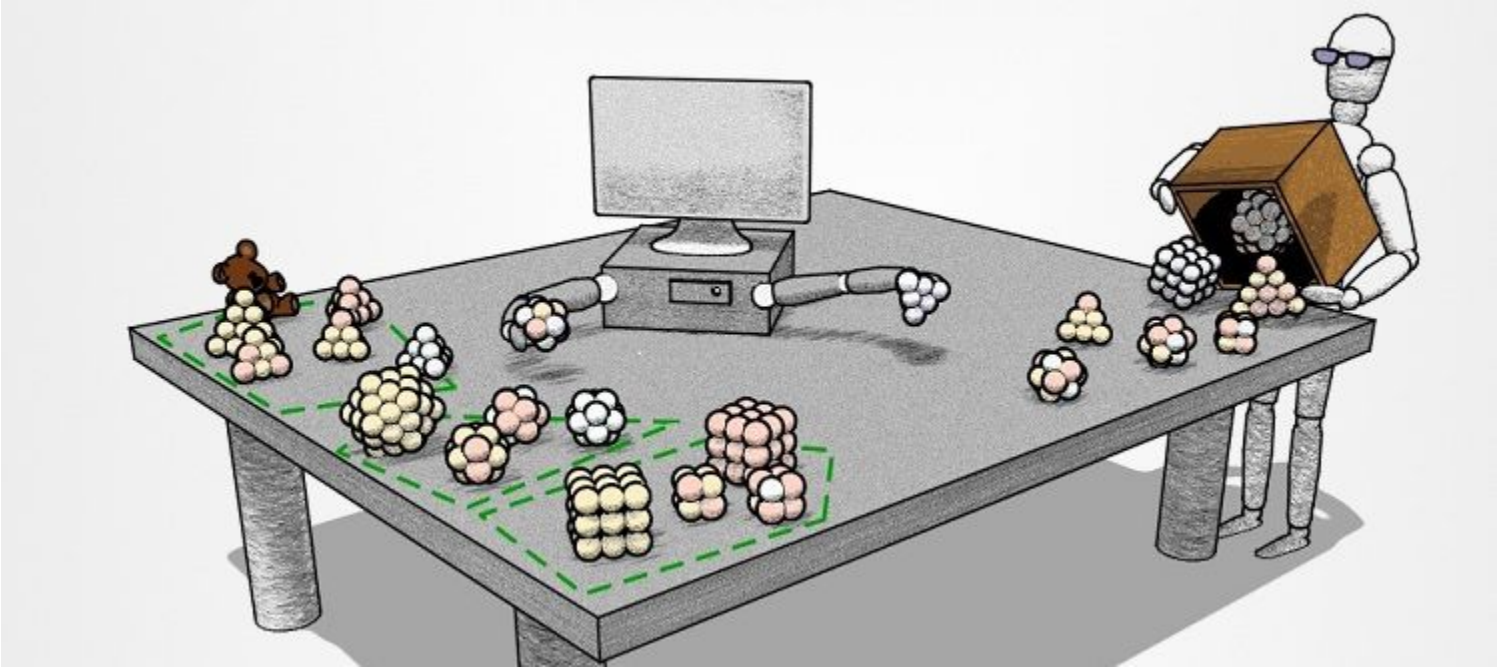
MACHINE-LEARNING IN CHEMISTRY

Yashasvi S. Ranawat

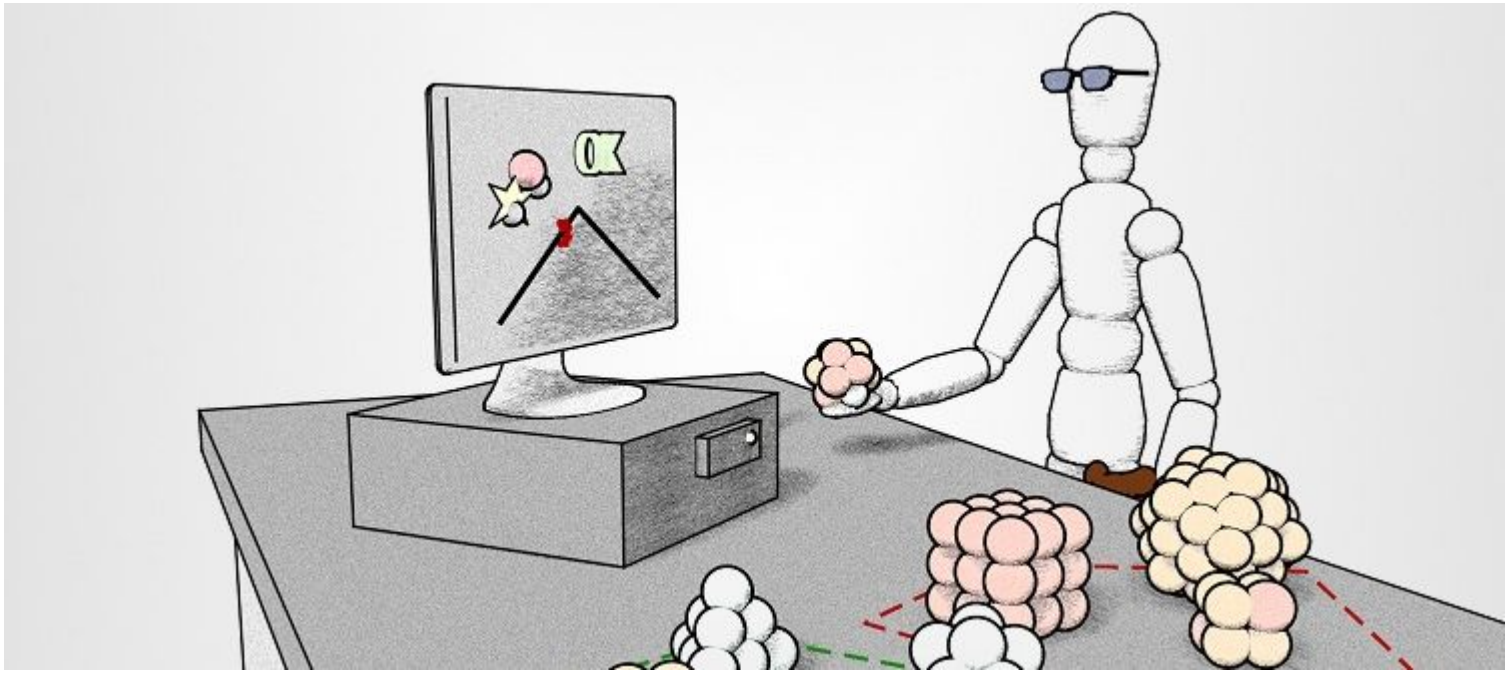
Filippo Federici

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

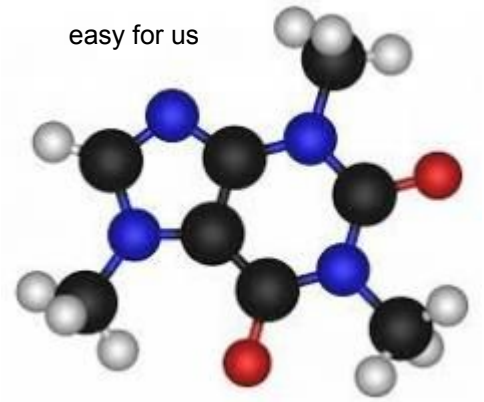
SUPERVISED LEARNING



- learns input → 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

```
0101101010100010111001000100011111110
```

Issues for ML:

- arbitrary size
- arbitrary order

Ideal features:

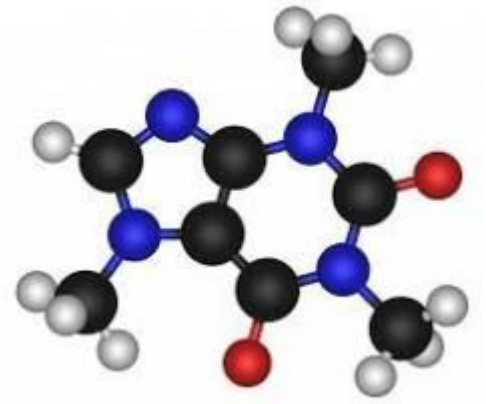
- general
- compact
- unique
- invariant *
- smooth
- fast

* 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:

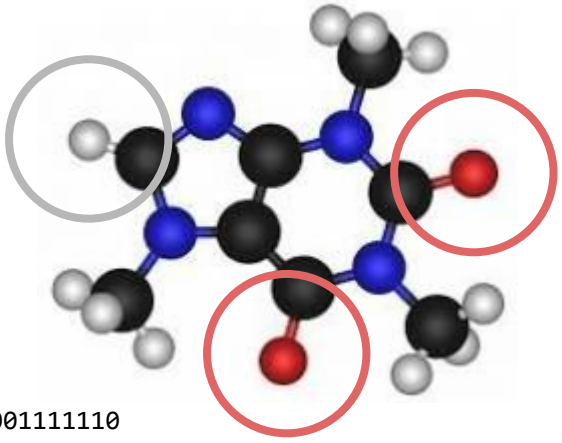
global descriptor



`01011010101000101110010001000111111O`

local/atomic descriptor

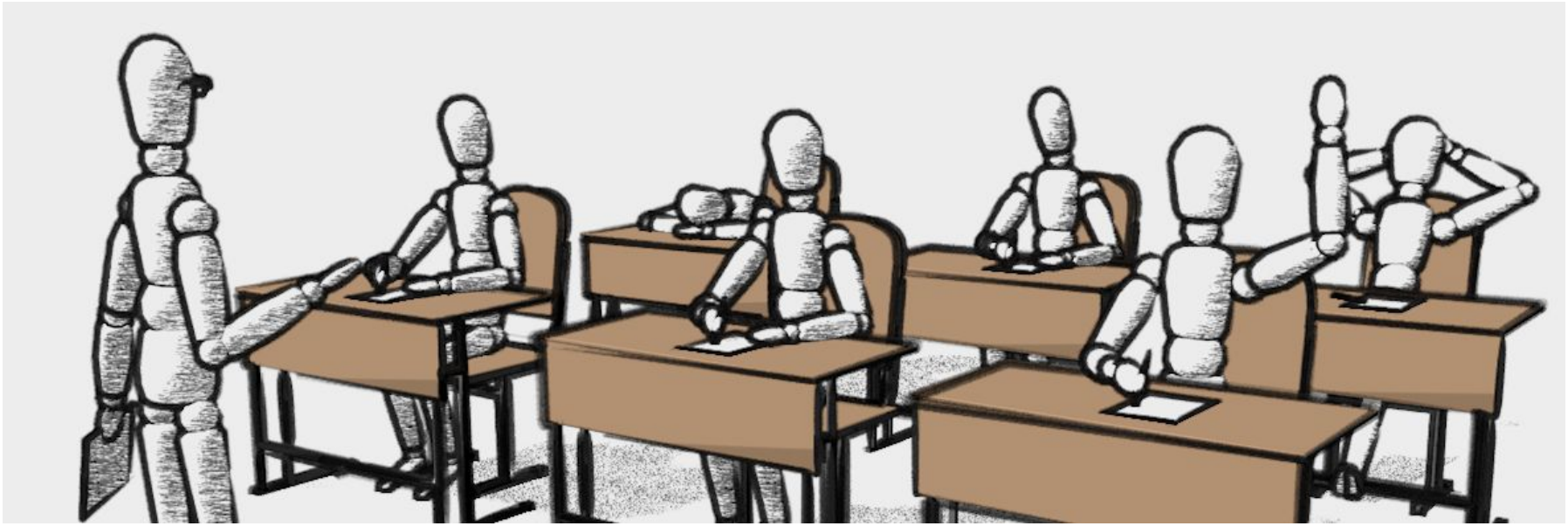
`11010001111000011001011111111O`



`11010001111000101111000011111110`

`01011010101000101111000011111110`

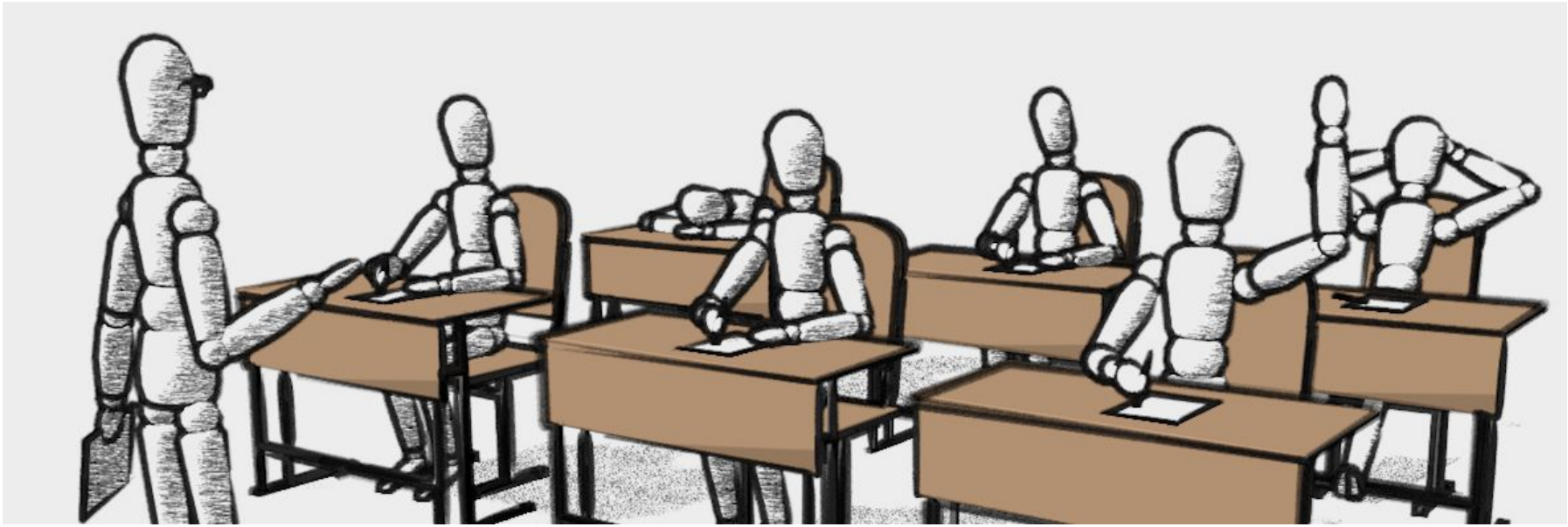
DESCRIPTORS FOR CHEMISTRY



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

SUPERVISED ML METHODS

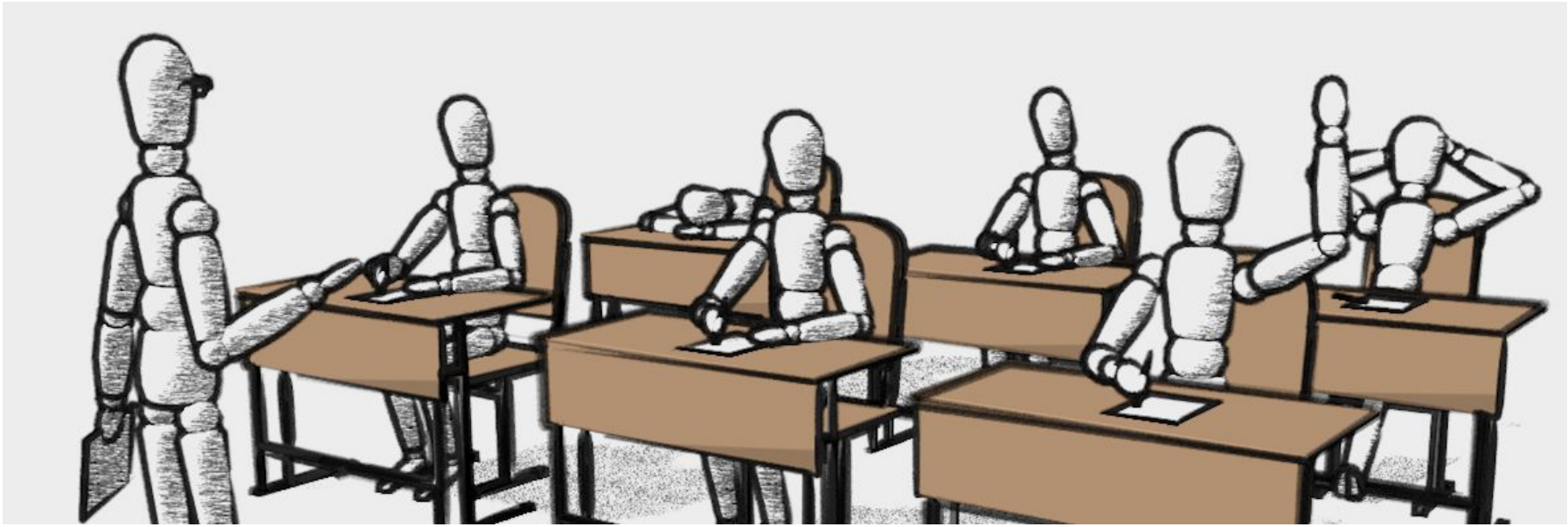
KERNEL RIDGE REGRESSION



- KRR - TotalEnergy.ipynb

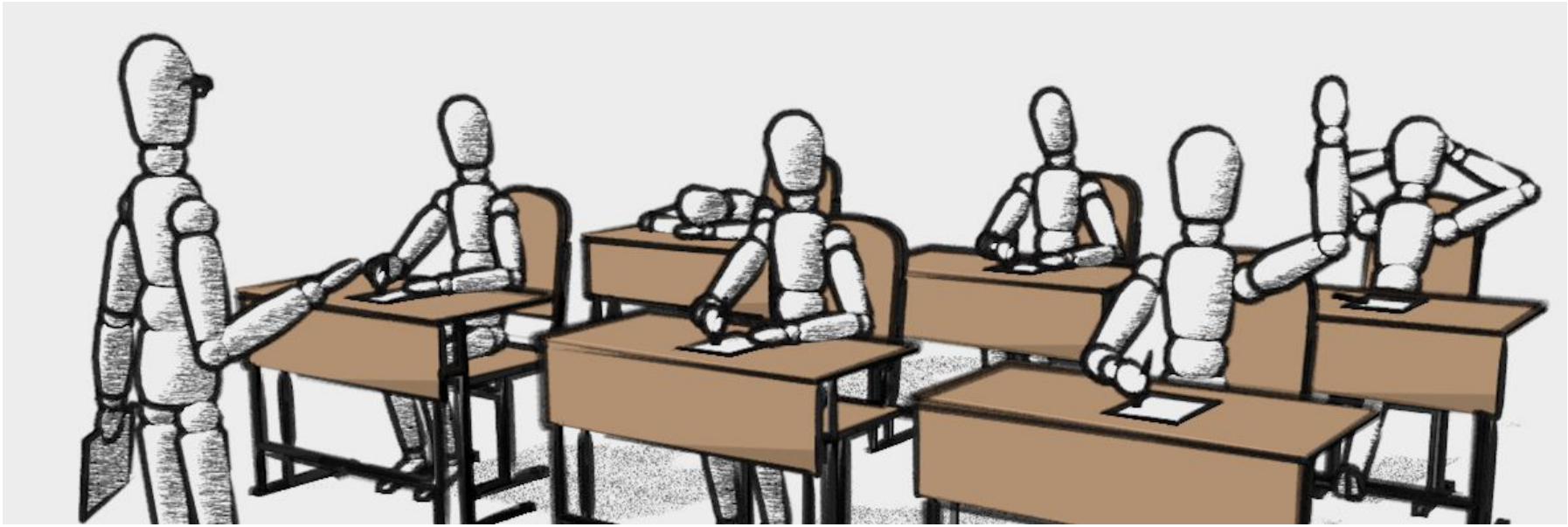
SUPERVISED ML METHODS

NEURAL NETWORKS



1. NeuralNetwork - Intro.ipynb
2. ACSF-Dimer.ipynb
3. NeuralNetwork - TotalEnergy.ipynb
4. NeuralNetwork - AtomicCharges.ipynb

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