

Mapping Chemical Performance Space

Chemical	Screen P1	Screen Q1	Screen R1	Screen P2
Standard				
Best				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				

Matrix Completion

- n chemicals, d (rate \times weed), $A \in \mathbb{R}^{n \times d}$
- Approximate A by a rank k matrix X
- $\min \|A - X\|_{Fro}$ over $rank(X) \leq k$, subject to $A_{ij} = X_{ij}$ for all (i, j) not blank.
- the norm is defined to be

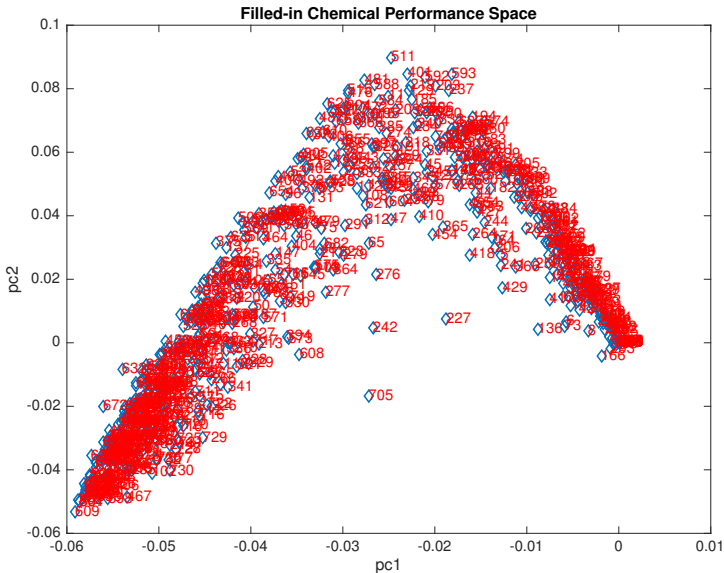
$$\|A - X\|_{Fro}^2 = \sum_{(i,j) \text{ not blank}} |A_{ij} - X_{ij}|^2$$

- set $X = BC$, $B \in \mathbb{R}^{n \times k}$, $C \in \mathbb{R}^{k \times d}$

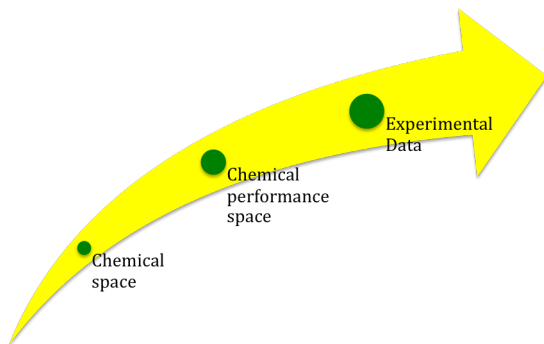
Filled in values

Data =							Data =						
NaN	NaN	NaN	NaN	NaN	NaN	NaN	0	0	0	0	NaN	0	NaN
NaN	90	90	90	100	90	90	0	0	0	0	NaN	0	30
NaN	NaN	NaN	NaN	100	NaN	NaN	0	NaN	0	0	NaN	0	NaN
NaN	90	90	100	100	90	100	0	10	0	0	NaN	0	60
NaN	NaN	NaN	100	100	NaN	100	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	100	100	NaN	100	NaN	NaN	NaN	NaN	NaN	NaN	NaN
FilledIn =							FilledIn =						
0	110	21	43	0	9	103	0	0	0	0	0	0	1
58	90	90	90	100	90	90	0	0	0	0	-449	0	30
60	100	94	92	100	66	99	0	26	0	0	-346	0	30
84	90	90	100	100	90	100	0	10	0	0	-319	0	60
91	97	103	100	100	103	100	1	32	1	3	-340	1	32
101	94	99	100	100	99	100	6	36	17	15	-206	13	39

Building the Chemical Performance Space



Things to Do



- Setting a prior for Not Missing At Random e.g. Bayesian Networks.
- Incorporating prior into matrix completion algorithm.
- Non-linear mappings e.g. Gaussian Process Latent Variable Model.
- Incorporating Chemical Structure Information e.g. Gaussian Process based on molecular similarity measures.