Output Correction in HEP using DPGMM

Adil Omari[1],
Roberto Días-Morales[2],
and Juan J. Choquehuanca-Zevallos[2],

[1]Universidad Autonoma de Madrid

[2]Universidad Carlos III de Madrid

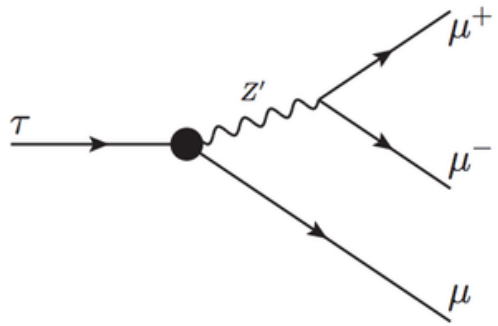
NIPS : Applying (machine) Learning to Experimental Physics Workshop

December 11th, 2015

Outline

Introduction

- In the recent days, machine learning has been shown to be an indispensable tool in several fields, among them in the field of high energy physics.
- A main contribution from machine learning field to HEP is to help to discover the structure of matter, and so, we could list some examples:
- In addition, important forward steps in this direction are the Higgs Boson Machine Learning Challenge [1] and the recently ended Flavours of Physics: Finding $\tau \rightarrow \mu\mu\mu$ challenge [2], both hosted by Kaggle.



Flavours of Physics Challenge

- Challenge description
  - Pysics Beyod Standard Model (BSM) allows new theoretical considerations, such as neutrino oscillations, that allows flavour violations that are not supported in classical models.
  - This challenge is a quest of searching the $\tau \rightarrow \mu\mu\mu$ decay.

Datasets

- Datasets
  - A set of real and simulated data from the LCHb experiment is used to train models.
  - three separate sets are used to test classifiers. They are:
    - test data set
    - Agrement data set
    - Correlation data set
- Metrics:
  - Weighted Area Under the ROC Curve (WAUC): general score to measure models
  - Cramer-vom Mises (CvM) test: to ensures estimations are not correlated with mass estimation.
  - Kolmogorov-Smirnov (KS) test: to ensure small discrepancies between real and simulated output distributions.

Proposal

Our approaches:

- first approach: all features, good WAUC, fails on the KS test.
- second approach: remove some features, worst WAUC, pass the test.
- so, why don't combine them?

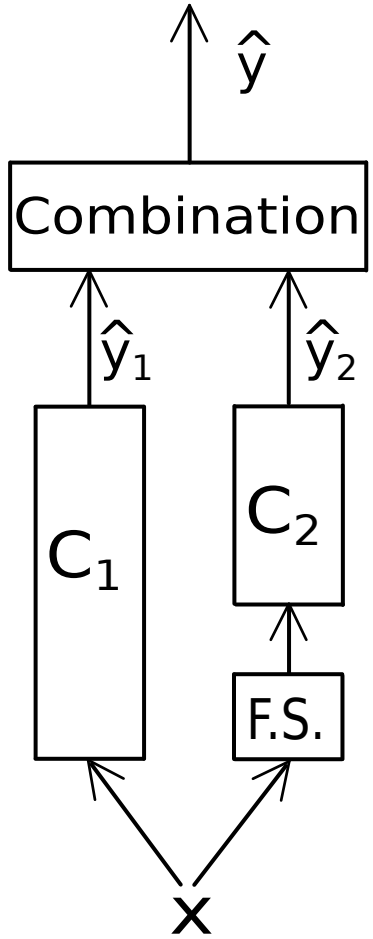


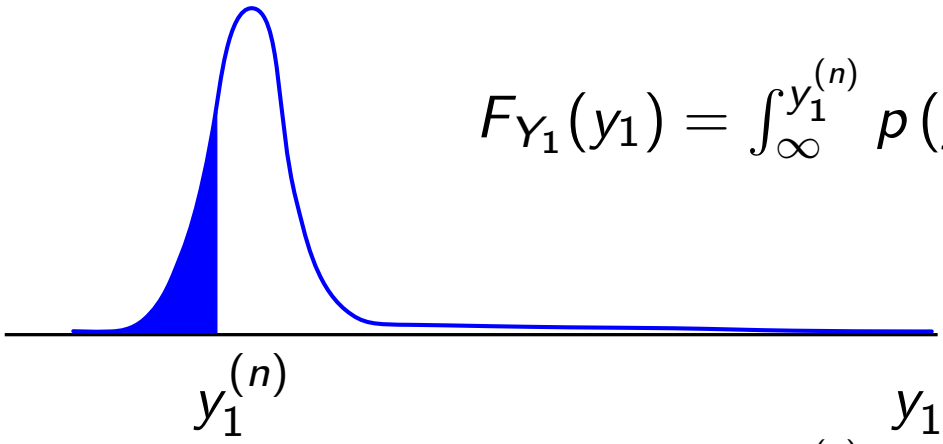
Figure: Combination of the two approaches

Combining heterogenous classifiers

- Problem: classical combination rules (i.e. weighted sums or products) do not help to accomplish the tests requirements (in particular KS test).
- we consider to main corner stones:
  - A good separability of classes is desired: the main idea is to maintain discriminative power of first block.
  - Meeting KS requirement is mandatory: for that, the idea is to detect regions in output space in where second block can act, helping to meet KS test of first block.

- Our proposal is to use the folling rule:

$$\hat{y}^{(n)} = \beta_r y_1^{(n)} + (1 - \beta_r) F^{-1}_{Y|C1}\left(F_{Y|C2}\left(y_2^{(n)}\right)\right) \quad (1)$$



Density estimation.

- Density estimations of $p(y^{(n)}|C2)$ and $p(y^{(n)}|C1)$ can be overcome by several algorithms, but many of them needs exhaustive hyperparameter searches.
- In general, we can utilize a GMM:

$$p\left(y|\pi,\mu,\Sigma\right) = \sum_{i=1}^{K} \pi_i \mathcal{N}\left(y|\mu_i,\Sigma_i\right) \tag{2}$$

where K is the number of components of the mixture and $\pi_i \in [0,1]$ are called the mixing proportions satisfying $\sum_i \pi_i = 1$ for $i = \{1,..,K\}$.

Results

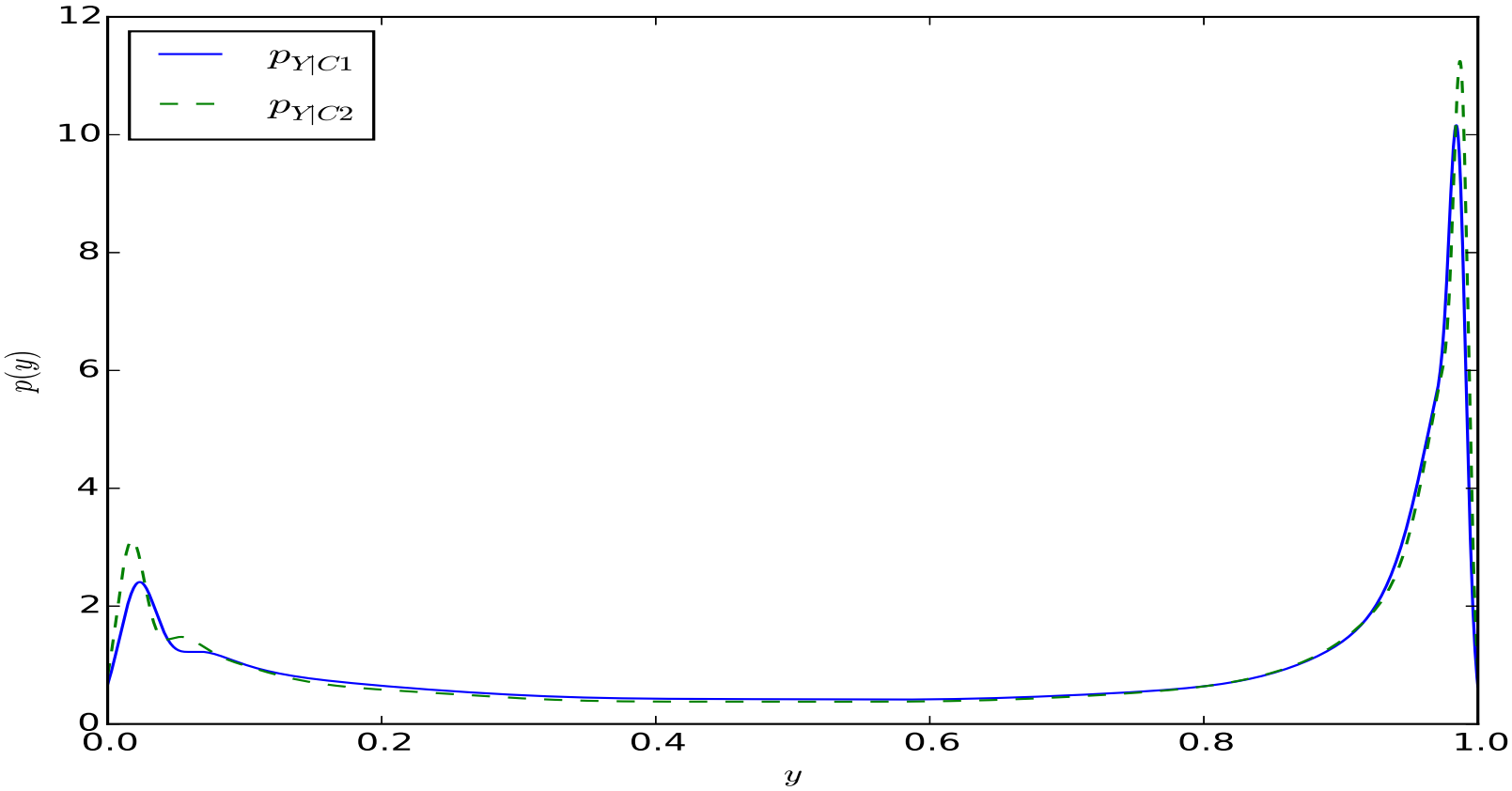


Figure: output (train data) pdf estimations for blocks 1 and 2 with DPGMM algorithm.

Results

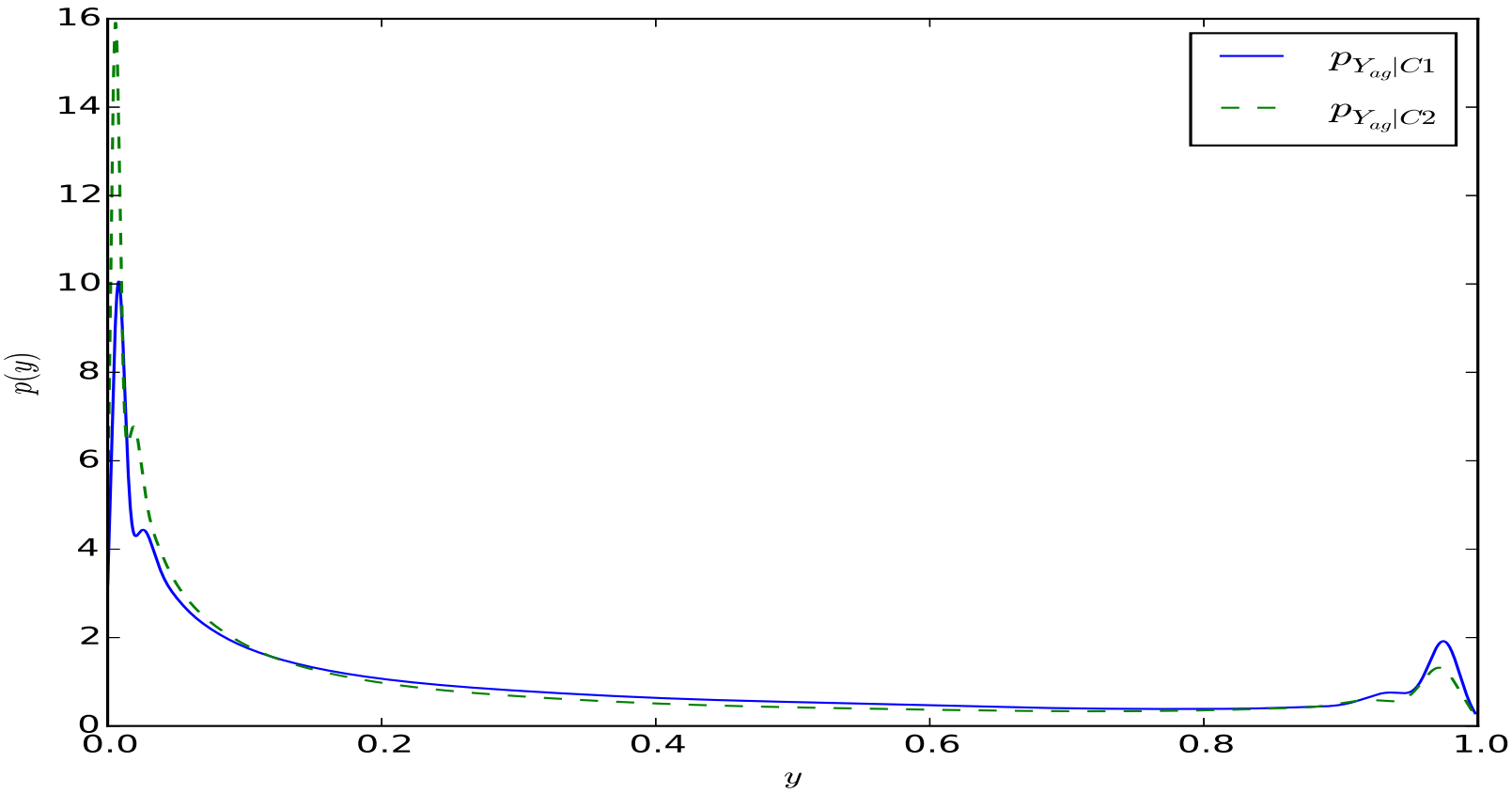


Figure: output (check agreement data) pdf estimations for blocks 1 and 2 with DPGMM algorithm.

Results

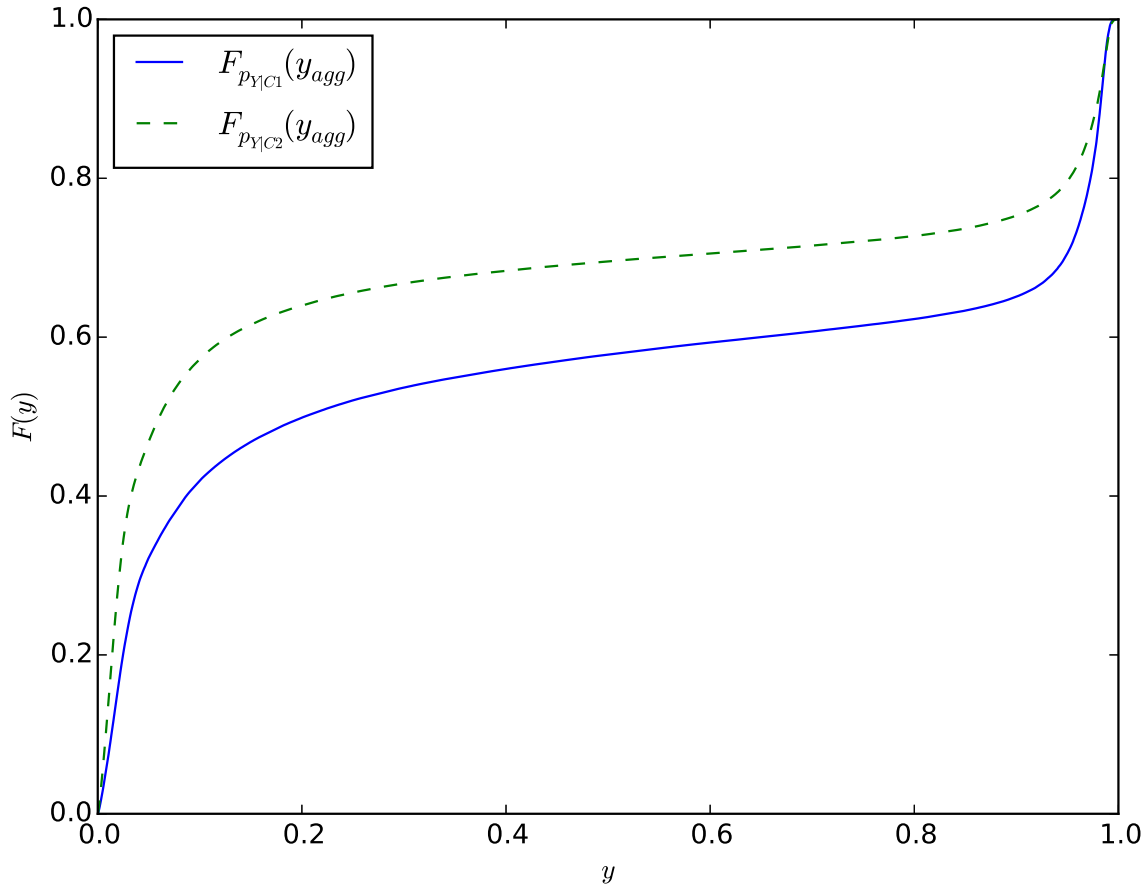


Figure: Cumulative pdf estimations of check agreement data output for blocks 1 and 2.

Results

Results on Validation dataset:

	WAUC	CvM test	KS test
Block 1	0.9881	0.001 (True)	0.202 (False)
Block 2	0.9828	0.001 (True)	0.084 (True)
Fusion	0.9830	0.001 (True)	0.086 (True)

Tabla: Results on validation data set.

Results on Test dataset:

	WAUC
Block 1	-
Block 2	0.988411
Fusion	0.98872

Tabla: Results on test data set.

Conclusions and future works

- Conclusions
  - We have presented a method to combine classifiers with different output pdf.
  - We have tested this combining method in Flavours of Physics Challenge.
  - Our results show that improvements can be obtained when applying our approach.

- Future works
  - Applying the proposed method in a general combination for heterogeneous ensemble.
  - Replace the combining weight β_r by a function on $\mathbf{x}$, $\beta(\mathbf{x})$ using algorithms like [3] [4].

Acknowledgements

My gratitude to:

- The organizers.
- Centro de Computación Científica (CCC) at UAM.
- Spain's grants TIN2013-42351-P and S2013/ICE-2845 CASI-CAM-CM.
- The Cátedra UAM–ADIC in Data Science and Machine Learning.

References

- "HiggsML. Higgs Boson Machine Learning Challenge," 2014. [Online]. Available: http://www.kaggle.com/c/higgs-boson
- "Flavours of Physics Challenge," 2015. [Online]. Available: https://www.kaggle.com/c/flavours-of-physics
- A. Omari and A. R. Figueiras-Vidal, "Feature combiners with gate-generated weights for classification," *IEEE Trans. Neural Networks and Learning Sys.*, vol. 24, pp. 158–163, 2013.
- A. Omari and A. R. Figueiras-Vidal, "Post-aggregation of classifier ensembles," *Information Fusion*, vol. 26, pp. 96–102, 2015.