EPFL CH-457

Philippe Schwaller

Invalid Date

Table of contents

Preface

AI4Chemistry course

The Artificial Intelligence (AI) for Chemistry course will be taught in Spring 2023. It is a course with a lot of hands-on exercises. Experience in Python programming and machine learning (ML) will help you to get up to speed quickly, but we will try to make it as accessible as possible.

We will make use of Google Colab to run the code directly in your browser, as there is zero configuration required and we will have access to GPUs free of charge.

A lot of the examples and ideas in this course are taken from the open-source community, which we will properly reference.

Contributors

This course is being created by the LIAC team. Many thanks to all the TAs:

- Victor Sabanza Gil
- Junwu Chen
- Theo Neukomm

And all the old TAs participating to the creation of the notebooks:

- Bojana Ranković
- Andres CM Bran
- Jeff Guo

Tentative content

- Python Crash Course & essential libraries (matplotlib, numpy, pandas)
- Cheminformatics toolkits (rdkit)
- Introduction into data science
  - Supervised machine learning (regression, classification)

  - Unsupervised machine learning
  - Data and standardisation

- Deep Learning for Chemistry
  - Property prediction models
  - Inverse Design[1]
  - Reaction prediction and retrosynthesis[2]

- Advanced topics in AI for Chemistry
  - Bayesian optimisation for chemical reactions

Exercises

Week	Topic	Link to Colab
1	Python and Jupyter Pandas Plotting data Intro to RDKit	
2	Supervised ML	
3	Introduction to Deep Learning Graph Neural Network GNN example - chemprop	
4	Dimensionality reduction Clustering Pd dimers discovery by kMeans	
5	De novo molecule generation (VAE)	
6	De novo molecule generation (SMILES-LSTM)	
7	Chemical reactions prediction: Template-free methods	
8	Retrosynthesis: Template-based methods	
9	Atom mapping Reaction fingerprints	
10	Bayesian Optimisation	
11	Model deployment: Git(hub)	

Week	Topic	Link to Colab
12	Model deployment: Streamlit Guest lecture: AlphaFold2	

The solutions can be found in this GitHub repo. Don't forget to leave a star, if you find it useful.

Inspiration

The cheminformatics and ML for chemistry have a lively open source community. Here is a collection of inspirational blogs and webpages, from which we discuss examples:

Cheminformatics / ML for Chemistry

- Andrew White's Deep Learning for Molecules & Materials book
- MolSSI Education Resources
- Greg Landrum's RDKit blog
- Esben Bjerrum's Cheminformania
- iwatobipens' blog
- Rocío Mercado's dl-chem-101
- Jan H. Jensen's Machine Learning Basics
- Pat Walter's Practical Cheminformatics With Open Source Software
- Kjell Jorner's blog

AI for Science

- Lewis Tunstall's Deep Learning for Particle Physicists
- Summer school on Statistical Physics & Machine learning organised by Florent Krzakala and Lenka Zdeborova, EPFL

ML & Data Science

- Practical Deep Learning
- MIT's Intro to Deep Learning
- Aurelien Geron's Hands-on Machine Learning
- Lewis Tunstall's Introduction to Data Science
- Natural Language Processing with Transformers by HuggingFace

Check them out and don't forget to leave a star on GitHub and follow the authors on Twitter, if you like the content. Those blogs and webpages have all helped me during the creation of this course (and also before, when I was learning about ML for Chemistry).

Tweets

< tweet pschwllr 1629098793399472130 >

References

1. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361**, 360–365 (2018).
2. Schwaller, P. *et al.* Machine intelligence for chemical reaction space. *Wiley Interdisciplinary Reviews: Computational Molecular Science* e1604 (2022).