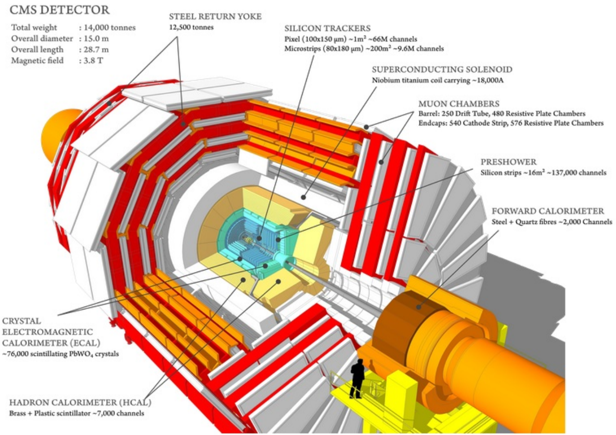
Presentation draft

Konstantinos Papadimos

The CMS Experiment overview

The CMS detector at the LHC



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).

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

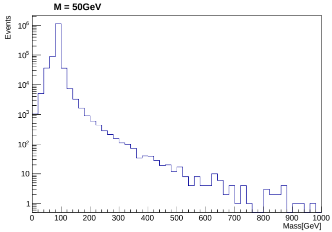
$\phi \in [0, 2\pi]$the azimuthal angle, and $\eta \in [-\infty, +\infty]$ is defined as:

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

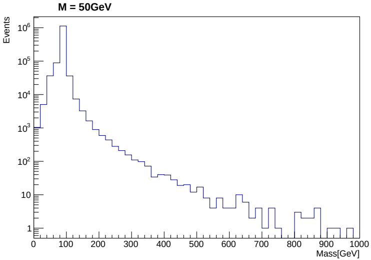
Decays & Resonances

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

- Detectable Decay Products $\rightarrow$ Resonance



- Non Detectable Decay Products $\rightarrow$ 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.

BDT 1: 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!

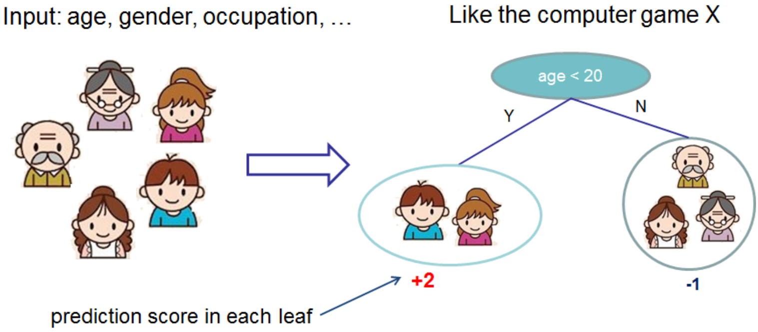
but what is this model?

- 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 2a: Boosted decision trees

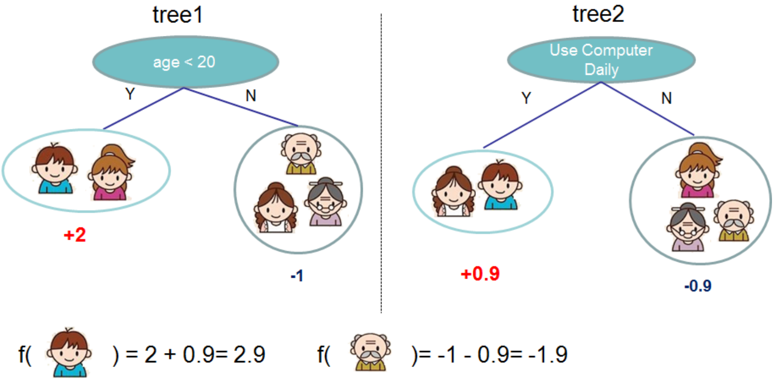
In this study the model of choice is Boosted Decision Trees(BDT).

- 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)



BDT 3a: Signal from Background Separation

In our case:

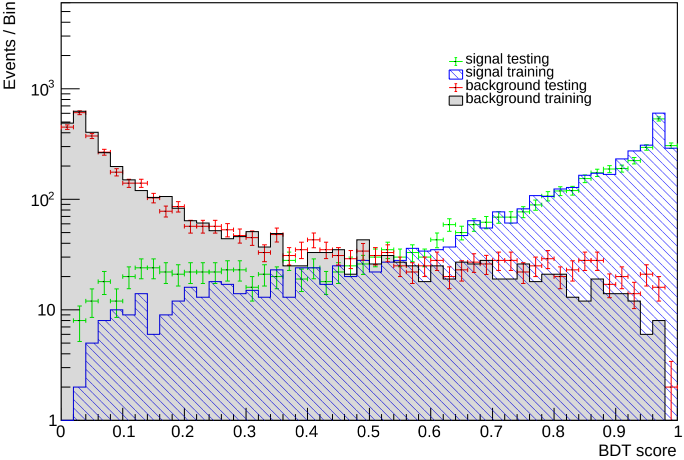
- Signal: a resonant decay Y->xx
- Background: a non resonant process

How to separate them?

- Plot the number of Signal and Background events per BDT score –> BDT histogram

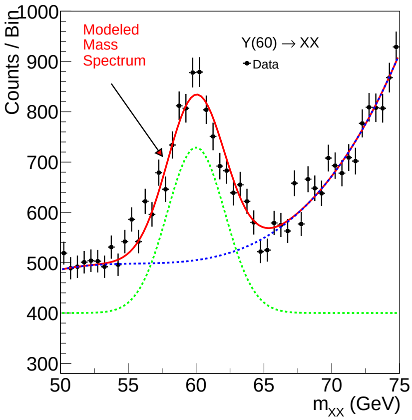
BDT 3b: Signal from background separation

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



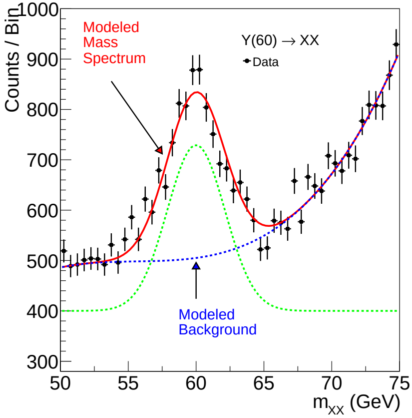
Fit based signal from background separation

Fit the mass spectrum . . .



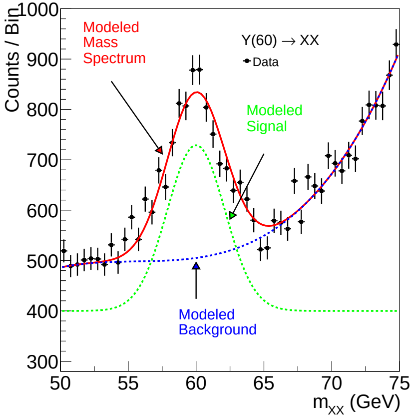
Fit based signal from background separation

. . . 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:

$$O = \int_I observation(x)dx \tag{3}$$

$$B = \int_I bkg(x)dx \tag{4}$$

$$S = O - B \tag{5}$$

Statistical interpretation of results

Are the signal events we counted, statistically significant?

- We use the following metric:

$$\text{Significance} = \frac{Signal}{\sqrt{Background}} \tag{6}$$

- The selected regions of interest both in BDT and Fit based analysis, are those that maximize the significance.

Searches for $Y \rightarrow XX$

Search for heavy $Y \rightarrow XX$

- Mass range from 100GeV up to 300GeV

Search for light $Y \rightarrow XX$

- Mass range from 50GeV up to 70GeV

The $Y \rightarrow XX$ 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
- Use a resonant process for signal
- Separate signal from background
- Apply energy scale uncertainties to signal
- Separate again
- Compare the nominal case with the smeared cases

The $Y \to XX$ channel: Background

- Drell-Yan process

The $Y \rightarrow XX$ channel: Signal

- W Φ –> ll

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 $\sigma = x/100$

Search for light $Y \to XX$

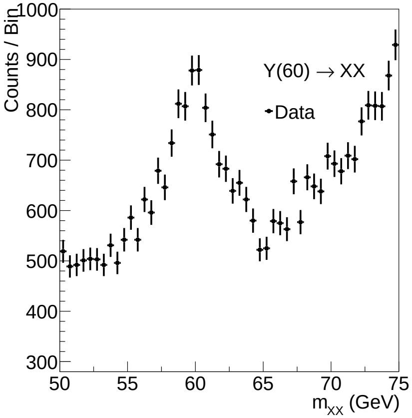
We will study the following smearing cases:

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

The number of events of the application set are quite low –> smearing has a significant effect real quick

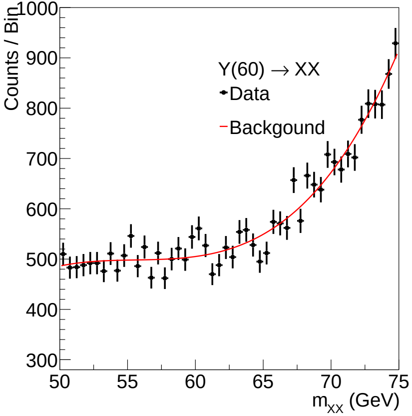
Fit based approach: The application set

- Signal events: 3K
- Background events: 30K



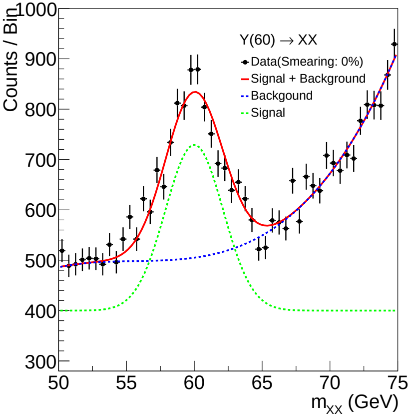
Fit based approach: Background Fitting

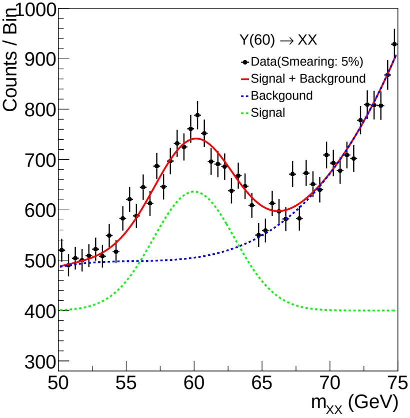
- To simplify things a bit, we fit the background sepratelly
- The background shape is kept constant throughout the fits
- Shape: $\alpha + \beta x + \gamma x^2 + \delta x^3$



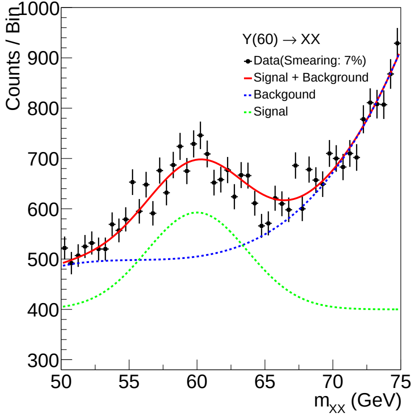
Fit based approach: Signal Fitting

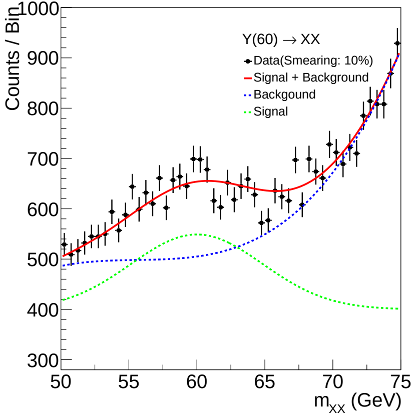
Then we proceed and fit the signal





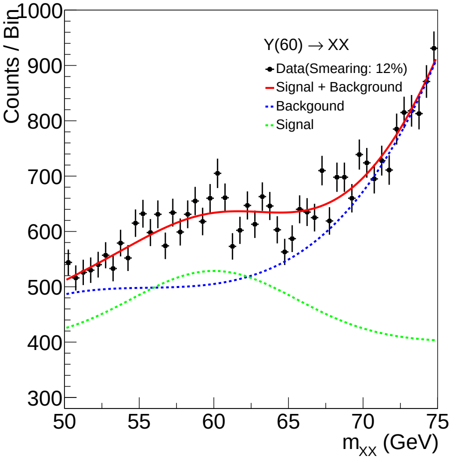
Fit based approach: Signal Fitting





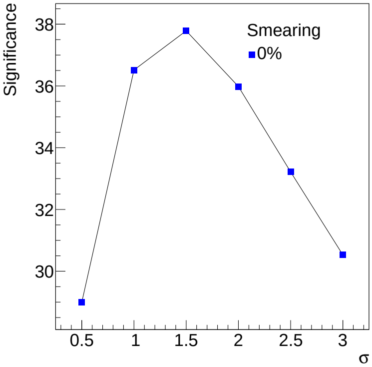
Fit based approach: Signal Fitting

Any further smearing will make the signal indistiguishable!



Fit based approach: Signal from background separation

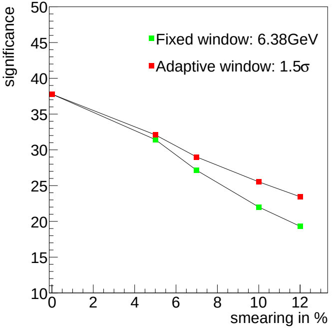
Working in the nominal case, we find the region that yields the best significance, by scanning the ranges $m = \pm\frac{n}{2}\sigma$, $n = 1, 2, 3, 4, 5, 6$



Fit based approach: Signal from background separation

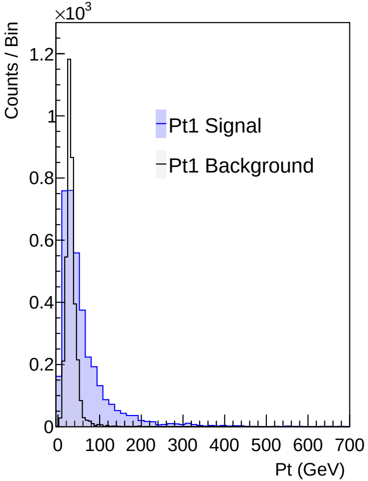
The region of interest that yields the best significance is the $\pm 1.5\sigma$. There are two ways to interpret this.

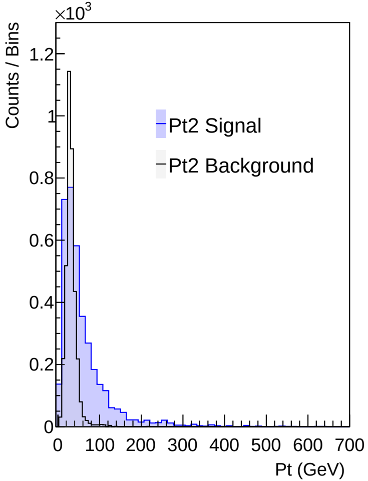
- interpret σ as the the spread of the nominal case –> fixed window
- interpret σ as the the spread of each cases –> adaptive window



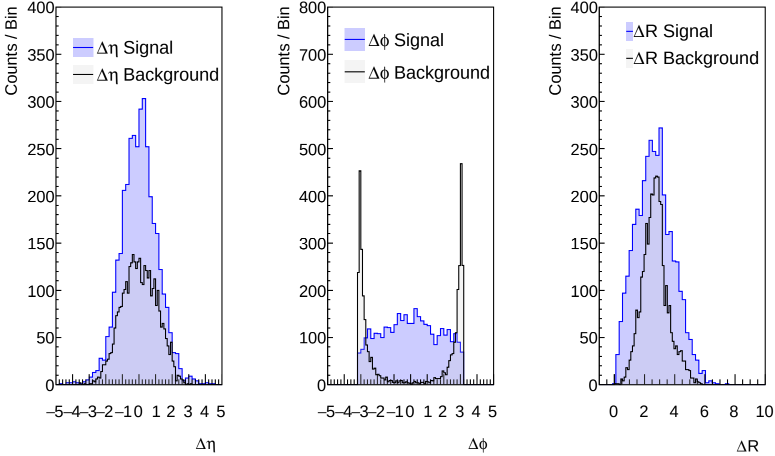
BDT approach 1: Feature space

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



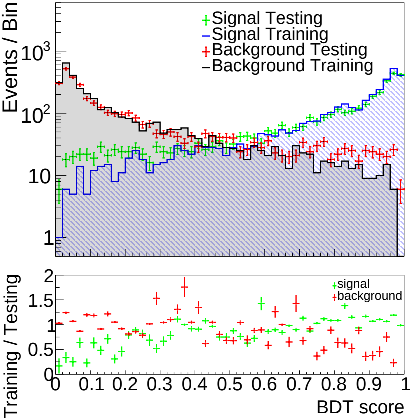


BDT approach1a: Feature space



BDT approach 2: The model

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



Results 1

Compare the BDT and FIt in terms of significance and robustness. Comment that even though fit based achieves a higher significance in the 0 smearing case, it is not as robust as bdt, it completely fails at extreme cases of smearing,. BDT is more robust

Results 2

Try to explain that bdt uses not only energy related features (Pts) but also geometrical ones, which do not get affected by smearing. Therefore, more stabillity to smearing. Nevertheless robustness does not mean greateer classification "power"(how many events got classified correctly and how manny didn't) –>Outlooks for better training methods in other to increase classification power.

Unused stuff

Welcome to the backup slides!

Resonance text

and therefore, the invariant mass calculation from the detected particles of such events will not result in a peak at the mass spectrum(Non resonant proces). Even though in decays where the poducts are detectable particles, the invariant mass calculation leads to a peak in the mass spectrum(resonant decays). In the present work we are interested in the later.