Machine Learning Modeling of Superconducting Critical Temperature

Valentin Stanev
University of Maryland, College Park

Machine Learning for Materials Research Bootcamp
August 2, 2018

Outline

- Short presentation explaining the idea, the data and the ML method
- Three notebooks
  - Quick look at the data
  - Classification models
  - Regression models

V. Stanev, C. Oses, A. Kusne, E. Rodriguez, JP. Paglione, S. Curtarolo, I. Takeuchi,
Machine learning modeling of superconducting critical temperature,
npj Computational Materials 4 (1), 29 (2018)

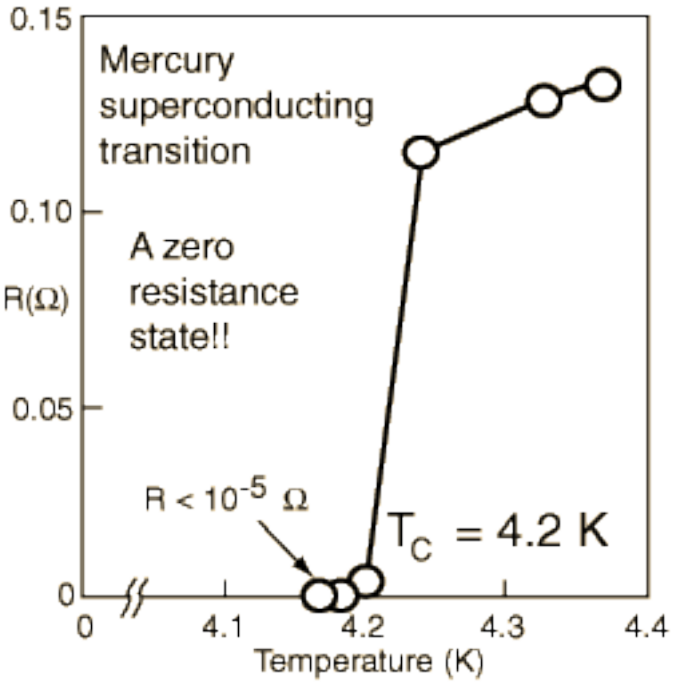
Motivation – what and why

Superconductivity – macroscopic quantum state (below critical temperature T_c)

Apply Machine Learning methods to superconductivity:

1) Better understand superconductivity in known materials – find connection between materials parameters and T_c
2) Search for superconducting materials?

We need data!



Critical temperature data

MatNavi SuperCon database – experimentally measured T_c (http://supercon.nims.go.jp/index_en.html)

~ 22,000 compounds in the database;

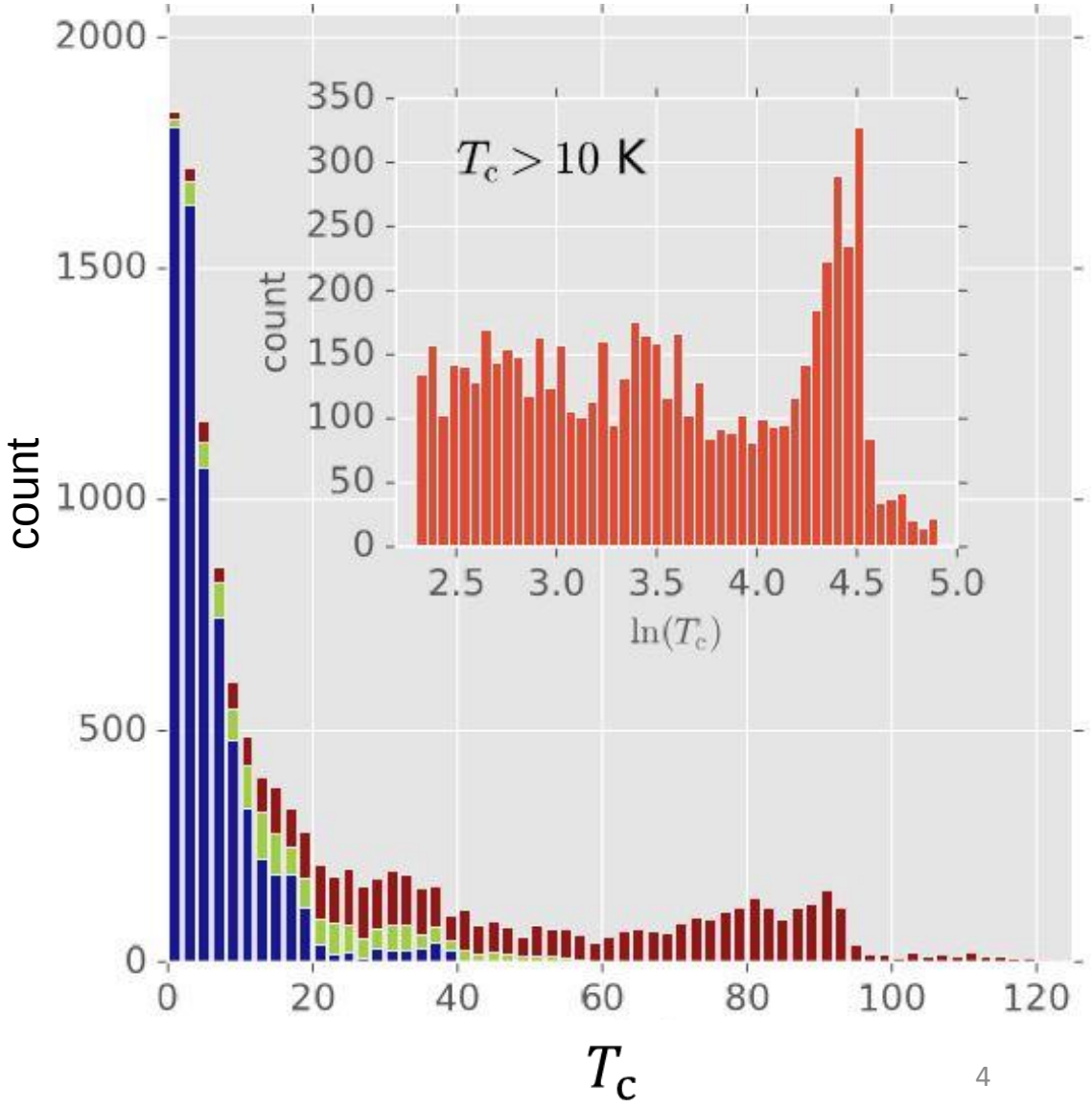
~16,400 after combining/removing the duplicates;

~12,400 with T_c plus about 4000 without T_c (what should we do with them?)

5700 cuprates (35%)

1500 Fe-based (9%)

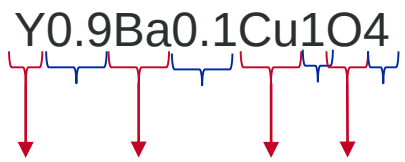
MatNavi
NIMS 物質・材料データベース



Predictors for T_c

We have only composition and T_c. Need some predictors!

Use Magpie to generate elemental features



elemental properties

composite property:
average, range, std, max, min

std(Electronegativity) = standard deviation of elemental electronegativities

A general-purpose machine learning framework for predicting properties of inorganic materials

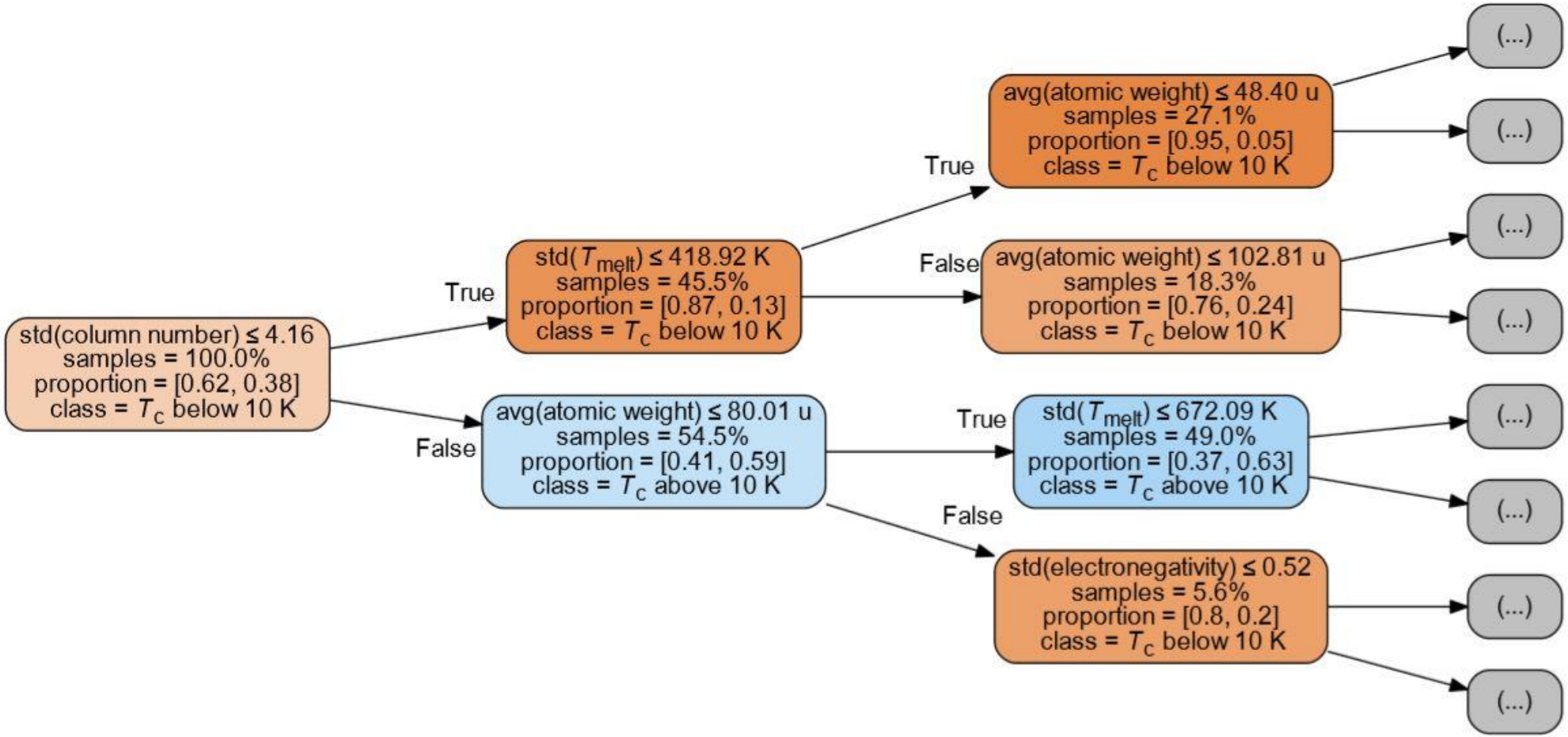
Logan Ward[1], Ankit Agrawal[2], Alok Choudhary[2] and Christopher Wolverton[1]

npj Computational Materials, (2016)

Atomic Number	Mendeleev Number[9]	Atomic Weight	Melting Temperature	Column
Row	Covalent Radius	Electronegativity*	# s Valence Electrons	# p Valence Electrons
# d Valence Electrons	# f Valence Electrons	Total # Valance Electrons	# Unfilled s States†	# Unfilled p States†
# Unfilled d States†	# Unfilled f States†	Total # Unfilled States†	Specific Volume of 0 K Ground State‡	Band Gap Energy of 0 K Ground State‡
Magnetic Moment (per atom) of 0 K ground state‡		Space Group Number of 0 K Ground State‡		

Random forest I

Now we have predictors and target (T_c). Let's create ML model.
Use decision trees – can be used for both regression and classification.



Random forest II

Now we have predictors and target (T_c). Let's create ML model.
Use decision trees – can be used for both regression and classification.

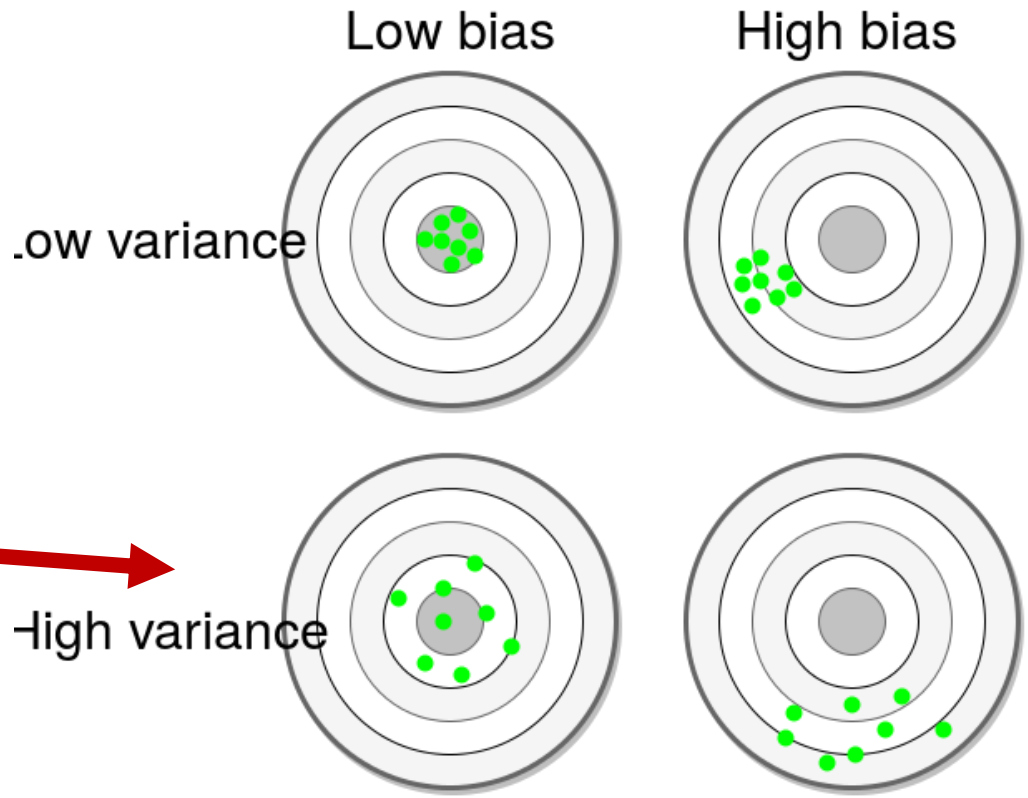
Pros:

- don't require pre-processing of the data
- can learn very non-linear functions
- can be inspected/interpreted

Cons:

- Easy to overfit – tend to be low bias/high variance

Use averaging to reduce the variance!
Bootstrap aggregation (bagging)

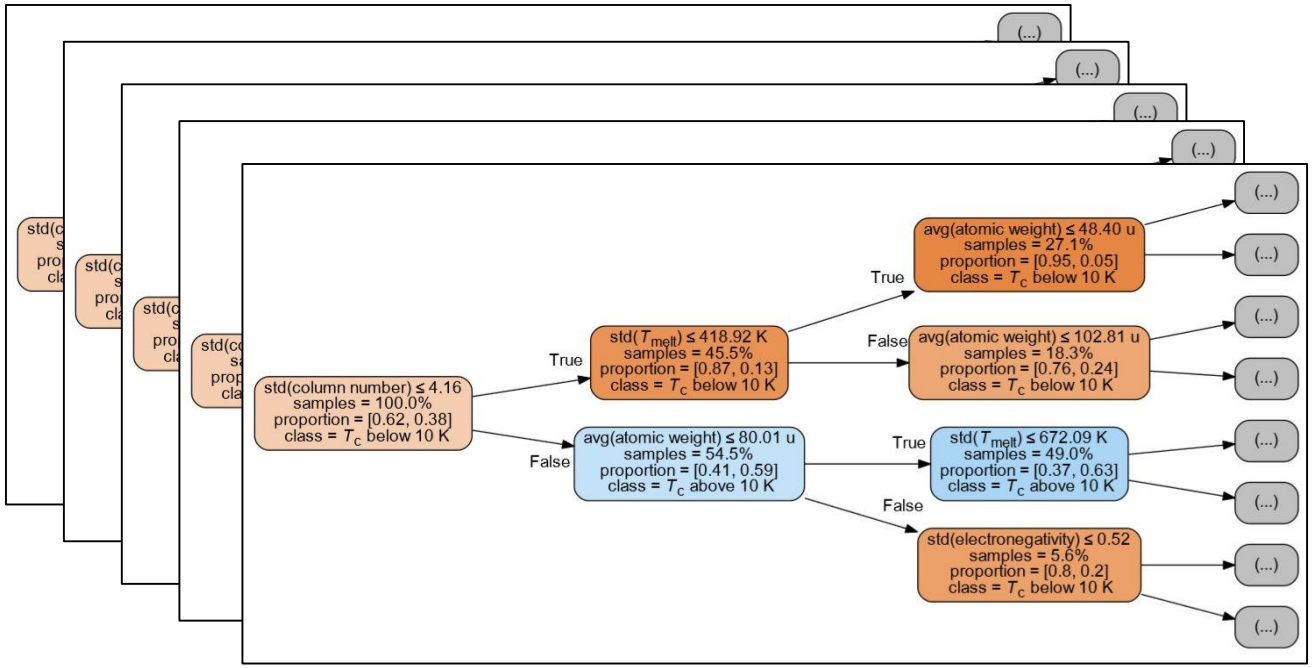


Random forest III

Random forest – combine many individual decision trees.

1. Each tree is grown on a bootstrapped sample of the data
2. At each split only a random subsample of the features is used - reduce correlation between the trees
3. Average the results of all the trees

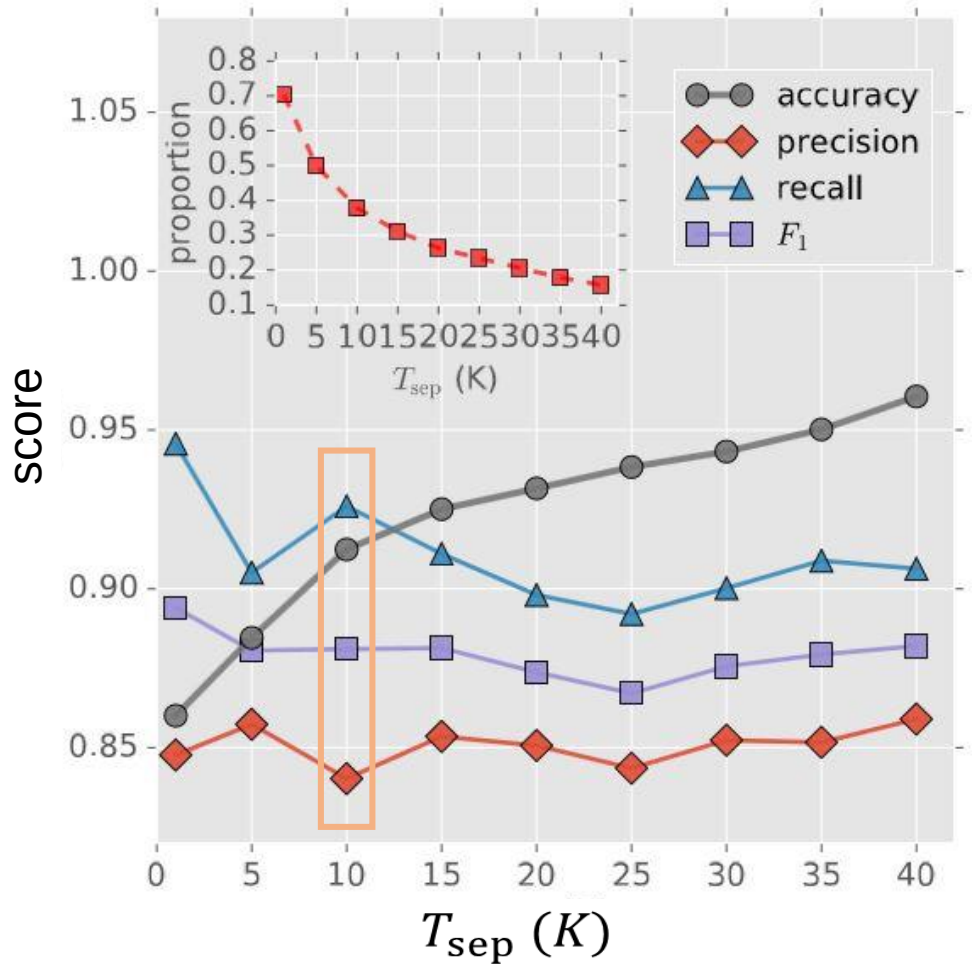
Accurate and easy to use algorithm.
Don’t require too much optimization.
Extra points – feature importance



Classification model for T_c

Separate materials in two classes: above-T_{sep} and below-T_{sep} materials (T_c bigger/smaller than T_{sep}).

Allows us to deal with compounds without T_c. But how to choose T_{sep}?



$$\text{accuracy} \equiv \frac{tp + tn}{tp + tn + fp + fn},$$

$$\text{precision} \equiv \frac{tp}{tp + fp},$$

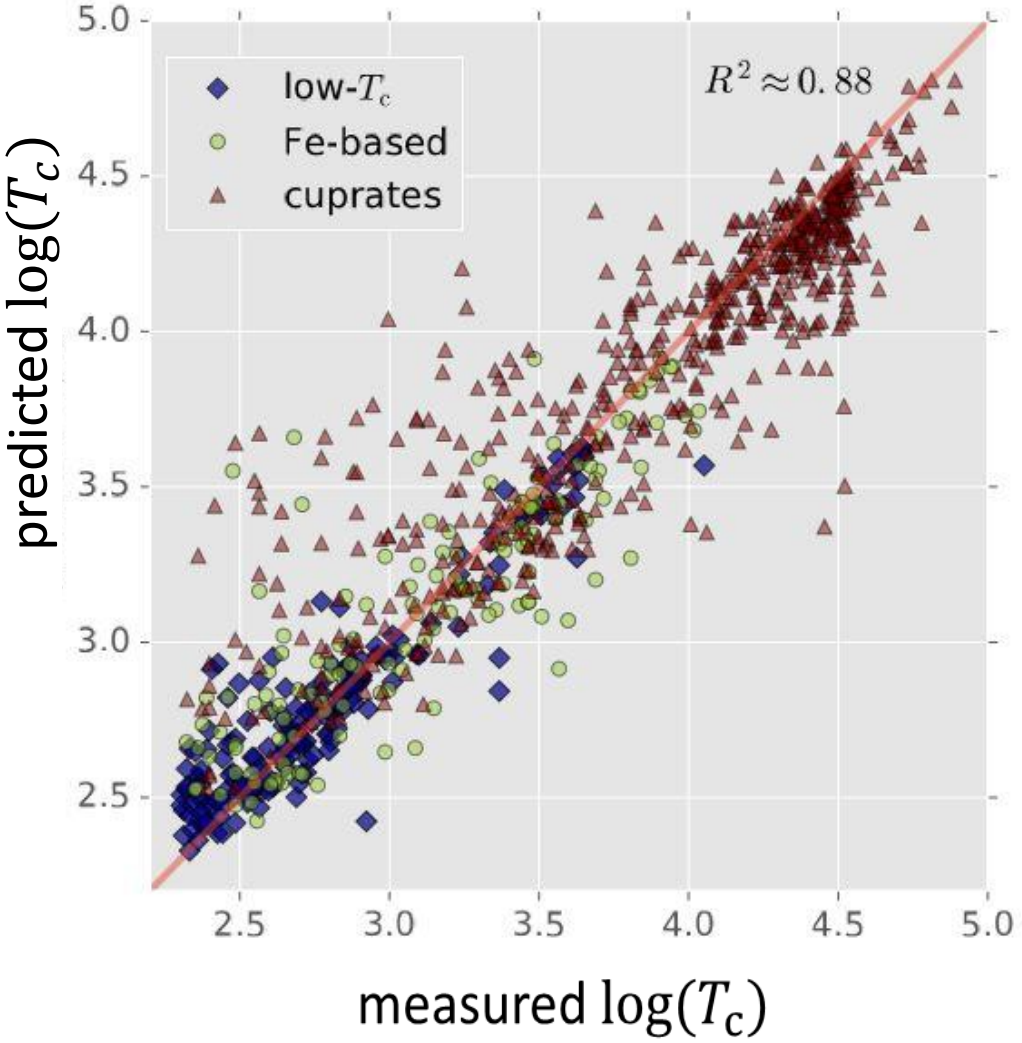
$$\text{recall} \equiv \frac{tp}{tp + fn},$$

$$F_1 \equiv 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}},$$

V. Stanev, C. Oses, A. Kusne, E. Rodriguez, JP. Paglione, S. Curtarolo, I. Takeuchi, npj Computational Materials 4 (1), 29 (2018)

Regression model for T_c

Separat the "T_c > 10 K" materials and fit regression model for T_c

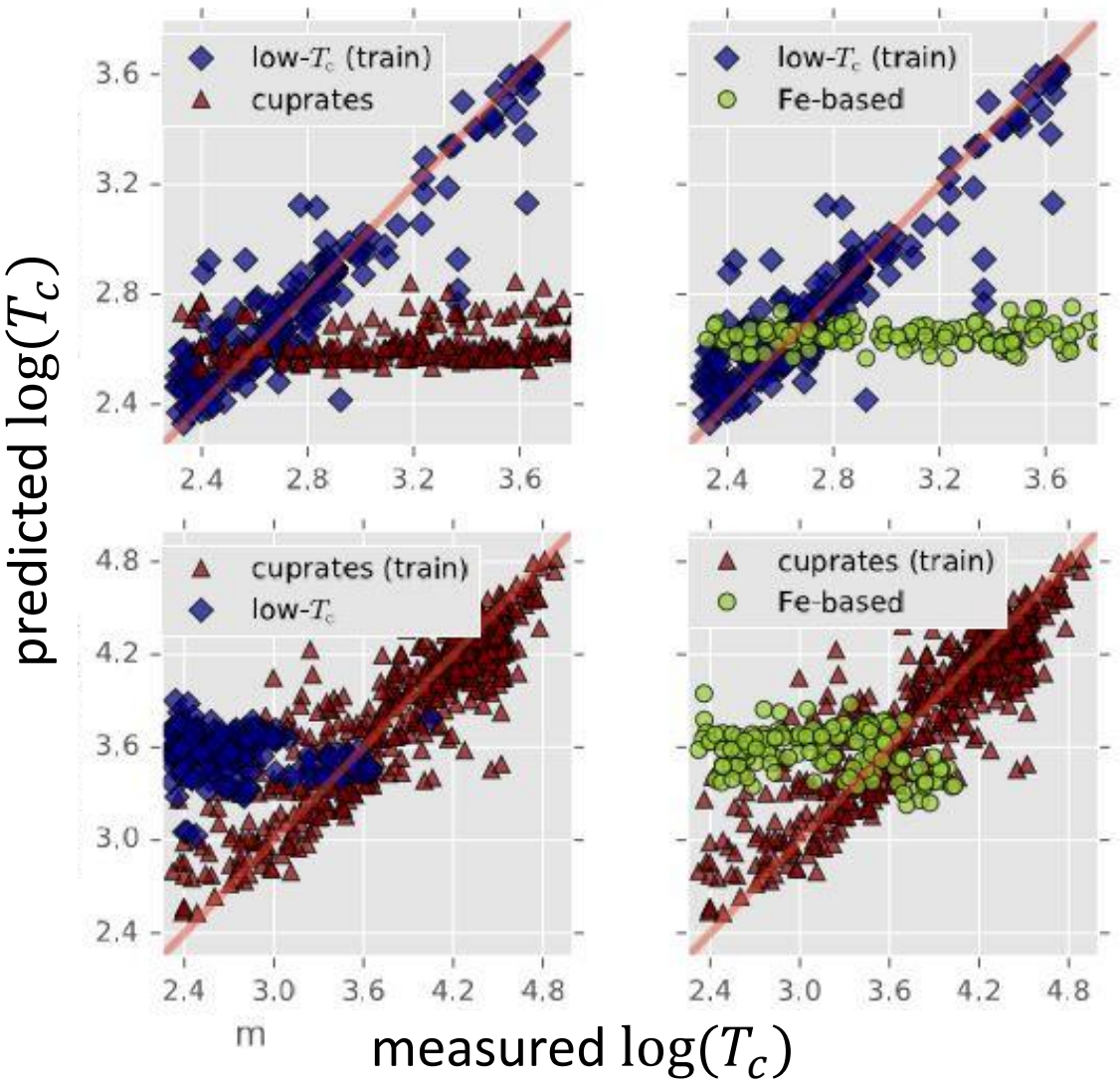


Single model for all families works very well!
But they have different physics…

V. Stanev, C. Oses, A. Kusne, E. Rodriguez, JP. Paglione, S. Curtarolo, I. Takeuchi, npj Computational Materials 4 (1), 29 (2018)

Regression model for T_c II

Separated the "T_c > 10 K" materials and fit regression model for T_c



The model is creating different branches for different groups.

Model trained on a single family - no predictive power for other families.

V. Stanev, C. Oses, A. Kusne, E. Rodriguez, JP. Paglione, S. Curtarolo, I. Takeuchi, npj Computational Materials 4 (1), 29 (2018)

Conclusion

- Machine Learning is valuable tool for study of superconductivity
- Use composite descriptors to train classification and regression models

Now let’s try it!