

IML Machine Learning Leaderboard

Stefan Wunsch, Sergei Gleyzer

September 29, 2016

Inter-Experimental Communication

"I achieved an AUC of 0.873 on the HIGGS dataset. With only 42 minutes CPU time!"

⇒ **Let's support this!**

Basic Problems

- ▶ Most machine learning (ML) methods have a **lot of hyper parameters** (especially upcoming **deep neural networks**)
- ▶ **Tedious** and **computational expensive** task to optimize and compare with other methods
- ▶ **Common** machine learning **benchmarks** most often **do not fit** our needs (Iris, MNIST, CIFAR, ...)
- ▶ **Does a specific model perform superior on different physics-related datasets?**

Idea: IML Machine Learning Leaderboard

CERN Machine Learning Leaderboard

Leaderboard

Datasets

Submit

About

Classification: HIGGS

Type	Measure	Score	Identifier	Reference
DNN	AUC	0.885	jsductor	https://arxiv.org/abs/1402.4735
BDT	AUC	0.861	fundub	Link to repository/paper/website

Classification: SUSY

Type	Measure	Score	Identifier	Reference
DNN	AUC	0.876	jsductor	https://arxiv.org/abs/1402.4735

Regression: F00

Prototype only: <https://iml.web.cern.ch/leaderboard>

Proposal

- ▶ Start **database** or **leaderboard** for **machine learning models** that incorporates **quantities of interest for Physicists**, e.g., performance measured in AUC or MSE, but as well training duration and used hardware
- ▶ Offer good **starting point to tackle machine learning problems in HEP**
- ▶ **Publish benchmark definitions** to ensure easy comparability (training and test split, performance measure, ...)
- ▶ **Encourage model sharing** with possibility to upload your setup with your submission
- ▶ Provide **plug-and-play examples with TMVA** in addition to submitted setups
- ▶ Host as **part of IML website**

⇒ Make it **simple, useful** and **fun to use**

Details

- ▶ **Submissions** will be managed with **GitHub** and YAML submission files
- ▶ **Shared model definitions and training configurations** will be hosted on GitHub as well
- ▶ **Trustworthiness of submissions** are ensured by allowing **multiple submissions of the same model** and displaying number of submissions and mean/variance/histogram of score (AUC, MSE, ...)
- ▶ **IML website** serves as graphical frontend for the database

Example submission file:

```
# Benchmark dataset (HIGGS, SUSY, ...)
```

```
dataset: HIGGS
```

```
# Score on test data
```

```
score:
```

```
  measure: AUC
```

```
  value: 0.86
```

```
...
```

Outlook

- ▶ Nothing happened so far
- ▶ Still collecting ideas
- ▶ Shall come along with collection of benchmark datasets published on IML website (work in progress)
- ▶ Hopefully becomes reality in the next few month
- ▶ **Do we want this? Is it useful? Any suggestions?**