MACHINE LEARNING IN PHYSICS

PROJECTS

HARRISON B. PROSPER

PHY6938

Project Description

The project has 3 parts:

1. Use a machine-learning model to solve one of the following classes of problem: classification, regression, anomaly detection, or an ordinary differential equation.

2. Document the project in a short paper using the cernrep LaTeX package.

3. Give a 10-minute presentation of your project during the last week of the semester.

Project Description

Problem You can either choose a problem from your field or tackle one of the following problems.

1. Jet image classification.

2. Nuclear properties modeling (AME 2020).

3. Solve the Friedmann equation for the ΛCDM model.

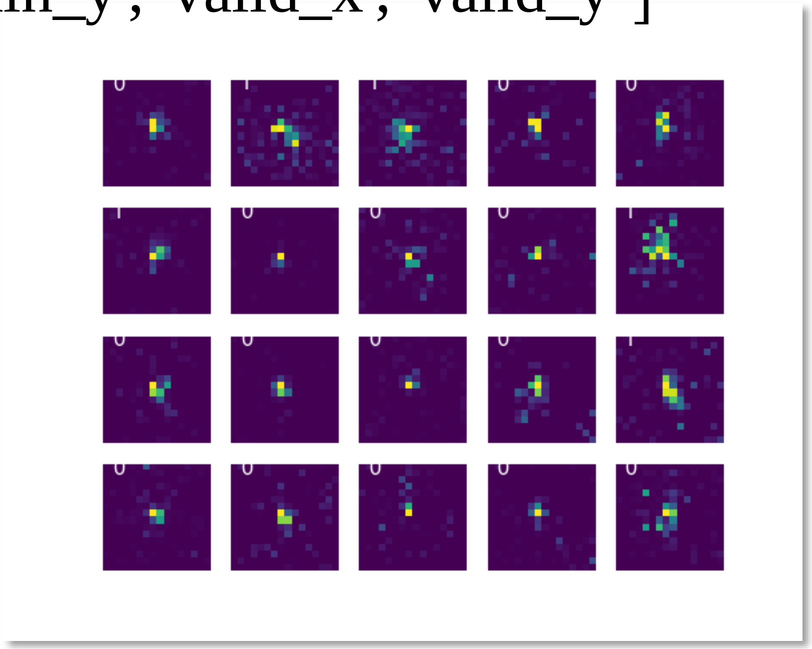
4. Anomaly detection using density ratio.

Jet Classification

Problem: Using 2D images of jets of particles, classify them into two classes: jets initiated by gluons or by quarks.

Dataset: Images of shape (1, 16, 16), which have been divided into: ['test_x', 'test_y', 'train_x', 'train_y', 'valid_x', 'valid_y’]

Filename: jets.h5



len(train_x): 50,000

len(valid_x): 5000

len(test_x): 5000

Nuclear Properties

Problem: Predict a nuclear property given N, Z, and A.

Dataset: 3,558 nuclei from the AME 2020 mass data[1].

	name	N	Z	A	Mexcess	eMexcess	bindingE	ebindingE	betaE	ebetaE	atomicM	eatomticM
0	n	1	0	1	8.071318	4.4e-07	0.0	0.0	0.782347	4e-07	1.008665	4.7e-07
1	H	0	1	1	7.288971	1.3e-08	0.0	0.0	-999	-999	1.007825	1.4e-08
2	H	1	1	2	13.135723	1.5e-08	1.112283	2e-07	-999	-999	2.014102	1.5e-08
3	H	2	1	3	14.949811	8e-08	2.827265	3e-07	0.018592	6e-08	3.016049	8e-08
4	He	1	2	3	14.931219	6e-08	2.57268	1.5e-07	-999	-999	3.016029	6e-08
5	Li	0	3	3	-999	-999	-999	-999	-999	-999	-999	-999
6	H	3	1	4	24.621129	0.1	1.720449	0.025	22.196213	0.1	4.026432	0.107354
7	He	2	2	4	2.424916	1.5e-07	7.073916	2e-07	-22.9	0.212132	4.002603	1.6e-07
8	Li	1	3	4	25.32319	0.212132	1.15376	0.053033	-999	-999	4.027186	0.227733
9	H	4	1	5	32.892447	0.089443	1.336359	0.017889	21.661213	0.091651	5.035311	0.09602

Filename: AME2020.csv

1. The AME 2020 atomic mass evaluation, Chinese Phys. C **45**, 030002 (2021) and Chinese Phys. C **45**, 030003 (2021).

Nuclear Properties

Problem: Predict a nuclear property given N, Z, and A.

Dataset: 3,558 nuclei from the AME 2020 mass data[1].

Energy Variables:

Mexcess, bindingE, betaE: in MeV

atomicM: relative atomic mass

Note: entries in the CSV file with -999 implies missing data.

Friedmann Equation

Problem: Solve the Friedmann equation using a PINN.

Dataset: To be generated by you!

Equation:

$$\frac{da}{dx} = a\sqrt{\Omega(a)}, \qquad x \equiv H_0 t$$

where for the ΛCDM model,

$$\Omega(a) = \frac{\Omega_M}{a^3} + \frac{(1 - \Omega_M - \Omega_\Lambda)}{a^2} + \Omega_\Lambda$$

Initial condition: $a(x = 0) = 0$.

Suggested domain: $(x, \Omega_M, \Omega_\Lambda) = (0, 1.5) \otimes (0, 1) \otimes (0, 1)$ with the constraint $\Omega(a) > 0$.

Anomaly Detection

Problem: Detect an unknown signal in a dataset.

Dataset: Data comprising a mixture of mostly background plus a small signal and another containing background only.

Anomaly Detector:

$$\frac{p_{data}(x)}{p_{bkg}(x)} > t$$

where t is the threshold above which you declare x to be a signal. You can assume that

$$p_{data}(x) = \epsilon p_{sig}(x) + (1 - \epsilon) p_{bkg}(x)$$

Writeup

Your Paper, which should be at most 5 pages, excluding references, must include the following elements and sections:

1. A Title and an Abstract
2. Introduction
3. Dataset Description
4. Model Description
5. Experiments
6. Summary

Ground Rules

1. Thursday class will be devoted to work on your project. You're strongly advised to show up!

2. You are free to help each other, but your paper must be your own.

3. You are free to use any resources, but you must cite all them.

4. You must document the data used, the architecture of your model, the training protocol and the results.