Course Syllabus

[VM571]

[Machine Learning in Molecular and Materials Sciences]

[2021 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. 4221
Office: Room 527
Office hour: Tuesday 2-3pm or by appointment

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

References (not required):

Deep Learning for the Life Sciences, 1st ed, by Bharath Ramsundar, Peter Eastman, Patrick Walters, Vijay Pande, 2019, ISBN-10 : 1492039837

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	NO.	lectures and Exams	Comments
1	1	Introduction and setting up	
	2	Introduction to Python & your first deepchem model and solubility	Tutorial# 1
2	3	Introduction to neural network & understanding datasets	Tutorial# 2
	4	More on neural network & introduction to MoleculeNet	Tutorial# 3
3	5	Basics of NumPy and molecular fingerprints	Tutorial# 4
	6	Introduction to convolutional neural networks & creating models with TensorFlow	Tutorial# 5
4	7	More on convolutional neural networks & molecules as graphs	Tutorial# 6
	8	Introduction to matplotlib & more on molecular featurization	Tutorial# 7
5	9	Building a neural net from scratch	
	10	Back propagation, splitter and hyperparameter tuning	Tutorial# 8-9
6	11	Introduction to panadas and pubchempy, working with experimental data, introduction to model interpretability	Tutorial# 10, 23
	12	Review	
7	–	Holiday	
	13	Introduction to scikit-learn, unsupervised learning, putting multitask learning to work, paper presentation from literature	Tutorial# 11
8	14	Unsupervised embeddings for molecules, model protein ligand interactions, generative adversarial networks	Tutorial# 13, 14, 16
	15	Paper presentation from literature	
9	16	Large-scale chemical screens, introduction to bioinformatics, synthetic feasibility scoring	Tutorial# 19, 21, 24
	17	Uncertainty in deep learning, transfer learning with ChemBerta, reinforcement learning, paper presentation from literature	Tutorial# 22, 25, 26
10	18	Review and possible catch-up	
	19	Final test	