Flavours of Physics: third place solution

NIPS, 2015

Josef Slavíček

Original and simplified solutions

This presentation describes the simplified solution:

- It contains all principal ideas
- It is easily understandable
- Runs faster
- Scores only slightly worse than the original one
- It is still sufficient to achieve the third place in the competition

Simplified solution architecture

Construct two classifiers:

- «Strong» - gradient boosted decision trees (GBDT)
- «Weak» - underfitted GBDT

Combine them into final model:

- $q * \text{Strong}^{\text{exponent}} + (1 - q) * \text{Weak}$
- q, exponent are parameters

«Strong» classifier

- input preprocessing (remove nSPDHits):
  - produce features
  - PCA for decorrelation and normalization (without dimensionality reduction)
- high weighted ROC score (~0.9993)
- fails both the Agreement (0.1479) and the Correlation tests (0.0751)

«Strong» classifier

- The huge failure in the Correlation test suggests that the GBDT probably found way to reconstruct particle mass and predictions are based on it
- Simplified solution doesn't use manually reconstructed particle mass as an input variable (while the original solution does)

«Weak» classifier

- construct it without data preprocessing
- passes both the Agreement and the Correlation tests
- any classifier easily passed the both tests can be used as a weak classifier (even random one)
- the original solution uses small neural nets (NN)

Final model

- For **exponent** = 1 formula calculates weighted arithmetic average:
  - prediction passes the both tests for the small values of **q** (~0.1)
  - achieve relatively high ROC score of 0.9968 (~7th place)
- For **exponent** >> 1
  - much bigger values of **q** can be used still passing the tests
  - achieve the same ROC score as the «Strong» classifier alone

Exponent trick and Agreement test

- On the training and test data predictions of the «Strong» classifier are very close to either zero or one
- On the Agreement data predictions of the «Strong» classifier are very close to zero
- These «almost zero» predictions are distributed in such way that the Agreement test fails
- For big values of **exponent** the «Weak» predictions dominate in the final formula

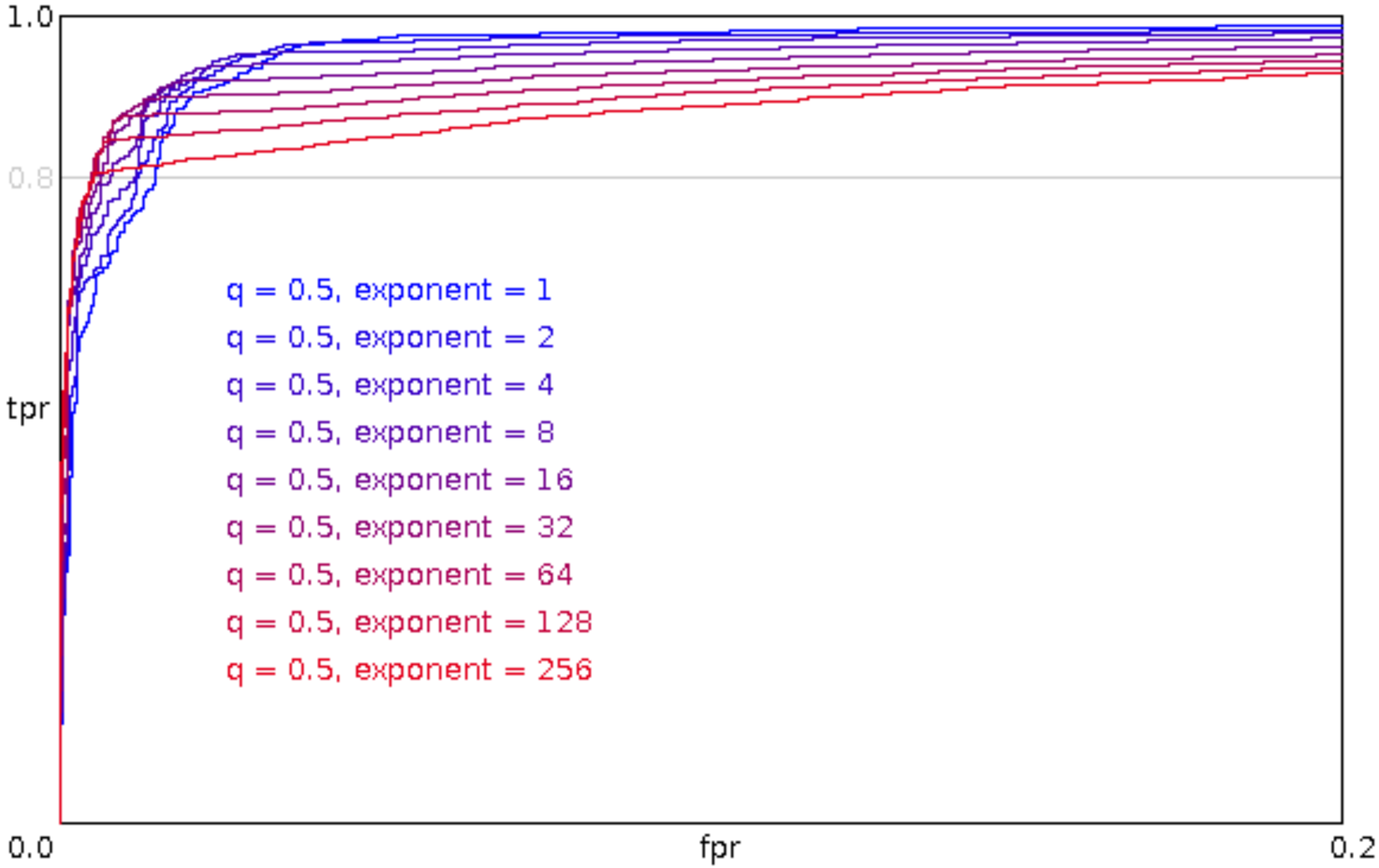
Exponent trick and Correlation test

- The mechanism is similar
- Predictions for the Correlation data are not always close to zero
- For more than 92% data predictions will be less than 0.98
- **Exponent** = 256 (used in the competition) gives the «Weak» predictions domination to pass the Correlation test

Exponent trick and evaluation metric

- Shape of ROC curve is determined by the order of data points sorted by prediction values
- Because of weighted AUC, only data points with the biggest predictions are important
- The strong classifier's predictions will dominate for big **exponent** during AUC score calculation even for smaller values of **q**

Exponent trick and evaluation metric



Original solution

- simple estimator for mass reconstruction
- estimated mass is used as additional feature for «Strong» classifier
- «Weak» classifier is implemented as a bunch of several small neural networks (NN):
  - single hidden layer of 64 neurons
  - all neurons logistic
  - learned by backpropagation

Conclusions, words of defense

- Presented solution is scored well in the competition but it's probably useless — it wouldn't work in practice
- Unfortunately, instead of doing something useful, presented solution only found simple way to bypass the tests introduced by competition organizers
- I'm sad for it and I didn't it intentionally

Useful links

- Competition description
- Detailed description of the simplified version of the third place solution
- Source code of the simplified version of the third place solution, runnable as a script in the Kaggle environment
- Description of both the original and the simplified versions of the third place solution