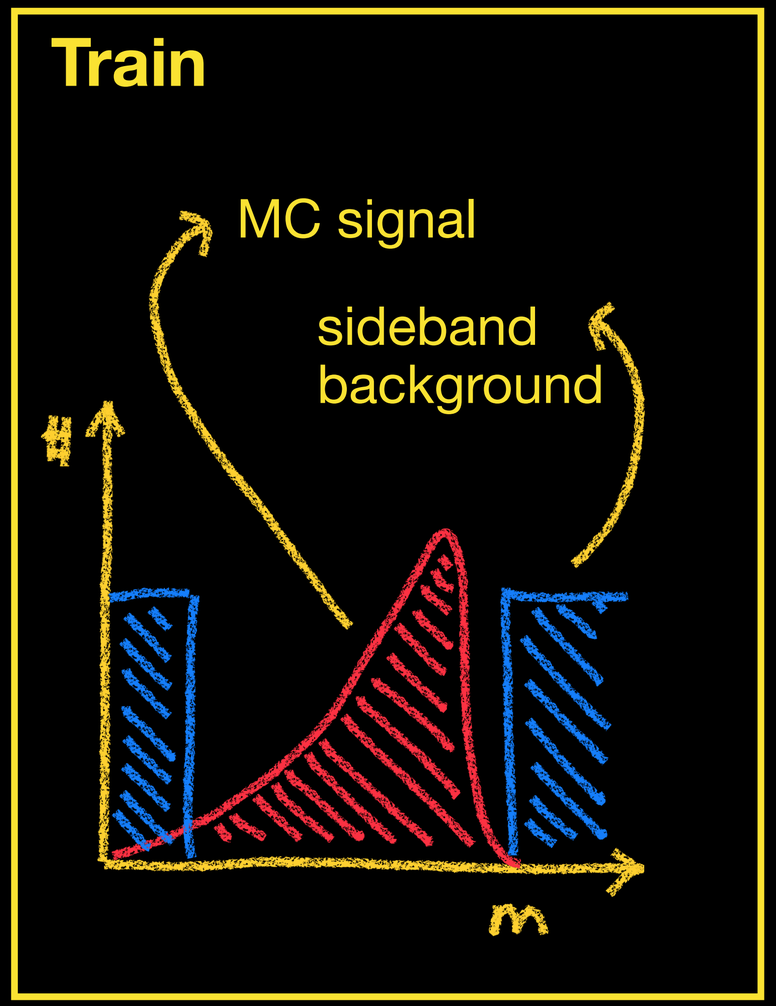


I had a problem perfectly suited for ML…

- Rare signal, plenty of background
  - 500.000 proton-proton collisions per second
  - ~0.7% produce a beauty quark pair
  - A few tens of % of these hadronize to a $B^{\pm}$ meson
  - ~0.006% of these decay in «my» channel
- Several discriminating variables
- Strong correlation between them
- **Solution: Boosted Decision Tree (BDT)**

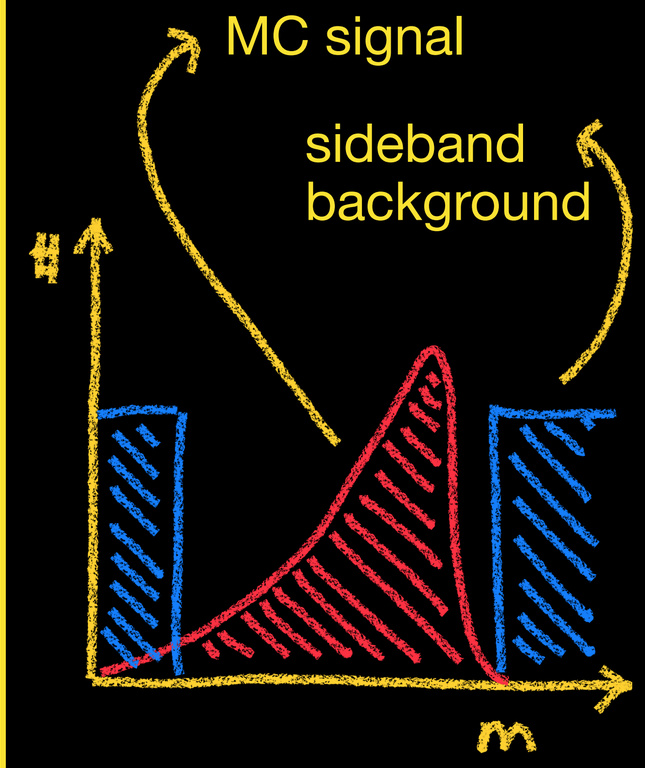


It worked. Until it didn’t.



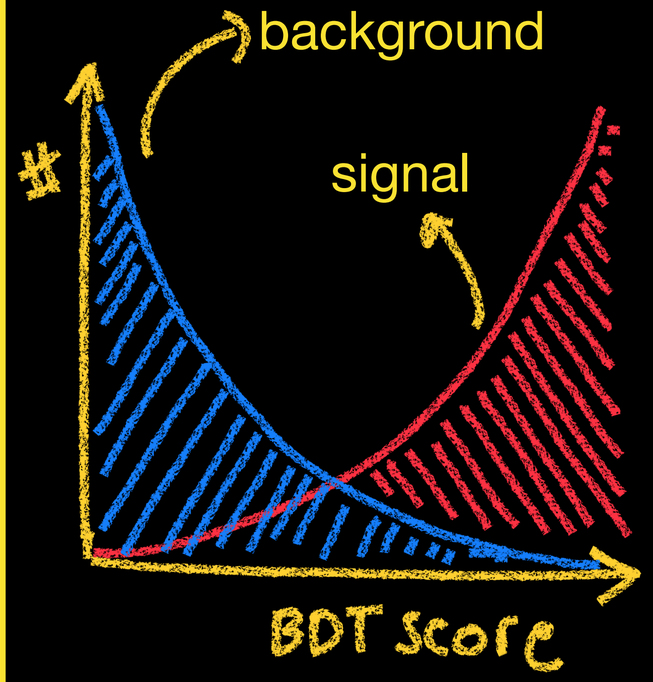
It worked. Until it didn't.

Train



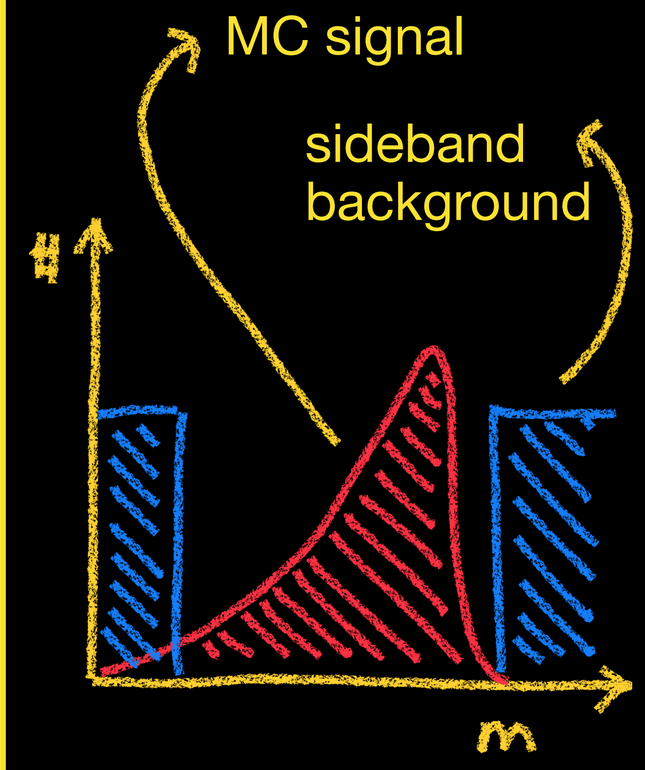
Promise

AUC = 0.99, large expected significance gain



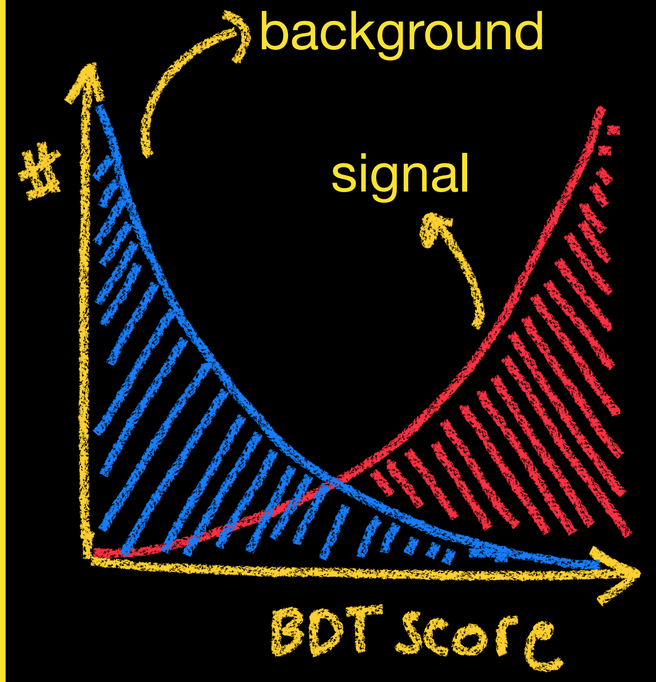
It worked. Until it didn't.

Train



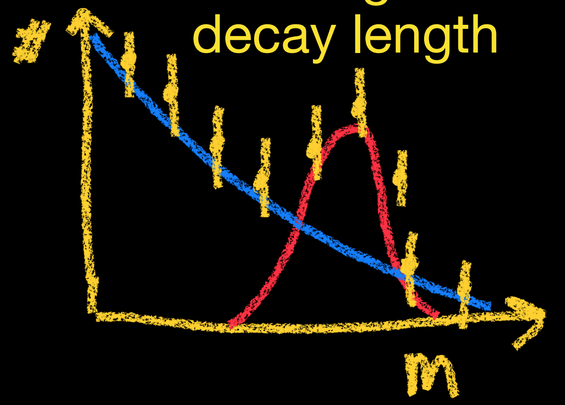
Promise

AUC = 0.99, large expected significance gain



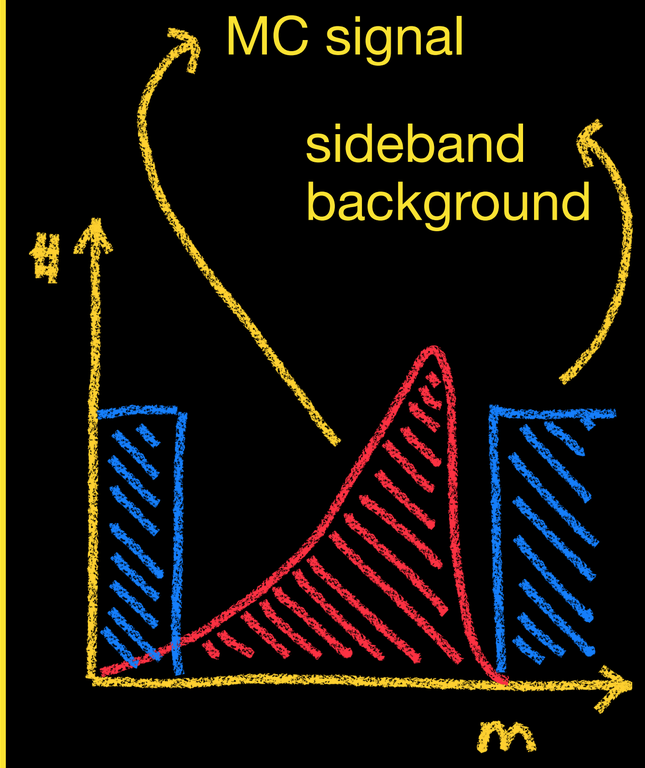
Reality

Rectangular cuts on decay length



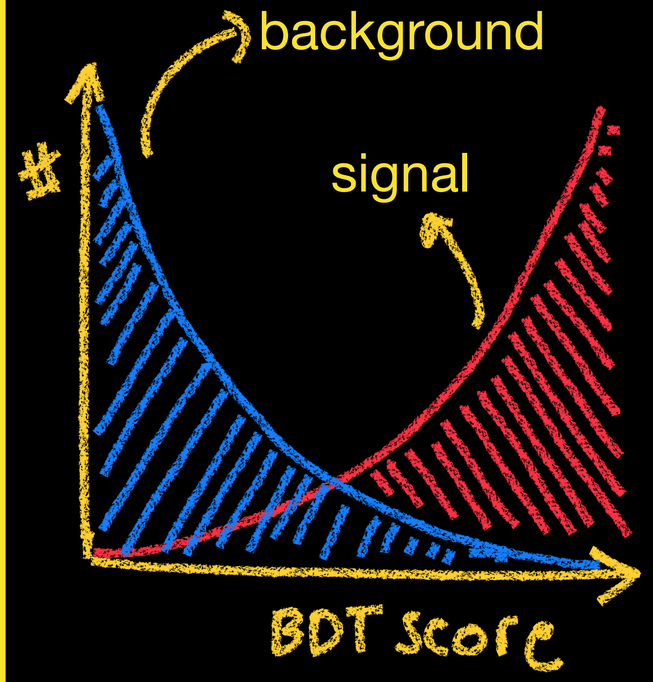
It worked. Until it didn't.

Train



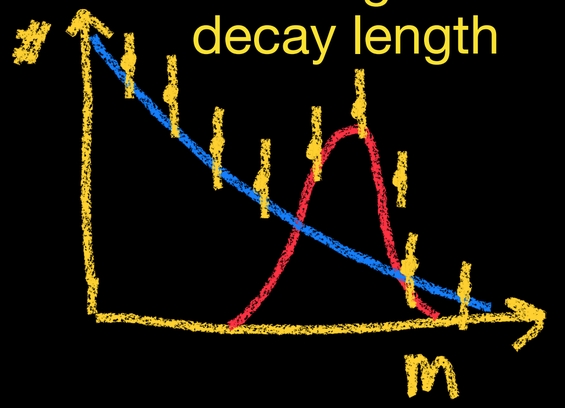
Promise

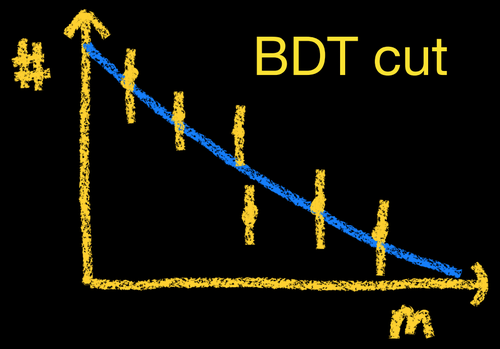
AUC = 0.99, large expected significance gain



Reality

Rectangular cuts on decay length





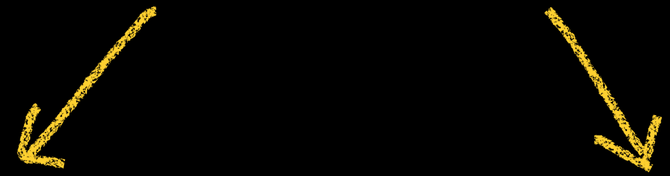
BDT works on test sample,
but fails on data

BDT works on test sample,
but fails on data

Signal/background
definition

Is MC signal
representative?

BDT works on test sample,
but fails on data



Signal/background
definition

Is MC signal
representative?

BDT works on test sample,
but fails on data

Signal/background
definition

?

Is MC signal
representative?

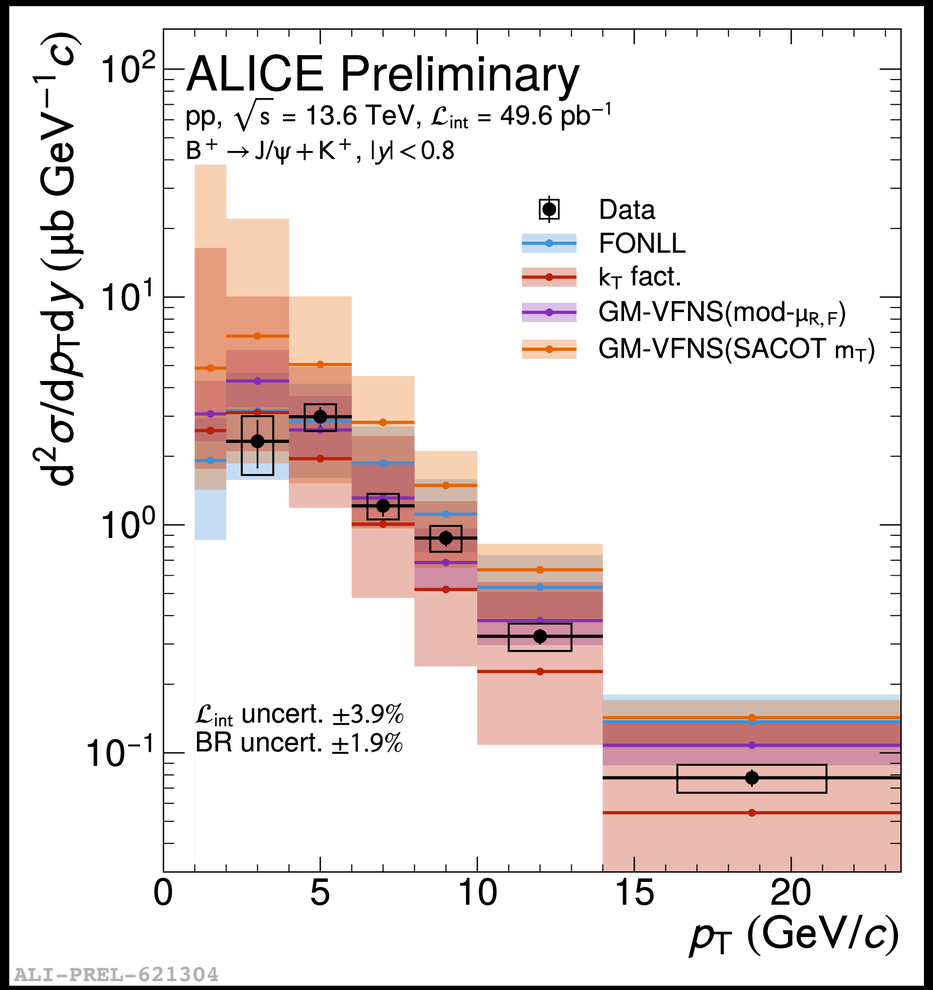
?

Rectangular cuts
work

PhD to finish

Was I right to stop?
What would you have tried next?

Backup



$$\sigma_{pp}^{H} = \sum_{a,b=g,q,\bar{q}} f_{a/A}(x_1,\mu_F^2) \otimes f_{b/B}(x_2,\mu_F^2) \otimes \hat{\sigma}_{ab\to Q\bar{Q}}(x_1,x_2,\mu_F^2,\mu_R^2) \otimes D_{Q\to H_Q}(z,\mu_F^2)$$