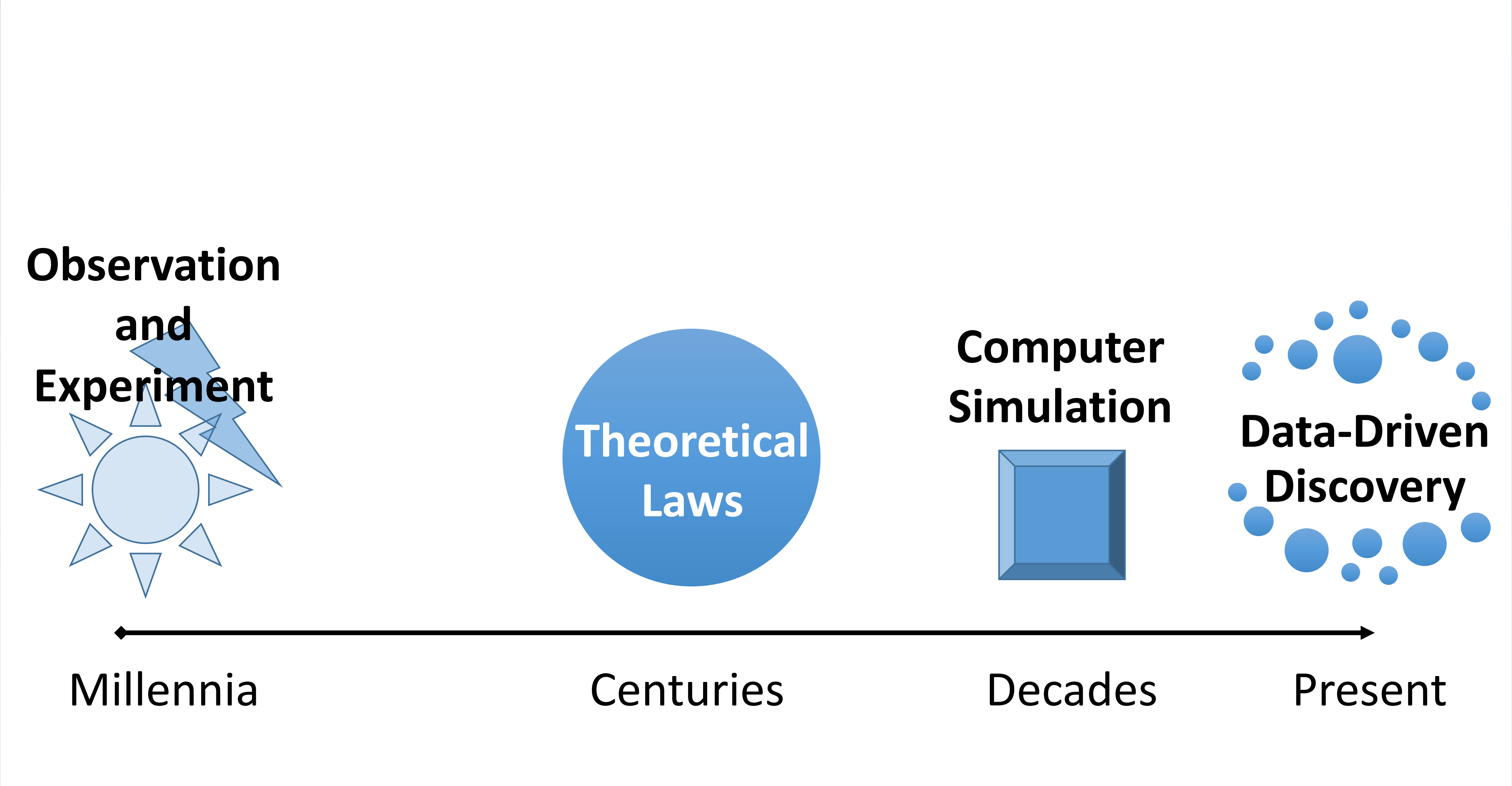


MACHINE LEARNING FOR PHYSICAL SCIENCES

CSE 5095-006 Spring 2019

Qian Yang

COURSE OVERVIEW

Application of machine learning to materials science, chemistry, physics

- What types of problems can ML uniquely address?
- What are the challenges and opportunities?
- Principles of good applied ML for scientific problems
- Needed theoretical developments for scientific ML

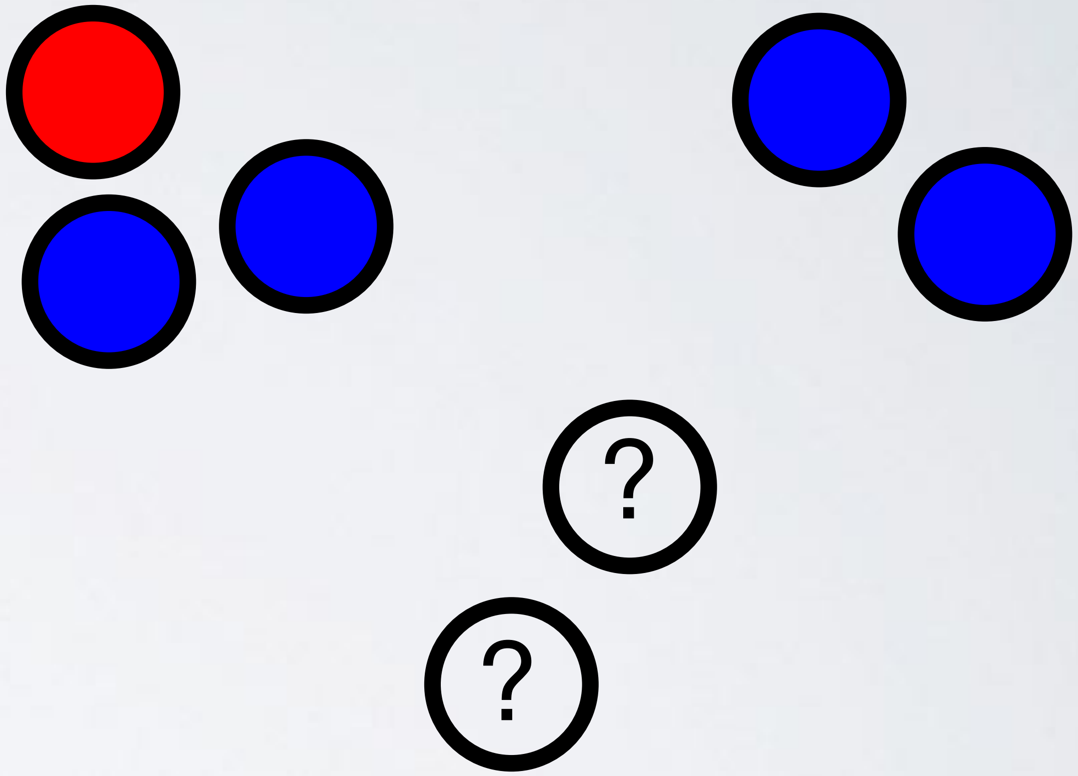
An objective of this class is to bring together your combined expertise in CSE and other engineering fields to explore creative solutions to interdisciplinary problems.

Part I: Machine Learning

- Scientific data in the ML setting
- Evaluating model performance
- Feature engineering
- Deep-learning based strategies
- Interpretable ML

Part II: Scientific Applications

- Scientific databases
- Property prediction for molecules and crystals
- Enabling faster molecular dynamics simulation
- Scientific imaging
- *Interests of the class.*



Check class website for reading assignments:

https://qianyanglab.github.io/teaching/cse5095_mlmaterials.html

Course assessments will be submitted via HuskyCT.

Assessments

- Project (50%) - teams of 2-4 students; final paper and flash talk; project suggestions and milestones will be distributed next week
- Quizzes (20%) - in weeks 4-8, there will be a short weekly quiz on the previous week's lecture (sample ungraded quiz will be given in week 3)
- Paper Review (30%) - in weeks 9-13, we will review papers as a class; each group of 2-3 students will be responsible for leading the discussion on one paper

Logistics

Lecture T/Th, 9:30am-10:45am, KNS 201

Office Hours Th, 11:00am-12noon, ITE 259

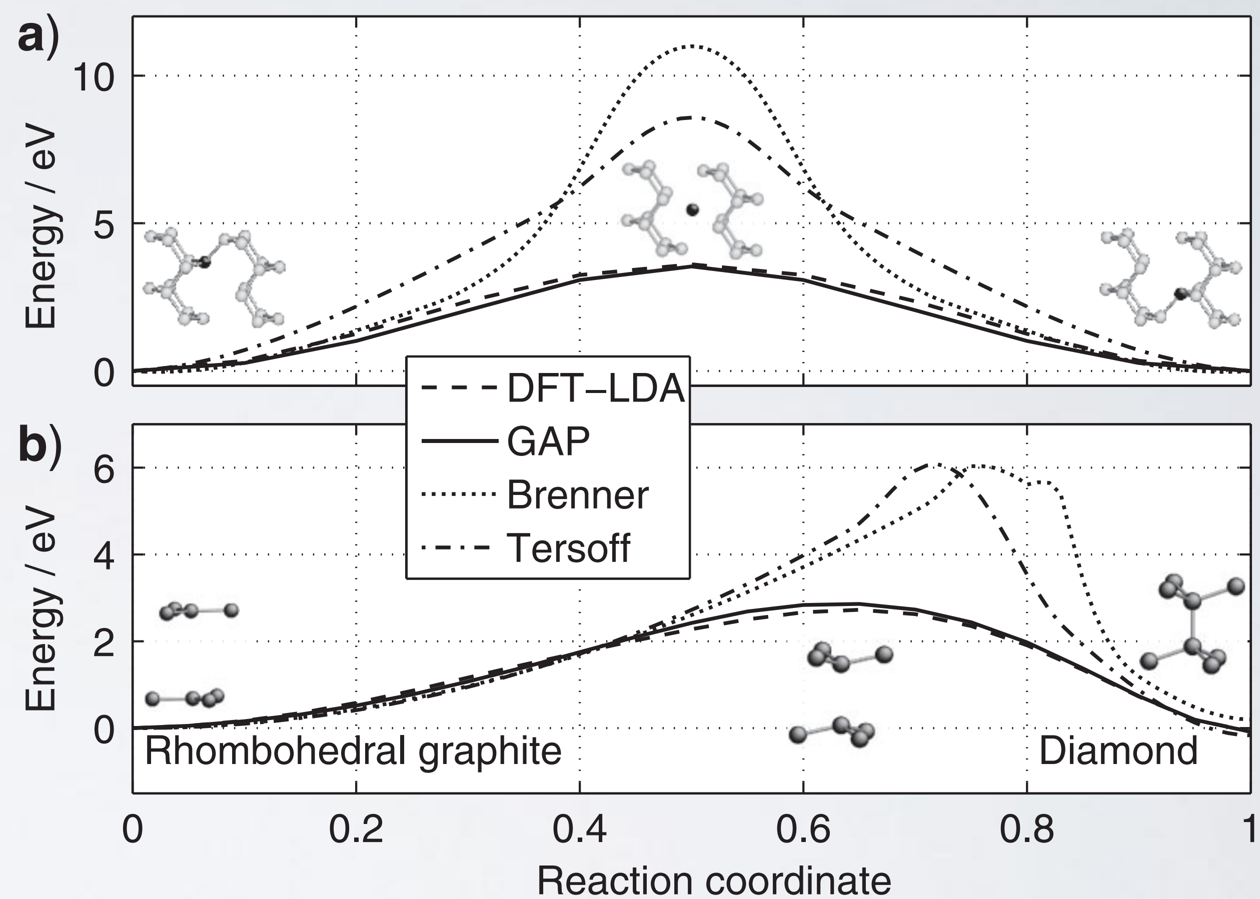
and by appointment

Start thinking about final project ideas and teams.

What kind of scientific problems is Machine Learning (not) good for?

THEMES

- Speeding up existing computational methods without decreasing accuracy
  - ML of expensive components
  - Model reduction

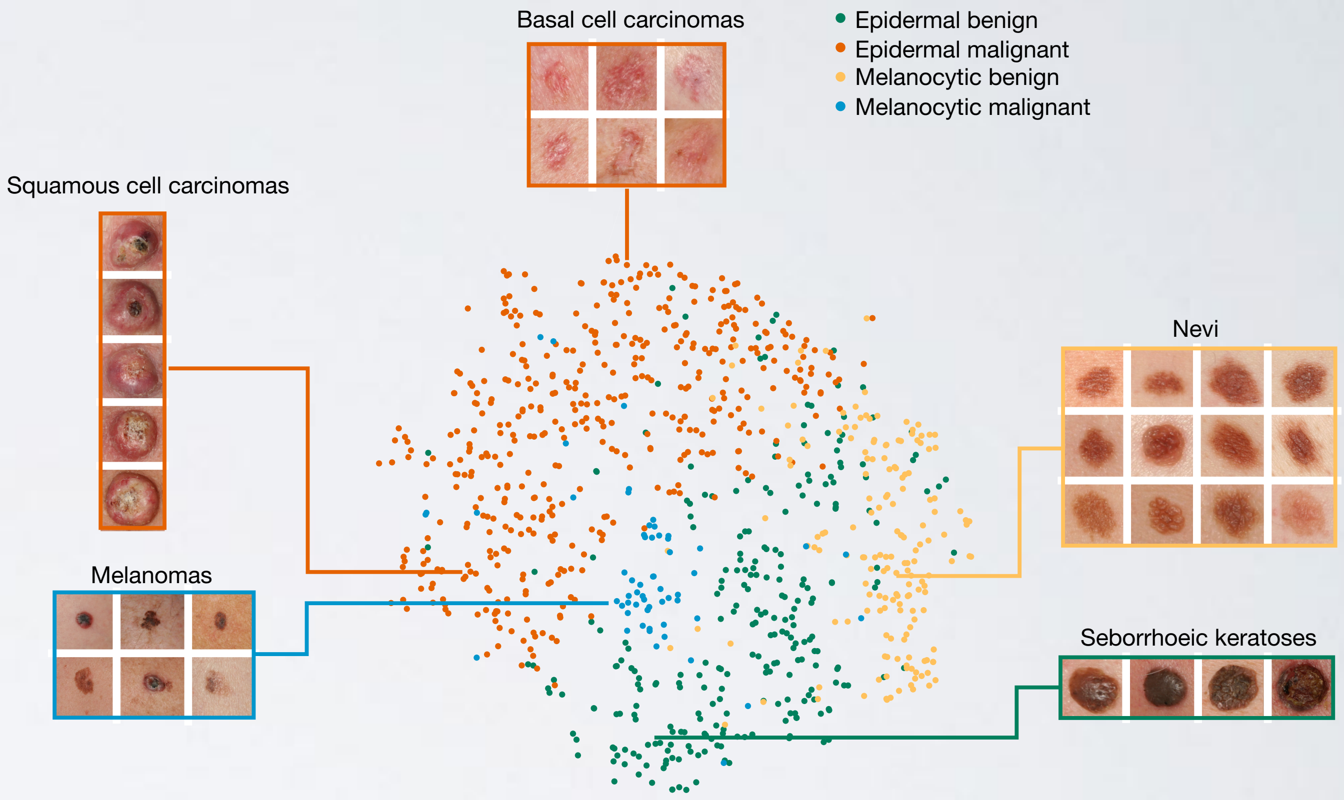


ML energy potentials for faster, more accurate atomistic simulation

Bartok, A. P., Payne, M. C., Kondor, R., and Csanyi, G., Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Physical Review Letters*, doi:10.1103/PhysRevLett.104.136403 (2010)

THEMES

- Automation: doing what humans could do, but faster and better

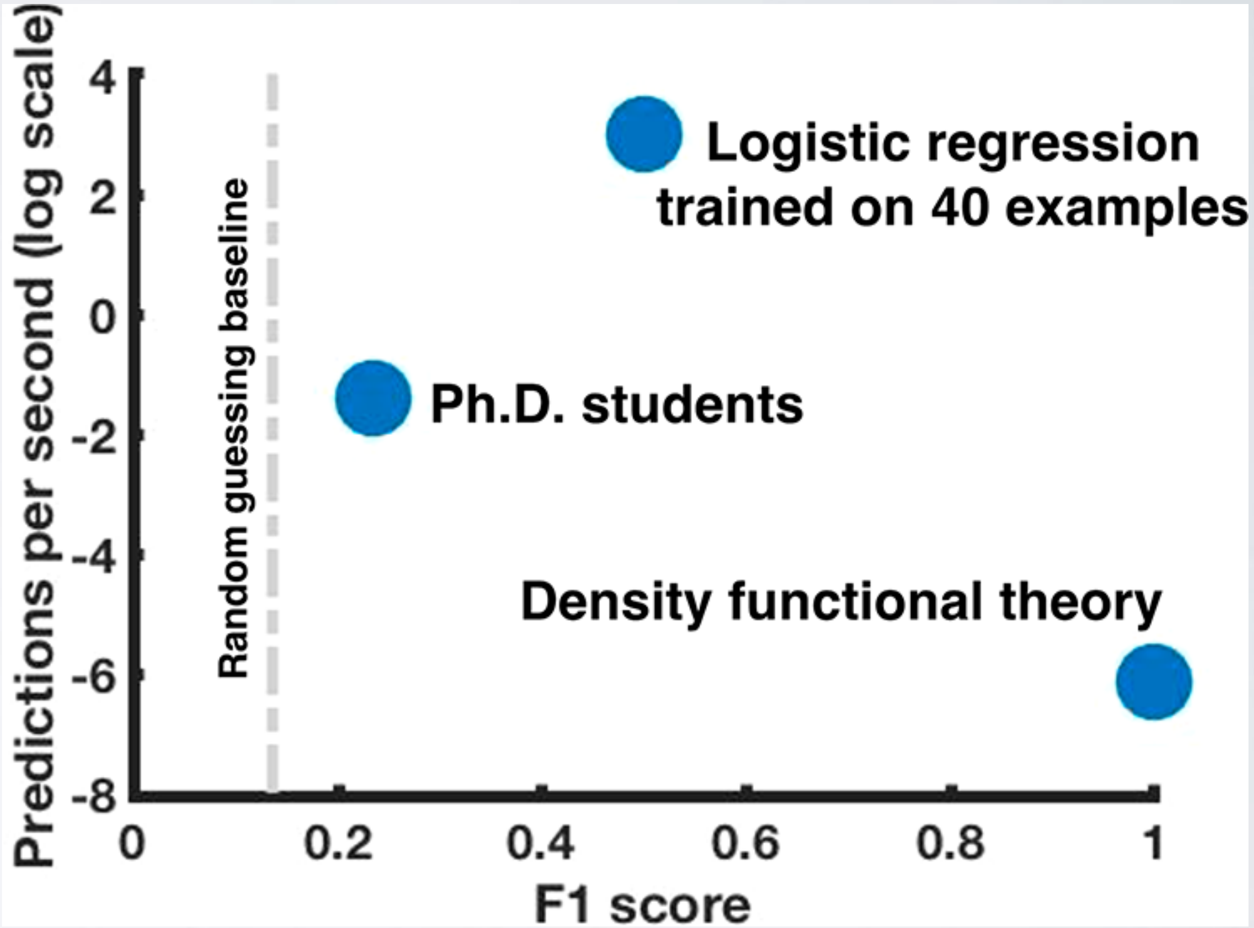


Computer vision for disease detection

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S., Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, doi:10.1038/nature21056 (2017)

THEMES

- Structure-property prediction



Predictions of superionic Li conductors at room temperature.

Sendek, A.D., Cubuk, E.D., Antoniuk, E.R., Cheon, G., Cui, Y., and Reed, E. J., Machine Learning-Assisted Discovery of Solid Li-Ion Conducting Materials. *Chemistry of Materials*, doi:10.1021/acs.chemmater.8b03272 (2018)

WHAT MAKES SCIENTIFIC MACHINE LEARNING HARD?

- Distribution of Data
- Feature Engineering
- Model Performance Requirements
- Interpretability
- Others?

ADVANTAGES IN SCIENTIFIC MACHINE LEARNING

- Problems are often highly structured
- We can often generate new data in a principled way (experiments, computation)
- Existing physical models provide initial guesses and meaningful constraints

Next Class: a whirlwind tour through ML algorithms.

Please email me with a quick response to the following before lecture Thursday (subject line: cse5095spring19survey):

1. What is your background in linear algebra, probability & statistics, and machine learning?
2. How strong is your programming background? (scale 1-3, strongest = 3)
3. What scientific applications of ML are you interested in/would you like to see explored in this class?

INTRODUCTIONS