Bagging, boosting, and stacking

Ensembling → combine different models into 1

Bagging

$y, \vec{X}$

y_1, X_1 $\quad y_2, X_2$ $\quad y_3, X_3$

f_1 $\quad f_2$ $\quad f_3$

$f(X) = g(f_1, f_2, f_3)$

Boosting

$y, \vec{X}$

$f_1(X) \approx y$

$e_1 = y - f_1$

$f_2(X) \approx e_1$

$e_2 = e_1 - f_1$

$f_3(X) \approx e_2$

$f = g(f_1, f_2, f_3)$

Stacking

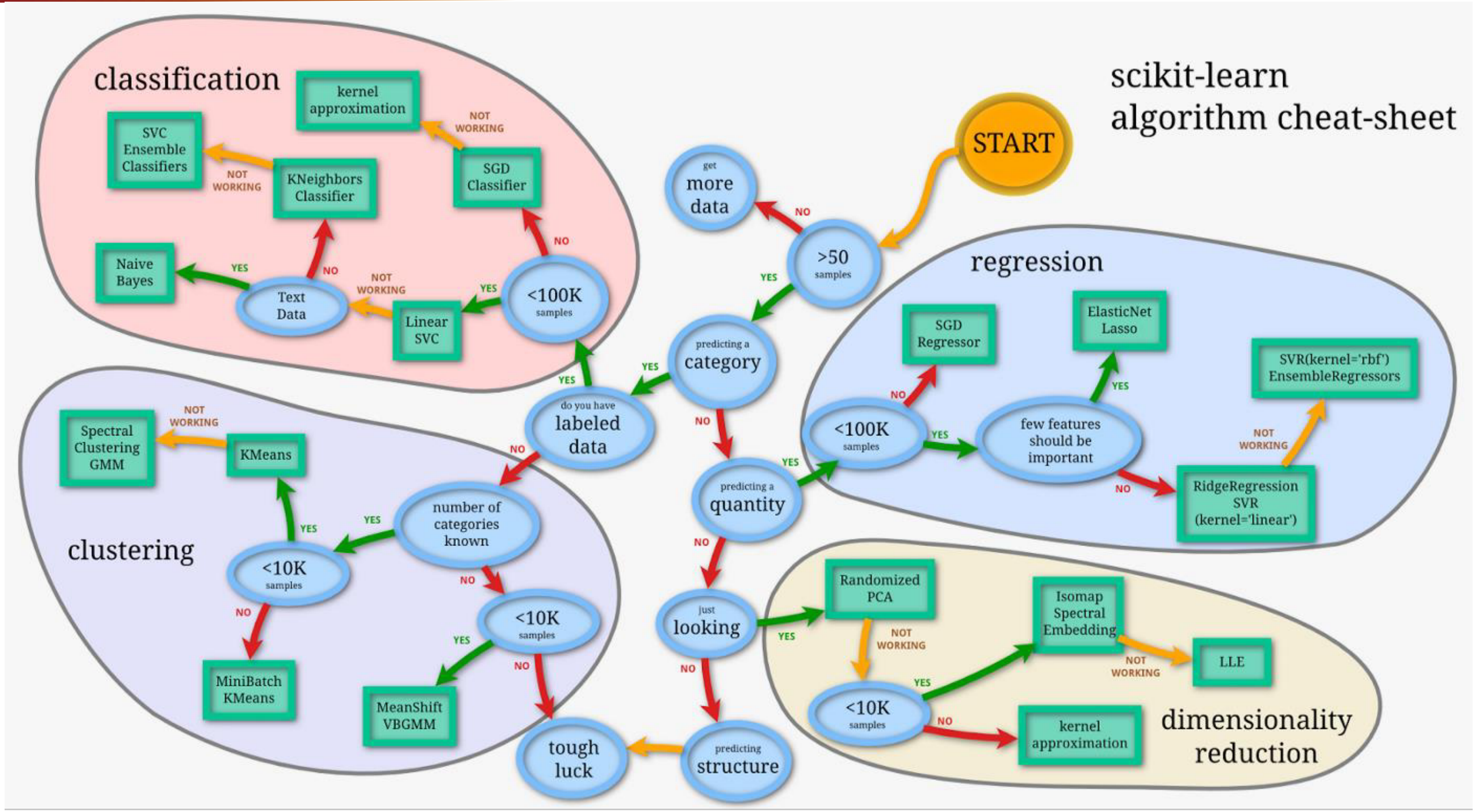
y, X

$f_1(x)$ $\quad f_2(x)$ $\quad f_3(x)$

$f = g(f_1, f_2, f_3)$

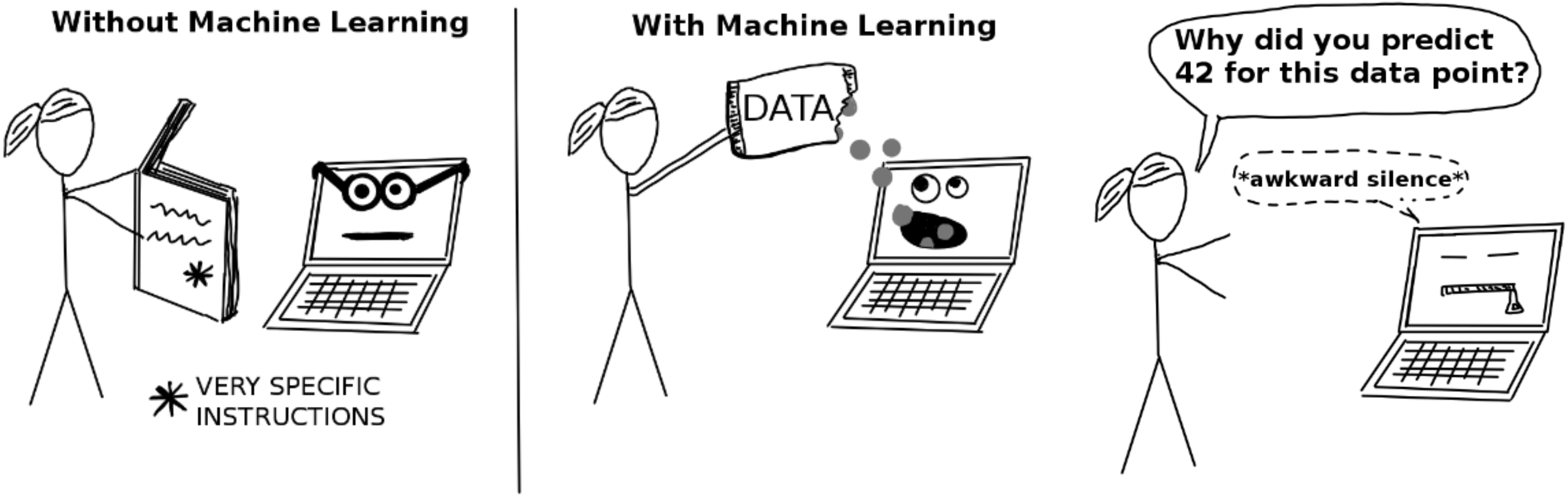
Which model to use??

CEMS
Chemical Engineering
& Materials Science



General approaches to interpretable ML

CEMS Chemical Engineering & Materials Science

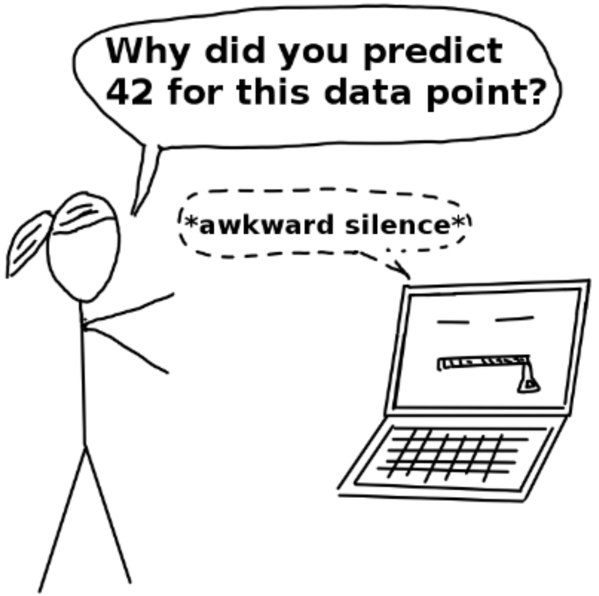


Why do we care about interpretability?

Reliability ⇒ small Δx_i shouldn't produce large $\Delta \hat{y}_i$

Causality ⇒ as we change x_i, can we anticipate $\Delta \hat{y}_i$

Trust ⇒ adoption $\propto$ understanding.



Taxonomy of interpretability

CEMS
Chemical Engineering
& Materials Science

Intrinsic $\Rightarrow$ models are simple,
so we interpret directly.

Post-hoc $\Rightarrow$ train any "black box" model
and analyze predictions

Local $\Rightarrow$ why was a certain prediction made?

Global $\Rightarrow$ how does the model work?

Intrinsic interpretability in linear regression

Chemical Engineering & Materials Science

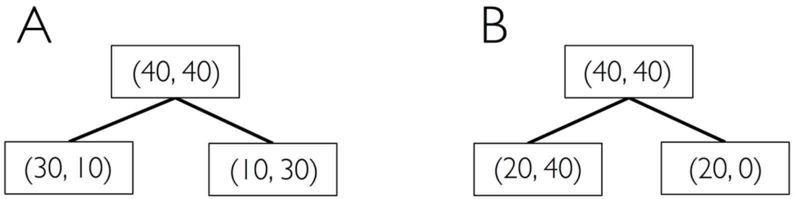
$$y \approx \hat{y} = f(X) = w_0 x_0 + w_1 x_1 + \cdots \qquad w = w: \min_w ||\hat{y} - y||_2^2$$

Features with largest "effect"

are most important

effect $\equiv |w_i x_i|$

Intrinsic interpretation of decision trees



Features that ↓ impurity the most are most important

From model-specific to model-agnostic methods

Model-agnostic methods

↳ doesn't matter model type

some model $\hat{y} = f(\vec{x})$

≈

interpret f

One global method: permutation importances

CEMS
Chemical Engineering
& Materials Science

Idea: how much worse does my model get if a feature is "shuffled"?

1) Train a model, $\hat{y} = f(\vec{x})$

2) Compute error, $e_0 = \|y - \hat{y}\|_2^2$

3) For feature, x_j in $\vec{x}$,

   a) Shuffle x_j

   b) Compute error, e_j using $f(\vec{x})$

4) Importance $= e_j - e_0$

Fisher, Aaron, Cynthia Rudin, and Francesca Dominici. "All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously." http://arxiv.org/abs/1801.01489 (2018)

One global method: permutation importances

Compute on training data or validation data?

training ⇒ how much is x^j leveraged during training?

validation ⇒ how important is x^j for generalization to new data?

Fisher, Aaron, Cynthia Rudin, and Francesca Dominici. “All models are wrong, but many are useful: Learning a variable’s importance by studying an entire class of prediction models simultaneously.” http://arxiv.org/abs/1801.01489 (2018)

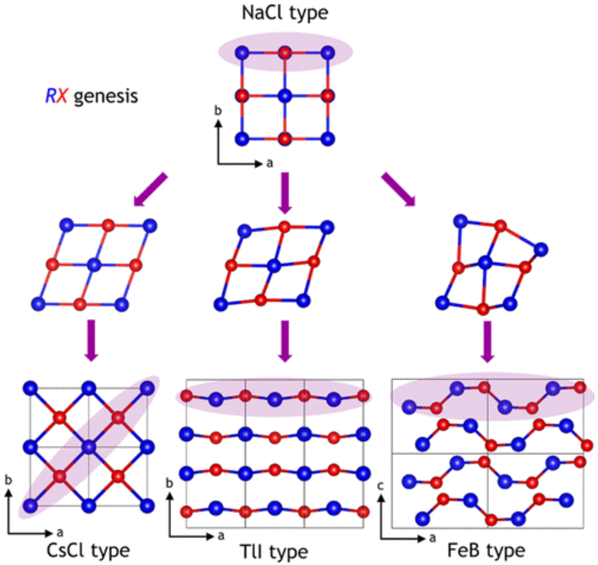
One global method: permutation importances

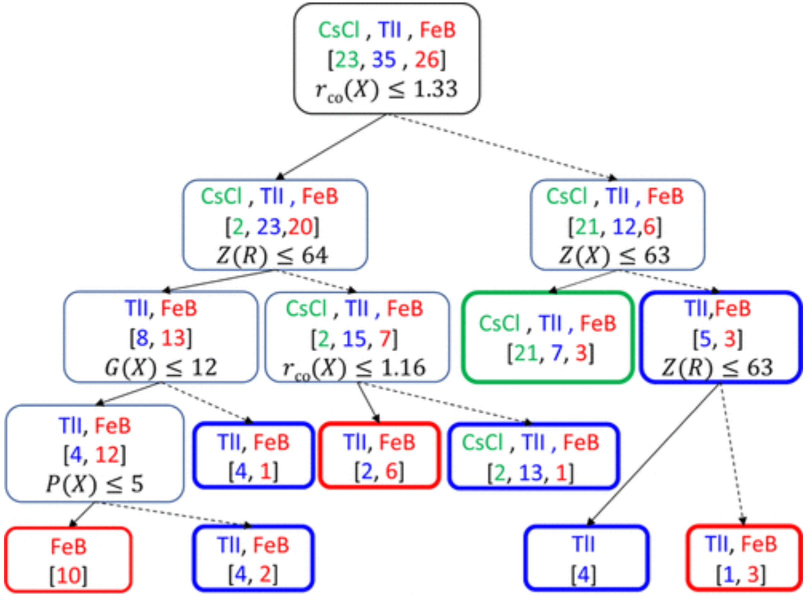

CHEMISTRY OF MATERIALS
pubs.acs.org/cm Article

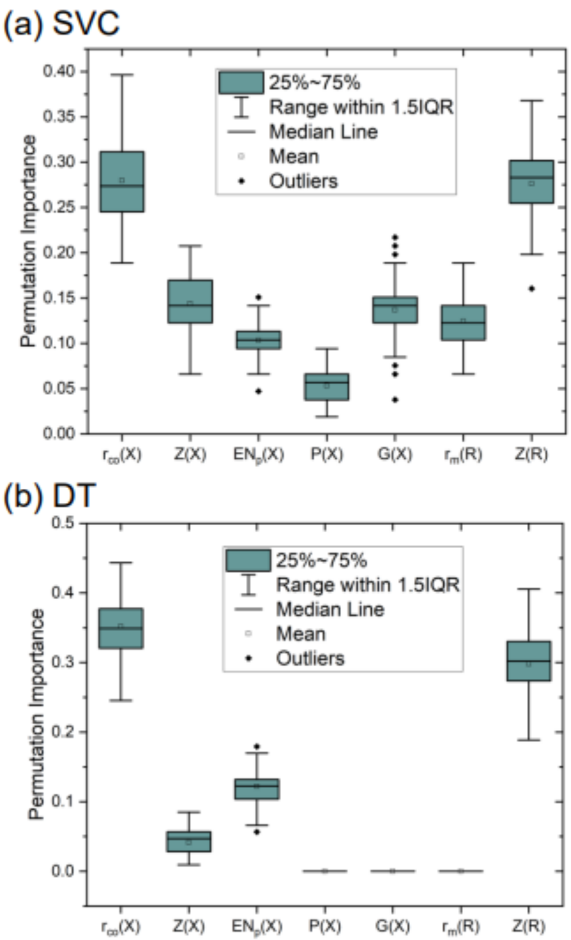
Revealing Hidden Patterns through Chemical Intuition and Interpretable Machine Learning: A Case Study of Binary Rare-Earth Intermetallics *RX*

Volodymyr Gvozdetskyi,* Balaranjan Selvaratnam, Anton O. Oliynyk, and Arthur Mar*

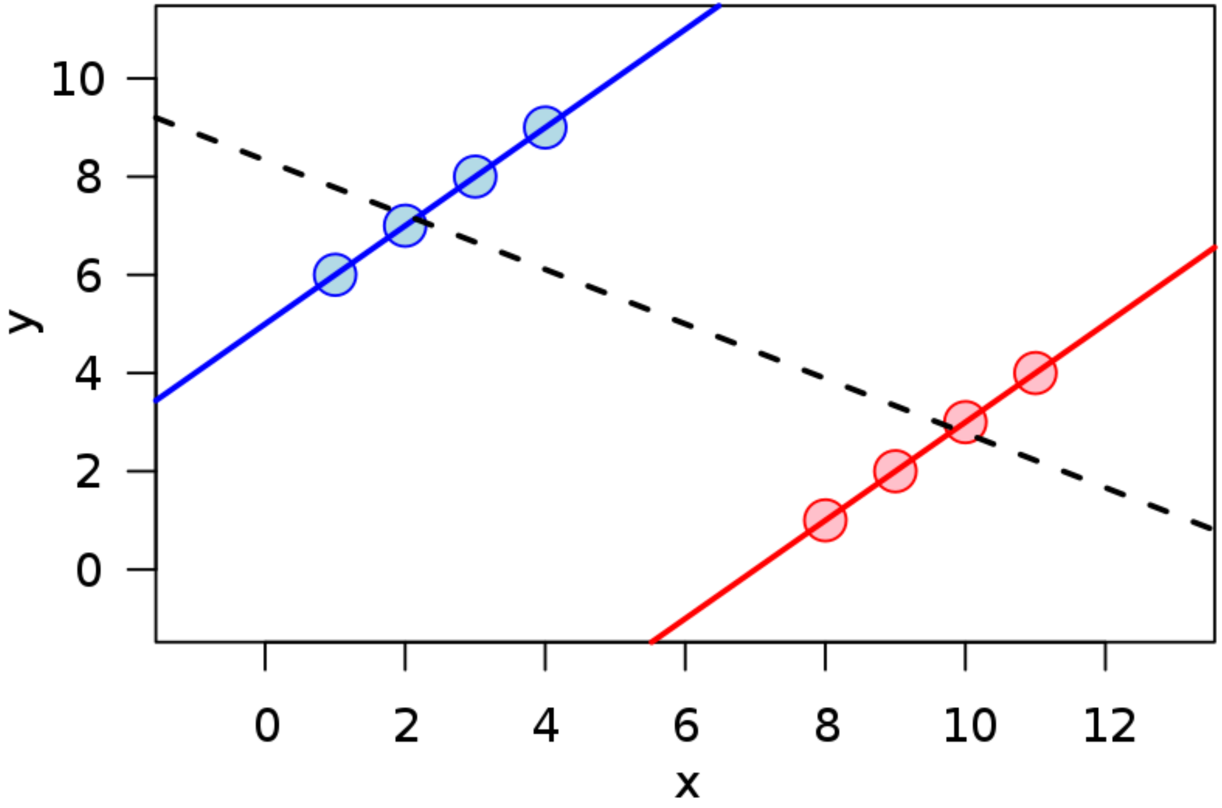
Cite This: *Chem. Mater.* 2023, 35, 879–890 Read Online





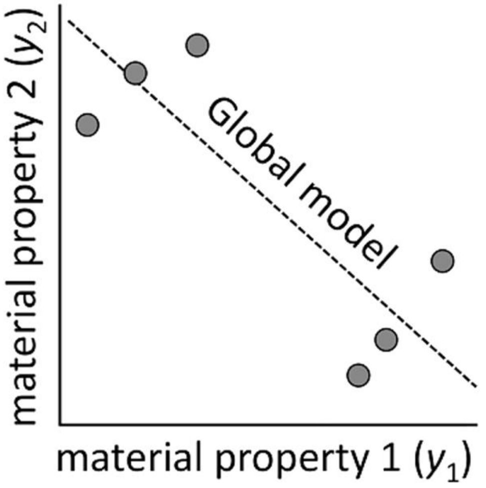


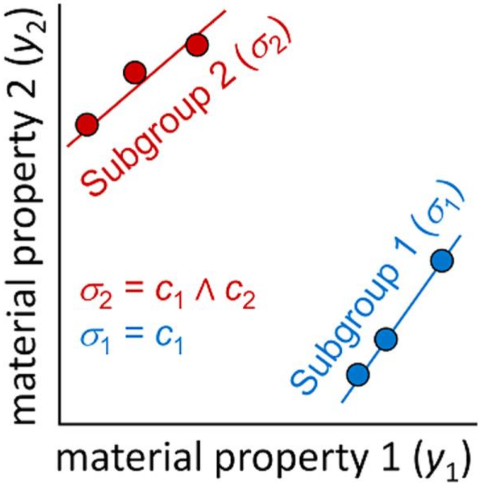
From global to local methods

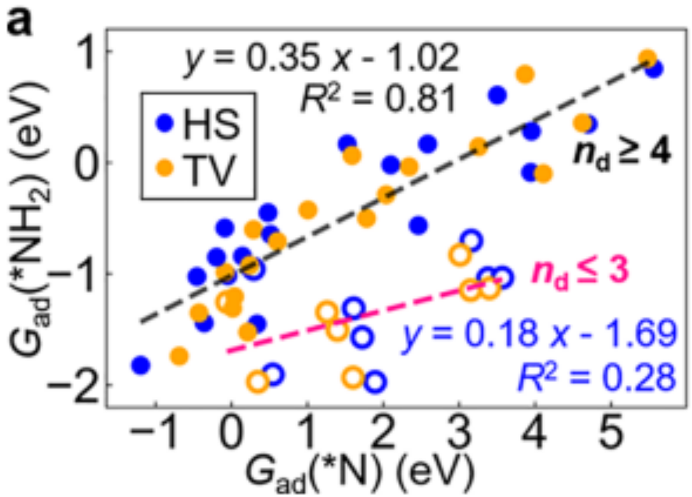


https://en.wikipedia.org/wiki/Simpson's_paradox

Our data may not be globally interpretable







 Goldsmith et al., New Journal of Physics, 2017, 19, 013031 | Li et al., ACS Catalysis, 2021, 11, 7906

Local version of surrogate: LIME

LIME ⇒ Local Interpretable Model-Agnostic Explanation

Given: $\hat{y} = f(\vec{x})$

1) Select probe point, $\vec{x}_i$
2) Randomly generate M new points
3) Predict $\hat{y}$ for these M points
4) Train linear model on $\hat{y}_M$
   weighting $\mathcal{L}$ by proximity to $\vec{x}_i$
5) Interpret this simple model.

Local version of surrogate: LIME

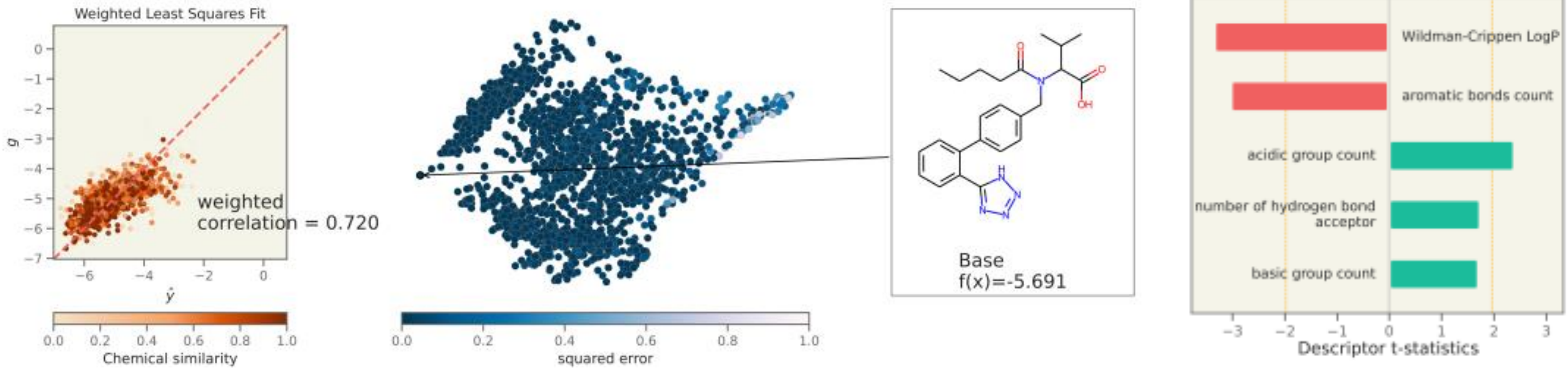
CEMS
Chemical Engineering
& Materials Science

EXPLAINING MOLECULAR PROPERTIES WITH NATURAL LANGUAGE

A PREPRINT

Heta A. Gandhi
Department of Chemical Engineering
University of Rochester
Rochester, NY, 14627
hgandhi@ur.rochester.edu

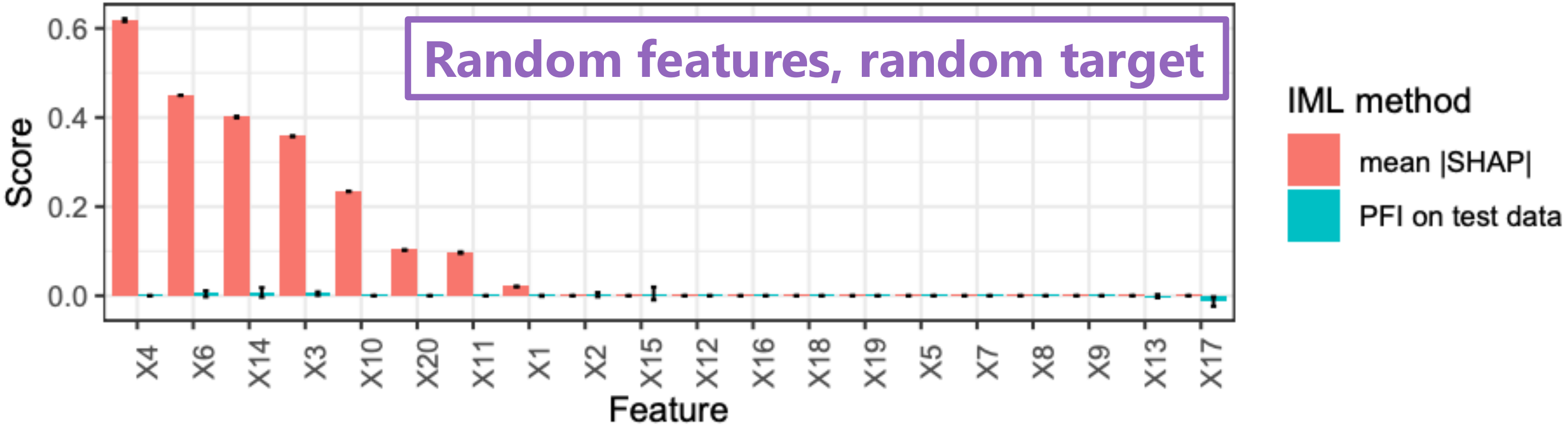
Andrew D. White*
Department of Chemical Engineering
University of Rochester
Rochester, NY, 14627
andrew.white@rochester.edu



Pitfalls of interpretable ML

CEMS
Chemical Engineering
& Materials Science

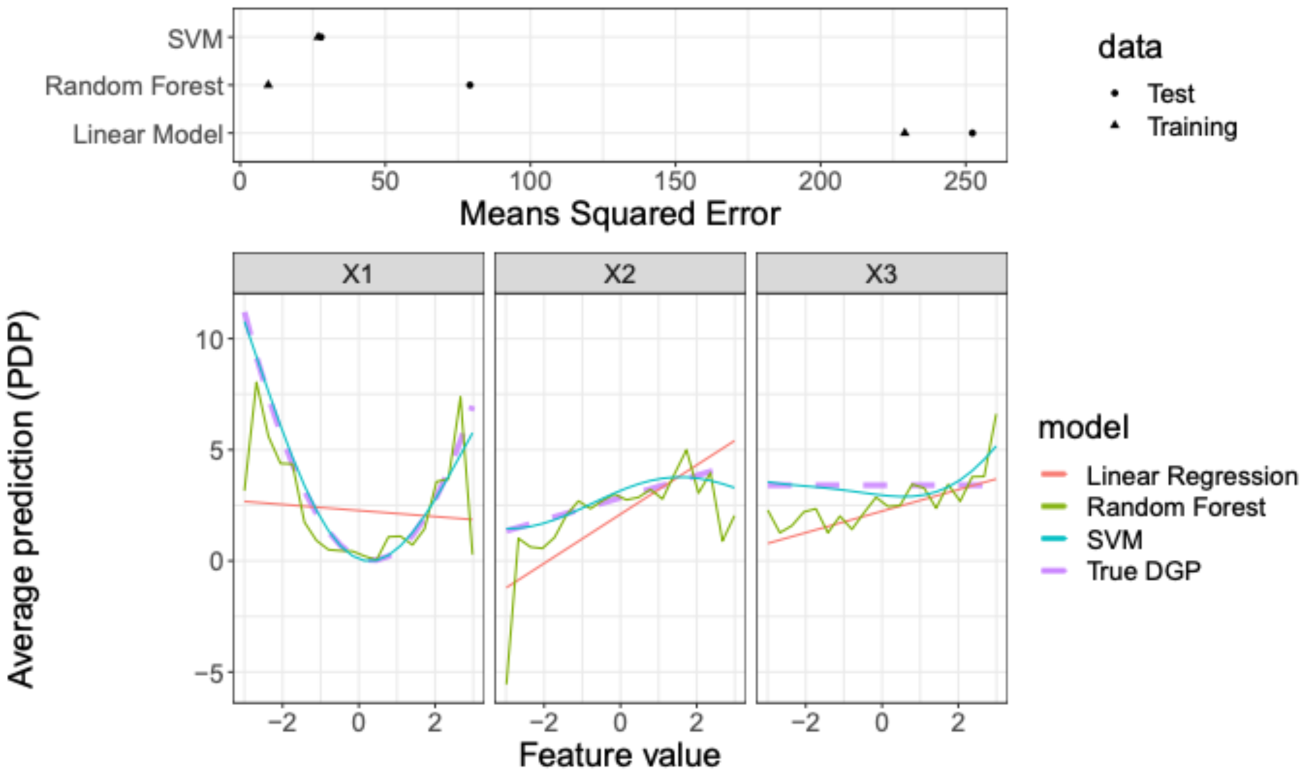
1. Assuming one method will always work



Molnar et al., 2020, arXiv:2007.04131

Pitfalls of interpretable ML

2. Bad model generalization



Molnar et al., 2020, arXiv:2007.04131

Pitfalls of interpretable ML

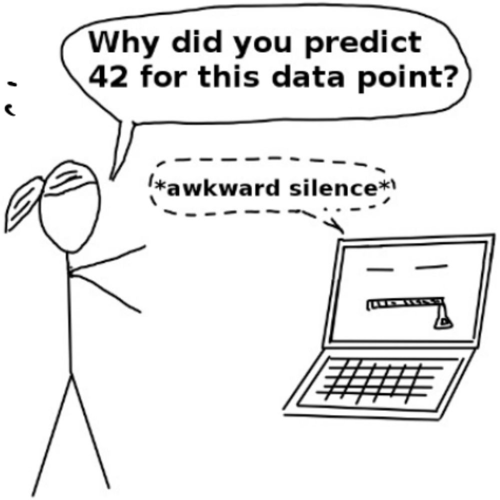
3. Unnecessary complexity

Recall why we care about interpretability:

Reliability → small change in x_i shouldn't lead to a large change in $\hat{y}_i$

Causality → as we change x_i, can we anticipate change in $\hat{y}_i$

Trust ⇒ adoption ∝ understanding



Intrinsic interpretability is always preferred!

Molnar et al., 2020, arXiv:2007.04131