How to create powerful machine learning projects in astronomy

J. Buchner, Max Planck Institute for extraterrestrial Physics, Giessenbachstr, 85748 Garching, Germany
S. Fotopoulou, HH Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, BS8 1TL, UK

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS

OVERVIEW

Background

- Large, freely available, well-maintained data sets make astronomy a popular playground for machine learning (ML)
- ML projects are becoming very popular

  yet room for improvement towards robust insights in ML and/or physics
- Lack of a guideline document for establishing workflows that critically verify, characterize and calibrate machine learning models

Goal

- Audience:
  - astronomers who are novices in machine learning
  - data scientists who are novices in astronomy
- collection of guidelines for setting up ML projects to make them
  - likely useful for science
  - less frustrating
  - reduce time for scientist & computer
  - likely to lead to robust insights

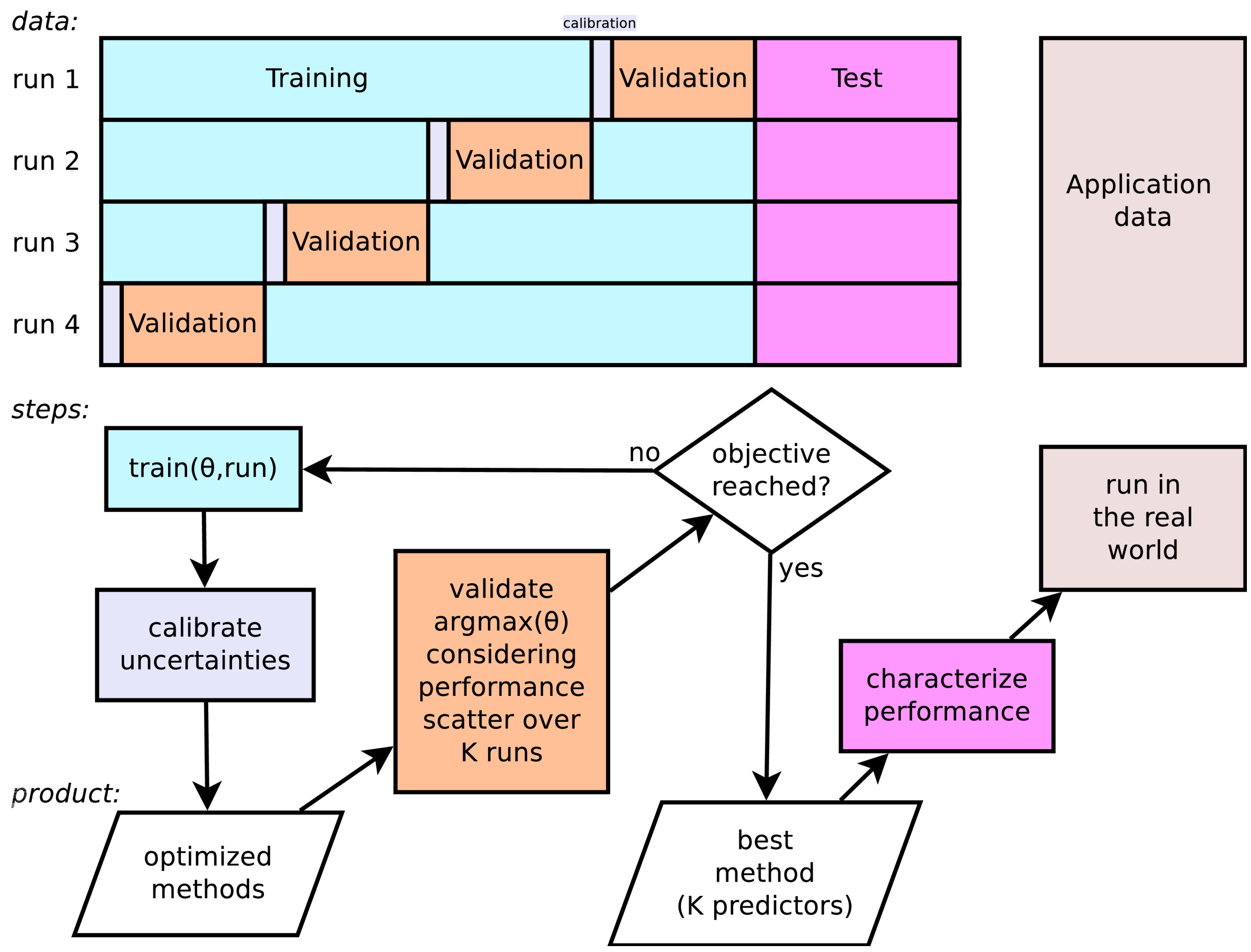
CONTENT

The guidelines are independent of the adopted ML techniques and focus on project management and workflows.

- Your new project
  - Have an objective
  - Have a good objective
  - Define a finish line
    - Comparing classifiers
    - Comparing regressors
    - Comparing unsupervised methods
  - Create a test harness
- Work environment
  - Understand your role
    - Interface with the domain experts
    - Open the black box
  - Separate the person and the work
- The starting point
  - Establish a `dummy' baseline
  - Establish a historical baseline
  - Establish an initial baseline
  - Be robust to sample variations
  - Negligible improvements
- Being useful for science
  - Embrace uncertainty
  - Produce uncertainties
  - Beware that methods have priors
  - Vary your prior
  - Take care of covariate shift
  - One possible workflow
- Human learning
  - What has the machine learned?
    - Ablation studies
    - Ablation studies of the model
    - Interrogate the model
  - Communicating
  - Have stakes, make predictions
- Recommendations

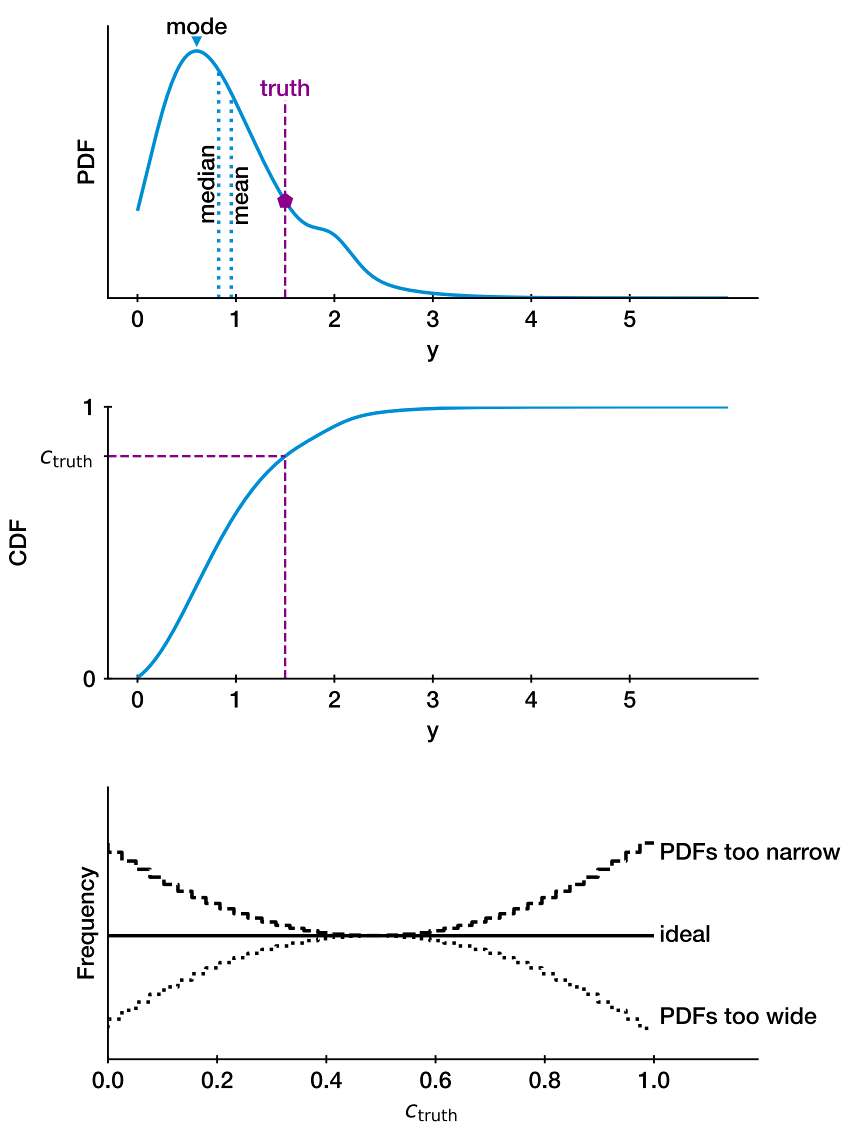
The recommendations have been influenced by projects with students, and discussions at conferences including ML-IAP2021 in Paris

A POSSIBLE WORKFLOW



A workflow for predicting with trustworthy uncertainties. A small portion of the training data is held out to calibrate the uncertainties. This can be considered part of the training. Instead of training once, K-folding (here K=4) trains with different seeds and data partitions. On the validation data set, the performance and its variance are measured, and the best hyper-parameters θ are selected. Among methods that are equally good within the error, we can use a secondary criterion like interpretability or speed to select one. If the objective seems to be reached, the performance is finally characterized with the yet unseen test data set. The model is delivered and applied to the real world data.

CALIBRATION OF UNCERTAINTIES



From the predicted probability density function (blue curve in the top panel) of one validation sample, the probability of the true value (purple) can be read off (ptruth). The higher the product of these probabilities over the entire validation sample, the more informative the prediction.

The middle panel shows the corresponding cumulative probability function, from which the cumulative probability of the true value can be read off (ctruth).

The histogram of ctruth values over the validation sample is illustrated in the bottom panel. If the PDFs are overly narrow, the ctruth values are frequently near 1 or 0 (dashed). Ideally, the values show a uniform distribution. Based on Hamill (2001)