WELCOME TO APPLIED AI FOR MATERIALS ENGINEERING

Logan Ward
Asst. Computational Scientist
Argonne National Laboratory

7 February 2022

Today's Poll

Checking on where the blind spots are

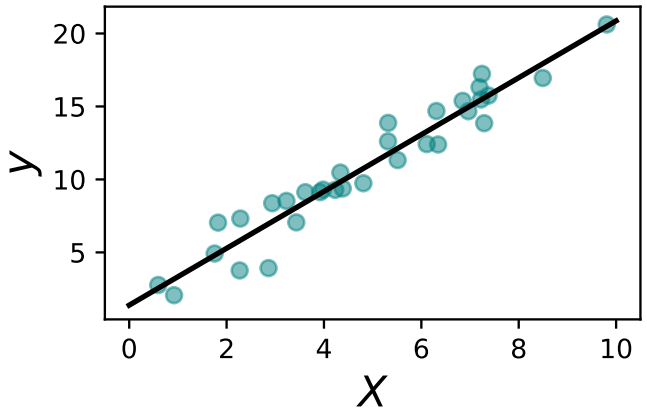
Feed back from poll

- Still ~40% of the class struggling with the homework

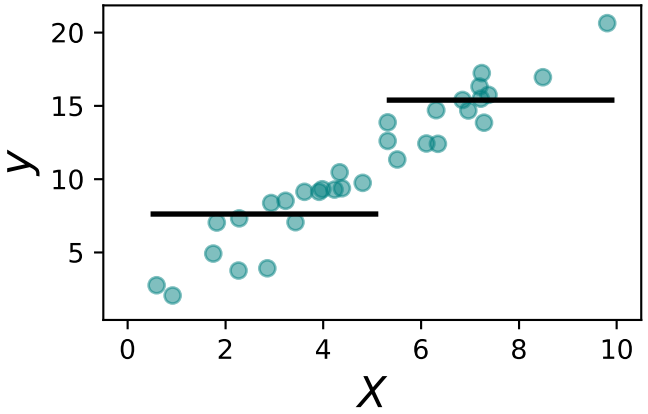
How do linear regression and decision trees select features?

Linear features are important in both models

Linear regression can find linear relationships



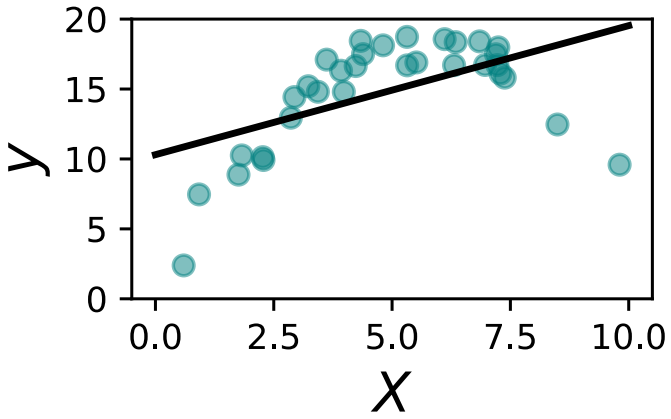
Decision trees can split into two populations



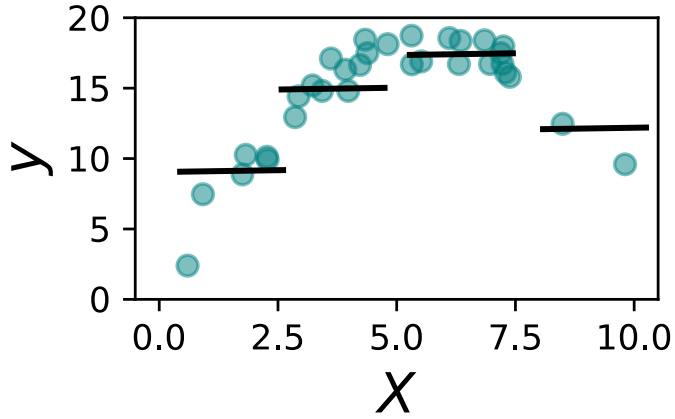
How do linear regression and decision trees select features?

Linear features are important in both models

Linear regression cannot find nonlinear relationships



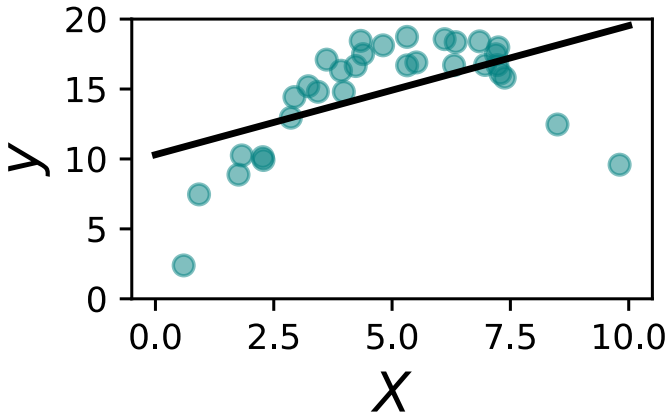
Decision trees can split into many populations



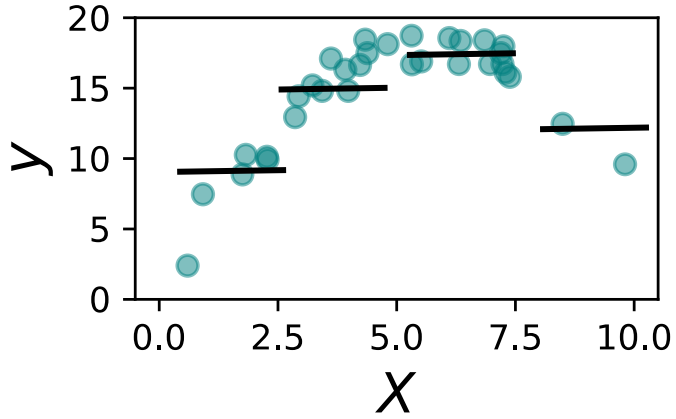
How do linear regression and decision trees select features?

Nonlinear features are only seen in

Linear regression cannot find nonlinear relationships



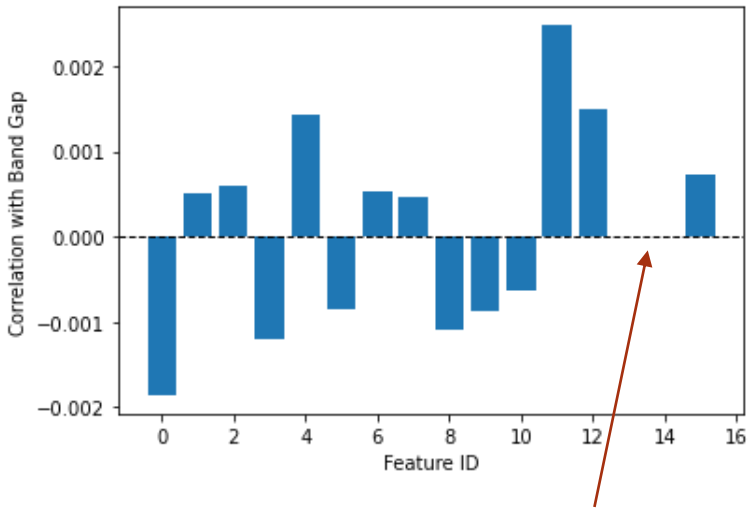
Decision trees can split into many populations



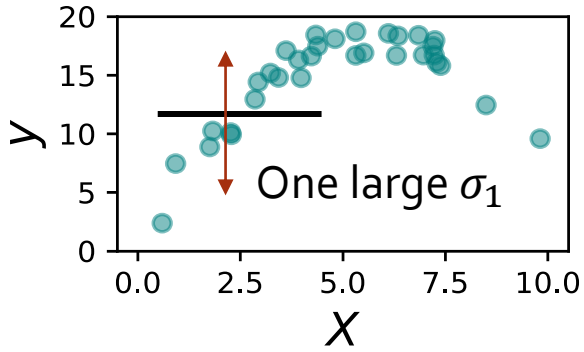
How about feature importance?

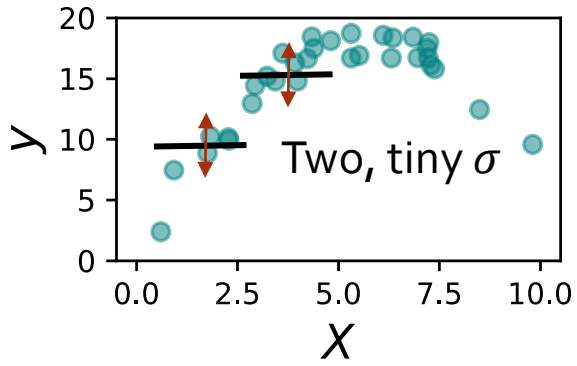
LASSO and RF do things different ways

More important features in LASSO have greater (normalized!) coefficients



Not important = 0





Each time a feature is used, it gets a score

$$Score = \sigma_1 - (\sigma_{2a} + \sigma_{2b})$$

<u>Your HW question:</u> What kind of features get a better score with each method?

Any questions about lecture? Or course in general?

Finding new crystalline compounds: A history

An important and contentious subject

1. Many types of machine learning
2. Many claiming others are fundamentally flawed

Short version for non-chemists:

1. Many materials are crystals
2. Crystals only form at special compositions
3. We still find $O(100)$ new crystals each year
4. People think there are many more left
5. Combinatorics makes random searches impossible

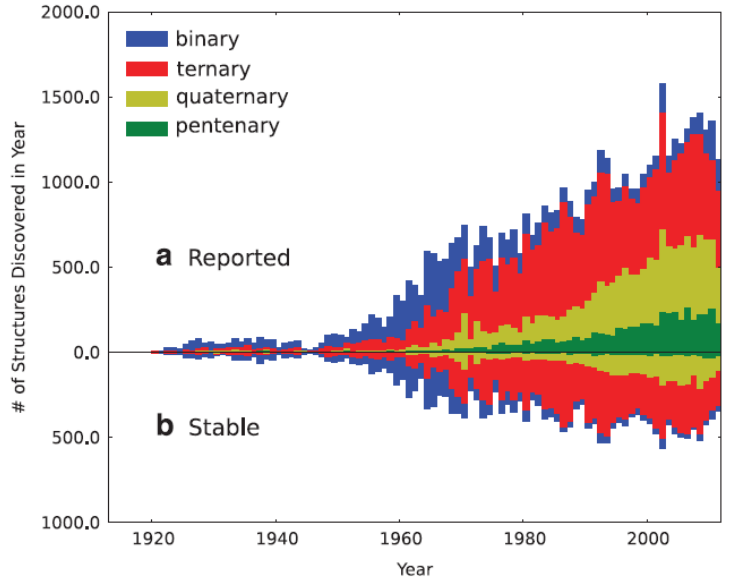


Figure 4. (**a**) The total number of compound discovery within the ICSD by year. (**b**) The number of *stable* ($T = 0$ K) compound discovery in the ICSD, where the stability is assesed by the OQMD energies. The year for a structure corresponds to the earliest publication year for ICSD entries at that given structure's composition.

Let’s step through a history

2006-: Compounds by analogy

2014-: Compounds by predicting energy from composition

2016-: Compounds by predicting energy from composition *and structure*

2020-: The debate continues to rage (without me)

Next Steps

- Next lecture topic: Interatomic potentials
- Watch it and post questions/discussion topics on Canvas
  - The more time you give me, the more stuff I can prepare
- First homework due Feb 3
  - We'll have in class time to work on it Tuesday too
- Going to give you the next homework (two options) starting Thursday