

# Active Learning: A Machine Learning Algorithm for Improved Test Selection

Better Results with Fewer Tests

Lyu Jian Lu Advised By Dr. Hua Wang

# A brief introduction to Machine Learning Terms for AM

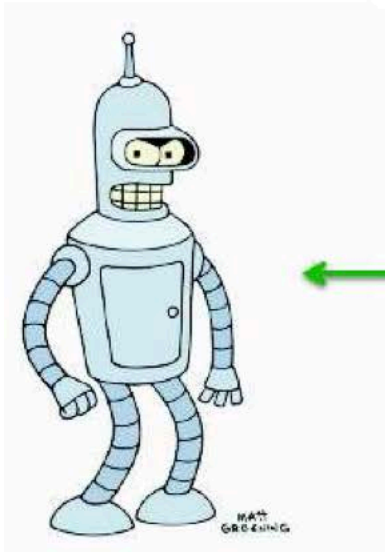
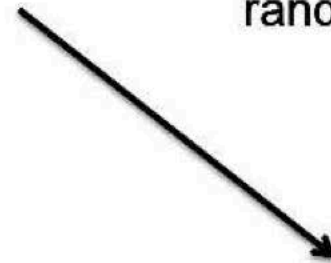
- **Feature** is a control variable, such as laser speed, laser power, or powder size.
- **Response** is a property of a sample, for example, its hardness, toughness or stiffness.
- **Labeled Data** is data where features are paired with responses.
- **Unlabeled Data** considers only the features.
- **Classifier** is a way to group data based on similar features, similar responses, or both.

# Passive Learning relies on an expert to choose which samples to label



raw unlabeled data  
 $x_1, x_2, x_3, \dots$

random sample



**Passive Learner**  
Train a Classifier

labeled training instances

$\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle, \langle x_3, y_3 \rangle, \dots$



**Expert/Oracle**  
Analyze Experiments  
to determine labels

# Active Learning helps choose which samples should be tested (labeled)



inspect the  
unlabeled data

raw unlabeled data  
 $x_1, x_2, x_3, \dots$

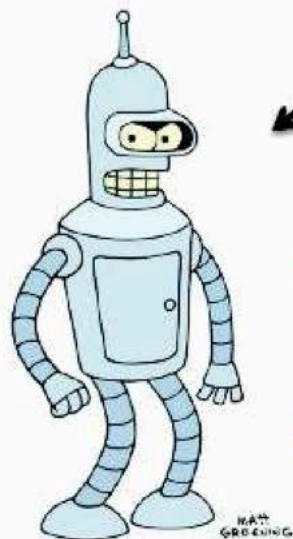
request labels for selected data

$\langle x_1, ? \rangle$

$\langle x_1, y_1 \rangle$

$\langle x_2, ? \rangle$

$\langle x_2, y_2 \rangle$



**Active Learner**  
Train a Classifier

**Expert/Oracle**  
Analyze Experiments  
to determine labels

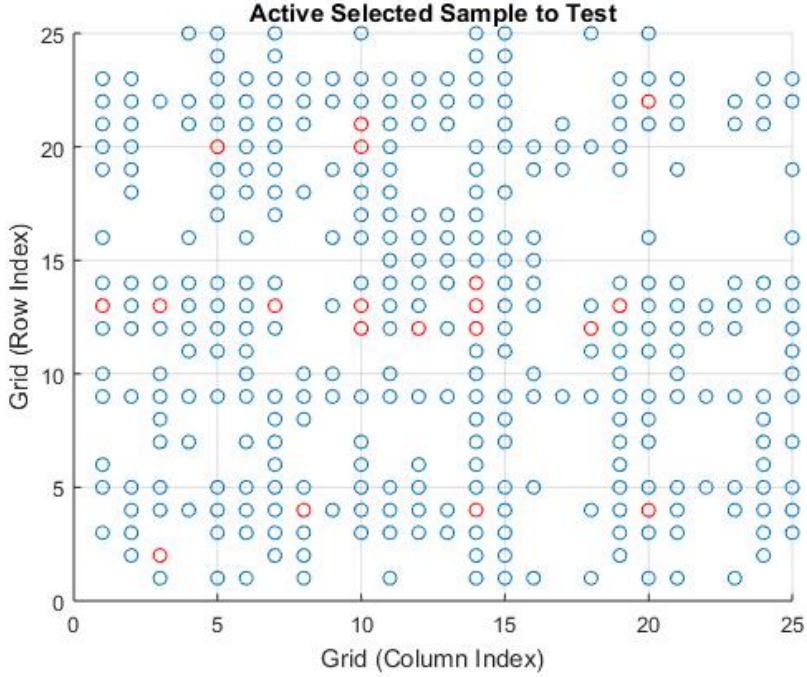
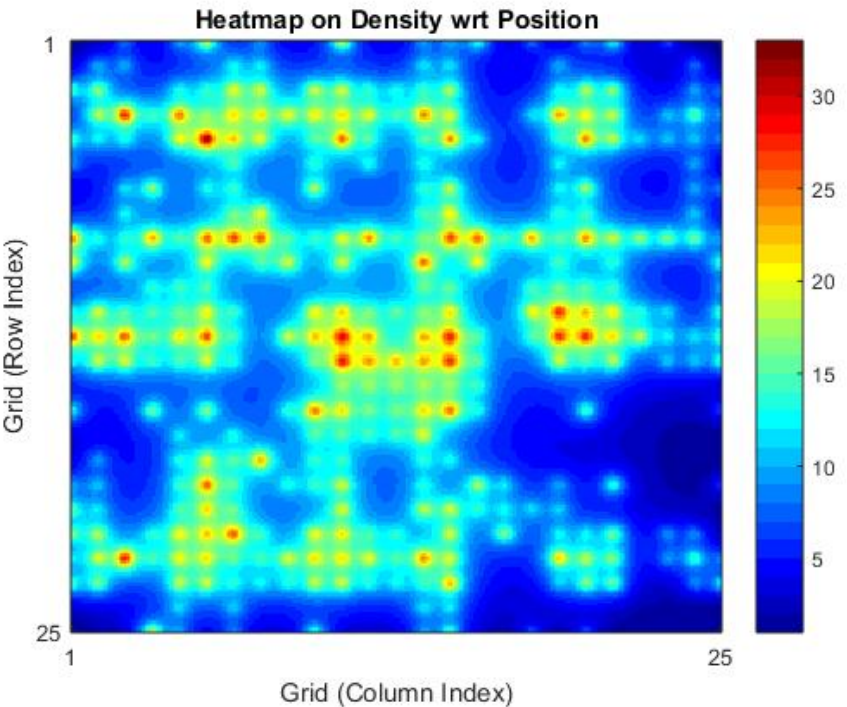
# Apply Active Learning to ADAPT

## Dataset

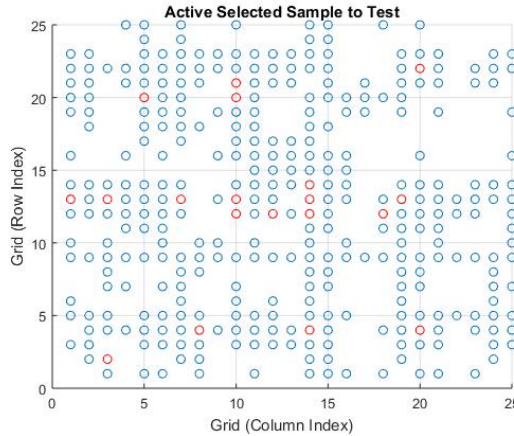
	Polar Angle	Azimuth Angle	Grid (Column)	Grid (Row)	...	Laser Speed	Laser Power	
Sample 1	45	90	Y	12		800	160	
Sample 2	0	0	F	13	...	800	160	
Sample 3	45	180	F	10		800	160	→ $x_i$
Sample 4	90	45	M	15	...	920	136	
Sample 5	45	270	K	20		920	136	→ $x_j$
Sample 6	90	0	D	25		800	136	
Sample 7	0	0	B	22	...	800	160	
Sample 8	45	0	L	9		800	160	
Sample 9	45	90	T	12		680	184	→ $x_k$
...	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
Sample n-1	90	90	C	14		800	160	
Sample n	45	0	P	9	...	800	160	

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \xrightarrow{\text{Actively Select Samples to detect}} \tilde{X} = \begin{bmatrix} x_i \\ x_j \\ \vdots \\ x_k \end{bmatrix}$$

# Apply Active Learning to ADAPT Dataset



# Our Goal: Representatively sample the features that make the data unique



$$J = \min \|X - W\tilde{X}\|_2^2 \quad \longrightarrow \text{Representation Error}$$

$$s.t. \text{rank}(W) < C \quad \longrightarrow \text{Limit the number of Selected Samples}$$

Where  $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$  denotes the original dataset samples

$\tilde{X} = \begin{bmatrix} x_i \\ x_j \\ \vdots \\ x_k \end{bmatrix}$  denotes the selected data samples to be tested

$W$  denotes the importance with each selected samples

$C$  denotes the maximum number of samples we select



# Use $\ell_{21}$ -norm to select the tests

$$J = \|X - W\tilde{X}\|_2^2 + \lambda \text{trace}(W)$$

Where  $\|M\|_2^2 = \begin{bmatrix} m_{11} & \dots & m_{d1} \\ \vdots & \ddots & \vdots \\ m_{1n} & \dots & m_{dn} \end{bmatrix}_2^2 = \sum m_{1i}^2 + m_{2i}^2 + \dots + m_{di}^2$

$$\frac{30^2}{5^2 + 5^2 + 5^2 + 5^2 + 30^2} = 90\%$$

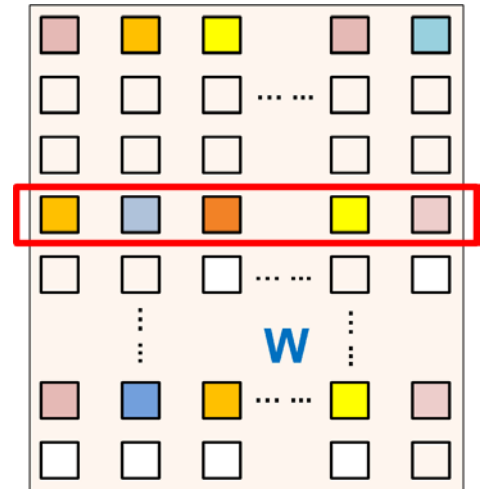
Sensitive to Outliers

$$\frac{30}{5 + 5 + 5 + 5 + 30} = 60\%$$

$$J = \|X - WX\|_{2,1} + \lambda \|W\|_{2,1}$$

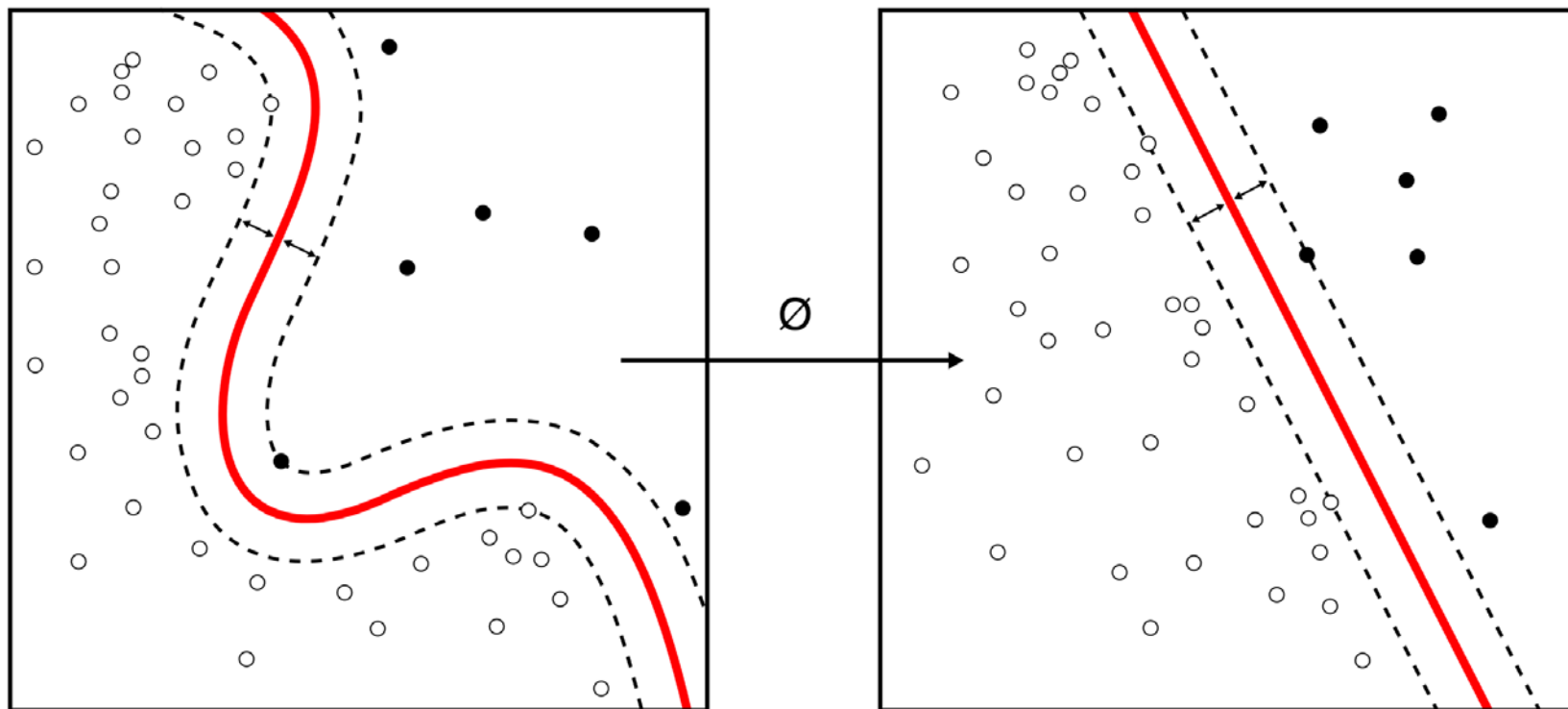
Either select an entire row, or discard an entire row

Where  $\|M\|_{2,1} = \begin{bmatrix} m_{11} & \dots & m_{d1} \\ \vdots & \ddots & \vdots \\ m_{1n} & \dots & m_{dn} \end{bmatrix}_{2,1} = \sum \sqrt{m_{1i}^2 + m_{2i}^2 + \dots + m_{di}^2}$



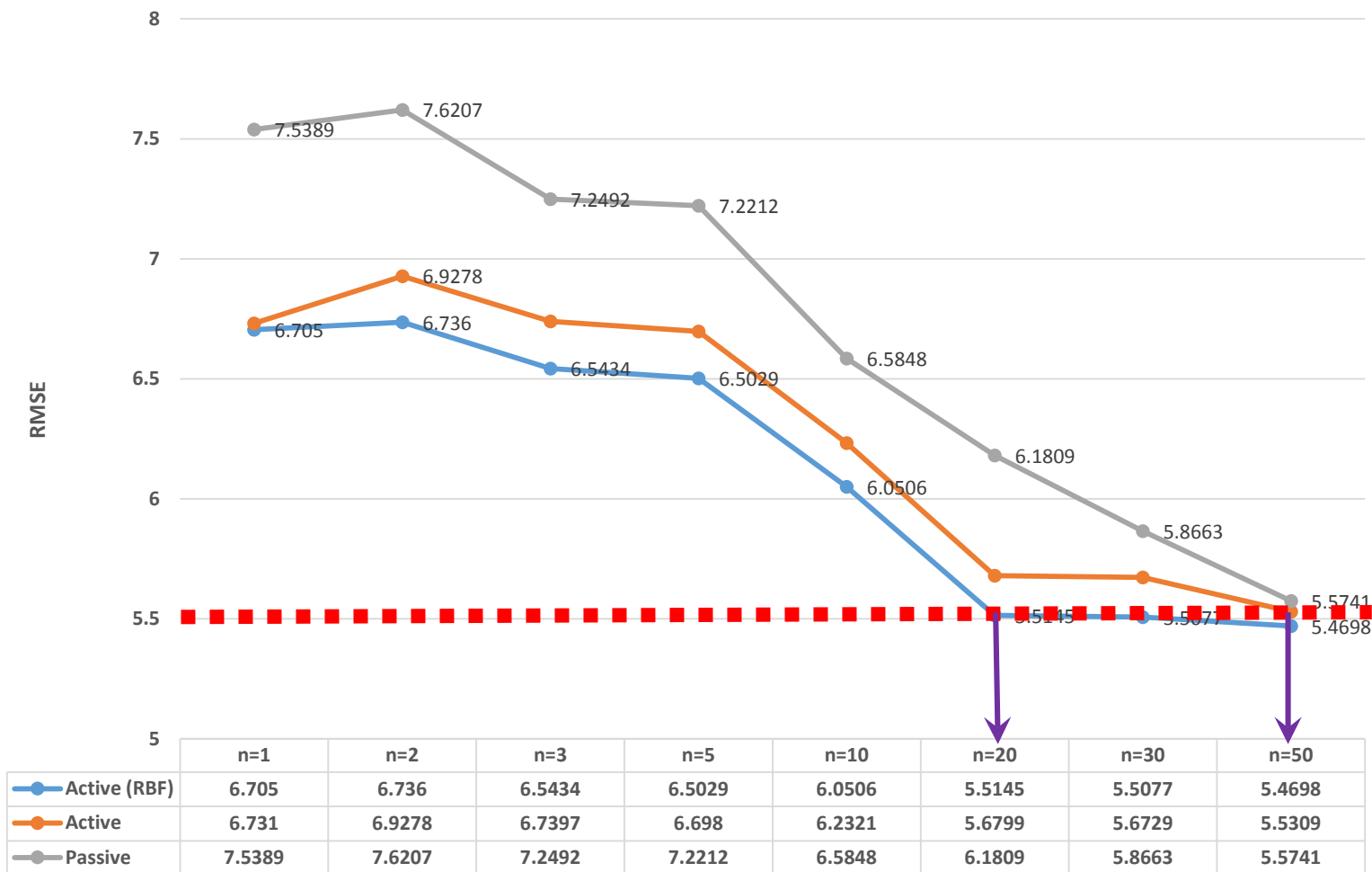


# Kernel Trick to capture the nonlinear relationship



$$J = \|\phi(X) - W\phi(X)\|_{2,1} + \lambda\|W\|_{2,1}$$

# Experimental Results – Better results with Fewer Tests



# Future Work – Incorporate Existence Response to Better Selection Model

	Polar Angle	Azimuth Angle	Grid (Column)	Grid (Row)	...	Laser Speed	Laser Power	Pore Size	Young's Modulus	Yield Strength	...	Ductility
Sample 1	45	90	<i>Y</i>	12		800	160	75.03	217	651		?
Sample 2	0	0	<i>F</i>	13	...	800	160	174.17	?	?	...	?
Sample 3	45	180	<i>F</i>	10		800	160	76.44	?	?		?
Sample 4	90	45	<i>M</i>	15	...	920	136	185.88	70	494.7		?
Sample 5	45	270	<i>K</i>	20		920	136	?	?	?	...	?
Sample 6	90	0	<i>D</i>	25		800	136	?	?	?		?
Sample 7	0	0	<i>B</i>	22	...	800	160	58.5	?	?		?
Sample 8	45	0	<i>L</i>	9		800	160	146.72	100	346.4		?
Sample 9	45	90	<i>T</i>	12		680	184	?	?	?		?
...	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Sample n-1	90	90	<i>C</i>	14		800	160	87.22	130	678		15.4
Sample n	45	0	<i>P</i>	9	...	800	160	74.44	?	?		?

Control Variables

Sample Response

**Thank You!**