Increasing Higgs analysis accuracy with use of mass shape

Benno Kaech

ETH Zurich

June 12, 2020
Introduction

- Ask questions!
- Blue text → Backup
- Gauge symmetry vs short ranged weak force
- Higgs mechanism!
- Simplest description: Higgs field → Broken symmetry
- Physical field → excitations (i.e. particles)
- Existence would prove Higgs mechanism
- 2012 new particle discovered in LHC@Cern, 2013 discovery of Higgs announced
- Large Hadron Collider (LHC); world’s largest proton accelerator
- 4 larger detectors → Compact Muon Solenoid (CMS)
- Want mass independent classification
- Explore additional dimension during fit → Higher accuracy
Hypothesis testing

Counting experiments, events carry information

Figure of merit: significance $Z$

This is for discovery of a signal

Signal strength $\mu = \frac{\sigma_{obs}}{\sigma_{SM}}$

$\Rightarrow \begin{cases} 
\mu > > 1 & \text{disagreement} \\
\mu = 1 & \text{perfect agreement} \\
\mu << 1 & \text{disagreement}
\end{cases}$
- Higgs produced in Higgstrahlung
- Higgs decays to bottom quark pair
- 3 VHbb channels depending on decay W/Z boson (0, 1, 2 lepton channel) → Study 0 lepton channel
- \( Z = 4.8 \) in VHb\( \bar{b} \)
- Lifetime Higgs Boson \( 10^{-22} \) s → Reconstruct decay
- Reconstruction difficult → Monte Carlo simulation
- Multiple processes in collider
- Diboson contribution

Feynman diagram for the process investigated in my studies
Simulations compared to measured data
- Deep neural net (DNN) classifier
- Complexity demands DNN
- Fit $\mu$ on score
- Score $\rightarrow$ Probability for signal
- Use mass during classification $\rightarrow$ lose mass shape information
- Now: Expand fit, mass distribution shape
- Fit variables must not be correlated
Main investigation during first part of my studies

Seen in $m_{jj}$

**Sculpting**: background mass distribution resembles signal mass distribution

Excluding mass is not enough to stop this from happening

Solution: mass decorrelated/agnostic network

Quantize sculpting with $\chi^2$-distance

Scale Factors $\rightarrow$ additional uncertainty from mass correlation

Correlations are highly non linear!

In highest quartile of score, mass distribution of background looks like signal mass distribution
Methods of Decorrelation

- 3 methods, review simplest one
- Distance Decorrelation: Add DisCo coefficient $\rho_{\text{DisCo}}$ to loss function
- Like Pearson, but accounts non linear
- $\rho_{\text{DisCo}} = 0 \iff$ independent distributions
- Loss function minimized during training $\rightarrow$ Controls learning process
- Minimization of loss $\rightarrow$ Minimization of correlation
- **Adversarial decorrelation:** Score may not contain mass information
- **Disco Adversarial decorrelation:** Combination of both methods

Increasing Higgs analysis accuracy with use of mass shape
Diboson peak

- Same process, Z instead Higgs
- Leaves same signature as signal
- Peaks at different $m_{jj}$
- Mass decorrelated → Missclassify
- Used as criteria
- Sanity test: Exclude sample → Peak disappears
Results

- In depth optimization with **grid and random** searches
- Multiple figures of merit
- Multiple tests
- Removes sculpting completely
- Diboson peak visible, passes Sanity test
- Lose about 15% of classification accuracy
- This is expected!
- Gain back during fit
Insight in current Higgs analysis

Biggest experiment

Cutting edge statistical methods

Did not forget Physics

Successfully decorrelated neural net from a feature

Can be used in other cases

Now: 2D fit of the signal strength using mass and score

Never treated before!
\( \chi^2 \) defined as

\[
\chi^2 = \sum_i \frac{(n_i - m_i)^2}{\sigma_n^2 + \sigma_m^2}
\]

- \( n, m \) are the yields of the two histograms
- Note that a low \( \chi^2 \) is not a sufficient criteria for mass decorrelation
- For a \( \chi^2 > 10 \) we exclude a set network configuration

\( \chi^2 < 10 \), but still see slight sculpting

\( \chi^2 \rightarrow \) high \( \chi^2 \) → we clearly see sculpting
- If score independent from $m_{jj}$, it is not possible to estimate $m_{jj}$ of the score
- Add neural network to estimate the mass using just score as input
- Subtract the loss of this net from the loss function of the classifier:

$$L_{adv} = L_{clf} - \lambda L_{Reg}$$

- $\lambda$ is again used as a knob to regulate decorrelation and prediction accuracy
- Note that the minimization is not straight forward

$$\min_{\theta_{clf}} \max_{\theta_{reg}} L_{clf}(\theta_{clf}) - \lambda L_{Reg}(\theta_{reg})$$

- Mass regression network has few orders bigger learning rate
- Convergence issues
- More sensitive to small correlations in theory
Use both methods sequentially

- For m epochs:  \( L = L_{\text{crossentropy}} + \lambda_{\text{clf}} \rho_{\text{DisCo}} \)
- After m epochs:  \( L = L_{\text{crossentropy}} - \lambda_{\text{reg}} L_{\text{reg}} \)

- Note that \( \lambda_{\text{clf}}, \lambda_{\text{reg}} \) differ in few orders of magnitude
- Sensitivity to small correlations but still good convergence
Grid/Random searches

- Grid/Random Search: Optimize on set of hyperparameters
- Figure of merit: bin-by-bin Asimov significance applying the classification
- Use $\chi^2 > 10$ to exclude some configurations
- Grid search: all possible configurations of hyperparameters are tested
- Random searches converge more quickly
- Most optimal run: highest significance with "no sculpting" → Diboson peak

Example for a heatmap of the significance, this is from a grid search
Multiple grid searches, scanning \( \lambda \) and the total epochs \( n \)

For the best run a significance of 1.70 and the following mass distribution is obtained

At this point exclude method due to convergence issues in favor of DisCo

Note that we can not recognize the diboson peak
For DisCo Adversarial scanned $\lambda_{DisCo}$, $\lambda_{reg}$ and few others

Optimal run: significance of 1.88 obtained with the following mass distribution

Mass distribution if we include the Diboson Sample.
Note the excess of events around 90 GeV

Mass distribution if we remove the Diboson Sample

Distribution of scores for the Diboson sample and the signal sample

Increasing Higgs analysis accuracy with use of mass shape.