Weekly report

FANGYI GUO

Machine learning

Input sample

- Signal: VBF Higgs $H \rightarrow \gamma\gamma$, Nevents~50k
- Background: continuum $\gamma\gamma$+jets sample, Nevents~50k
- All events weight = 1 (XGBoost would return 0 if sample_weight is added)
- Training : test = 7:3

Input variables:
"pTt_yy","m_jj","DeltaEta_jj","DeltaPhi_yy_jj","Zepp","minDeltaR_y_j"

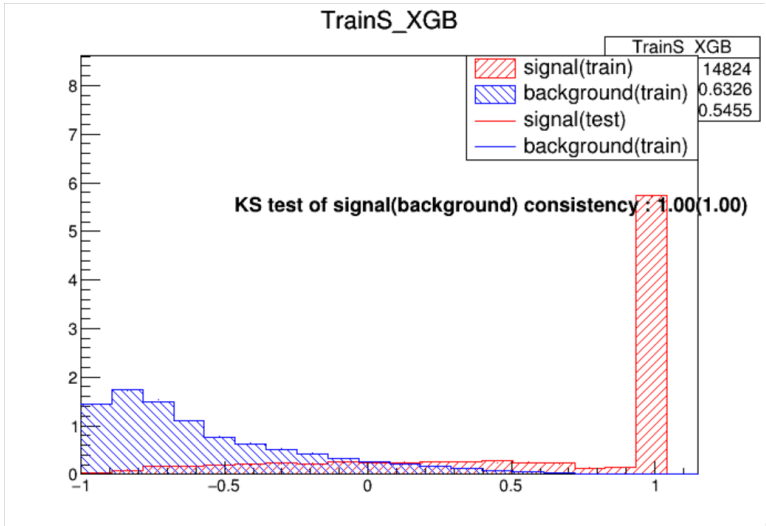
Training method: XGBoost, BDTG, AdaBoost

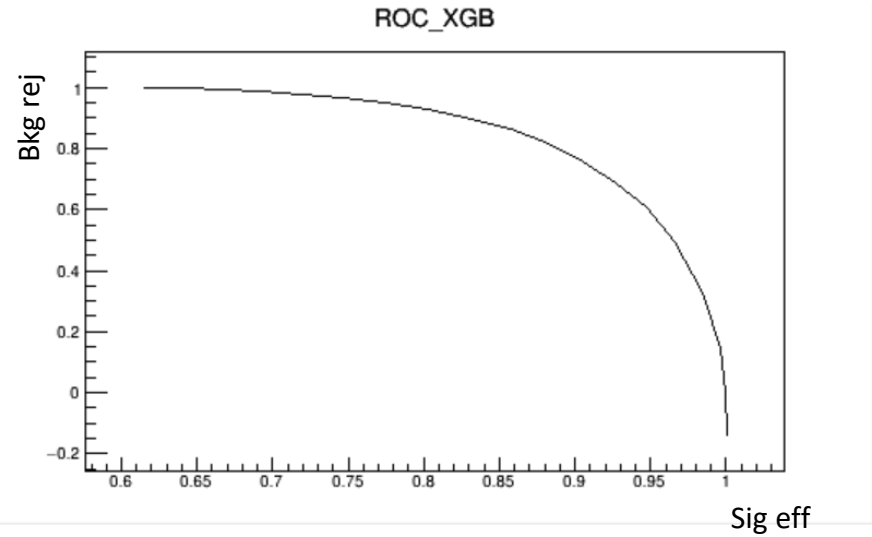
Tutorial(Chinese ML class):

https://www.bilibili.com/video/BV1vJ41187hk?p=85

http://www.peixun.net/view/1281.html

XGBoost



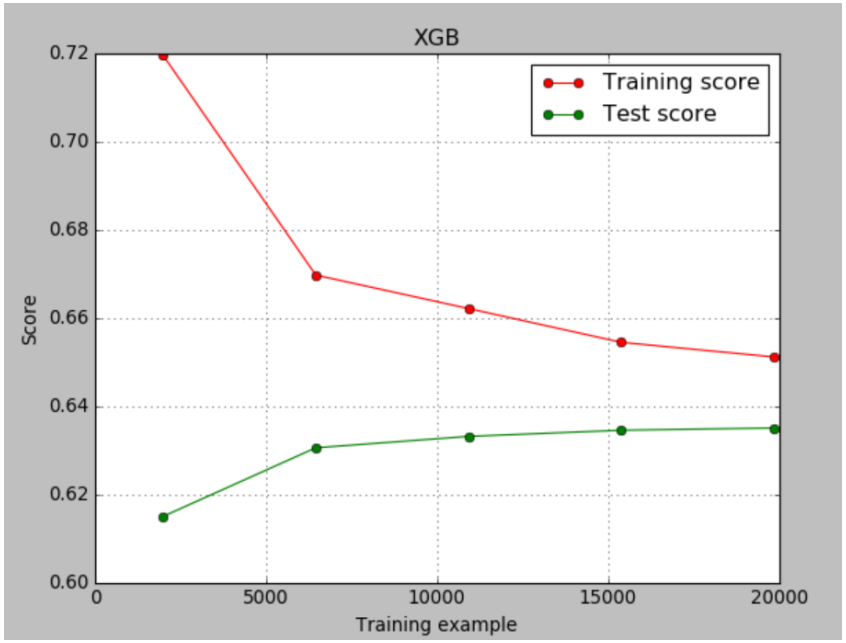


Training result: BDT distribution for signal/background, training and test(overtraining) and ROC.

XGBoost

Score of XGBoost (R^2):

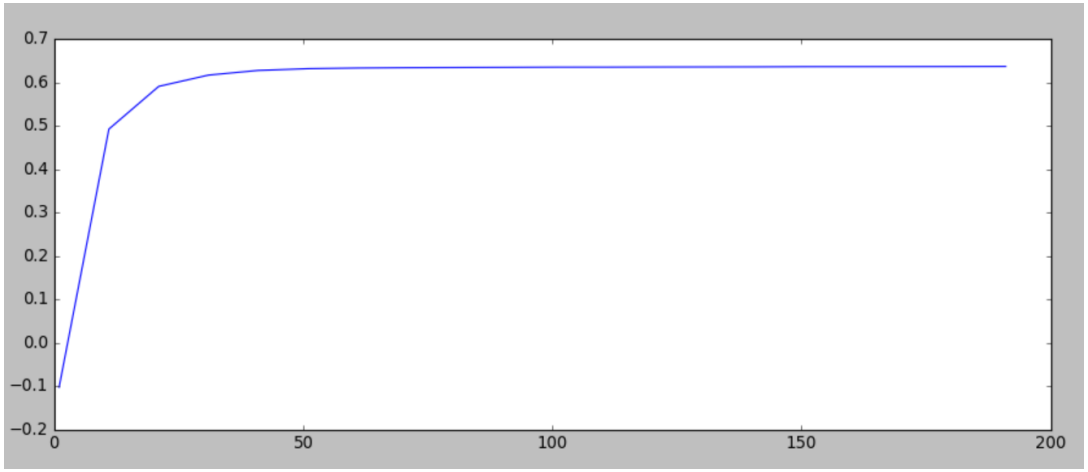
- Test sample: 0.634
- Cross validation(cv=3): [0.62977547 0.63824284 0.63629543]
- Learning curve: R^2 vs. training sample



A bit overtraining in N_{train}~20k.

XGBoost

Learning curve: R^2 vs. n_estimator (1, 201, 10)



Best n_estimator: 191
R^2: 0.6366

Grid Search for parameters tuning: n_estimator[100,1000], learning_rate [0.2, 0.5, 0.8]

```
[Parallel(n_jobs=1)]: Done  18 out of  18 | elapsed:  2.1min finished
{'n_estimators': 100, 'learning_rate': 0.2}
0.6360999613538862
```

Method Comparison

Learning score (R^2)

- XGBoost: 0.6345
- BDTG: 0.6346
- AdaBoost: 0.5888