Course Syllabus

[VM571]

[Machine Learning in Molecular and Materials Sciences]

[2023 Spring]

Course Description:

Machine learning has an increasing impact in molecular and materials sciences. On the one hand, machine learning provides new perspectives in how we record and analyze the structures and properties of molecules and materials. On the other hand, molecular and materials sciences are fertile grounds for the application of various machine learning techniques. This course intends to be an interdisciplinary bridge between data sciences and molecular/materials sciences. You will not only learn/review neural networks, convolutional neural networks, graphs, featurization, regression and classification, and other main machine learning techniques and concepts, but also learn/review molecular conformation, chirality, protein folding, symmetry groups, electron diffraction, and other techniques and concepts in molecular and materials sciences. Some familiarity with Python is recommended but not required. We will use DeepChem as the main package for the course.

Instructor:

Name: Wendong Wang
Email: wendong.wang@sjtu.edu.cn
Phone: 021-3420-6765 ext. 5271
Office: Room 527
Office hour: Tuesday 2-3pm or by appointment

Textbook (Author, Book Title, Publisher, Publication Year, ISBN):

We will use mostly online documentations and Jupyter notebooks:

Time and Location:

Monday (10am – 12pm) and Wednesday (9am – 11:40am)

Dong Xia Yuan (East Lower Hall) 403

Course Prerequisites:

Graduate standing. Upper-year undergraduate upon permission

Course Website:

https://sites.ji.sjtu.edu.cn/wwanglab/vm-571/

Grading Policy (Assignments %, Project, Exams, etc.)**:**

In-class tests and participation: 20%.

Homework: 30%

Paper presentation (groups of 2 – 3 students): 30%

Final exam: 20%

Honor Code Policy:

We follow the guidelines set out by the JI honor code:

https://www.ji.sjtu.edu.cn/academics/academic-integrity/honor-code/

Some more specific requirements:

You may discuss with your peers about the problems in the homework, but you must complete the problems on your own.

Paper presentations are group projects. You are encouraged to form teams with peers from the field of research different from your own. Presentations will be judged by both contents and clarity of delivery.

Teaching Schedule:

Week	L#.	Contents	Comments
1	1	Introduction, setting up Conda environment	
	2	Intro to Python & Jupyter lab, first deepchem model and solubility	ID1
2	3	Intro to neural network and Python, deepchem datasets	ID2
	4	Basics of Numpy, practical neural network, intro to MoleculeNet	ID3
3	5	Intro to Tensorflow, Simplified Molecular-Input Line-Entry System (SMILES) and molecular fingerprints	ID4, MML1
	6	Intro to CNN and pytorch, creating models with TensorFlow and Pytorch	ID5
4	7	OOP in python, intro to matplotlib, molecules as graphs	ID6
	8	CNN using keras, molecular structures and molecular descriptors	ID7, MML2
5	9	Building a neural network from scratch, splitters and hyperparameters	ID8 & 9
	10	Back propagation, intro to pandas, working with experimental data,	ID10
6	11	PubChem and PubChemPy, introduction to model interpretability	MLM2
	12	Review	
7	13	Intro to Sci-kit learn, intro to ML in materials sciences	MS1
	14	Intro to Sci-kit learn cont., paper representation	
8	15	Naïve Bayes classifiers, support vector machines, intro to bioinformatics	B1
	16	paper presentation	
9	17	Generative adversarial networks, protein deep learning	ID14, MP1
	18	Intro to Gaussian process, paper presentation	
10	19	Review and possible paper presentation	
	20	Final test	