(P)DNN vs inv Mass fit

What's new

- R33-24 e reader fixed for signal jets in the Resolved case
- Fair comparison of significance between inv Mass fit and score fit
  - See Alessandra & Giannini
- In the merged regime use also Maria Multi-class scores:
  - Classify large-R jets in H->bb, W/Z->qq, W/Z->jets, top, multijet

SIGNIFICANCE

- http://www.dmf.unisalento.it/~spagnolo/allow_listing/DBL/significance.html

Analysis Strategy

- complexity:
  - DNN score fit -> bkg score distribution depends on mass hypo under test
    - When deriving limits, for each mass hypo we need
      - A NN model
      - a template for the signal pdnn score
      - a template for the bkg pdnn score
  - (p)DNN score fit a single NN model (per macro region) but bkg score distribution still depends on mass hypo under test
    - When deriving limits, for each mass hypo we need
      - a template for the signal pdnn score
      - a template for the bkg pdnn score
- Invariant mass fit
  - When deriving limits we need a common bkg distribution + for each mass hypo under test
    - a template for the signal pdnn score
- Invariant mass fit after cut on (p)DNN score sculpt the invariant-Mass distribution

A NEW STRATEGY ??

- When computing a score of an event with a pDNN model,
  - We do:
    - score = NNOutput(mass-hypo; input variables) -> mass-hypo dependent

  - For each kind of signal (RSG/Radion/HVTWZ) and region/regime
    - Select N = 20-30 arbitrary values of the mass hypo (in the sensitivity range)
    - Compute $O = \sum_{i=1}^{N} NNOutput(\ m_i; \text{input var})$
      - ***O does not depend on m_i***
      - The distribution of O
        - for bkg events will peak at 0
        - For a signal event of mass close to *M* it will peak at 1 due to the components in the sum corresponding to $m_i \sim M$
          - The distribution for the signal depends on the signal mass, but is very similar for all values of signal masses (always a peak at 1, how much wide ?)
          -

A NEW STRATEGY ??

- For each kind of signal (RSG/Radion/HVTWZ) and region/regime
  - Select N = 20-30 arbitrary values of the mass hypo (in the sensitivity range)
    - Compute $O = \sum_{i=1}^{N} NNOutput(\ m_i; \text{input var})$
      - ***O does not depend on m_i***

- ***Estimate O for the background (once)***
- ***Estimate Oi for the signal at each (fully simulated) m_i value (with I=1, …k) we want to use in limit derivation (once for each mass)***
  - ***Interesting to see how similar/different Oi are for I = 1, …, K***

- ***A cut $O > O_{cut}$ selects signal events (what event the mass of the signal is) (*)***
- ***Events in the signal enriched region may be used for an invariant mass fit***
  - ***Sculpting ?***
    - ***Maybe limited (negligible) since signal of any mass is selected by (*)***