WATCHMAL

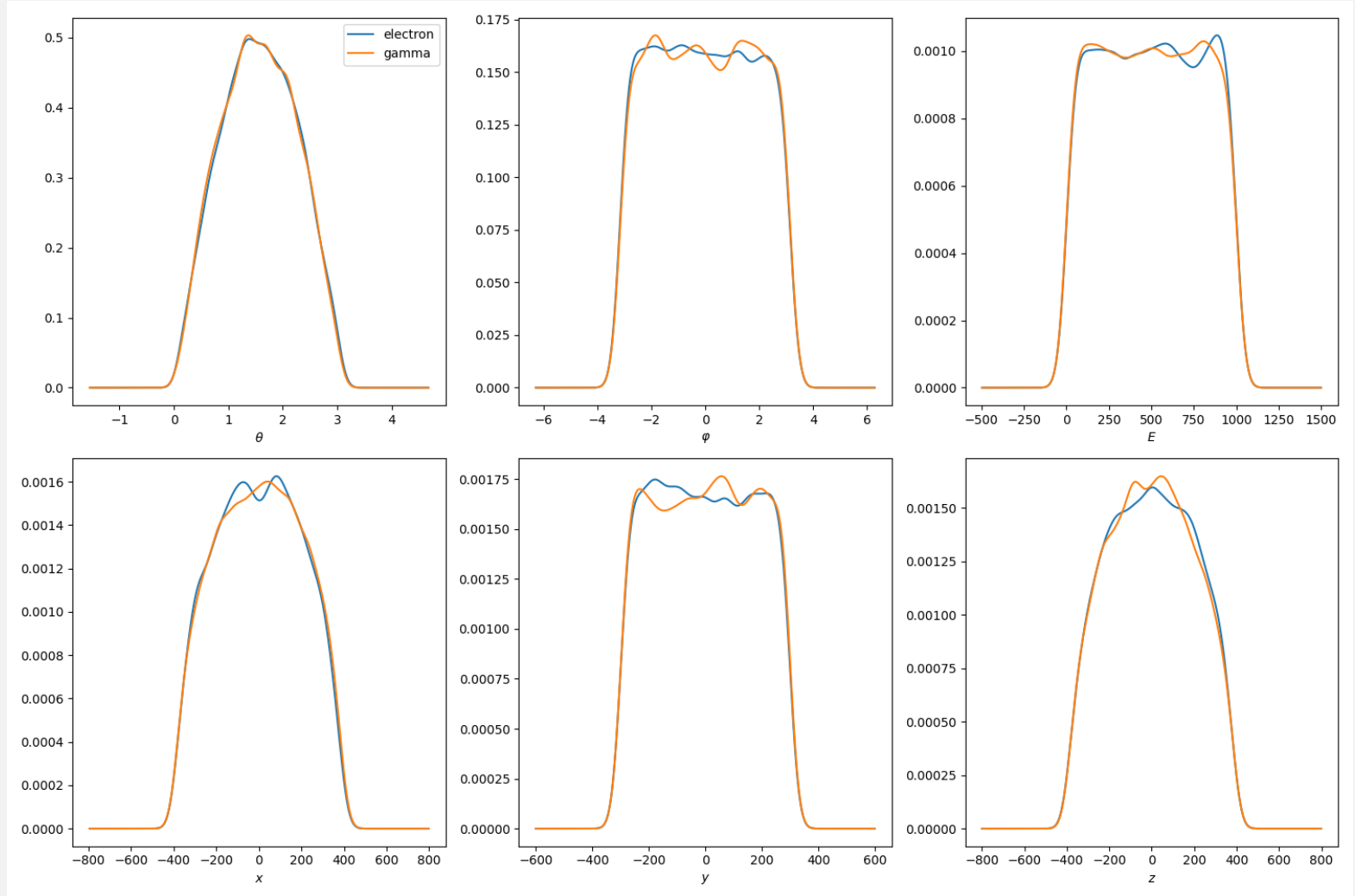## 3rd Report – Iñaki Erregue Alvarez-Buhilla

Discussed topics:

- Input exploration

- Time and Charge scaling
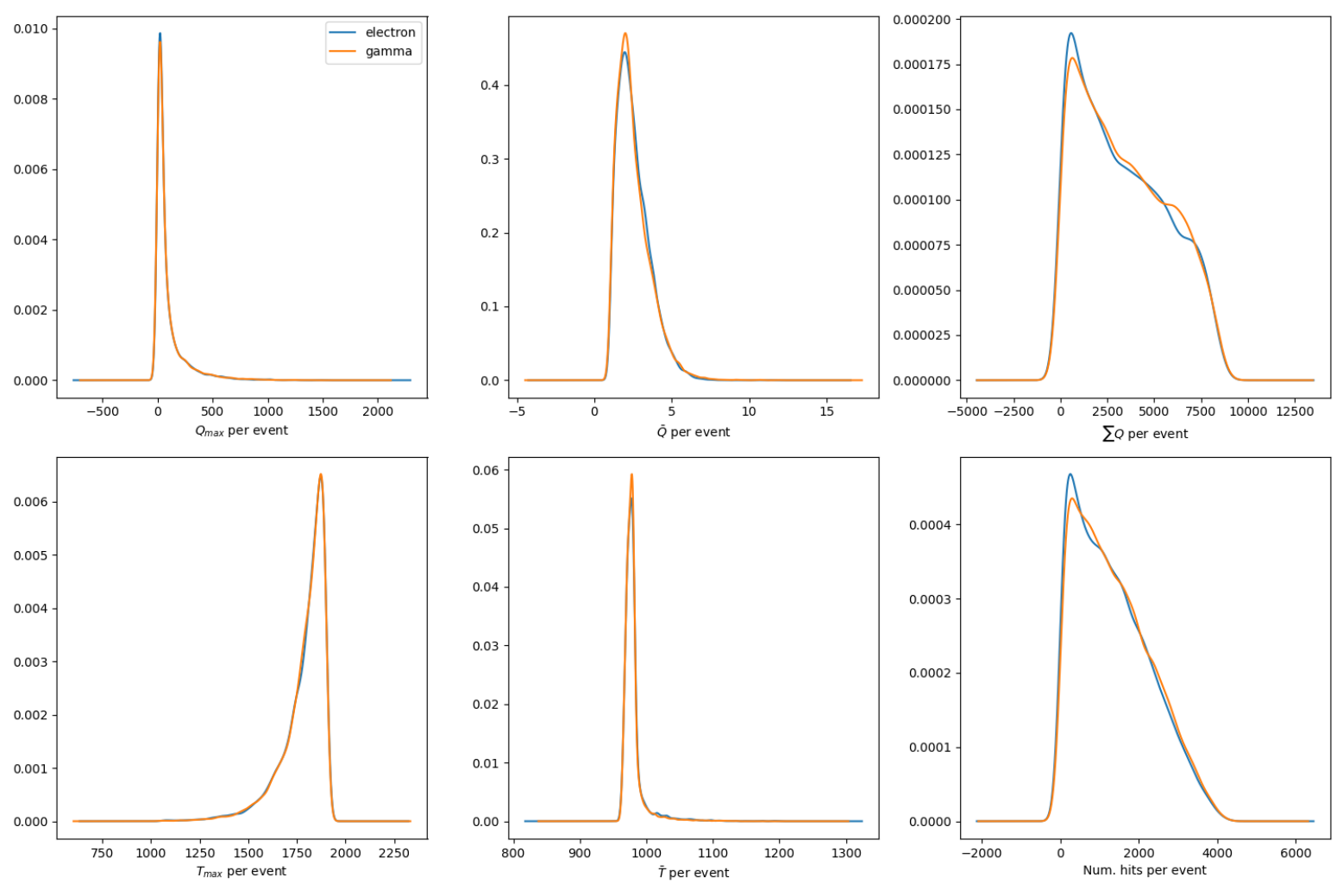
- Early Stopping

INPUT EXPLORATION
Kernel densities

- Now obtained directly from the main H5 file using indices

- Only displaying 2 classes: electrons (1) and gammas (0)

- Metadata distributions per event

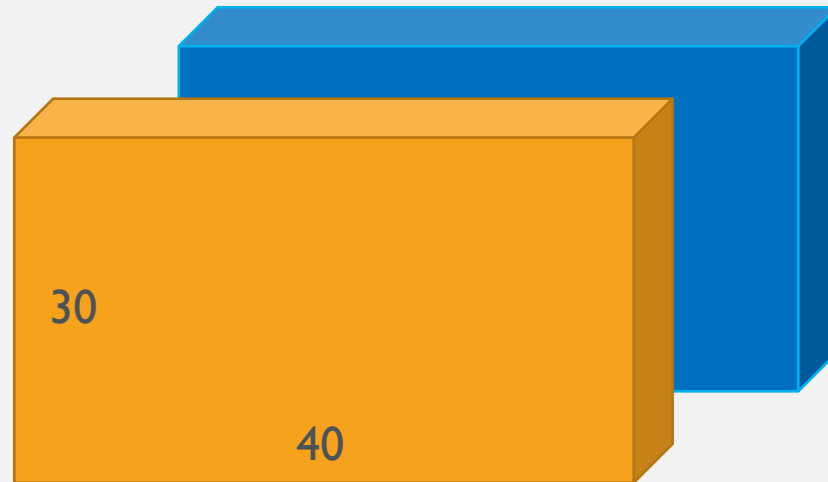- Hit, Charge and Time distributions per event

METADATA

SCALING

Charge data:

- 19 channels
- Values from 0 to 1530
- Mean of max charge per event: 90
- Mean value per hit: 2.5

Time data:

- 19 channels
- Values from 0 to 1900
- Mean of max time per event: 1700
- Mean value per hit: 1000

Very different scales, need to scale them!

SCALING

Two approaches:

- Sample scaling: simpler implementation

- Feature scaling: same for all events, more computational cost

Discussed methods:

- Standarization: $\bar{x} = (x - \mu)/\sigma$,   parameters are difficult to obtain for whole dataset (subsets?)

- Normalization: $\bar{x} = (x - x_{\{min\}})/(x_{\{max\}} - x_{\{min\}})$, $x_{\{min\}} \approx 0$ for Time and Charge

- Max. normalization: $\bar{x} = x/x_{\{max\}}$,   considered maximum: mean of maximum of subset of events

- Vector normalization: $\bar{x} = x/||x||$,   all values very close to 0

EARLY STOPPING
Added to quickly detect if an XP is overfitting
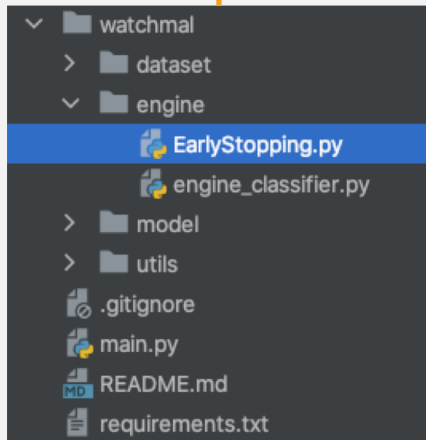

```python
"""MIT License..."""


import numpy as np



class EarlyStopping:
    """
    Early stops the training if validation loss doesn't improve after a given patience.
    Modified version with saving disable (EngineClassifier already does it).
    """
    def __init__(self, patience=7, verbose=True, delta=0, trace_func=print):
        """
        Args:
            patience (int): How long to wait after last time validation loss improved.
                            Default: 7
            verbose (bool): If True, prints a message for each validation loss improvement.
                            Default: False
            delta (float): Minimum change in the monitored quantity to qualify as an improvement.
                            Default: 0
            trace_func (function): trace print function.
                            Default: print
        """
```

watchmal
- dataset
- engine
  - EarlyStopping.py
  - engine_classifier.py
- model
- utils
- .gitignore
- main.py
- README.md
- requirements.txt

```yaml
1    epochs: 20
2
3    report_interval: 100
4    val_interval: 100
5    num_val_batches: 32
6
7    checkpointing: False
8
9    early_stopping:
10       patience: 3
11       delta: 0
12
13   data_loaders:
```

train_resnet.yaml

BACKLOG

This week:

- Familiarize Cedar Cluster

- Dockerize WatChMaL

- Run XPs on Cedar:

  - Only Q (19 channels)

  - Only Q (1 channels)

  - Only T (19 channels)

  - Only T (1 channels)

  - T+Q (19+19 channels)

  - T+Q (1+1 channels)

  - (Maybe changing scaling)