

Presentation draft

Kostas Papadimos

The CMS Experiment overview

Make a photo of the CMS detector and show where the various subdetectors are. Also explain what an event is

Coordinates at the CMS

The typical Pt Eta Phi coords etc.

Decays & Resonances

Explain that not every particle can be detected by the CMS detector(i.e neutrinos) and therefore, the invariant mass calculation from the detected particles of an event 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.

Calibration and energy scale uncertainties

Talk about the fact that the detector needs to be calibrated, and that calibration is done using resonances whose peak is well known. More over talk about the fact that usually MC results of the detector response usually deviates from the actual detector's response, thus one defines energy scale uncertainties. This is a good time to officially postulate the problem assessed in this work. $Y \rightarrow XX$ etc.

Explain what a bdt is. There is a nice example in XGBoost documentation.

Explain the train->test->Application workflow.

Talk about the output, explain what the BDT score is and what BDT histogram is. Discuss signal from background separation using bdt

Explain fit based signal from backgroun separation

Statistical interpretation of results

Talk about significance.

Energy scale uncertainties

How we implemented the smearing in our data set. How do we proceed from that, how many smearing cases.

BDT approach 1

Train Testing application set. Summarize the number of events.
Explain that in order to compare apples to apples, we will be analyzing the application set from now on.

BDT approach 2

Application summarize the results

Fit based approach 1

Show the mass spectrum that will be fitted

Fit based approach 2

discuss bkg fit is kept constant throughout the analysis. discuss signal fitting, show the plots(I will probably need more than one slide) at this part talk about the fact that after 20% the fit based technique fails.

Fit based approach 2

Present the significances.

Results 1

Compare the BDT and Flt 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 stability to smearing. Nevertheless robustness does not mean greater classification "power" (how many events got classified correctly and how many didn't) → Outlooks for better training methods in order to increase classification power.