
pandalyse Documentation

Simon Wehle

Oct 21, 2020

Contents

1	What is pandalyse	3
2	Installation	5
3	Usage	7
3.1	Selectors	7
3.2	Analysis	7
3.3	Trainer	8
4	Contents	9
5	Indices and tables	11

This is the documentation of **pandalyse**.

Note: This package is in the early stage of development.

CHAPTER 1

What is pandalyse

Pandalyse offers an analysis environment and tools for pandas. The main features are

- **Selector:** Define and store cuts in a Selector object: *cutted_df = selector(df)*
- **Trainer:** Train multiple mva with scikit-learn interface in the way: *trainer.fit(signal_df, background_df)*
- **Analysis:** Store and retrieve Selectors, Trainings, numpy-arrays and dataframes in predefined locations: *df = ana.data.get("MySignalData")*

CHAPTER 2

Installation

```
pip install pandalyse
```


3.1 Selectors

Selectors store cuts on columns of a *pandas.DataFrame*. All cuts are stored as a list of strings, which are applied with the AND condition.

Example:

```
import pandalyse

sel = pandalyse.Selector(['column1 > 0', 'column2 == 1'])

# Assume the existence of a pandas dataframe 'df' and 'second_df'
df_cutted_1 = sel(df)

sel.add_cut('column3 < 100')
df_cutted_2 = sel(second_df)

df_cutted_3 = sel(df, 'Temporary_Cut == 1')
```

3.2 Analysis

The Analysis is the central part of *pandalyse*. It consists of a *.pandalyse* file which contains information on folders where *pandas.DataFrames*, *pandalyse.Selectors*, *pandalyse.Trainer* and *numpy.arrays* are stored.

Example:

```
import pandalyse
import numpy as np

ana = pandalyse.analysis('path/to/(desired)/analysis/dir')
# ana = pandalyse.analysis() will use `pwd`
```

(continues on next page)

(continued from previous page)

```
# ...
# assuming the existence of a signal and background dataframe
ana.data.add(df_bkg, 'background')
ana.data.add(df_sig, 'signal')

# doing some calculations
ana.values.add(0.5, 'efficiency')
ana.values.add(np.arange(3), 'example_array')

print(ana.values.example_array/ana.values.efficiency)
# >> [0, 0.5, 1]

# ls path/to/(desired)/analysis/dir
# >> background.hdf signal.hdf efficiency.val example_array.val
```

3.3 Trainer

A *pandalyse.Trainer* can take a list of features of a dataframe and classifier with an *sklearn* interface methods can be added.

Example:

```
import pandalyse

ana = pandalyse.analysis()

tr = pandalyse.Trainer(['column1', 'column2'])
tr.add_method('bdt', some.sklearn_like.classifyer())
tr.add_method('nn', some.sklearn_like.classifyer2())

tr.fit(ana.data.get('signal'), ana.data.get('background'))

ana.trainigs.add(tr, 'first_training')
```

CHAPTER 4

Contents

CHAPTER 5

Indices and tables

- `genindex`
- `modindex`
- `search`