Machine learning Techniques in the Searches for Resonant Signatures at the LHC

Konstantinos Papadimos

Coordinates at the CMS

Given the solenoid geometry of the CMS detector, it is more convenient to use a spherical type of coordinates(r, ϕ, θ).

$$\begin{aligned} p_x &= P_T \cos\phi \\ p_y &= P_T \sin\phi \\ p_z &= P_T \sinh\eta \\ |\vec{P}| &= P_T \cosh\eta \end{aligned} \tag{1}$$

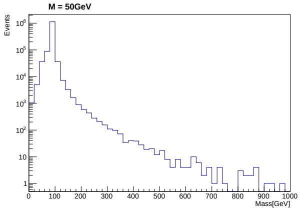
$\phi \in [0, 2\pi]$the azimuthal angle, and $\eta \in [-\infty, +\infty]$ is defined as:

$$\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{2}$$

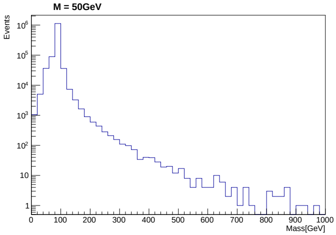
Decays & Resonances

Not every particle can be detected by the CMS detector(i.e neutrinos)

- Peak in the mass spectrum $\rightarrow$ Resonance



- Not a peak in the mass $\rightarrow$ Not a resonance



Calibration and energy scale uncertainties

- Calibration process adjusts energy scale and resolution to match well-known resonances (Z boson, J/psi meson) in data and simulation,
- Imperfect agreement due to subdetector complexities and nonlinear effects

How do analysis techniques respond to energy scale uncertainties ?

Our work will focus on the effects that energy scale uncertainties have, on a traditional fit-based analysis and a more modern Boosted Decision Tree-based analysis, using the generic diobject production process as the working example.

Classification techniques

In our case:

- Signal: a resonant decay Y->xx
- Background: a non resonant process

How to separate them?

- Boosted Decision Trees
- Fit based analysis

Searches for $Y \to XX$

Search for heavy $Y \to XX$

- Mass range from 100GeV up to 300GeV

Search for light $Y \to XX$

- Mass range from 50GeV up to 70GeV

The $Y \to XX$ channel

The specific characteristics(mass etc.) of each dataset is different but the main idea is the same

- Use a non resonant process for background –> Drell-Yan
- Use a resonant process for signal –> $W\Phi \to ll$
- Separate signal from background
- Apply energy scale uncertainties to signal
- Separate again
- Compare the nominal case with the smeared cases

Statistical interpretation of results

Are the signal events we counted, statistically significant?

- We use the following metric:

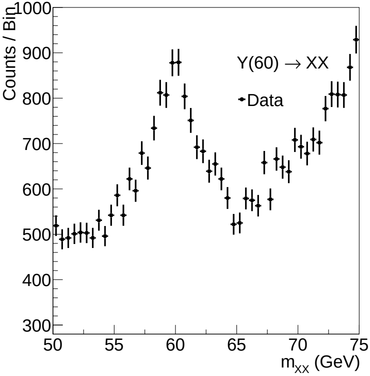
$$\text{Significance} = \frac{Signal}{\sqrt{Background}} \tag{3}$$

Search for light $Y \to XX$

We will study the following smearing cases:

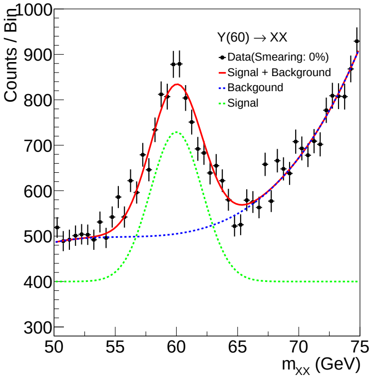
- 0%(Nominal case)
- 5%
- 7%
- 10%
- 12%

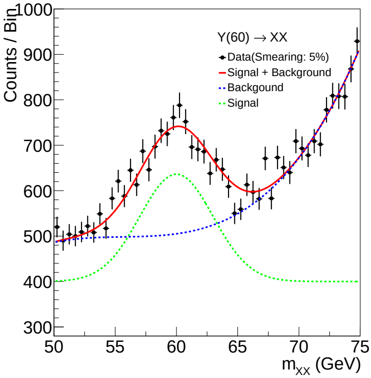
The working mass range is quite small –> smearing has a significant effect real quick



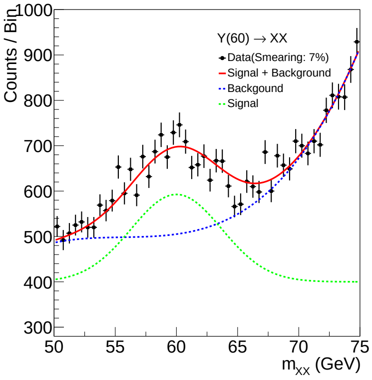
Fit based approach: Fitting

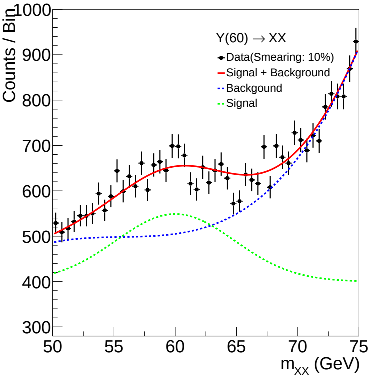
Then we proceed with the fits!





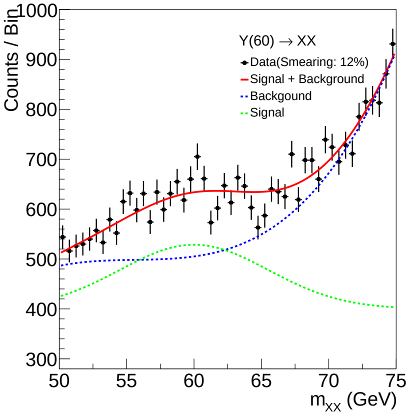
Fit based approach: Fitting





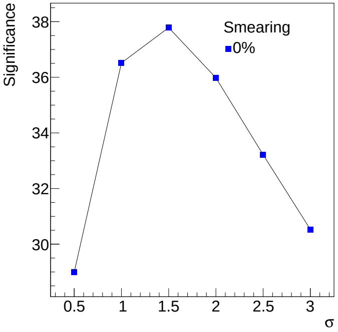
Fit based approach: Fitting

Any further smearing will make the signal indistiguishable!



Fit based approach: Signal from background separation

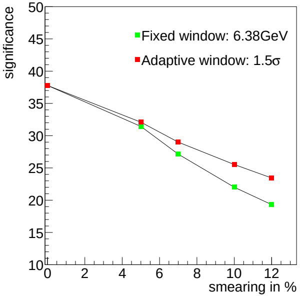
Working in the nominal case, we find the region that yields the best significance, by scanning the ranges $m = \pm\frac{n}{2}\sigma$, $n = 1, 2, 3, 4, 5, 6$



Fit based approach: Signal from background separation

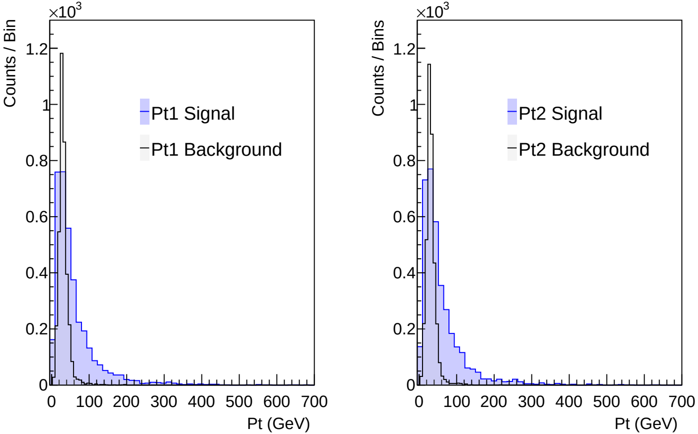
The region of interest that yields the best significance is the $\pm 1.5\sigma$. There are two ways to interpret this.

- interpret σ as the the spread of the nominal case –> fixed window
- interpret σ as the the spread of each cases –> adaptive window

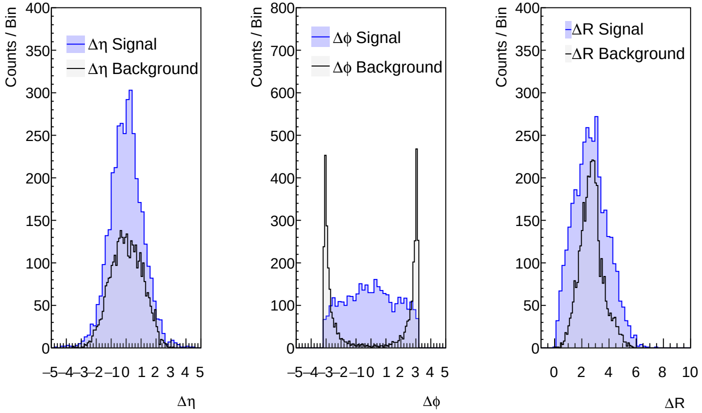


BDT approach: Feature space

What features of the dataset are best for the classification task?

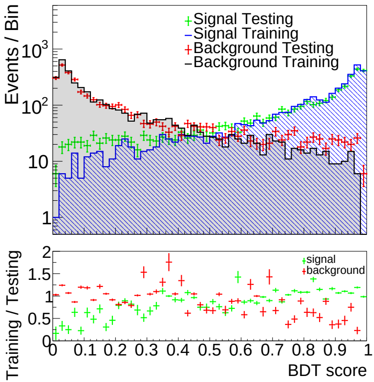


BDT approach: Feature space



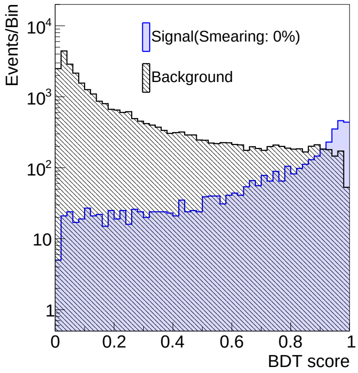
BDT approach: The model

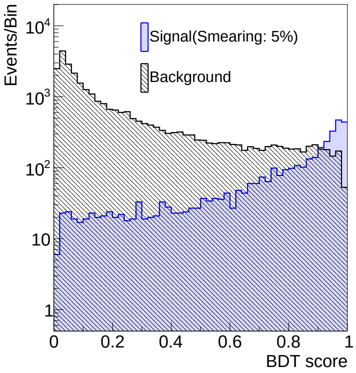
- Trained with approximately 3K events
- To examine overfitting we compare the ratio of training events to testing for each bdt score



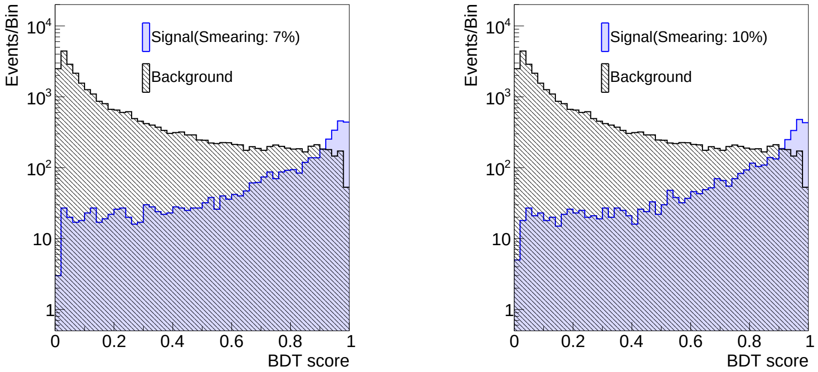
BDT approach: Application

Feed the application set to the BDT –> BDT plots

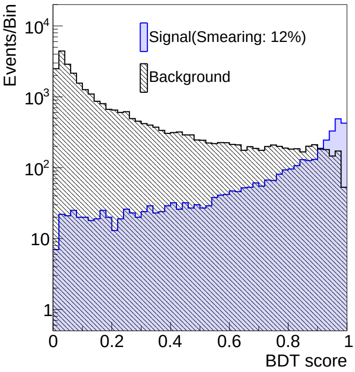




BDT approach: Application



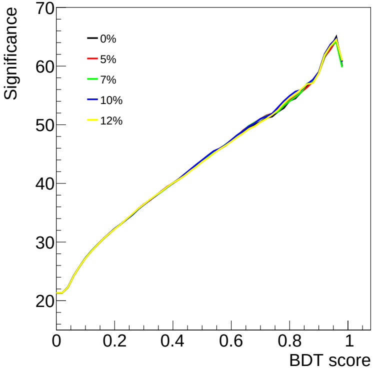
BDT approach: Application



BDT approach: Signal from background separation

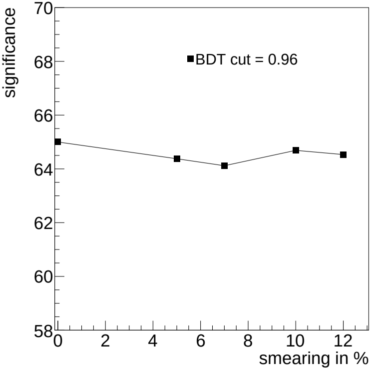
Where should we place the cut?

- Same philosophy as in the fit based search
- We scan the bdt range to find the best region of interest
- Best cut –> BDT score = 0.96.



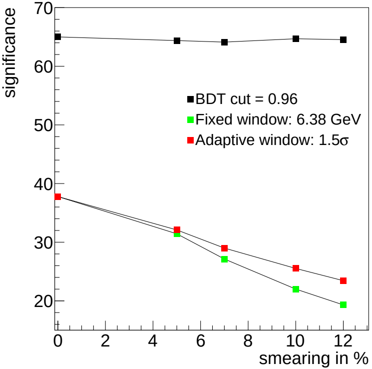
BDT approach: Signal from background separation

- The performance of the BDT remains invariant under energy scale uncertainties!



Synopsis

- BDT performs better than the fit based
- Remains invariant under smearing
- Performance of the fit drops



Search for heavy $Y \rightarrow XX$

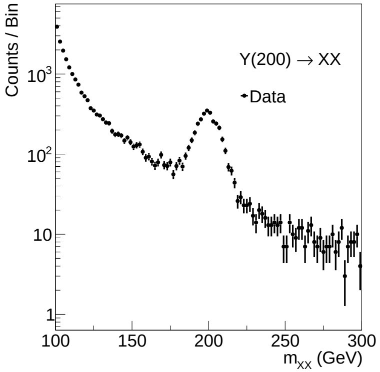
We will study the following smearing cases:

Medium to extreme cases

- 0%(Nominal case)
- 5%
- 10%
- 15%
- 20%

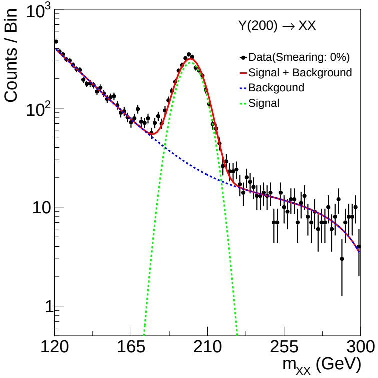
Plus some really extreme cases

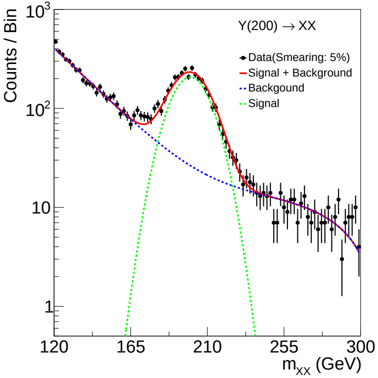
- 30%
- 40%
- 50%



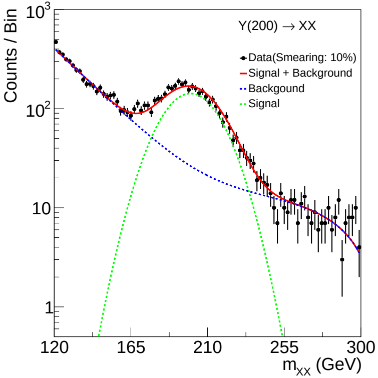
Fit based approach: Signal Fitting

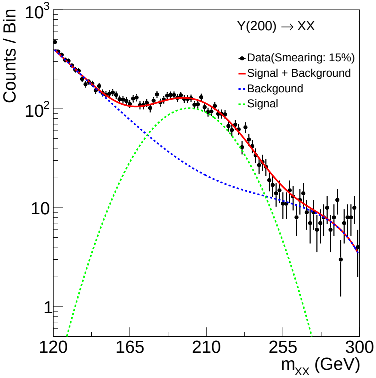
Then we proceed and fit the signal



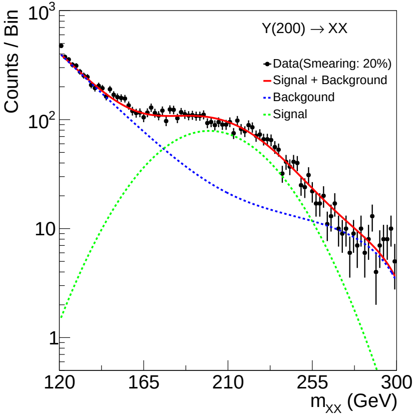


Fit based approach: Signal Fitting

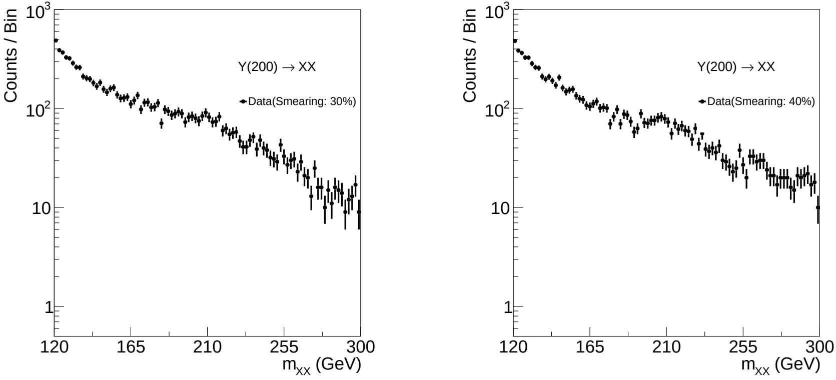




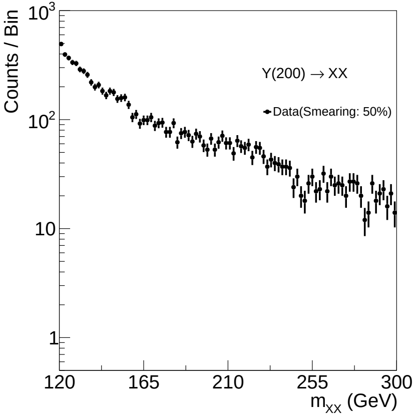
Fit based approach: Signal Fitting



Fit based approach: Signal Fitting

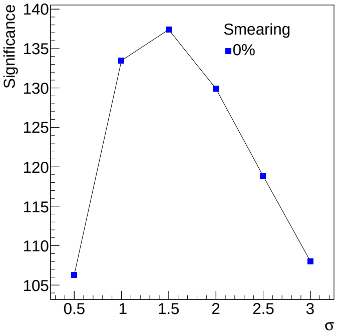


Fit based approach: Signal Fitting



Fit based approach: Signal from background separation

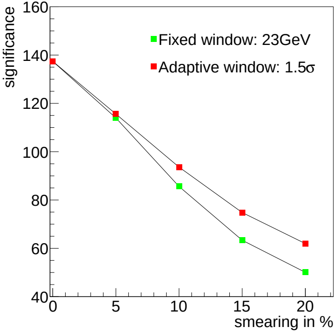
Working in the nominal case, we scan the ranges $m = \pm\frac{n}{2}\sigma$, $n = 1, 2, 3, 4, 5, 6$



Fit based approach: Signal from background separation

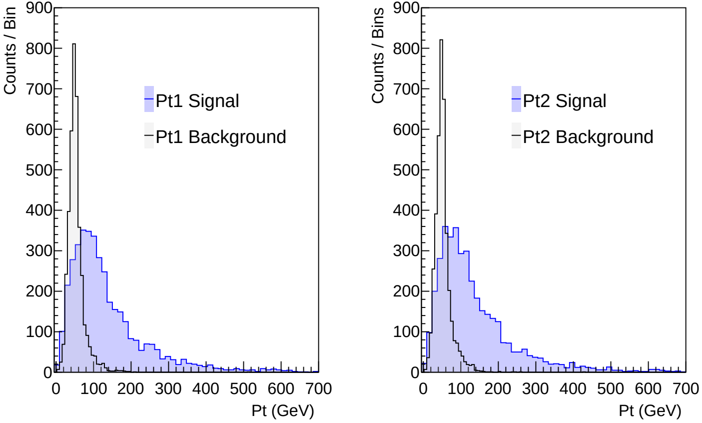
The best significance is in the $\pm 1.5\sigma$ range.

- fixed window
- adaptive window

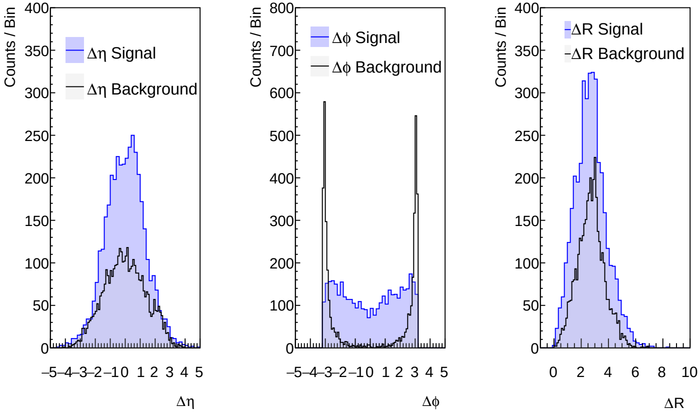


BDT approach: Feature space

We use the same feature space as with the light mass search

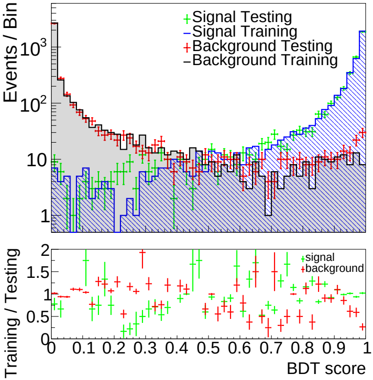


BDT approach: Feature space



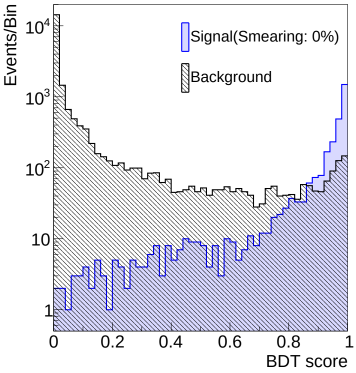
BDT approach: The model

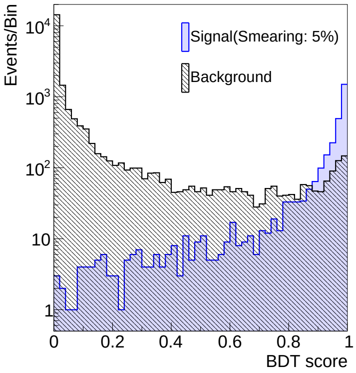
- Trained with approximately 3K events
- To examine overfitting we compare the ratio of training events to testing for each bdt score



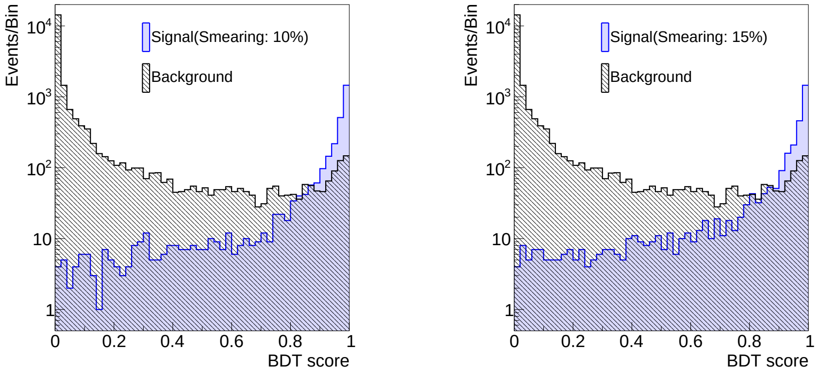
BDT approach: Application

Feed the application set to the BDT –> BDT plots

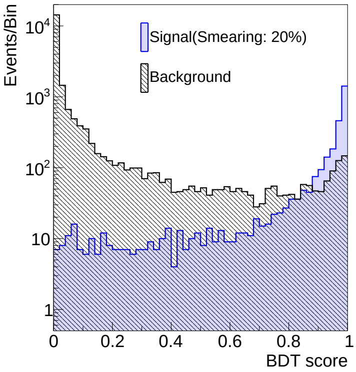


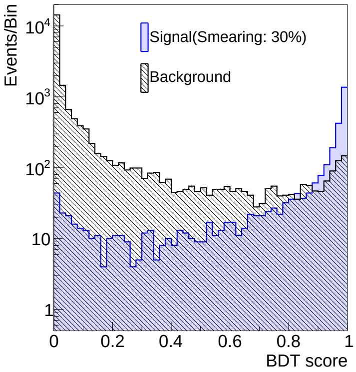


BDT approach: Application

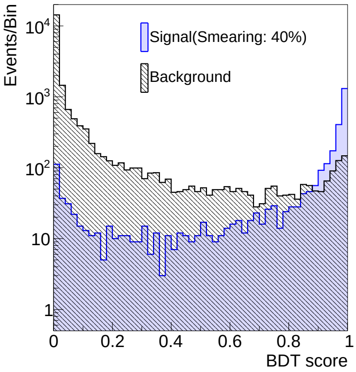


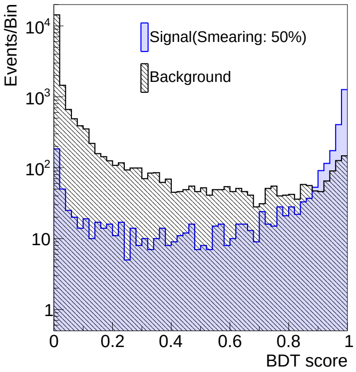
BDT approach: Application





BDT approach: Application

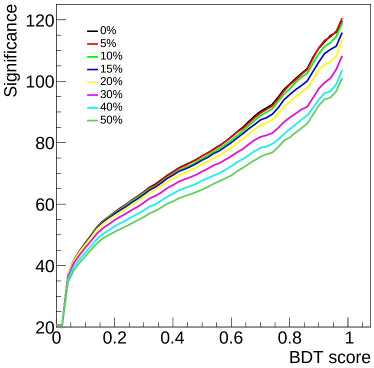




BDT approach: Signal from background separation

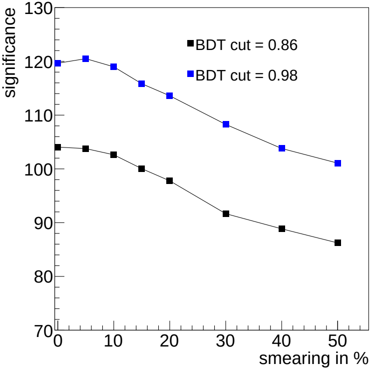
Where should we place the cut?

- We scan the whole bdt range to find the best region of interest
- Best cut –> BDT score = 0.98.
- This is rather tight, lets see what happens if we place a more relaxed cut at 0.86



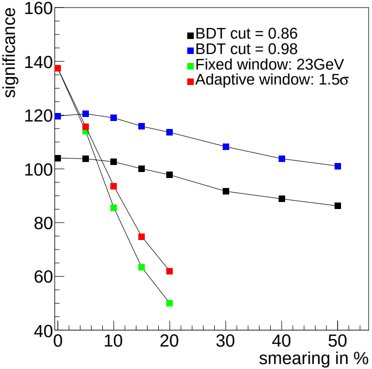
BDT approach: Signal from background separation

- The performance of the more relaxed cut is not as good as the best cut
- The bdt model is rather robust



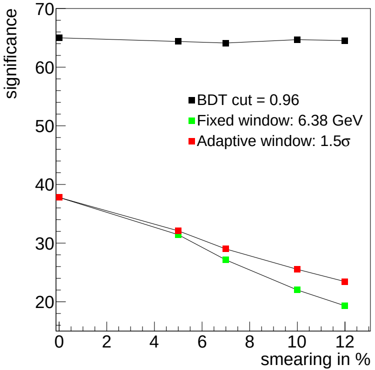
Synopsis

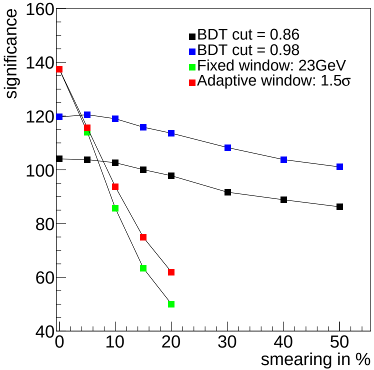
- The perfomance of the BDT and Fit are comparable when smeaing is mild
- Fit perfomance drops dramatically
- BDT is more robust



Results

- Light $Y \to XX$
- Heavy $Y \to XX$





Results

Overall, the BDT is more robust as it learns features that do not get affected by energy scale uncertainties

So is the BDT better?

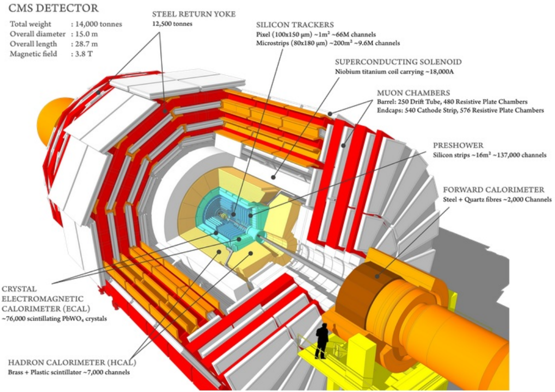
- No: A more carefull event selection can improve the performance of the fit based analysis
- yes: In the presence of energy scale uncertainties, the fit based analysis reaches a "breaking point"

Unused stuff

Welcome to the backup slides!

The CMS Experiment overview

The CMS detector at the LHC



Supervised Learning

Supervised learning:

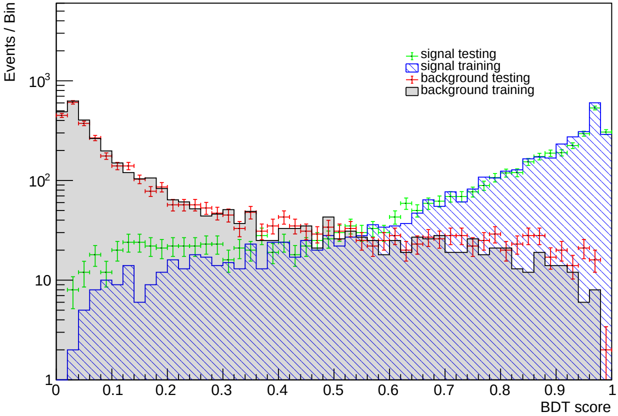
- The model is trained using training data
- The trained model is tested using testing data
- If we like the resulting model, we apply it!

but what is this model?

- A function that given the input feautres x, it returns the probability x beeing class A
- The goal of the training is to minimize the difference between the predicted output $y_i \in [0, 1]$ and the real output $\hat{y}_i = 0$ class B, or $\hat{y}_i = 1$ class A

BDT 3: Signal from background separation

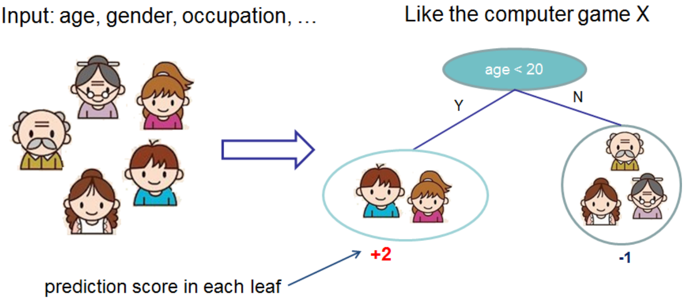
Where should we place the cut in order to accpet most most of the signal while rejecting most of background?



BDT 1a: Boosted decision trees

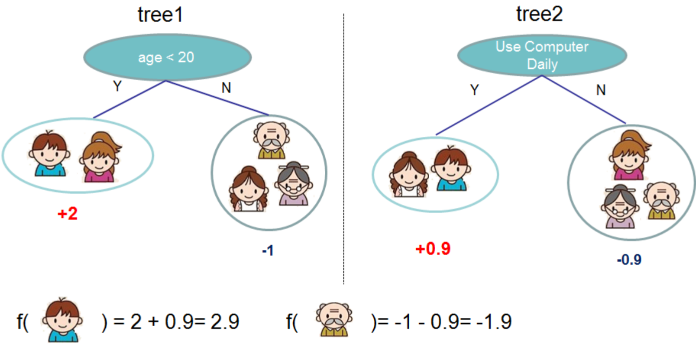
In this study the model of choice is Boosted Decision Trees(BDT).

- It classifies data using decision tree models



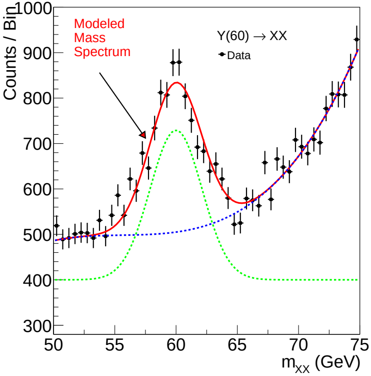
BDT 2b: Boosted Decision Trees

Usually only one tree is not power full enough –> Use more trees in additive manner(Boosting)



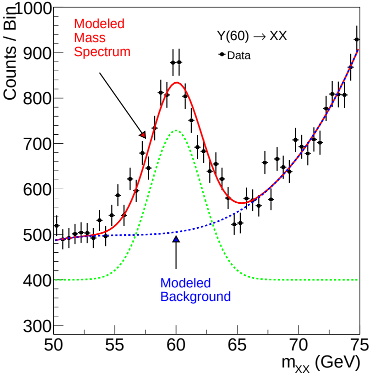
Fit based signal from background separation

Fit the mass spectrum . . .



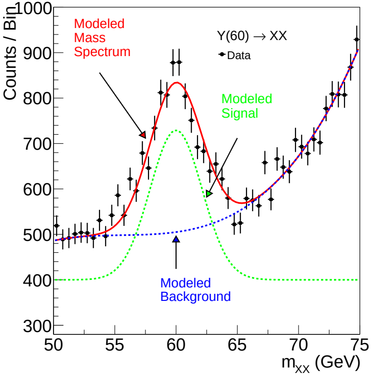
Fit based signal from background separation

. . . and decompose it to a background component . . .



Fit based signal from background separation

... and a signal component



Fit based signal from background separation

Then we can count the signal and background events, in a region of interest I:

$$O = \int_I observation(x)dx \tag{4}$$

$$B = \int_I bkg(x)dx \tag{5}$$

$$S = O - B \tag{6}$$

Energy scale uncertainties

To smear the data by x%,

- iterate over every signal event
- multiply each P_T by a number sampled from a Gaussian distribution of $\mu = 1$ and $\sigma = x/100$

The $Y \to XX$ channel: Background

- Drell-Yan process

The $Y \to XX$ channel: Signal

- W Φ -> ll

Search for light $Y \rightarrow XX$

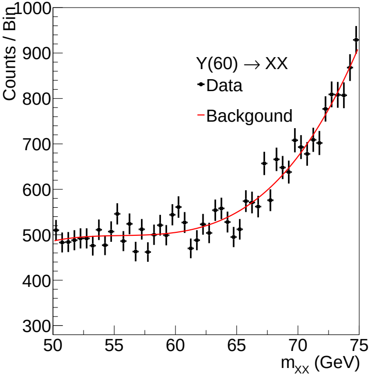
We will study the following smearing cases:

- 0%(Nominal case)
- 5%
- 7%
- 10%
- 12%

The working mass range is quite small –> smearing has a significant effect real quick

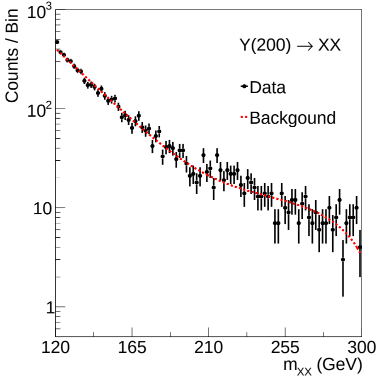
Fit based approach: Background Fitting light

- To simplify things a bit, we fit the background sepratelly
- The background shape is kept constant throughout the fits
- Shape: $\alpha + \beta x + \gamma x^2 + \delta x^3$



Fit based approach: Background Fitting

- The background shape is kept constant
- Shape: $\alpha + \beta x^{-1/2} + \gamma x^{-1} + \delta x^{3/2}$



Search for heavy $Y \to XX$

We will study the following smearing cases:

Medium to extreme cases

- 0%(Nominal case)
- 5%
- 10%
- 15%
- 20%

Plus some really extreme cases

- 30%
- 40%
- 50%