Higgs CP measurement potential with ML techniques and classifier evaluation with ROC and AUC

Jakub Zaremba, Kraków, 30.01.2019

Outline

1. Introduction
2. Possible Higgs CP state measurement using ML
3. Systematical error estimation using ML
4. Introduction to classifiers
5. Confusion matrix
6. ROC curves
7. AUC
8. Discussion

Introduction

In this talk i will be discussing results of:

1. **Potential for optimizing the Higgs boson CP measurement in H →$\tau\tau$ decays at the LHC including machine learning techniques** R. Józefowicz (Unlisted), E. Richter-Was (Jagiellonian U.), Z. Was (Cracow, INP), https://arxiv.org/pdf/1608.02609.pdf

2. **Deep learning approach to the Higgs boson CP measurement in $H\to\tau\tau$ decay and associated systematics** Elisabetta Barberio (Melbourne U.), Brian Le (Cracow, INP & Melbourne U.), Elzbieta Richter-Was (Jagiellonian U.), Zbigniew Was (CERN & Cracow, INP), Daniele Zanzi (Melbourne U.), Jakub Zaremba (Cracow, INP), https://arxiv.org/pdf/1706.07983.pdf

Possible Higgs CP state measurement using ML

- Possibility of mixed scalar-pseudoscalar Higgs state is still open
- Paper [1] discusses possibility of such measurement using decaj $H \rightarrow \tau\tau$ with subsequent tau decays into either rho$\rightarrow \pi\pi$ or a1$\rightarrow \pi\pi\pi$
- NN model uses open-source library tensorflow (M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al., Software available from tensorflow. Org 1(2015))

Possible Higgs CP state measurement using ML

Features/variables	Decay mode: $\rho^{\pm}-\rho^{\mp}$ $\rho^{\pm}\to\pi^0\,\pi^{\pm}$	Decay mode: $a_1^{\pm}-\rho^{\mp}$ $a_1^{\pm}\to\rho^0\pi^{\pm},\ \rho^0\to\pi^+\pi^-$ $\rho^{\mp}\to\pi^0\,\pi^{\mp}$	Decay mode: $a_1^{\pm}-a_1^{\mp}$ $a_1^{\pm}\to\rho^0\pi^{\pm},\ \rho^0\to\pi^+\pi^-$
$\varphi^*_{i,k}$	1	4	16
$\varphi^*_{i,k}$ and y_i, y_k	3	9	24
$\varphi^*_{i,k}$, 4-vectors	25	36	64
$\varphi^*_{i,k}$, y_i, y_k and m_i, m_k	5	13	30
$\varphi^*_{i,k}$, y_i, y_k, m_i, m_k and 4-vectors	29	45	78

Table 2: Dimensionality of the features which may be used in each discussed configuration of the decay modes. Note that in principle $y_i^{\pm}$, $y_k^{\mp}$ may be calculated in the rest frame of the resonance pair used to define $\varphi^*_{i,k}$ planes, but in practice, choice of the frames is of no numerically significant effect. We do not distinguish such variants.

Possible Higgs CP state measurement using ML

Features/variables	Decay mode: $\rho^{\pm}-\rho^{\mp}$ $\rho^{\pm}\to\pi^0\,\pi^{\pm}$	Decay mode: $a_1^{\pm}-\rho^{\mp}$ $a_1^{\pm}\to\rho^0\pi^{\mp}$, $\rho^0\to\pi^+\pi^-$ $\rho^{\mp}\to\pi^0\,\pi^{\mp}$	Decay mode: $a_1^{\pm}-a_1^{\mp}$ $a_1^{\pm}\to\rho^0\pi^{\pm}$, $\rho^0\to\pi^+\pi^-$
True classification	0.782	0.782	0.782
$\varphi^*_{i,k}$	0.500	0.500	0.500
$\varphi^*_{i,k}$ and y_i, y_k	0.624	0.569	0.536
4-vectors	0.638	0.590	0.557
$\varphi^*_{i,k}$, 4-vectors	0.638	0.594	0.573
$\varphi^*_{i,k}$, y_i, y_k and m_i^2, m_k^2	0.626	0.578	0.548
$\varphi^*_{i,k}$, y_i, y_k, m_i^2, m_k^2 and 4-vectors	0.639	0.596	0.573

Table 3: Average probability p_i that a model predicts correctly event x_i to be of a type A (scalar), with training being performed for separation between type A and B (pseudo-scalar).

Systematical error estimation using ML

- Models for a1 decay vary drastically, therefore question of their impact on measurement was posed
- This problem was explored in second paper

Systematical error estimation using ML

Features				STD	RχL	ALT	BBR
ϕ^*	4-vec	y_i	m_i				
$a_1 - \rho$ Decays							
✓	✓	✓	✓	0.604	0.604	0.603	0.603
✓	✓	✓	-	0.597	0.596	0.596	0.597
✓	✓	-	✓	0.604	0.604	0.604	0.604
-	✓	-	-	0.597	0.596	0.596	0.595
✓	✓	-	-	0.597	0.596	0.596	0.595
✓	-	✓	✓	0.593	0.593	0.593	0.593
✓	-	✓	-	0.582	0.579	0.580	0.578
$a_1 - a_1$ Decays							
✓	✓	✓	✓	0.567	0.563	0.564	0.564
✓	✓	✓	-	0.560	0.555	0.557	0.556
✓	✓	-	✓	0.568	0.564	0.566	0.566
-	✓	-	-	0.562	0.557	0.559	0.559
✓	✓	-	-	0.562	0.557	0.559	0.559
✓	-	✓	✓	0.547	0.546	0.547	0.545
✓	-	✓	-	0.537	0.534	0.535	0.533

Table 2. Area under ROC curve. NN trained with $a_1 - a_1$ decays of $\tau\tau$ system with standard CLEO current on exact MC sample. These NN are then tested on MC generated sample with alternative parameterisations of the hadronic currents.

Introduction to calssifiers

Classifier is as operator/program/working scheme that allows object to be classified as a member of specific group.

Terminology originated in medicine (healthy/ill classification) but is very useful for evaluation of algorithms recognizing patterns e.g. faces

Confusion matrix

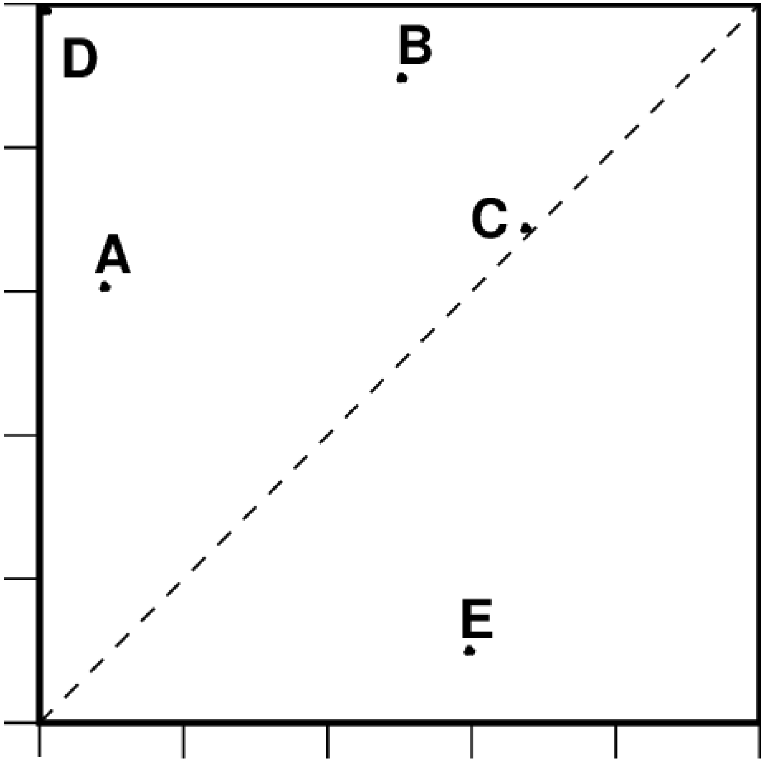
True values	Predicted values		
	-	+	
-	True negative	False positive	Cn
+	False negative	True positive	Cp
	Rn	Rp	N

- Traditionally confusion matrix was used to fully describe used classifier.

With values from confusion matrix we can define:
Accuracy: (Tp+Tn)/(Cp+Cn)
Sensitivity: Tp/Cp
Specificity: Tn/Cn
Precision: Tp/Rp

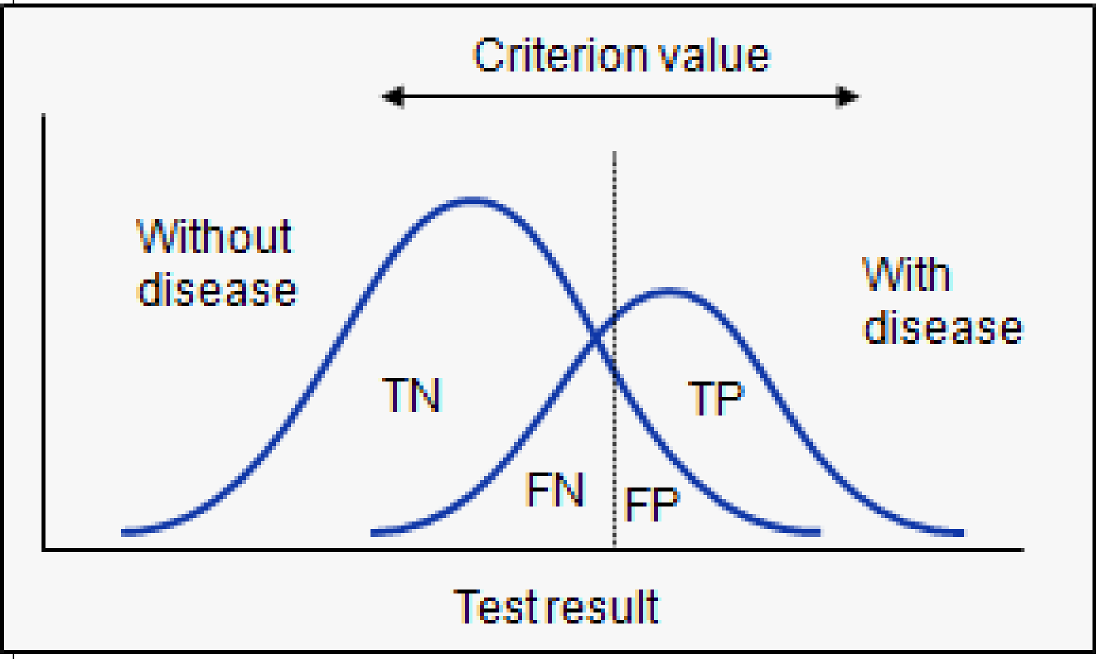
ROC curves

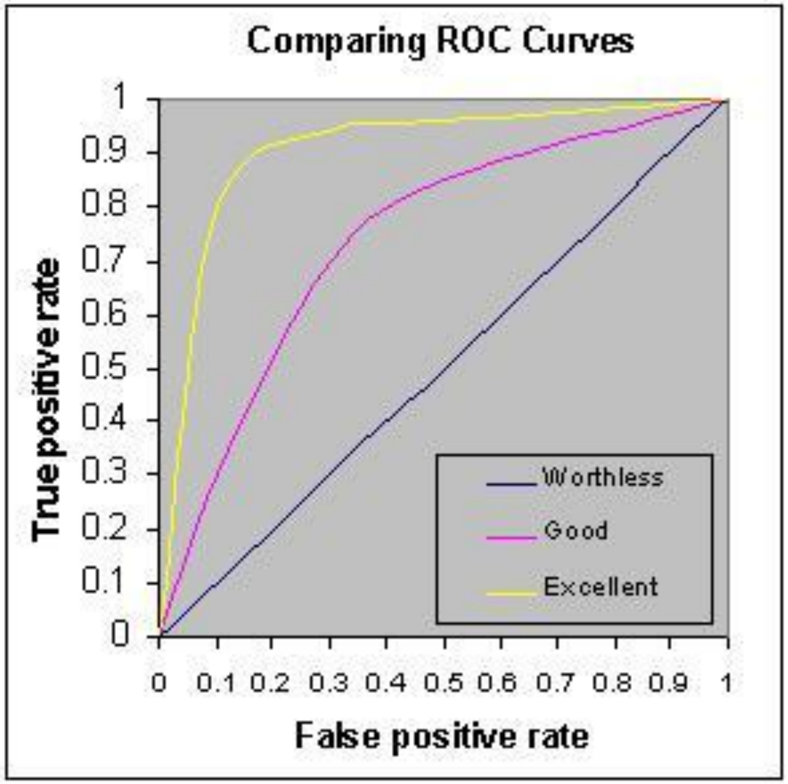
- Use of confusion matrix when we want to compare many classifiers may get… confusing.
- ROC (receiver operating characteristics) graphs are a nice solution
- For binary classifiers we get points in ROC space - Fp rate (x axis) vs Tp rate (y axis)



ROC curves

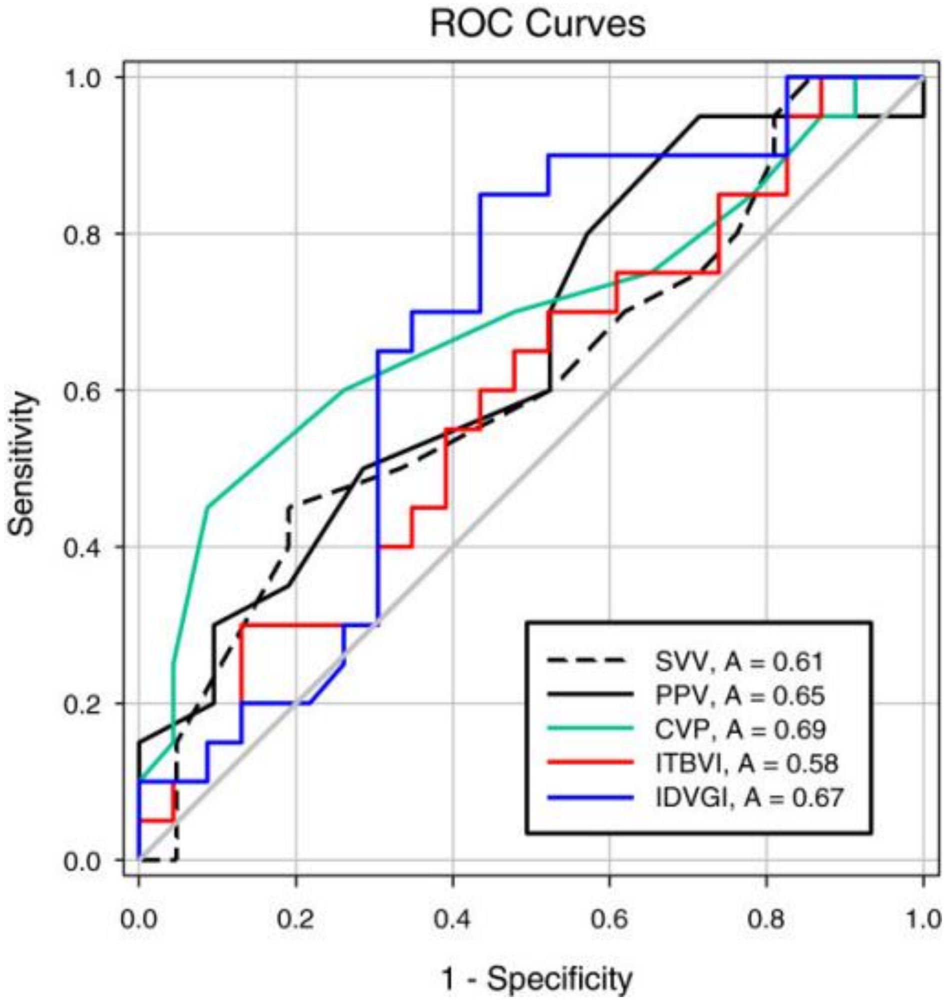
- ROC curves can be obtained for scoring clasifiers
- Classifiers emerging from NN by nature are scoring classifiers





AUC

- AUC = area under ROC curve



Discussion

- When there is more than two classes of objects ROC analysis is usually performed for each possible class as 1 vs all.
- It is possible to merge classifiers but approach is case specific.
- Note that ROC curve shows in fact how classifier ranks positive instances relative to the negative instances, so AUC can be interpreted as probability that positive instance will get higher score than negative one.

Discussion

- Higher AUC does not necessarily means classifiers is better for your purpose