

MACHINE-LEARNING IN CHEMISTRY

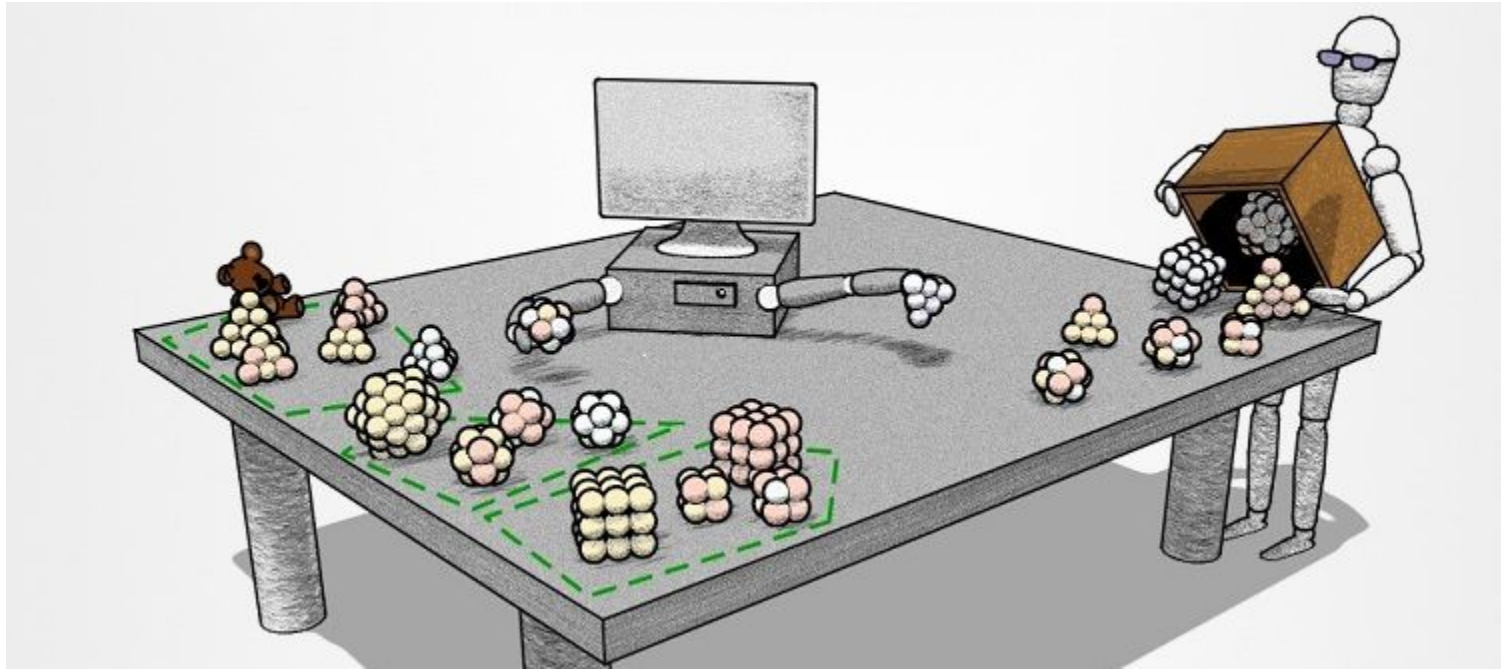
Yashasvi S. Ranawat



Filippo Federici

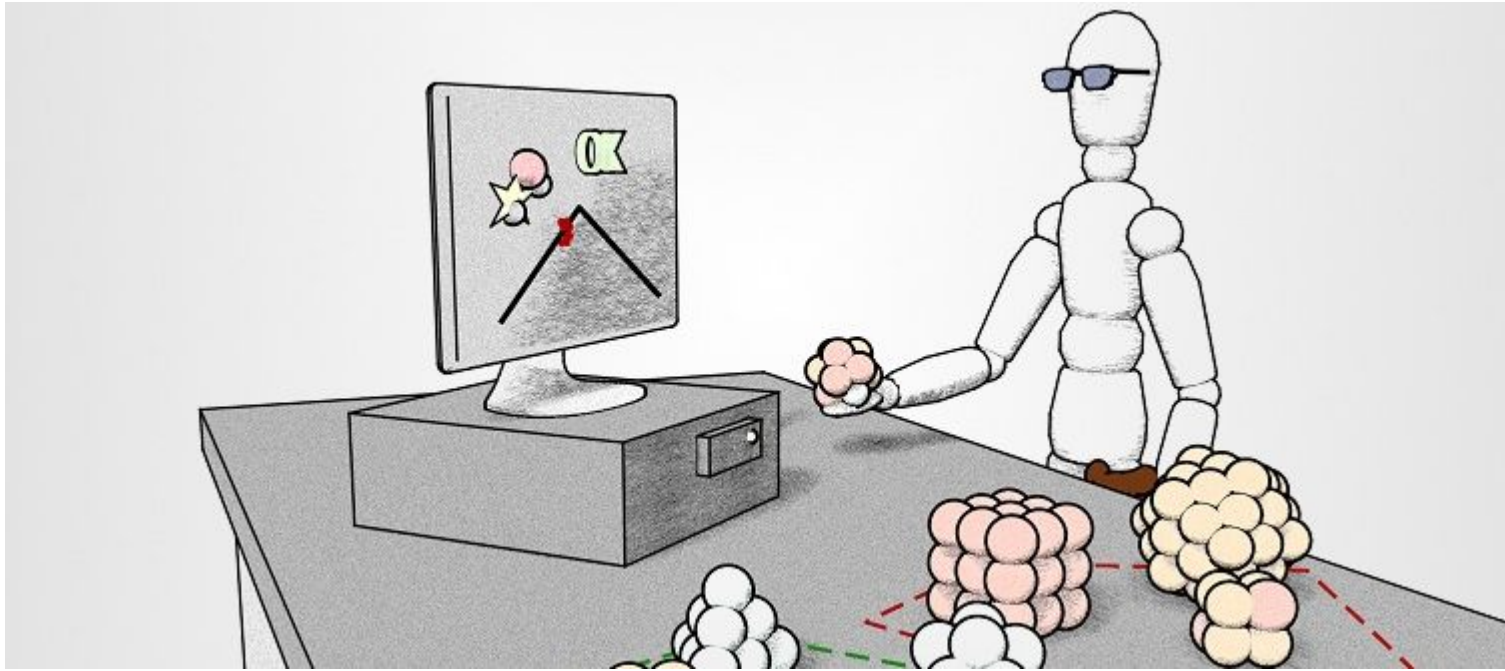


UNSUPERVISED LEARNING



- finds similarities in complex data records
- does not require knowledge of properties/outputs, only descriptors/inputs
- sensitive to the similarity measure
- requires the user to know how many classes to expect
- useful to reduce data dimensionality

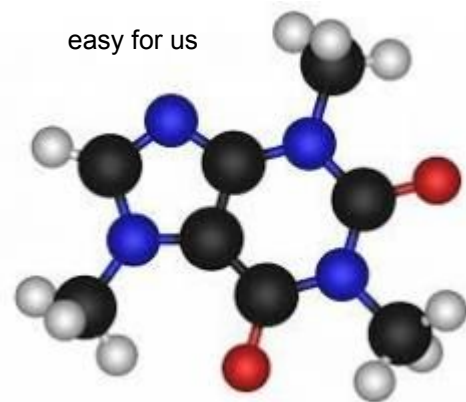
SUPERVISED LEARNING



- learns input \rightarrow output relation from examples
- training data is the limit
- useful for fast screening and classification

DESCRIPTORS FOR CHEMISTRY

ML methods need a computer-friendly way to input the atomistic system:



easy for cpu

010110101010001011100100010001111110

Issues for ML:

- arbitrary size
- arbitrary order

Ideal features:

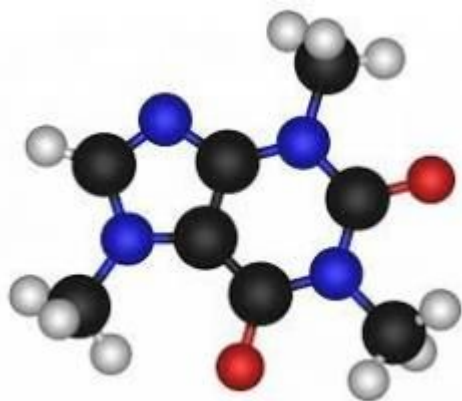
- general
- compact
- unique
- invariant *
- smooth
- fast

* invariants are determined by the physics of the quantity to predict from the descriptor!

DESCRIPTORS FOR CHEMISTRY

ML methods need a computer-friendly way to input the atomistic system:

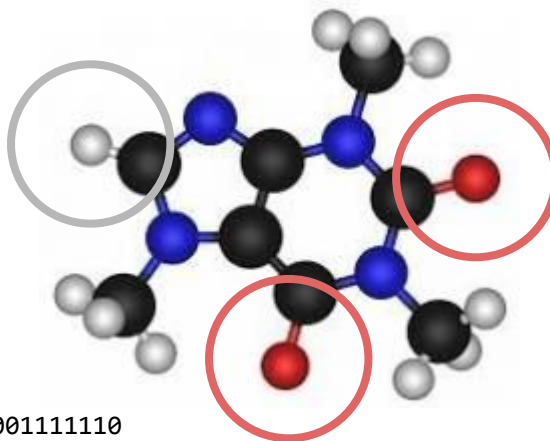
global descriptor



010110101010001011100100010001111110

local/atomic descriptor

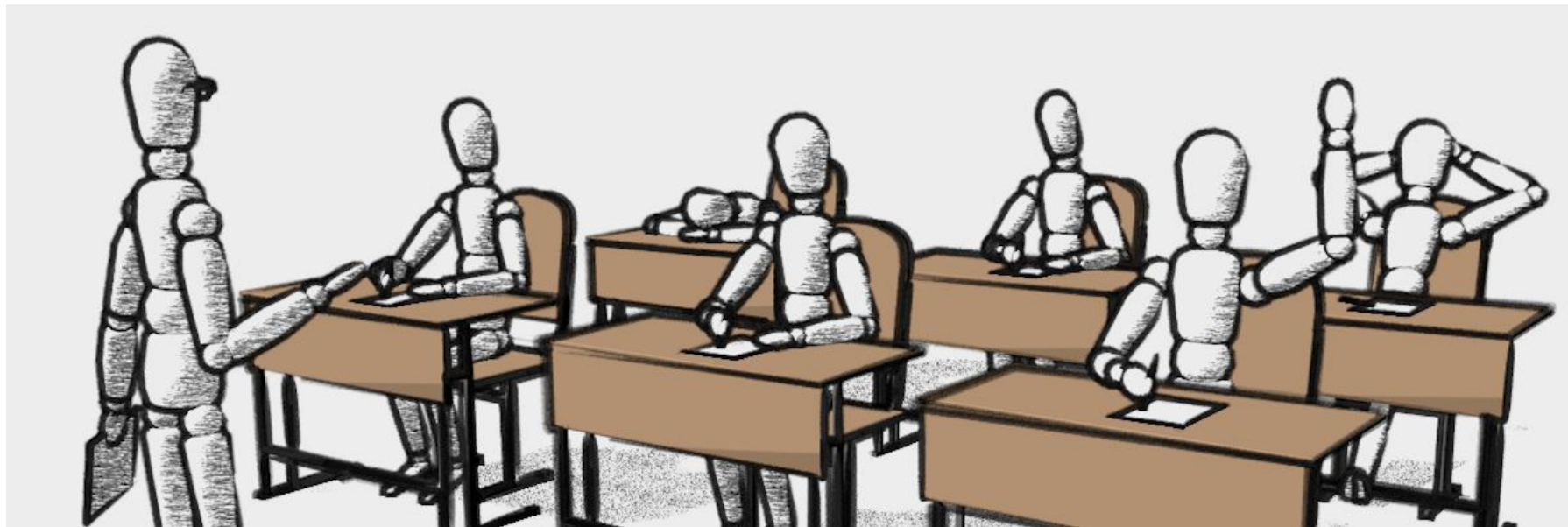
110100011110000110010111111110



1101000111100001011100001111110

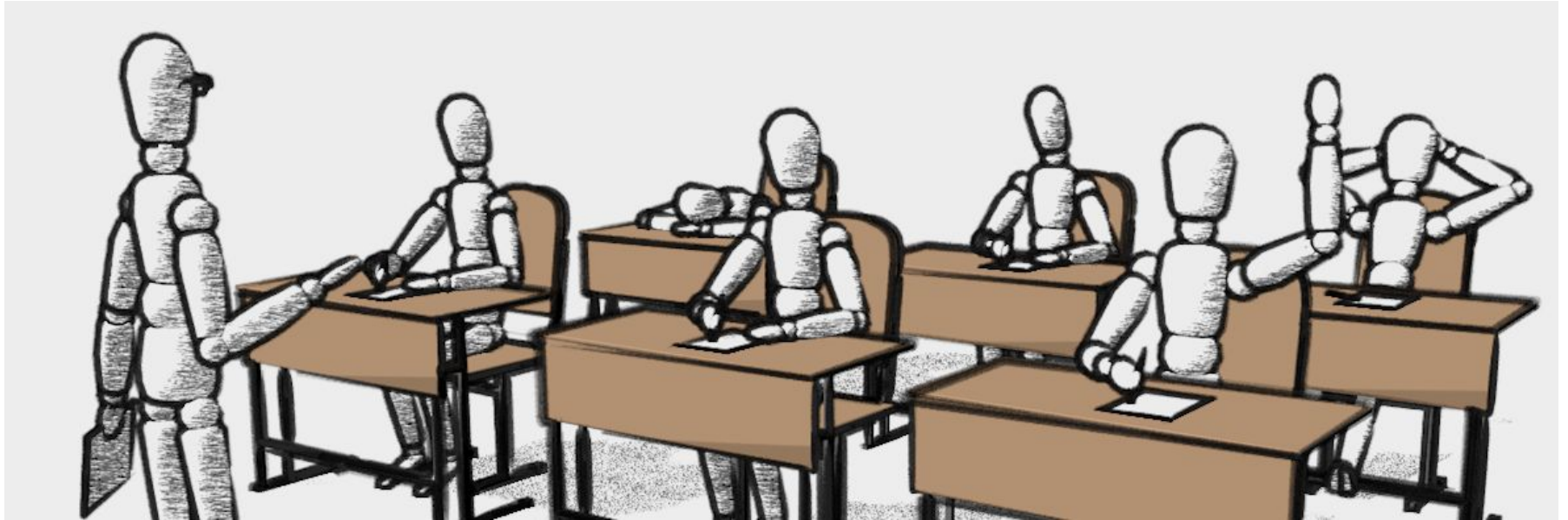
010110101010001011100001111110

DESCRIPTORS FOR CHEMISTRY



1. ACSF.ipynb
2. SOAP.ipynb
3. MBTR.ipynb
4. LMBTR.ipynb

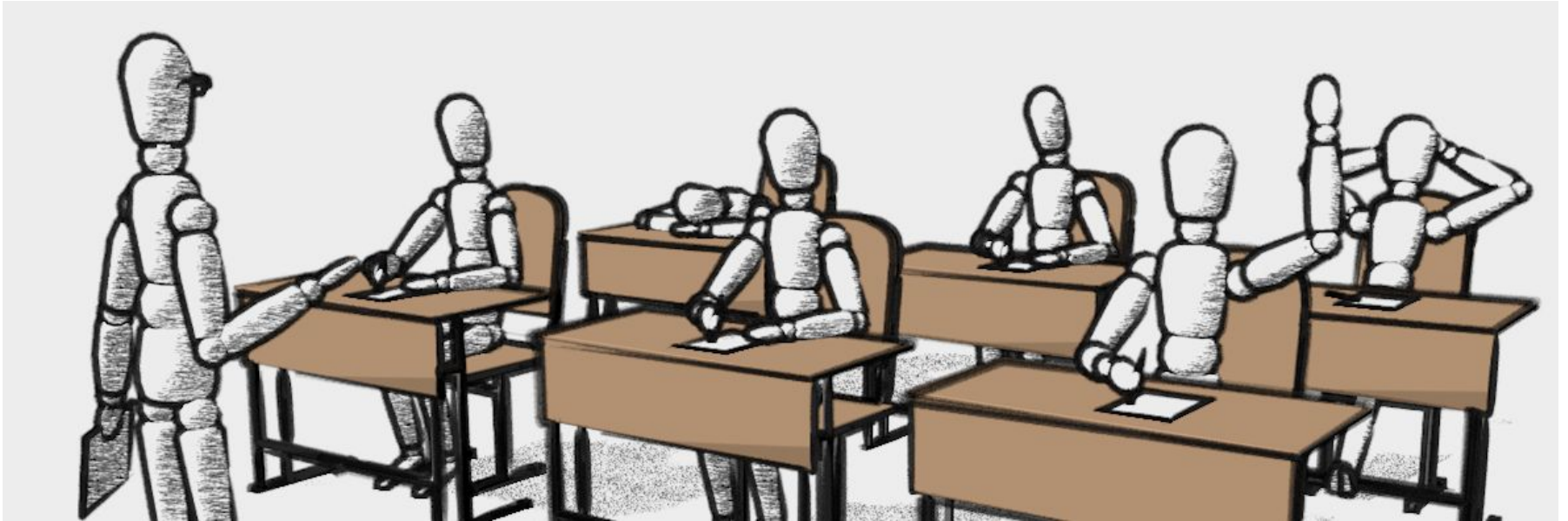
UNSUPERVISED ML METHODS



1. PCA.ipynb
2. tSNE.ipynb

SUPERVISED ML METHODS

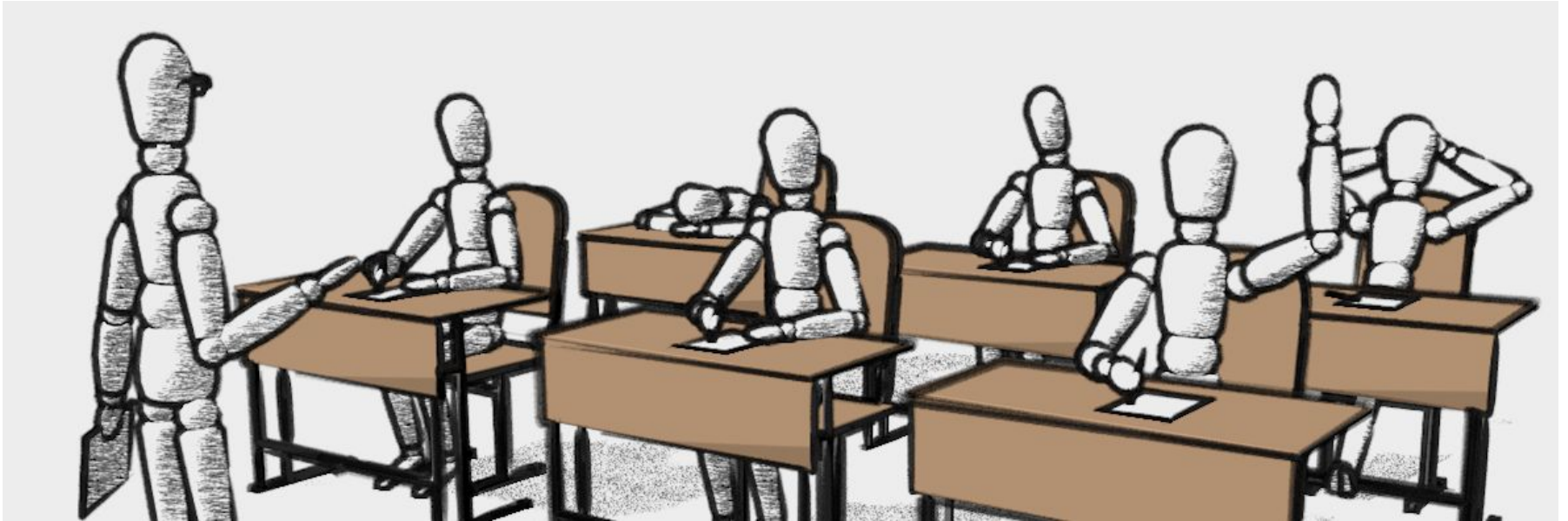
KERNEL RIDGE REGRESSION



- KRR - TotalEnergy.ipynb

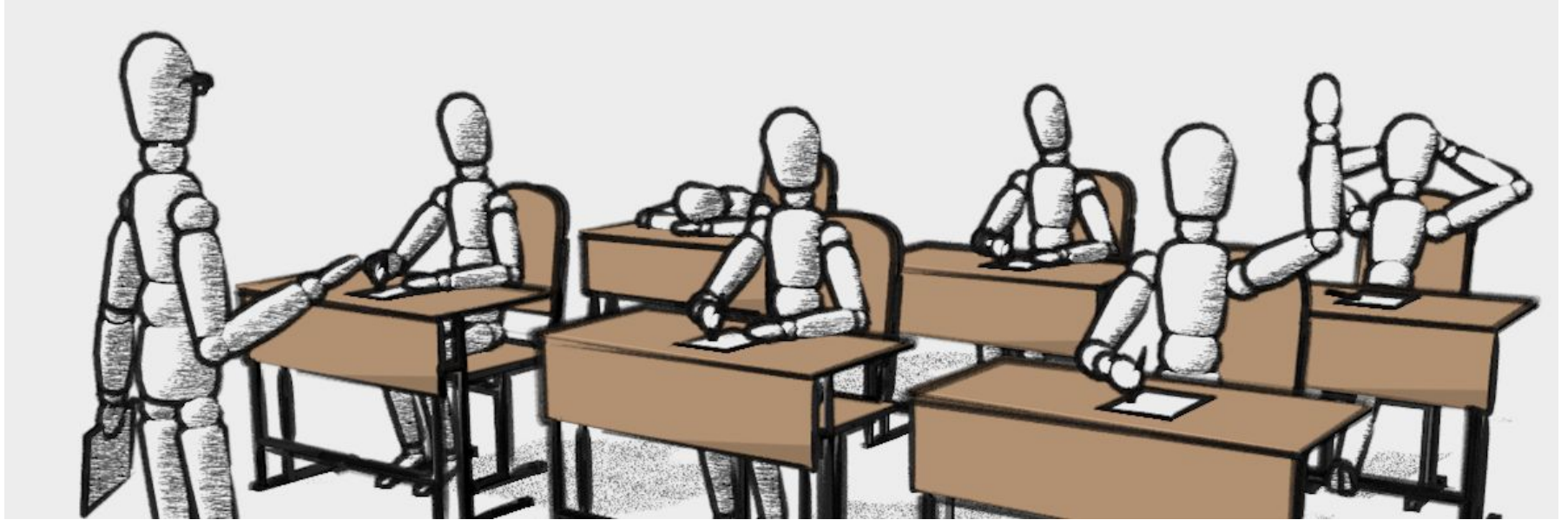
SUPERVISED ML METHODS

NEURAL NETWORKS



1. [NeuralNetwork - Intro.ipynb](#)
2. [ACSF-Dimer.ipynb](#)
3. [NeuralNetwork - TotalEnergy.ipynb](#)
4. [NeuralNetwork - AtomicCharges.ipynb](#)
5. [GenericAlgorithm.ipynb](#)

TAKEHOME



- all notebooks and data is in github: <https://github.com/fullmetalfelix/ML-CSC-tutorial>
 - notebooks require Jupyter python module
 - data in numpy array form
- useful goodies:
 - describe package: <https://github.com/SINGROUP/describe>
 - python package for creating machine learning descriptors for atomistic systems

