PCA

December 17, 2022

[1]:
```
%cd ../../../
```

```
/home/jan/FMF/masters
```

[2]:
```
saved = "ml_hep_sim/notebooks/article_notebooks/saved/"
```

[3]:
```
from ml_hep_sim.pipeline.blocks import *
from ml_hep_sim.pipeline.pipes import *
from ml_hep_sim.plotting.style import style_setup

import matplotlib.pyplot as plt

style_setup(seaborn_pallete=True)
```

[4]:
```
data_name = "higgs_sig"

x1 = ConfigBuilderBlock(
    override_config={
        "datasets": {
            "data_name": data_name,
            "data_params": {
                "subset_n": [10 ** 5, 10 ** 5, 10 ** 5],
                "batch_size": 1024,
                "rescale": None,
                "to_gpu": False,
                "shuffle_data": True,
            },
        },
    },
    model_name="",
)()
x2 = DatasetBuilderBlock()(x1)
x3 = ReferenceDataLoaderBlock()(x2)
x4 = PCARunnerBlock(save_data=False)(x3)
```

[5]:
```
pipe = Pipeline(pipeline_name=f"PCA_test_pipe", pipeline_path="ml_pipeline/")
```

[6]:
```
pipe.compose(x1, x2, x3, x4)
pipe.fit()
```

[6]:
```
<ml_hep_sim.pipeline.pipes.Pipeline at 0x7f1021f99910>
```

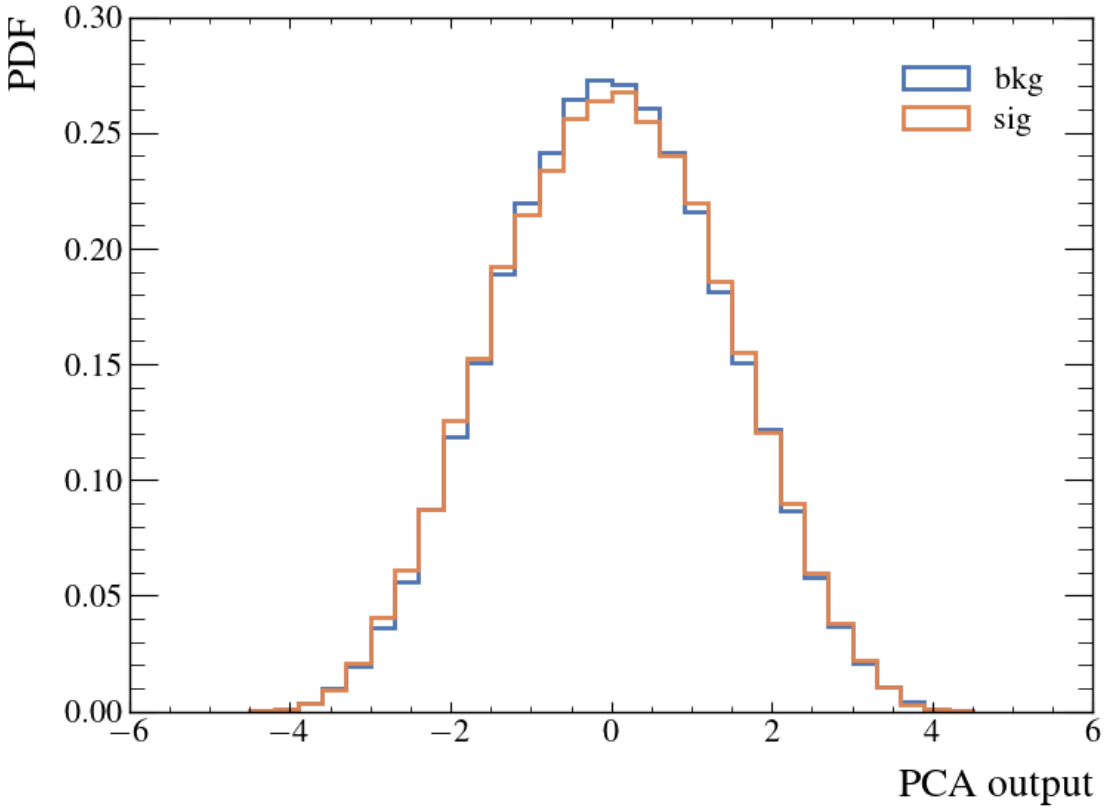
[7]:
```
# res = pipe.pipes[-1].results
# pickle_save(saved, f"PCA_test_{data_name}.p", res)
```

[8]:
```
save = ["PCA_test_higgs_bkg.p", "PCA_test_higgs_sig.p"]
```

[9]:
```
for s in save:
    plt.hist(pickle_load(saved, s), bins=30, histtype="step", density=True,␣
 ↪range=[-4.5, 4.5], lw=2)

plt.xlabel("PCA output")
plt.ylabel("PDF")
plt.legend(["bkg", "sig"])
plt.tight_layout()
plt.savefig(saved + "PCA_pdf.pdf")
```

```
WARNING:root:Loading from
ml_hep_sim/notebooks/article_notebooks/saved/PCA_test_higgs_bkg.p
WARNING:root:Loading from
ml_hep_sim/notebooks/article_notebooks/saved/PCA_test_higgs_sig.p
```



[ ]: