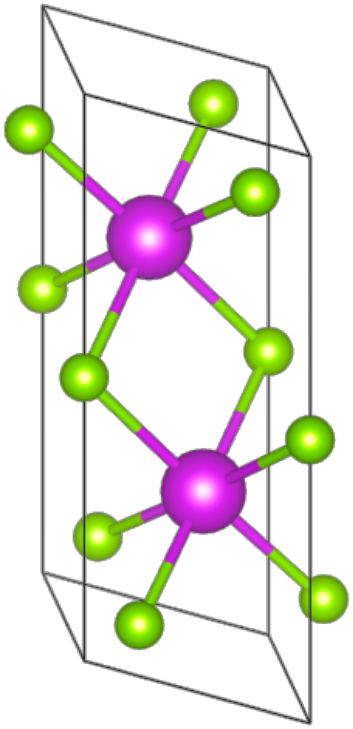
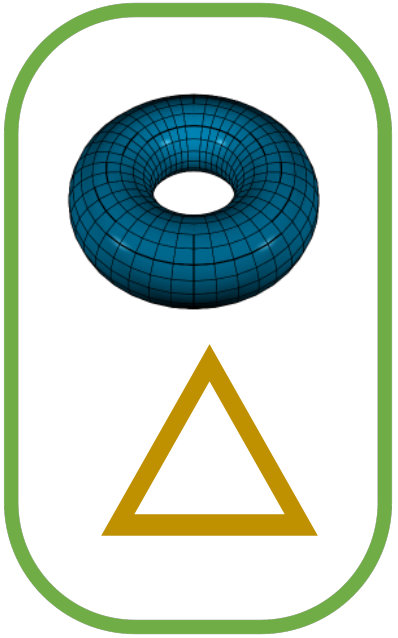


MaPPT

Materials **P**roject
Prediction **T**ool



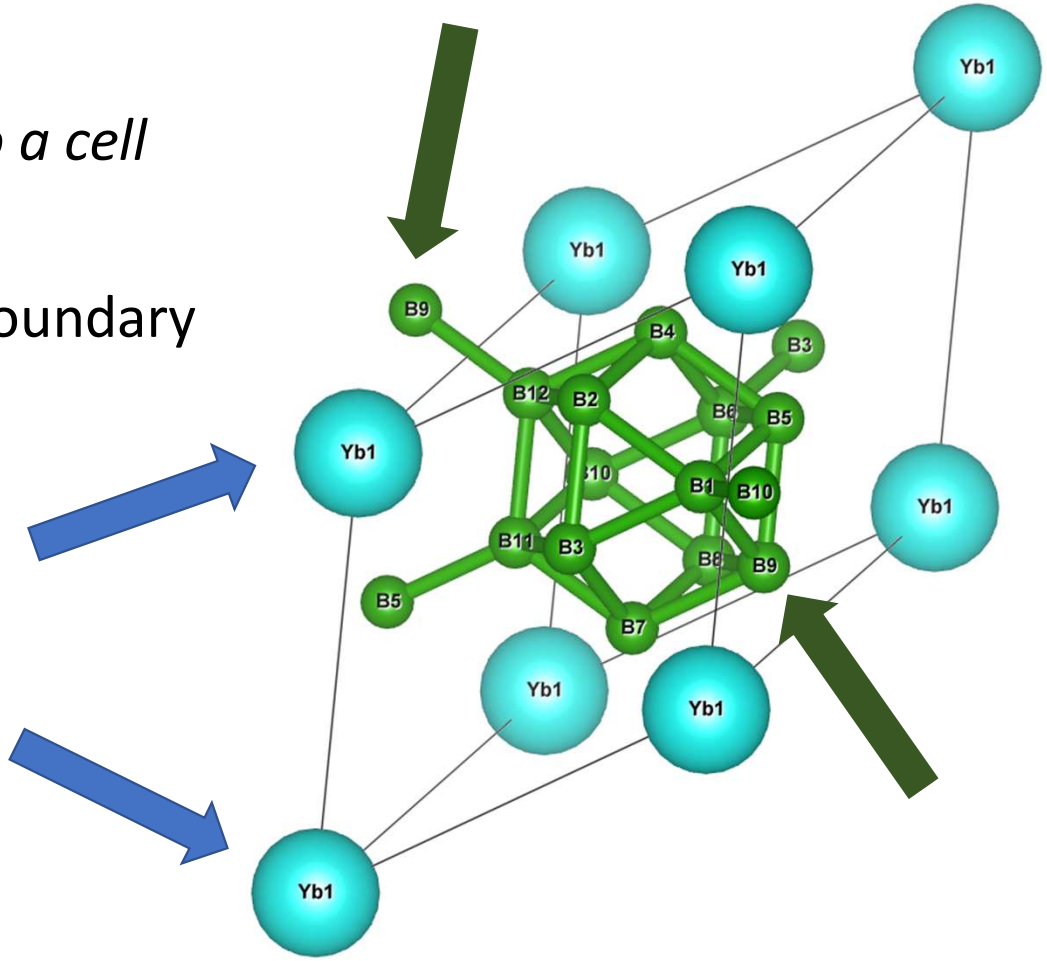
Given the crystal structure of any material, MaPPT can...

- Identify insulator vs. metal with 89% accuracy*
- Calculate the band gap with MAE = 0.565 eV*
- Identify non-trivial topology with 90% accuracy*

*When compared to first principles calculations

What is a crystal structure?

- A collection of atoms *confined to a cell*
- Physics terminology: “periodic boundary conditions”
- Ex: YbB_{12}



What is a band gap?

- Quantum mechanically, crystal structures = **periodic potentials**

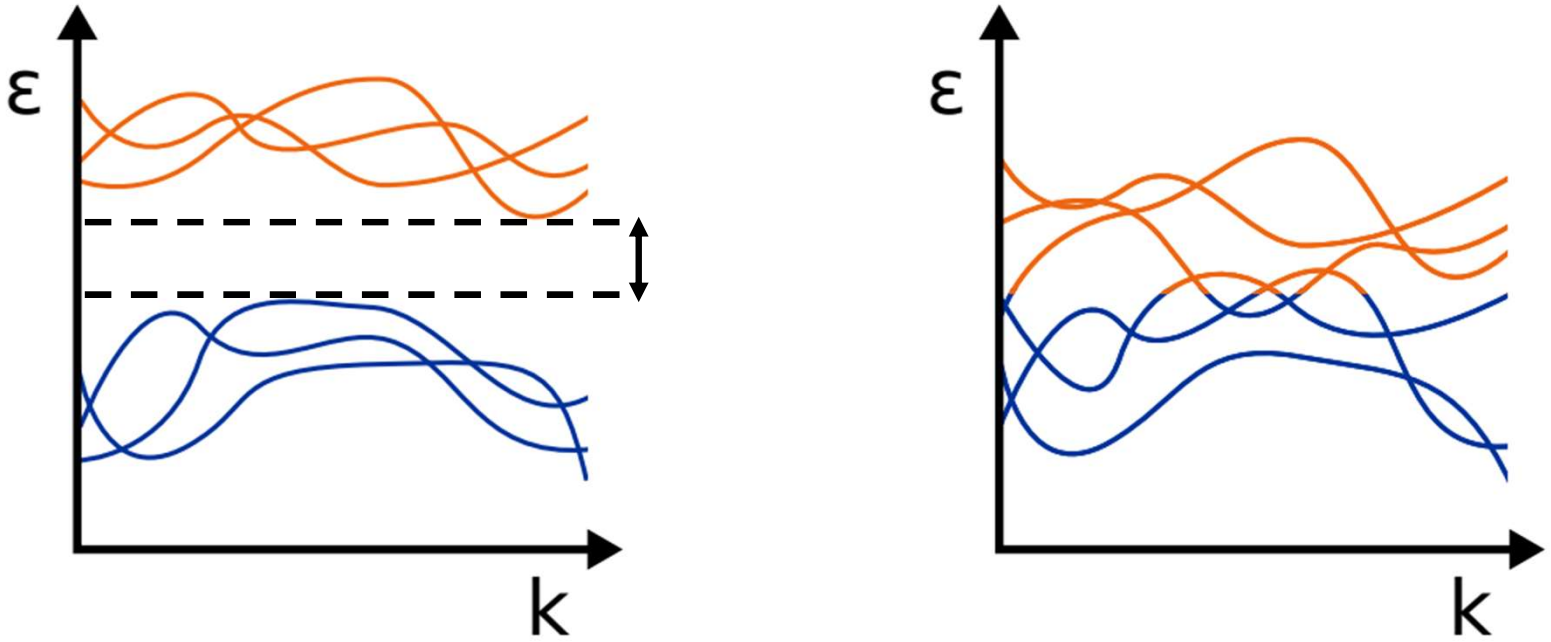
- Eigenvalue equation: $H(k)\left|\Psi_{n,k}\right\rangle = \varepsilon_{n,k}\left|\Psi_{n,k}\right\rangle$

Energy spectrum is called

band structure

Statistical mechanics: electrons occupy up to Fermi energy

$\Longrightarrow$ bands either full, empty, or partially filled



But why do we care?

Diagrams courtesy of Gresch and Soluyanov, see http://z2pack.ethz.ch/

Because it can be shown[1] that **only partial filled bands contribute to conduction.**

$$\text{gap} > 0 \implies \text{insulator}$$

$$\text{gap} = 0 \implies \text{metal}$$

[1]Ashcroft and Mermin, *Solid State Physics*

What is topology?

$$H(k)\left|\Psi_{n,k}\right\rangle = \varepsilon_{n,k}\left|\Psi_{n,k}\right\rangle$$

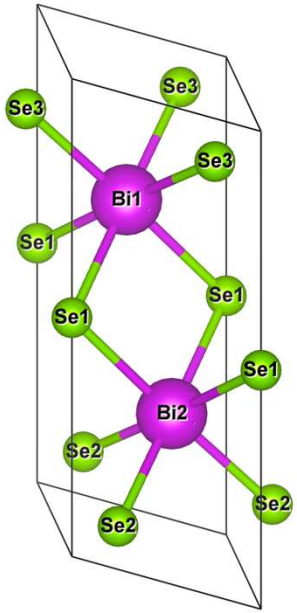
It turns out that materials have additional *physical* properties **not captured by** $\varepsilon_{n,k}$

Considering the "shape" of each $\left|\Psi_{n,k}\right\rangle$ gives rise to a **classification scheme**[1]

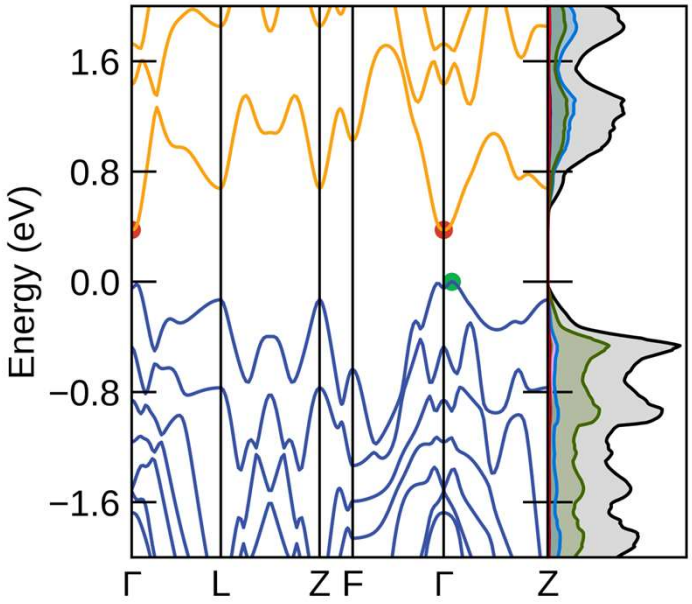
[1]Gresh and Soluyanov, http://z2pack.ethz.ch/doc/2.2/other_material.html

What are first principles calculations?

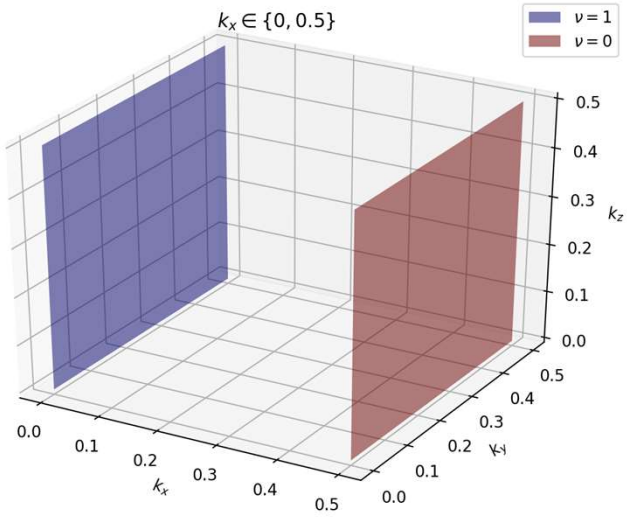
- Density functional theory (DFT) to get band gap
- Wannier charge center evolution to get topology



$\varepsilon_{n,k}$



$|\Psi_{n,k}\rangle$



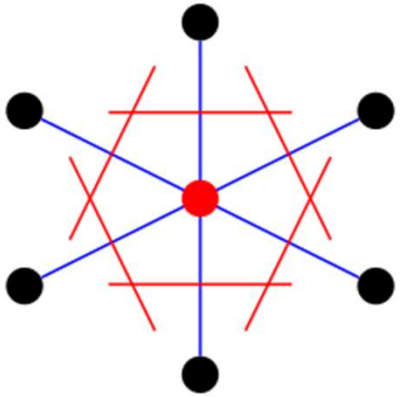
What about ML?

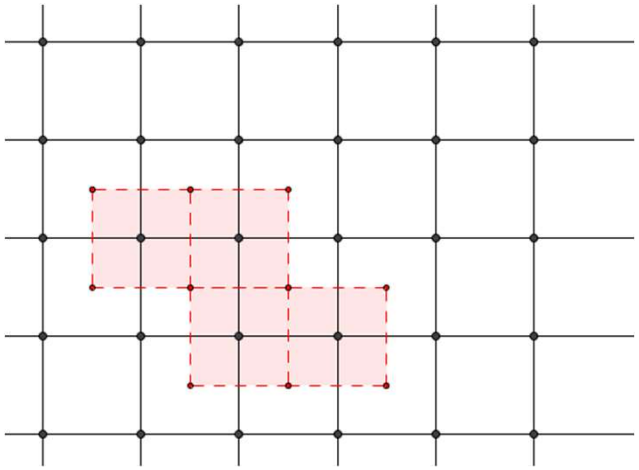
The current disadvantages of DFT have been well-documented

- magnetic materials
- *f*-electron materials
- the N^4 problem

But *featurizing* a crystal structure for ML is not so straightforward.

Voronoi tessellations





Phys. Rev. B 96, 024104 (2017)

$$CN = \frac{\left(\sum_n A_n\right)^2}{\sum_n {A_n}^2}$$

Diagrams courtesy of https://en.wikipedia.org/wiki/Wigner-Seitz_cell

The dataset

- Input: *pymatgen* structure object
- Outputs:
  - PBE band gap (accessible through API)
  - Topology (requires HTML scraping)
- Features: *matminer* preset
- 76,891 materials
  - 45,540 insulators
  - 31,351 metals

ML strategy

Need three models

- Two classifiers (insulator/metal, trivial/non-trivial)
- One regressor (band gap)

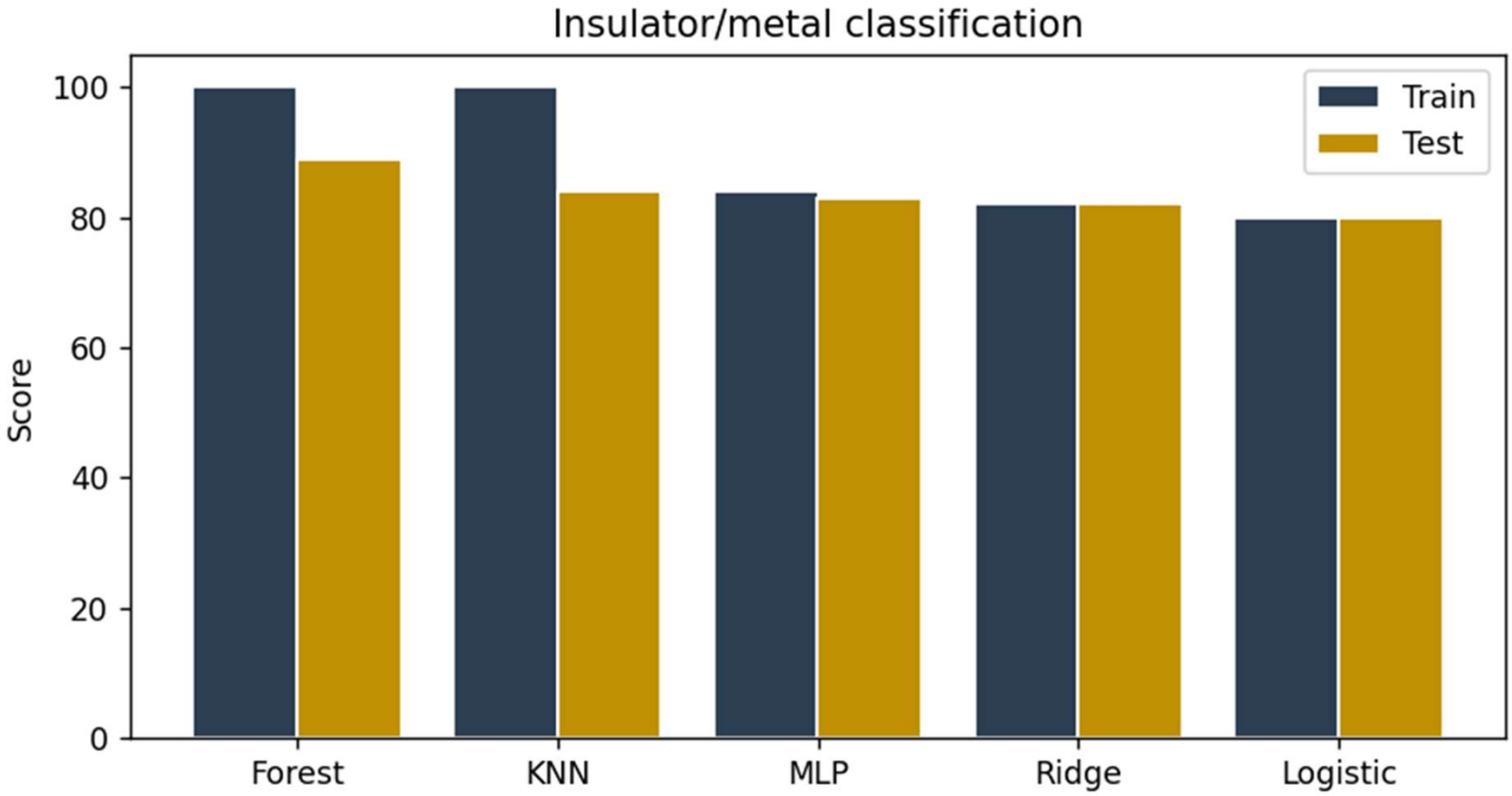
85/15 train-test split

- Training data: Pearson cutoff of 0.75

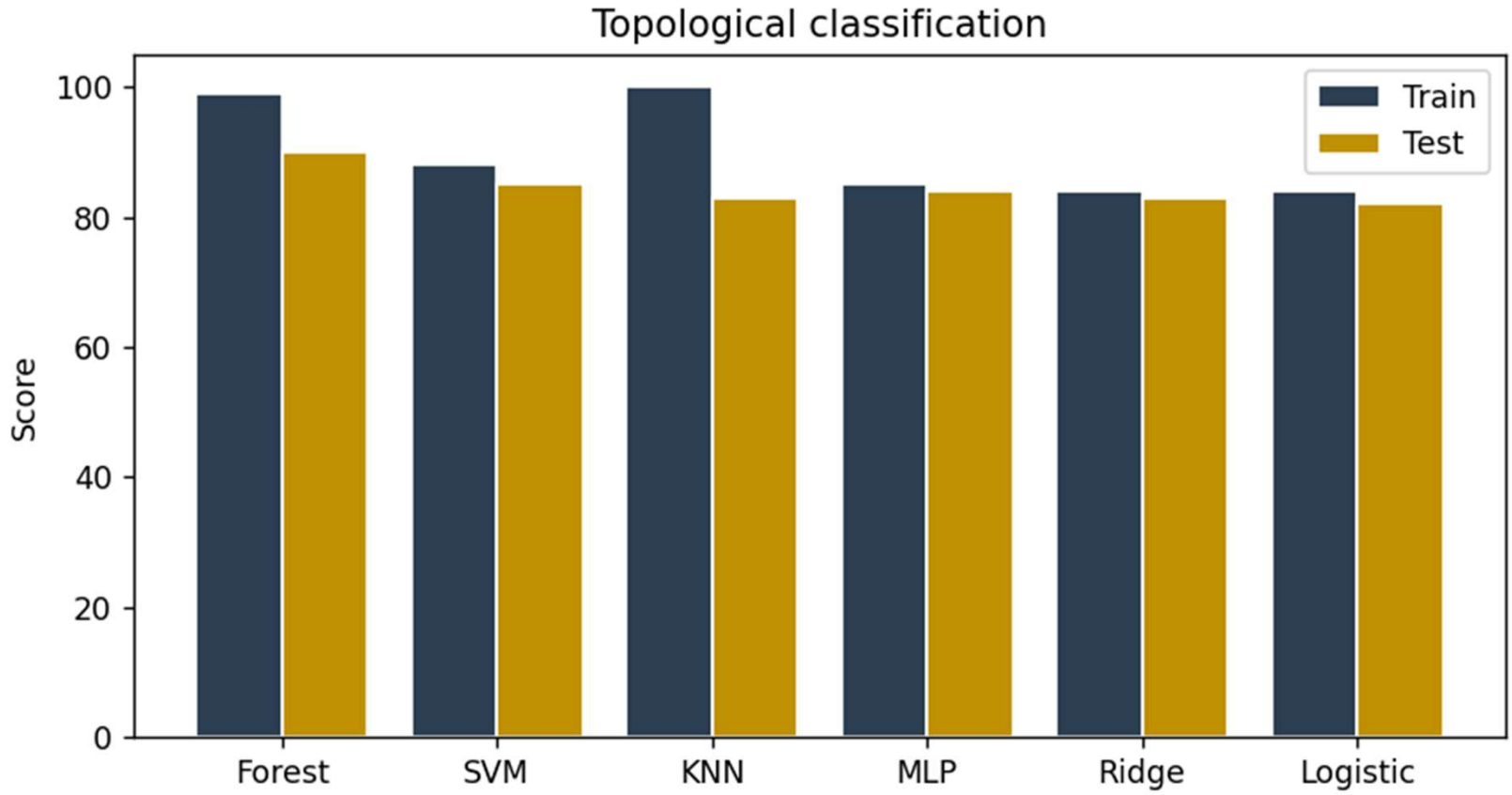
Consider linear and non-linear models

Standard hyperparameter tuning

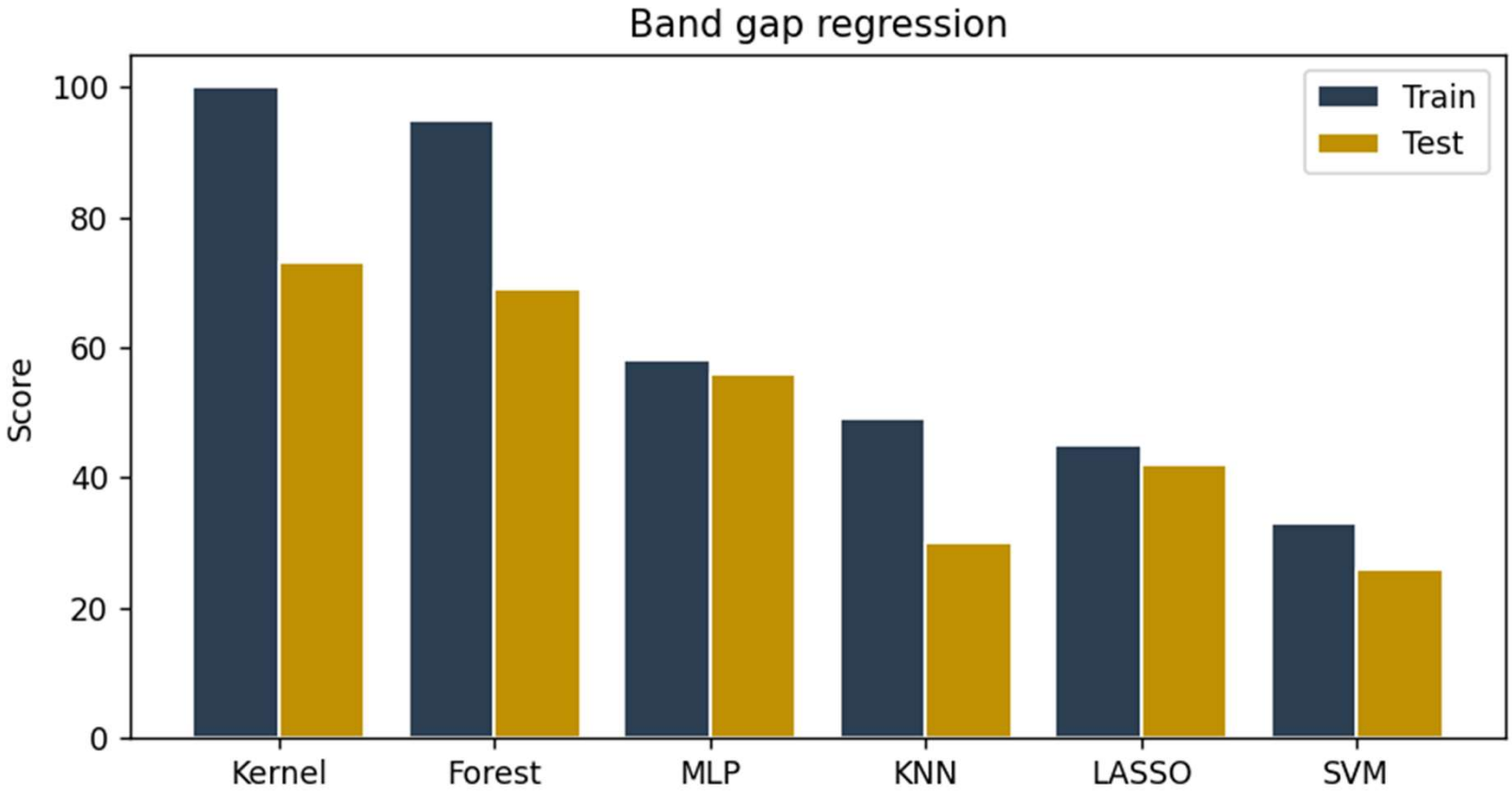
- Check accuracy converges w.r.t number of iterations



Random forest: 1500 trees, 4 min. samples, 50 max. depth

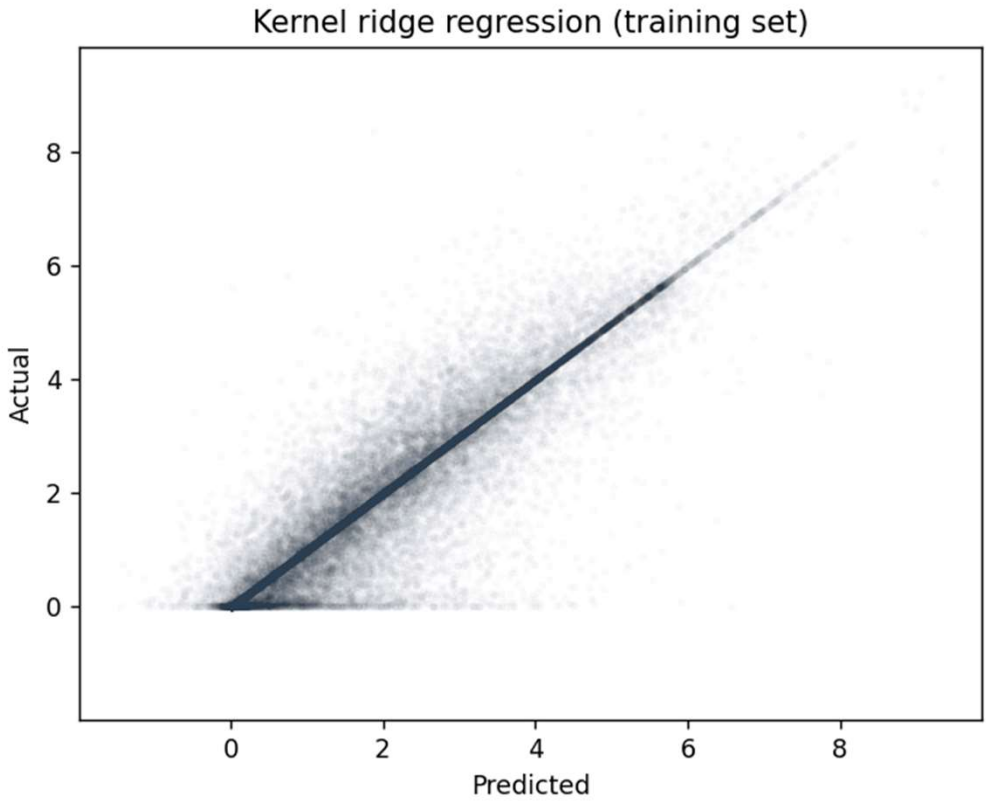


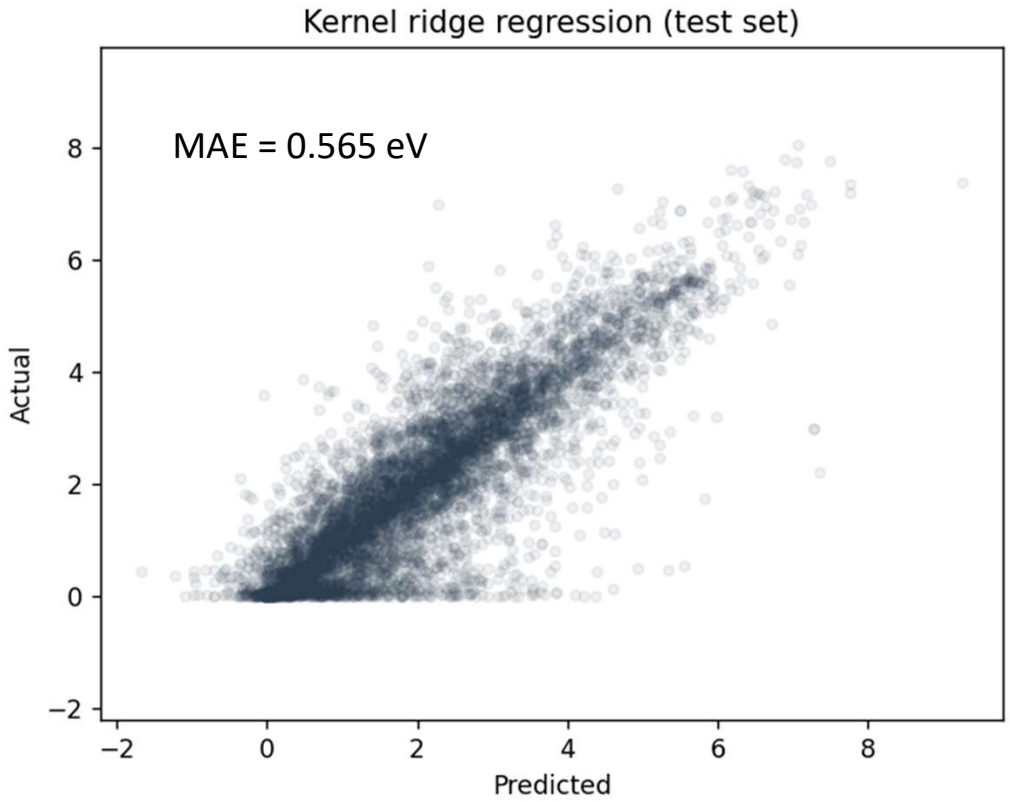
Random forest: 200 trees, 6 min. samples, no max. depth



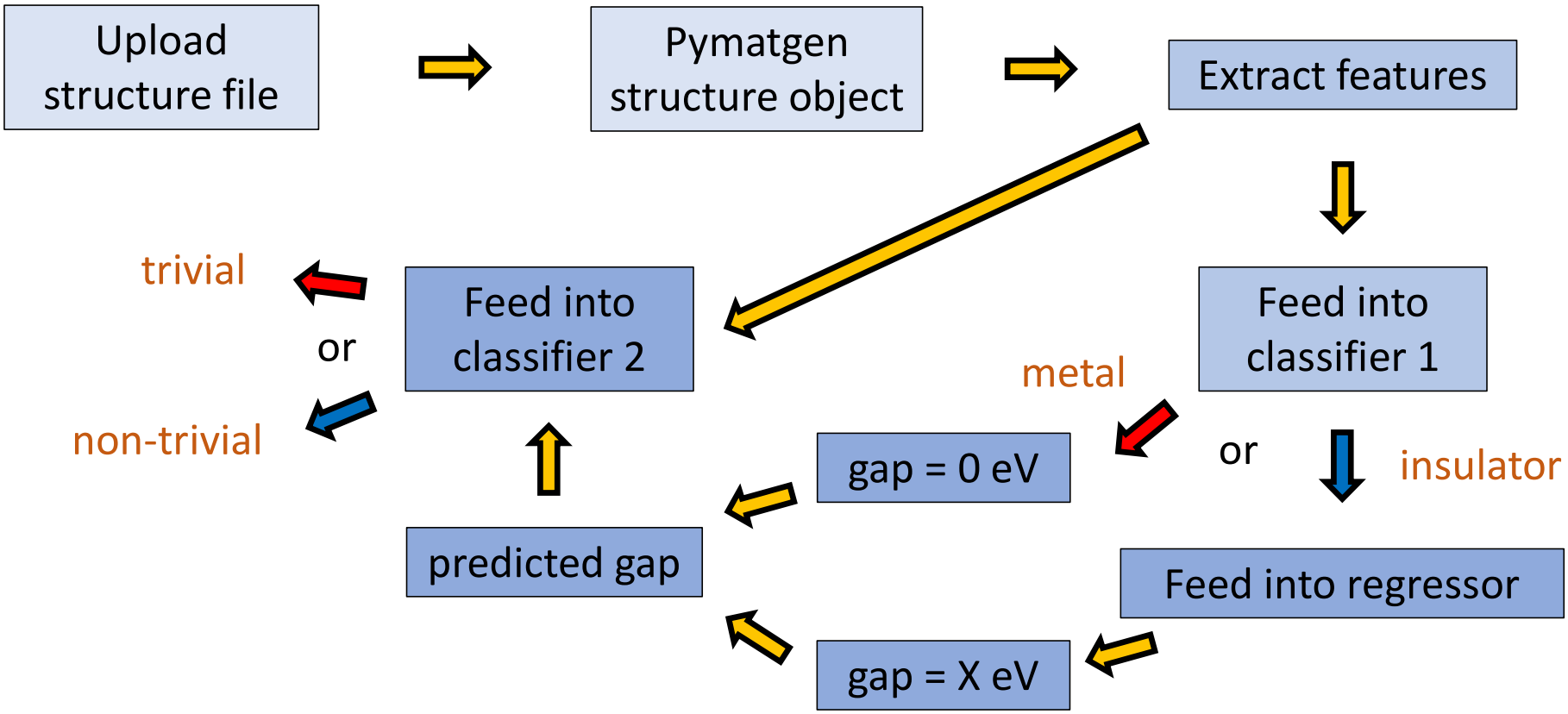
Kernel ridge: Laplacian, $\alpha = 0.0001$, $\gamma = 0.0017$

Band gap (eV)


Kernel ridge regression (training set)
Actual
Predicted
0
2
4
6
8


Kernel ridge regression (test set)
MAE = 0.565 eV
Actual
Predicted
−2
0
2
4
6
8

Demonstration


Upload structure file
Pymatgen structure object
Extract features
Feed into classifier 1
Feed into classifier 2
trivial
or
non-trivial
metal
or
insulator
gap = 0 eV
predicted gap
Feed into regressor
gap = X eV

Outlook

- Match the accuracy of DFT
  - Explore deep learning techniques

- DFT development: predict experimental values instead
  - Promising databases already exist[1]

- Predict full topological phases

[1] https://htem.nrel.gov/#/about

Thank you!