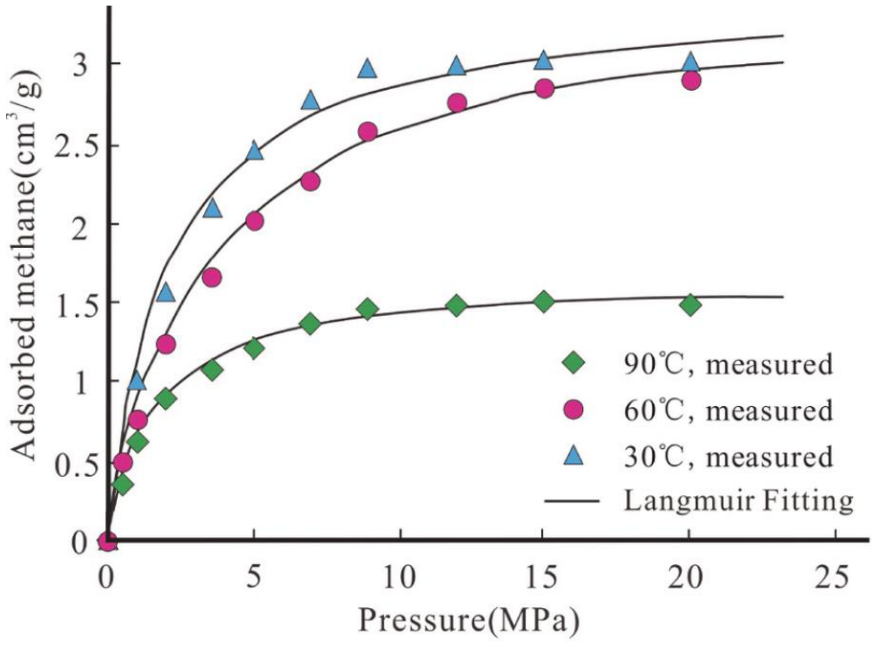
The prediction of Gas Isotherms by the Structures of Metal-Organic Frameworks

Group 6:
Wenqin You
Zhe Liu
04/17/2018

Background: Gas Adsorption

- Adsorption is one of the most essential processes in chemical engineering.
- $Adsorption\ isotherms = \frac{the\ amount\ of\ the\ adsorbate\ adsorbed}{mass\ adsorbent}$
- $Gas\ isotherms = f(Temp.,\ Pres.,Adsorbent,Adsorbate)$



Background: Metal-Organic Frameworks

- MOFs are good candidates for gas adsorption:
  - High surface area, porosity, and functionality
- Applications for adsorptive separation:
  - Energy issue: CH_4 uptake
  - Global warming issue: CO_2 capture
  - Industrial example: C_2H_4/C_2H_6 separation

The search for the optimal MOF requires aggressive screening of a variety of MOFs

Motivation of the prediction of the isotherm

Experimental Measurement

- Measurement takes long time
- Synthesis of new MOFs : >12h
- Accuracy depends on exp. conditions

Grand canonical Monte Carlo(GCMC)

- Couple hours
- Require accurate force field

Prediction by machine learning

- A few seconds/minutes
- Rapid screening
- Advice for material design

Goal: Isotherm = f(structure info …) without exp. or GCMC

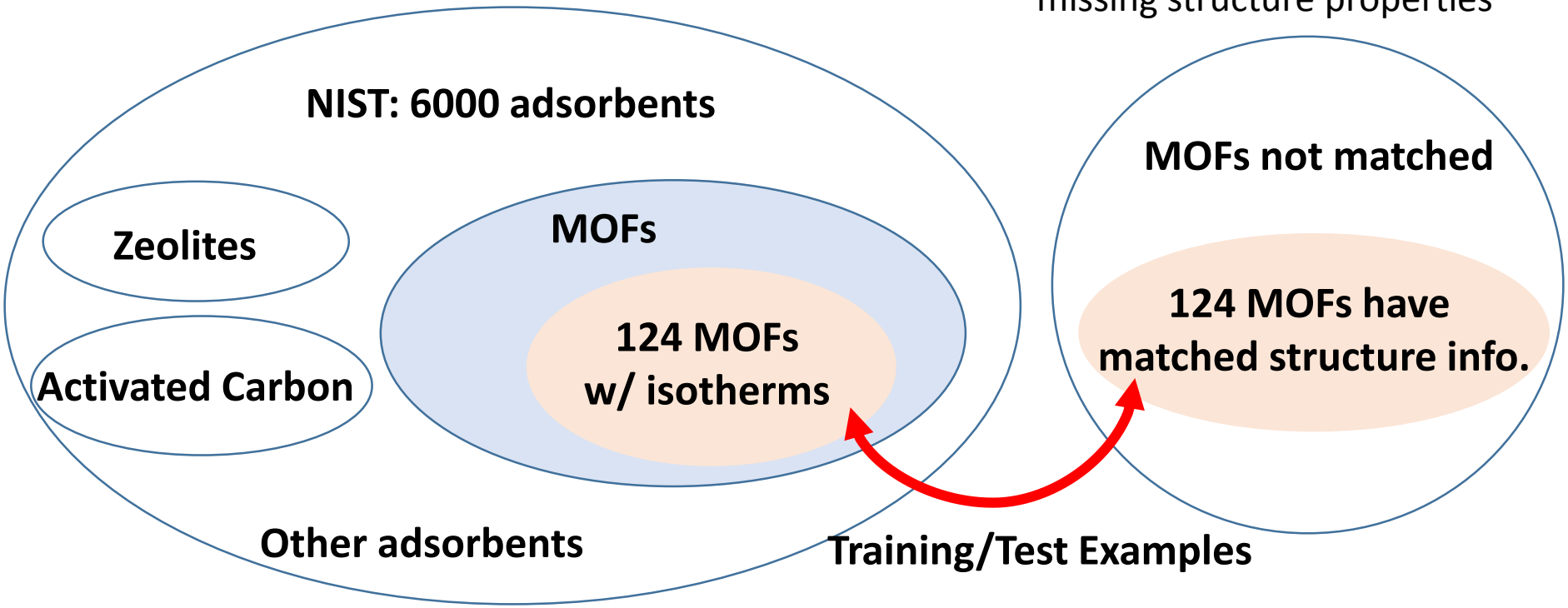
Workflow for Data Cleaning and Retrieval

Isotherms in NIST adsorption database

1. Remove non-MOFs, such as Zeolites and AC
2. Convert pressure unit and adsorption unit

Structure information in computational-ready experimental (CoRE) MOFs database

1. Remove MOFs with missing structure properties



Workflow for CoRE MOFs Database

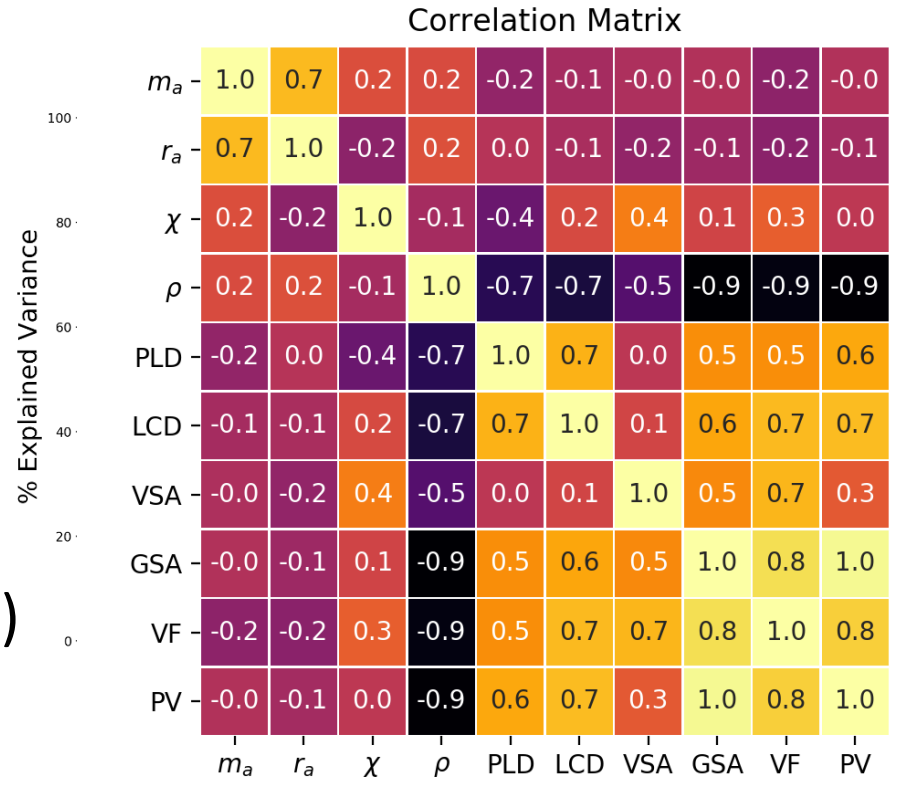
Define the feature vectors

- Each MOF can be represented by PLD, LCD, SA, VF and density etc.
- Metals type as numerical variables

Principle Component Analysis

- Calculate the explained variance
- Minimum # of principle components

- Metal type:
  - Atomic mass (m_a)
  - Atomic radius(r_a)
  - Electronegativity(χ)
- Density (ρ)
- pore-limiting diameter (PLD)
- largest-cavity diameter (LCD)
- Volumetric Surface area (VSA)
- Gravimetric Surface area (GSA)
- Pore volume (PV)
- Void fraction (VF)



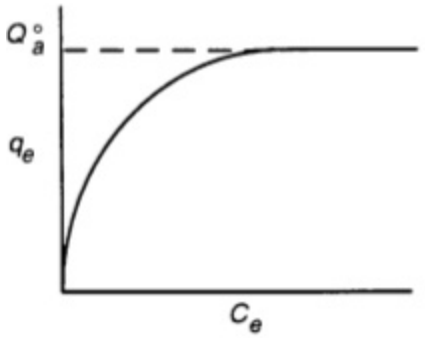
Workflow for NIST Adsorption Database

Linear Regression on Isotherms

- Curve-fit isotherm data

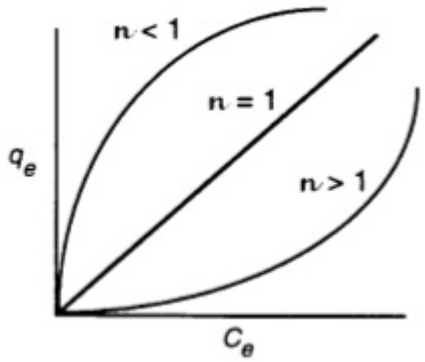
Isotherm Type Assignment

- Assign model based on R2 value
- Keep R2 higher than 0.95 (cutoff value)



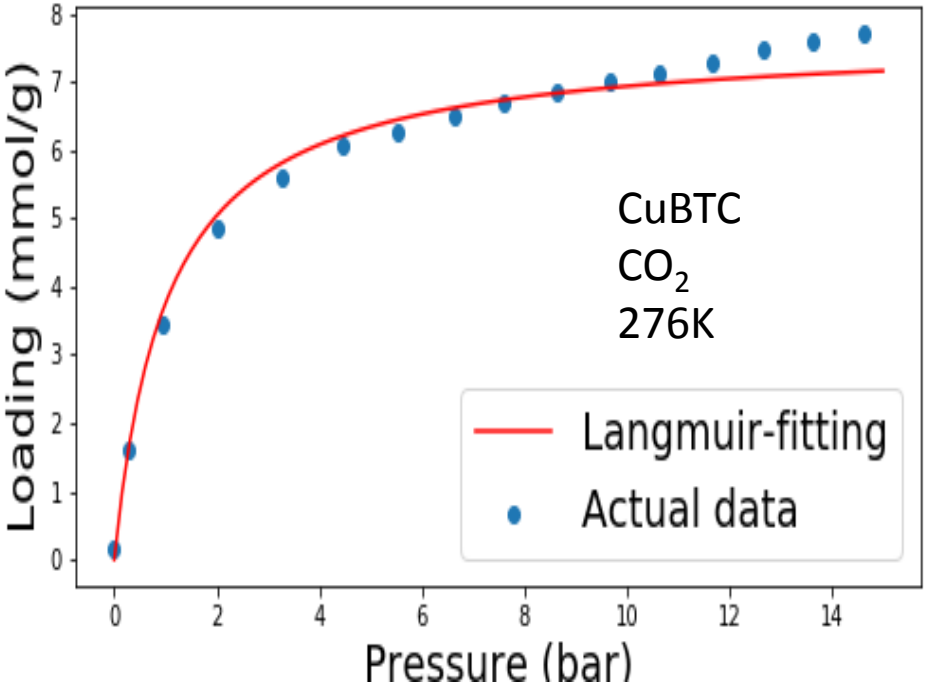
Langmuir: (Q_0, b)

$$n_{ad} = \frac{Q_0 bP}{1 + bP}$$



Freundlich: (K, n)

$$n_{ad} = KP^{\frac{1}{n}}$$



Workflow for machine learning

Results from NIST:
MOFs labeled with model, and corresponding parameters

Machine Learning

Results from CoRE MOFs
MOFs and their structure properties after PCA

Final Model

- Classification:

$$Model = f(PLD, LCD, SA, VF \ldots)$$

- Regression:

Langmuir:

$$Q_0 = f(PLD, LCD, SA, VF \ldots)$$

$$b = f(PLD, LCD, SA, VF \ldots)$$

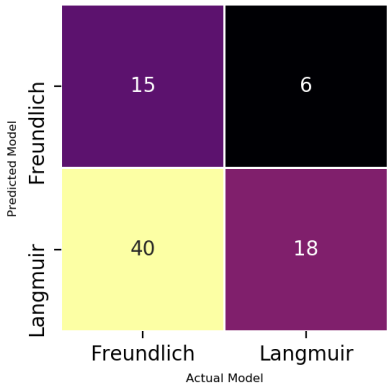
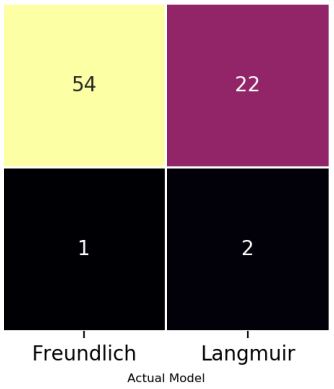
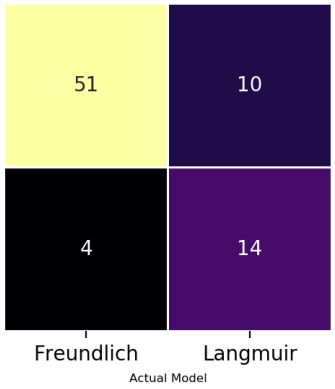
Freundlich:

$$K = f(PLD, LCD, SA, VF \ldots)$$

$$n = f(PLD, LCD, SA, VF \ldots)$$

Result 1: Isotherm classification for N_2

$$Isotherm\ Model = f(PLD, LCD, SA, VF \ldots)$$

Methods	Logistic	LDA	KNN	Kernel SVC
				
Accuracy	41.8%	70.9%	82.3%	83.5%
Cross Validation	24.1%	67.0%	82.3%	74.7%

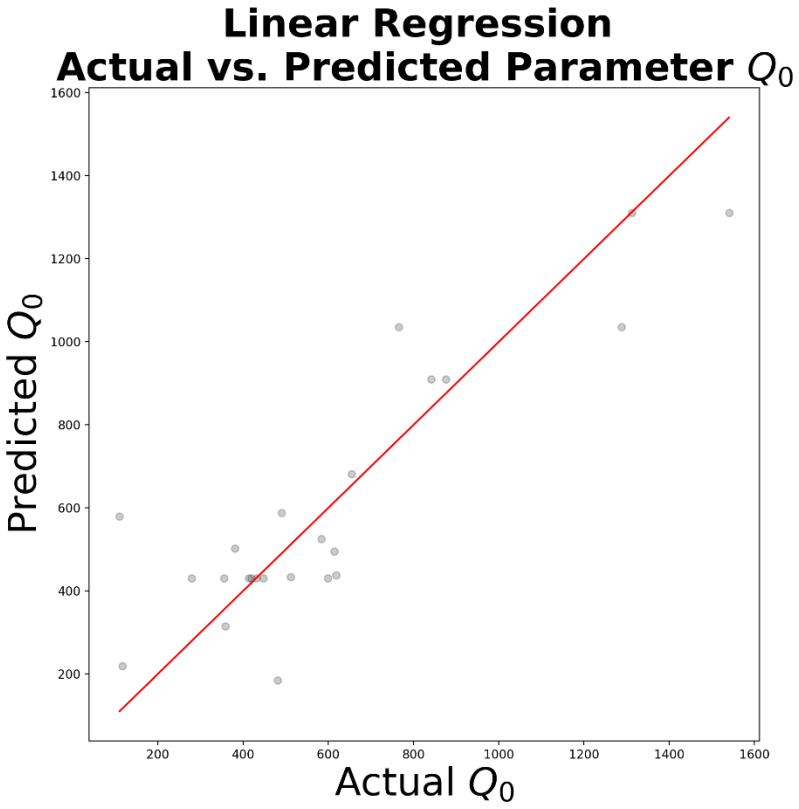
Result 2: Regression of N_2 by Linear Regression

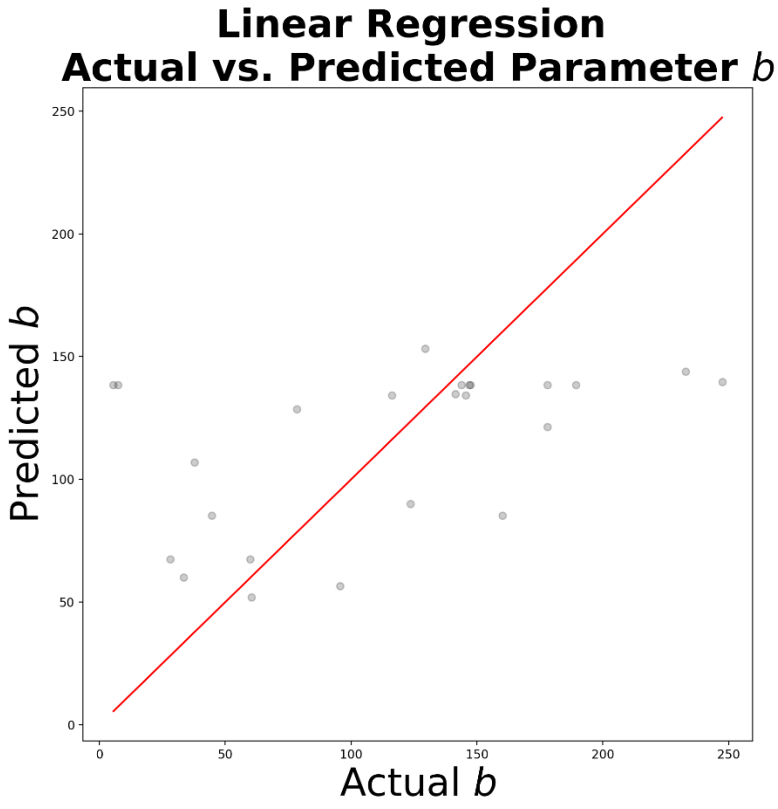
Langmuir:

$$Q_0 = f(PLD, LCD, SA, VF \ldots)$$

$$b = f(PLD, LCD, SA, VF \ldots)$$

Parameters by LR	R2	CV mean relative error
Q_0	0.78	1.02
b	0.24	3.02





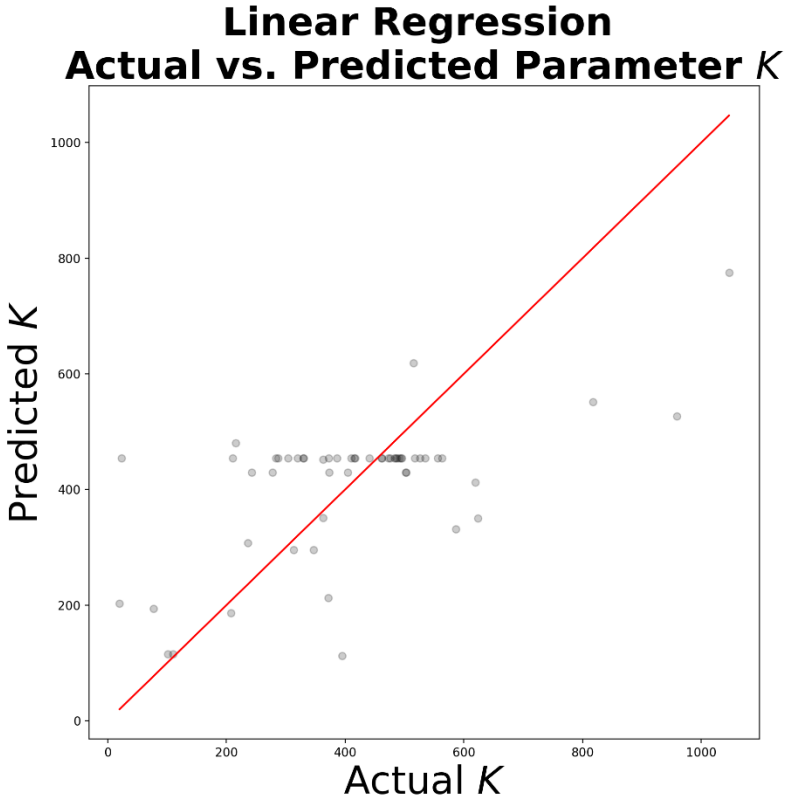
Result 2: Regression of N_2 by Linear Regression

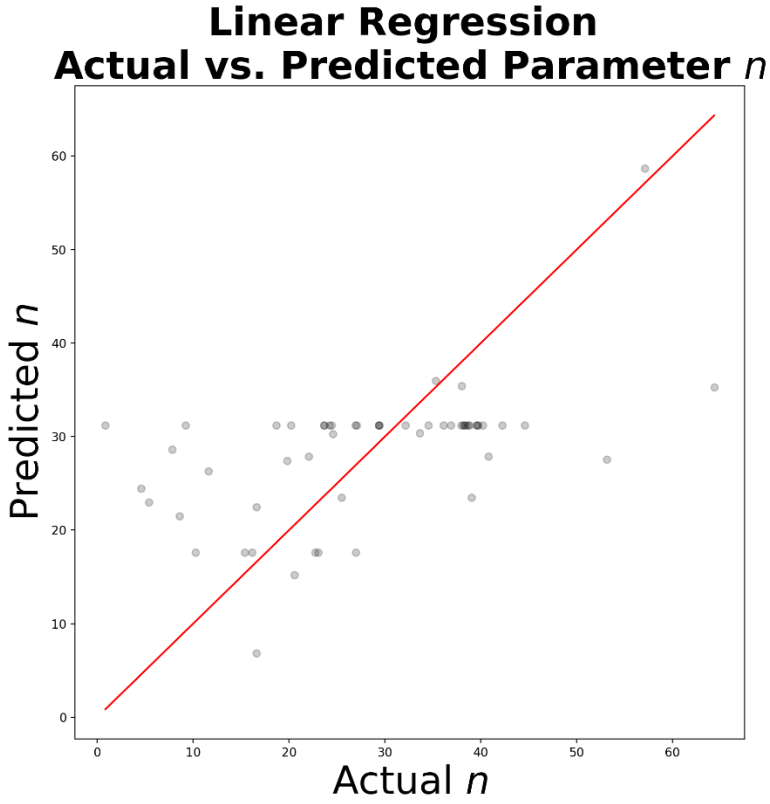
Freundlich:

$$K = f(PLD, LCD, SA, VF \ldots)$$

$$n = f(PLD, LCD, SA, VF \ldots)$$

Parameters by LR	R2	CV mean relative error
K	0.38	0.85
n	0.29	1.22





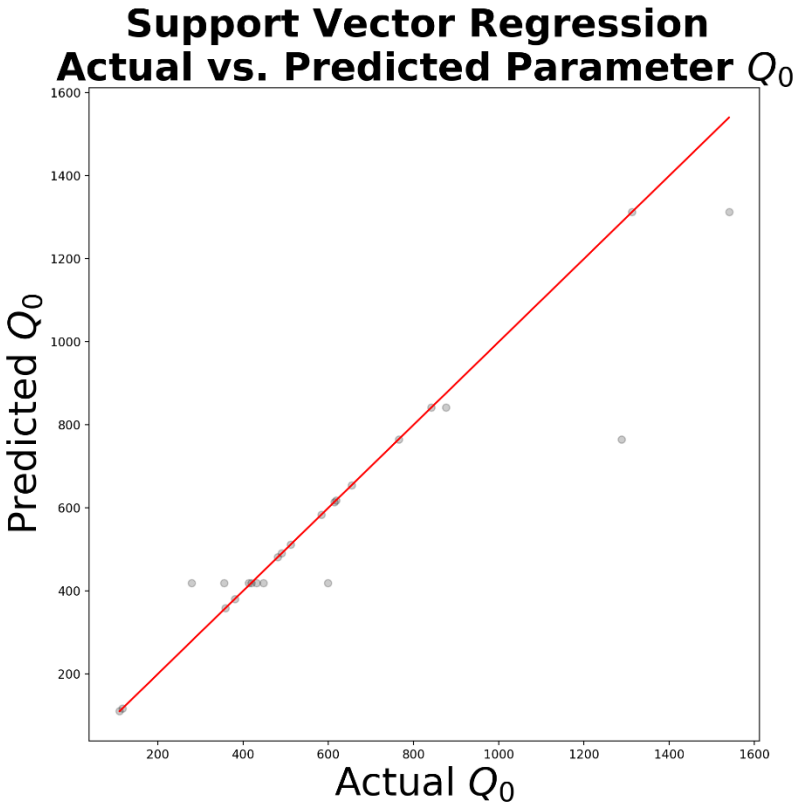
Result 3: Regression of N_2 by Kernel SVR, RBF

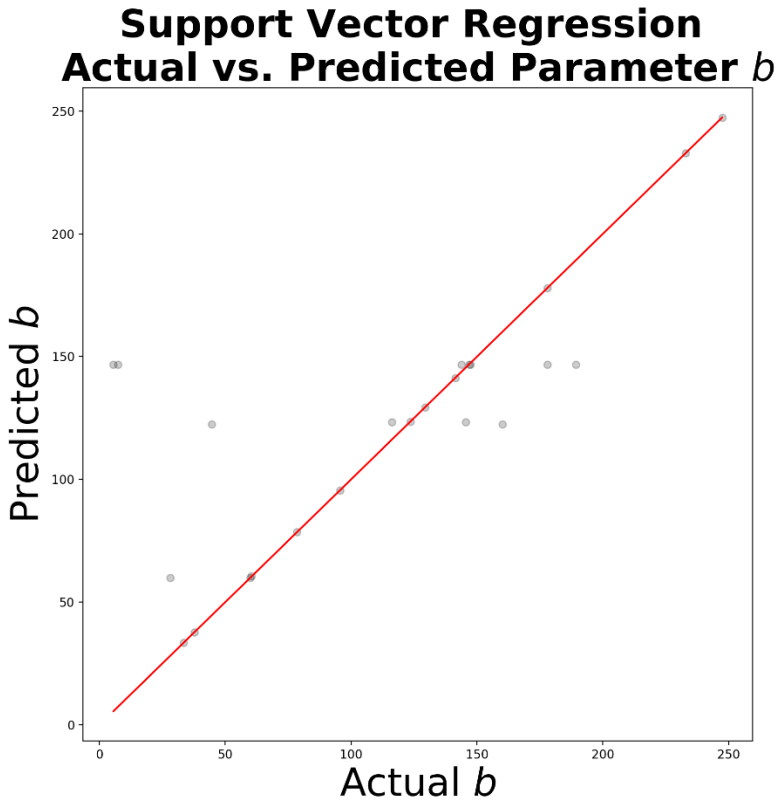
Langmuir:

$$Q_0 = f(PLD, LCD, SA, VF \ldots)$$

$$b = f(PLD, LCD, SA, VF \ldots)$$

Parameters by SVR	R2	CV mean relative error
Q_0	0.87	0.53
b	0.53	2.44





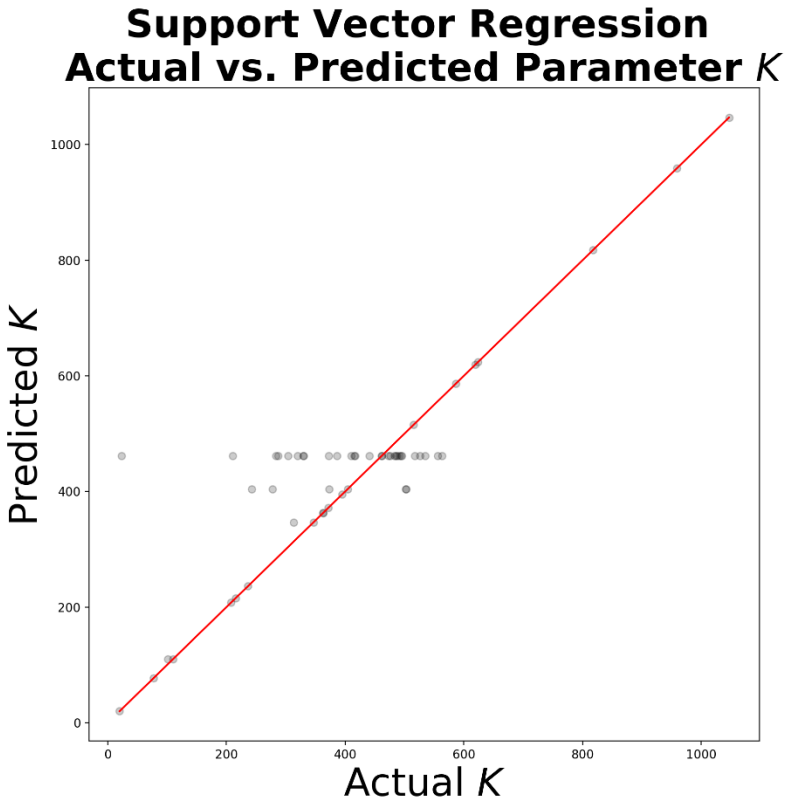
Result 3: Regression of N_2 by Kernel SVR, RBF

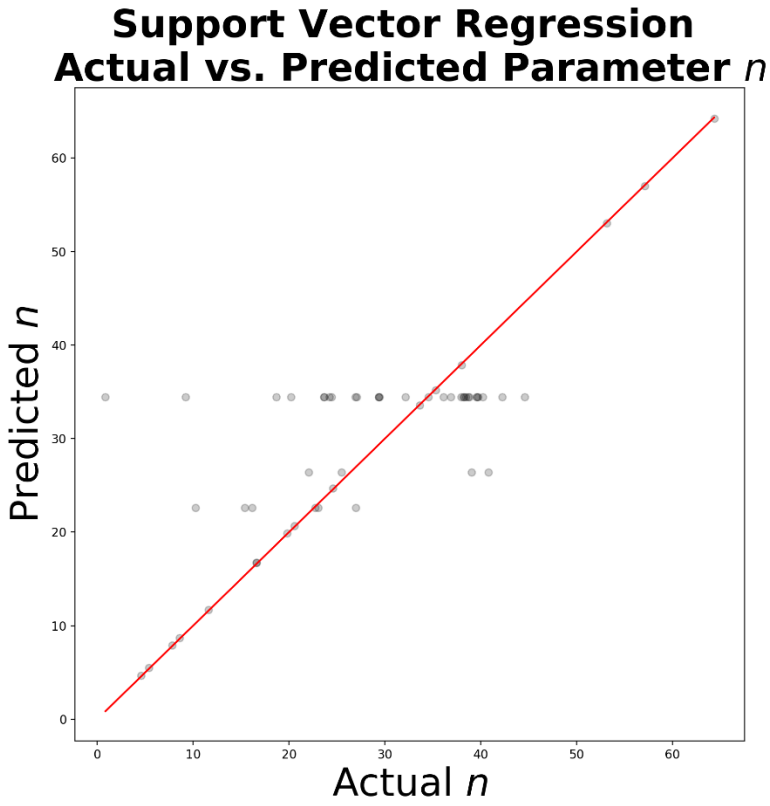
Freundlich:

$$K = f(PLD, LCD, SA, VF \ldots)$$

$$n = f(PLD, LCD, SA, VF \ldots)$$

Parameters by SVR	R2	CV mean relative error
K	0.75	1.10
n	0.60	1.35



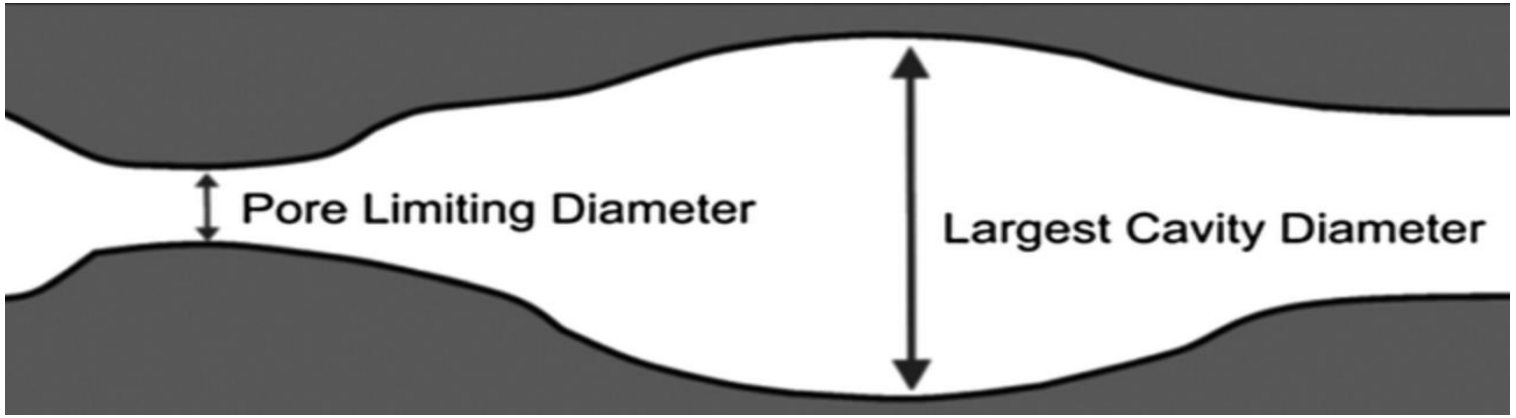


Discussion and Conclusions

- Classification of isotherm models:
  - Tried Logistic, LDA, KNN, SVC method to classify
  - Both KNN and SVC work well. (KNN:82.3% mean accuracy, CV)
- Regression of isotherm parameters for each model:
  - Tried linear regression and support vector regression with RBF
  - SVR works well with higher R2 and lower CV mean relative error
- Analysis on Data source :
  - Same MOF has different isotherm results
  - Training examples are not large enough (124 matched MOFs)
- Future plan:
  - Adding more isotherm types during the model assignment
  - Gather more matched MOFs manually to increase the training set
  - Consider the chemical information of MOFs as the feature vector: use more variables to describe metals and organic linkers

Questions

Backup: Adsorbent Structural Descriptors



- Density (ρ) g/cm3
- Volumetric Surface area (VSA) m2/cm3
- Gravimetric Surface area (GSA) m2/g
- Pore volume (PV): cm3/g
- Void fraction (VF)
- pore-limiting diameter (PLD): Å
- largest-cavity diameter (LCD): Å
- Metal type:
  - Atomic mass (m_a)
  - Atomic radius(r_a)
  - Electronegativity(χ)

Backup: These authors also explored the trade-off between volumetric capacity and gravimetric capacity. The results show a concave downward relationship between volumetric and gravimetric storage capacities.