Seyed Mohamad Moosavi, Kevin Maik Jablonka & Berend Smit

ChE-609

Basics of machine learning for chemistry and materials science

seyedmohamad.moosavi@epfl.ch

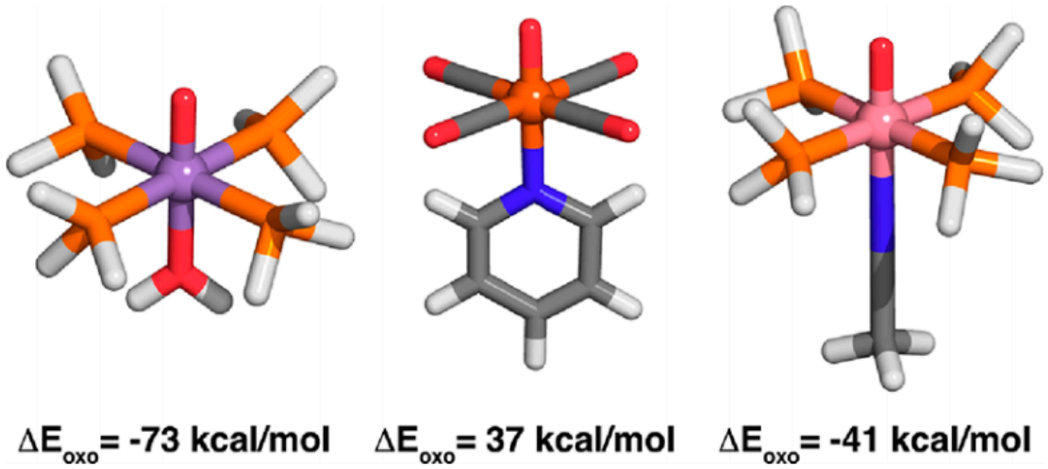
@SMohMoosavi

Interpreting the model

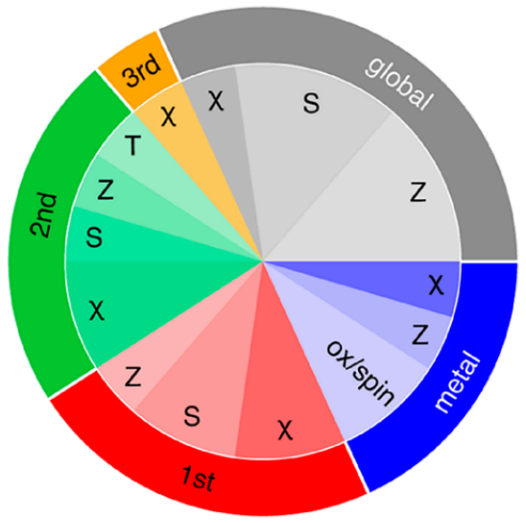
We would like to know what drives a phenomena (curiosity)

More practical:
Material design rules

Single metal catalyst:
variation in metal and ligands

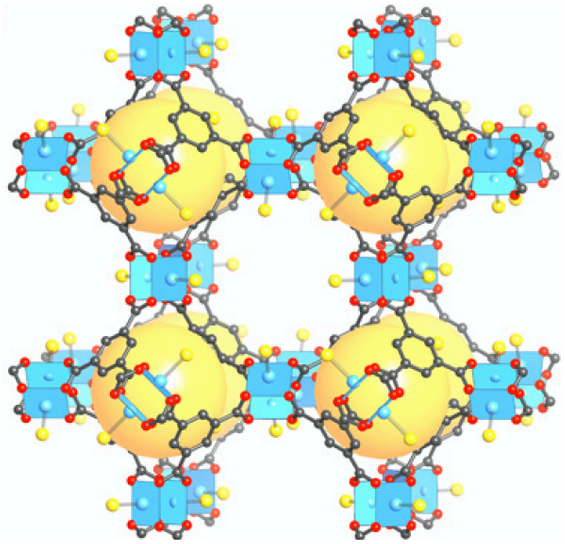


Importance of variables can help to understand which factors are important, e.g., type of atomic properties and global vs local

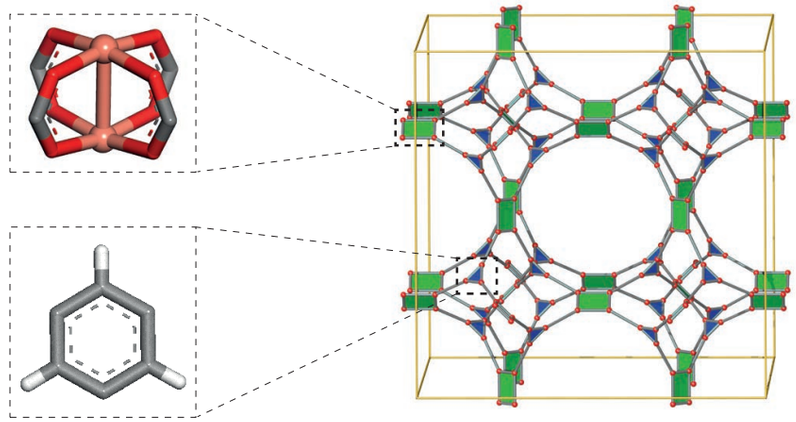


Interpreting the model (||)

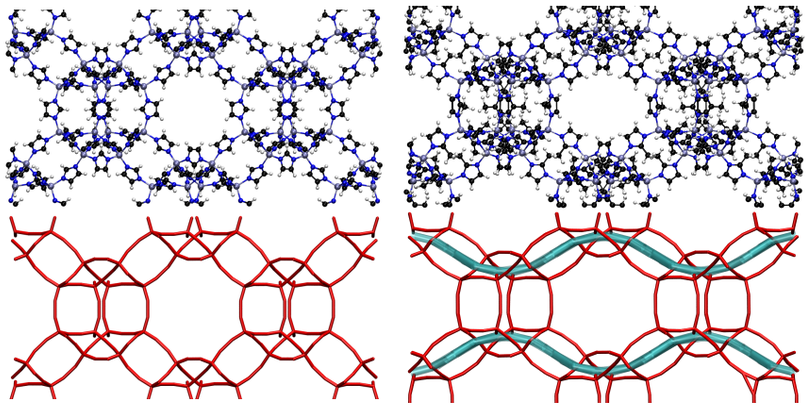
Make better models / representations
example: the case of mechanical properties of MOFs



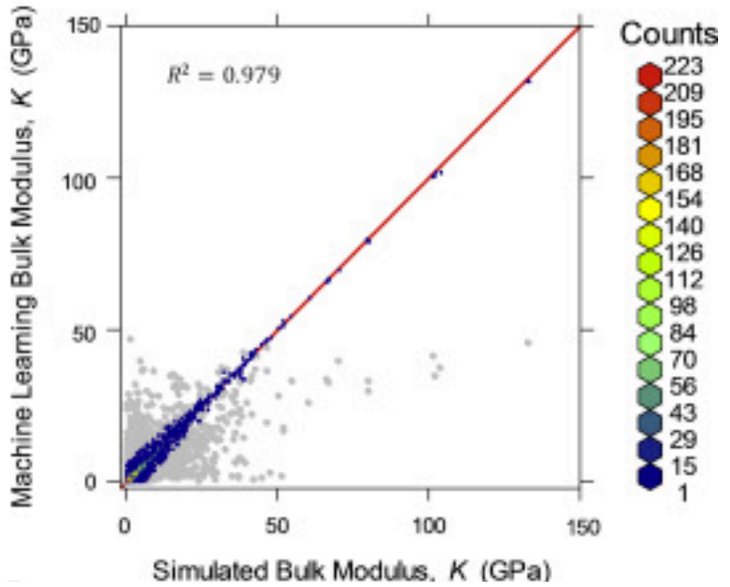
Simplified representation



It was shown that the underlying net is the most important factor for mechanical stability



Include net in representation



Moosavi, et al. ACS Central Science (2018)
Moghadam et al. Matter (2019)

Permutation importance

0. Build the model

$$X = \begin{pmatrix} 1 & x^{1,1} & x^{1,1} & \dots & x^{1,n} \\ 1 & x^{2,1} & x^{2,1} & \dots & x^{2,n} \\ 1 & \vdots & \ddots & & \vdots \\ 1 & x^{m,1} & x^{m,1} & \dots & x^{m,n} \end{pmatrix} \quad y = \begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^m \end{pmatrix} \qquad \hat{y}_{ML} = f(X), \mathscr{L}(y, \hat{y}_{ML})$$

1. Estimate the model error: $e^{\text{orig}} = \mathscr{L}(y, \hat{y}_{ML})$

2. For each feature j=1,..,n:
   - Generate permuted feature matrix for feature (j):

Shuffle

$$X_j^{\text{perm}} = \begin{pmatrix} 1 & x^{1,1} & \dots & x^{1,j} & \dots & x^{1,n} \\ 1 & x^{2,1} & \dots & x^{2,j} & \dots & x^{2,n} \\ 1 & \vdots & \dots & \vdots & \ddots & \vdots \\ 1 & x^{m,1} & \dots & x^{m,j} & \dots & x^{m,n} \end{pmatrix}$$

   - Estimate the error for the permuted feature matrix: $e_j^{\text{perm}} = \mathscr{L}(y, f(x_j^{\text{perm}}))$

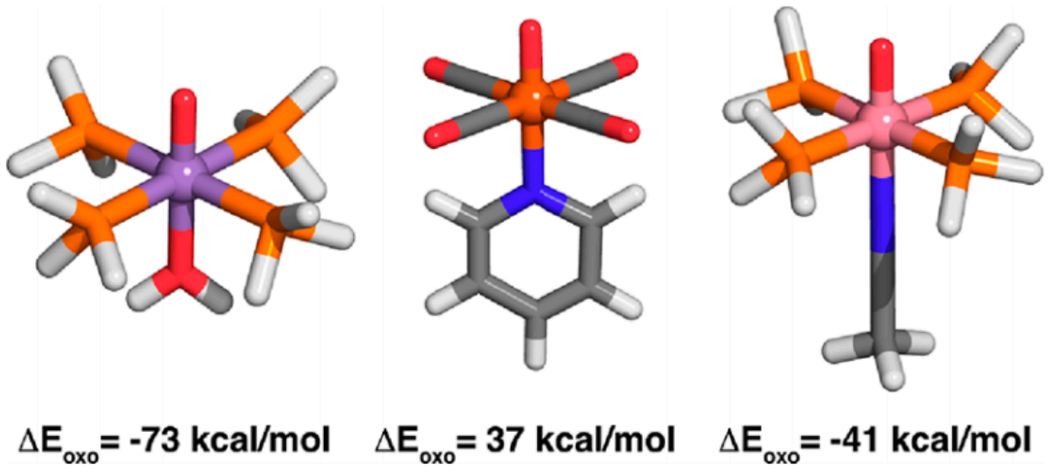
   - Estimate the error for the permuted feature matrix

$$\text{FI}_j = e_j^{\text{perm}} / e_j^{\text{orig}} \text{ or } e_j^{\text{perm}} - e_j^{\text{orig}}$$

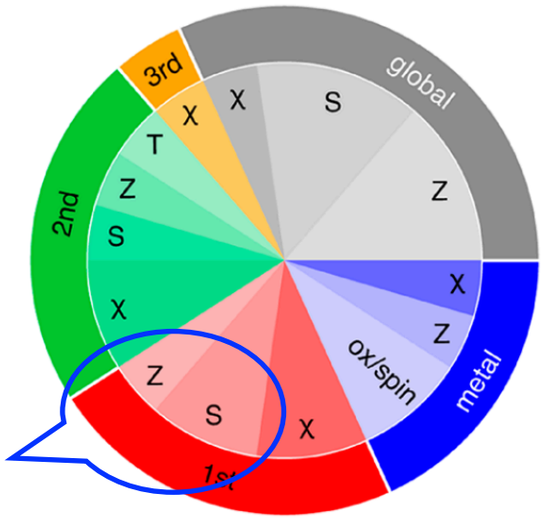
Permutation importance

- What happens if we have correlated features?

Permutation might make unphysical test cases



Split the importance between the correlated features



S = covalent radii
Z = nuclear charge

Summary of model interpretation

Feature importance is a way to interpret the model

- Get chemical insight
- Make better models

Permutation importance is a way to get the value of features in model predictions but one needs to be cautious to not over-interpret these numbers, e.g., when the features are correlated