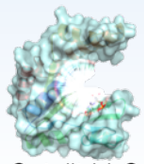
DataWknds.

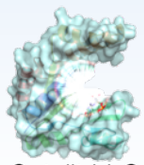
REGRESSION



Catalit LLC

REGRESSION

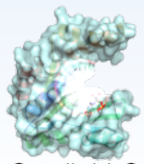
	CONTINUOUS	CATEGORICAL
SUPERVISED	?	?
UNSUPERVISED	?	?

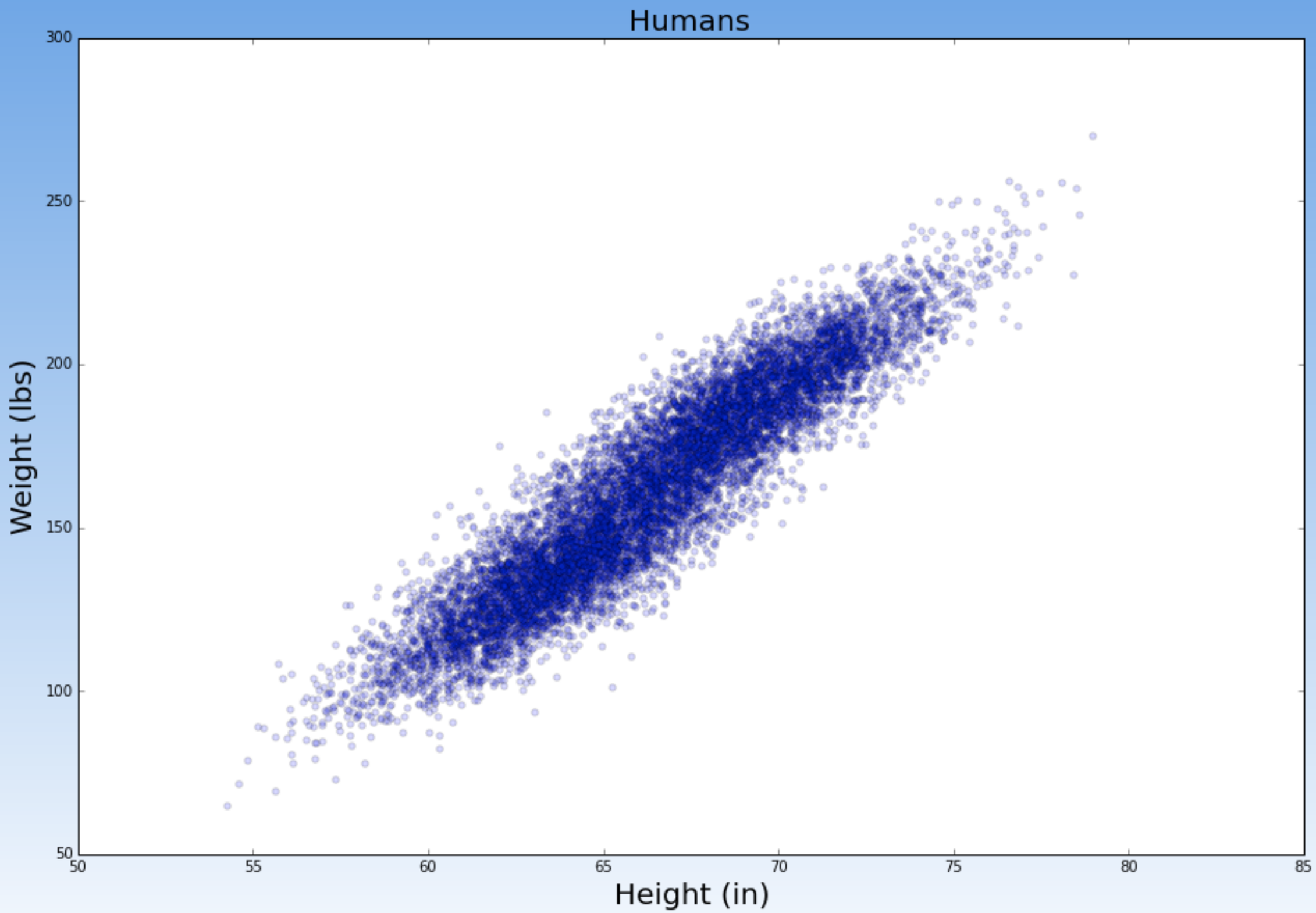


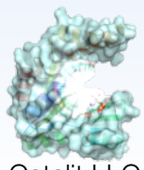
Catalit LLC

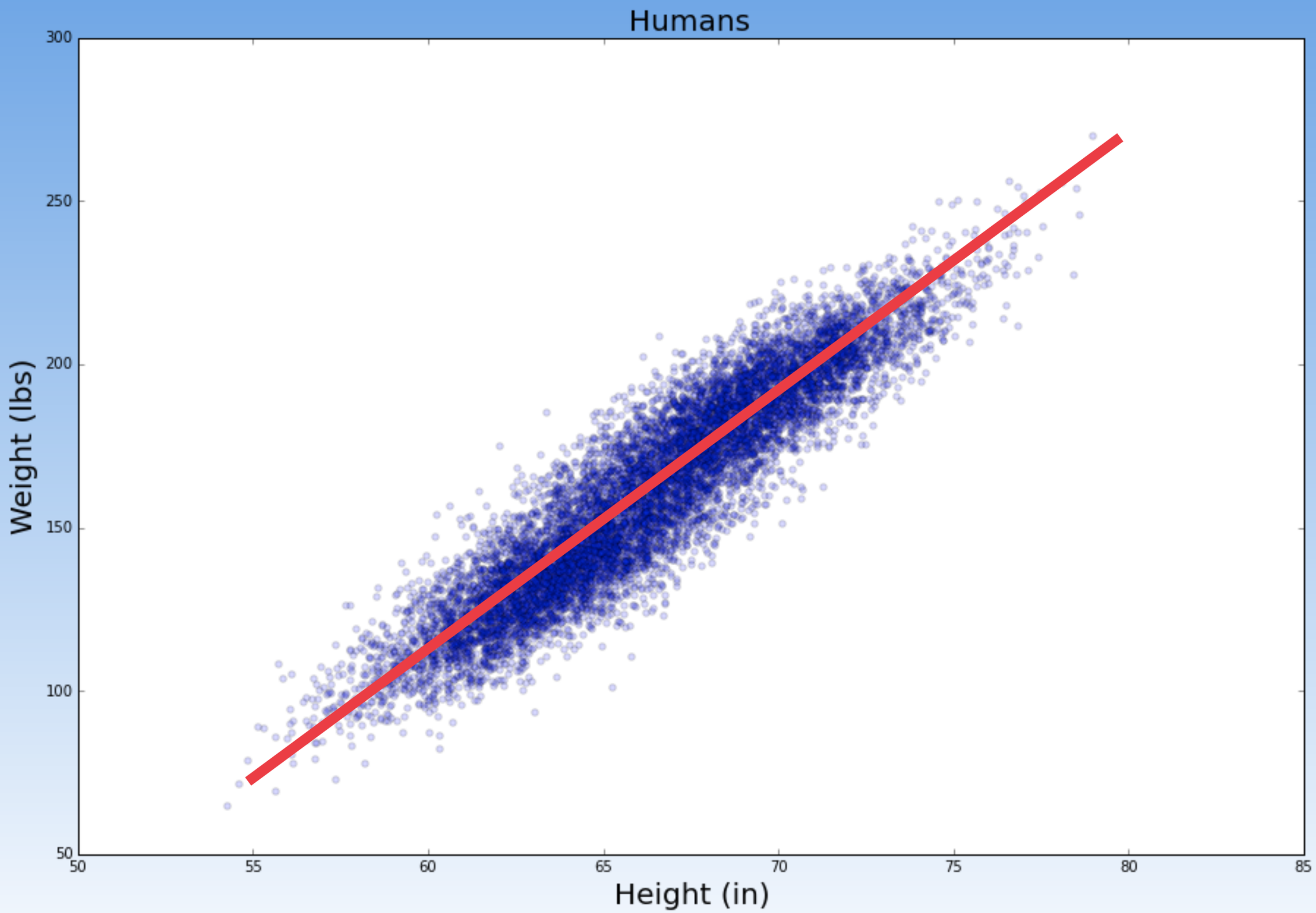
REGRESSION

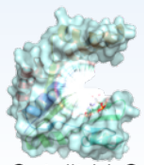
	CONTINUOUS	CATEGORICAL
SUPERVISED	REGRESSION	CLASSIFICATION
UNSUPERVISED	DIMENSION REDUCTION	CLUSTERING



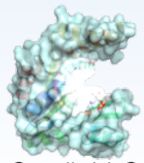

Humans
Weight (lbs)
Height (in)
300
250
200
150
100
50
50
55
60
65
70
75
80
85




Humans
Weight (lbs)
300
250
200
150
100
50
50
55
60
65
70
75
80
85
Height (in)



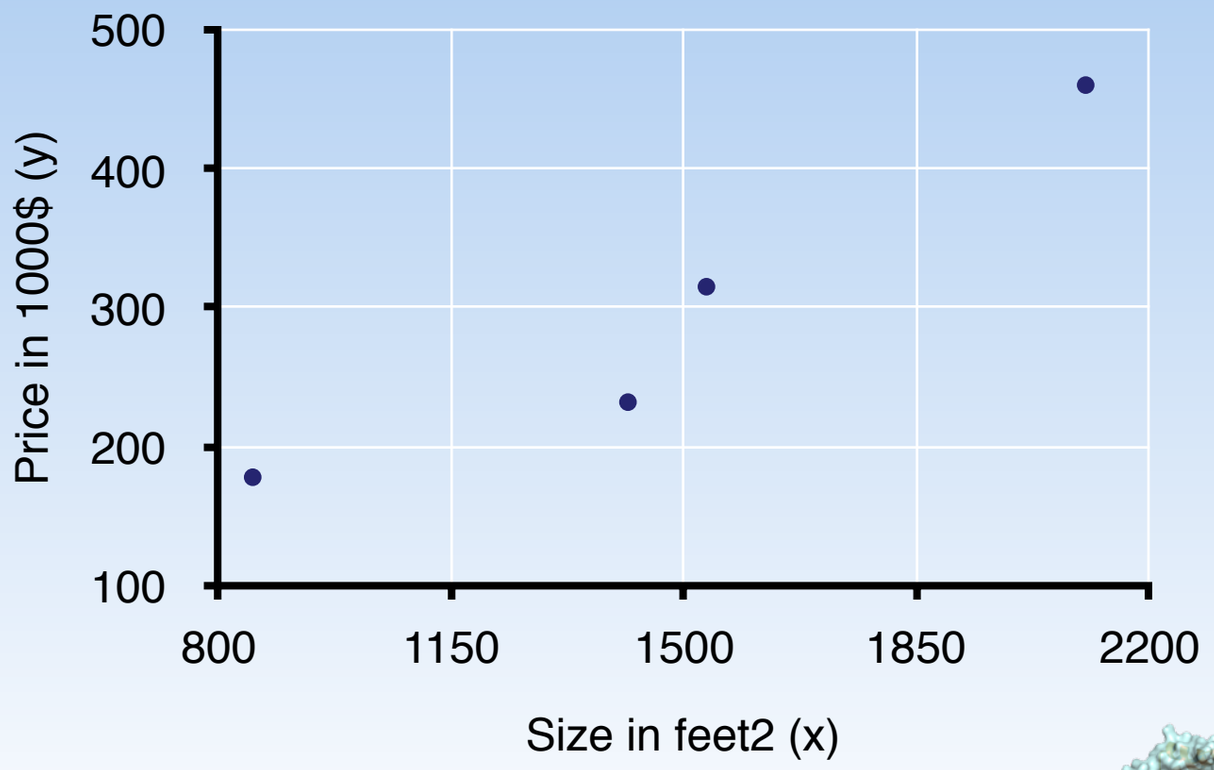
HOW DOES REGRESSION WORK?



Catalit LLC

Size in feet² (x)	Price in 1000$ (y)
2104	460
1416	232
1534	315
852	178
…	…

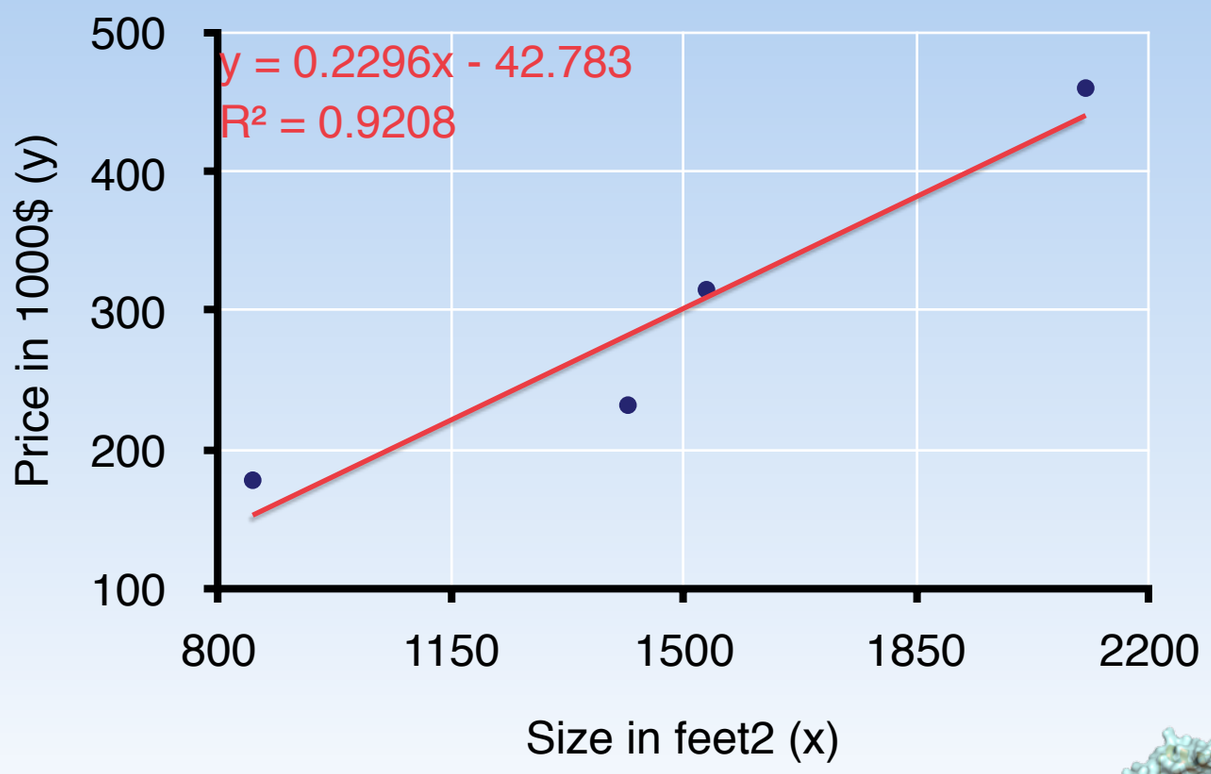
y = func(x)



Hypothesis

$$y = h(x) = \beta_0 + \beta_1 x$$

Size in feet² (x)	Price in 1000$ (y)
2104	460
1416	232
1534	315
852	178
…	…



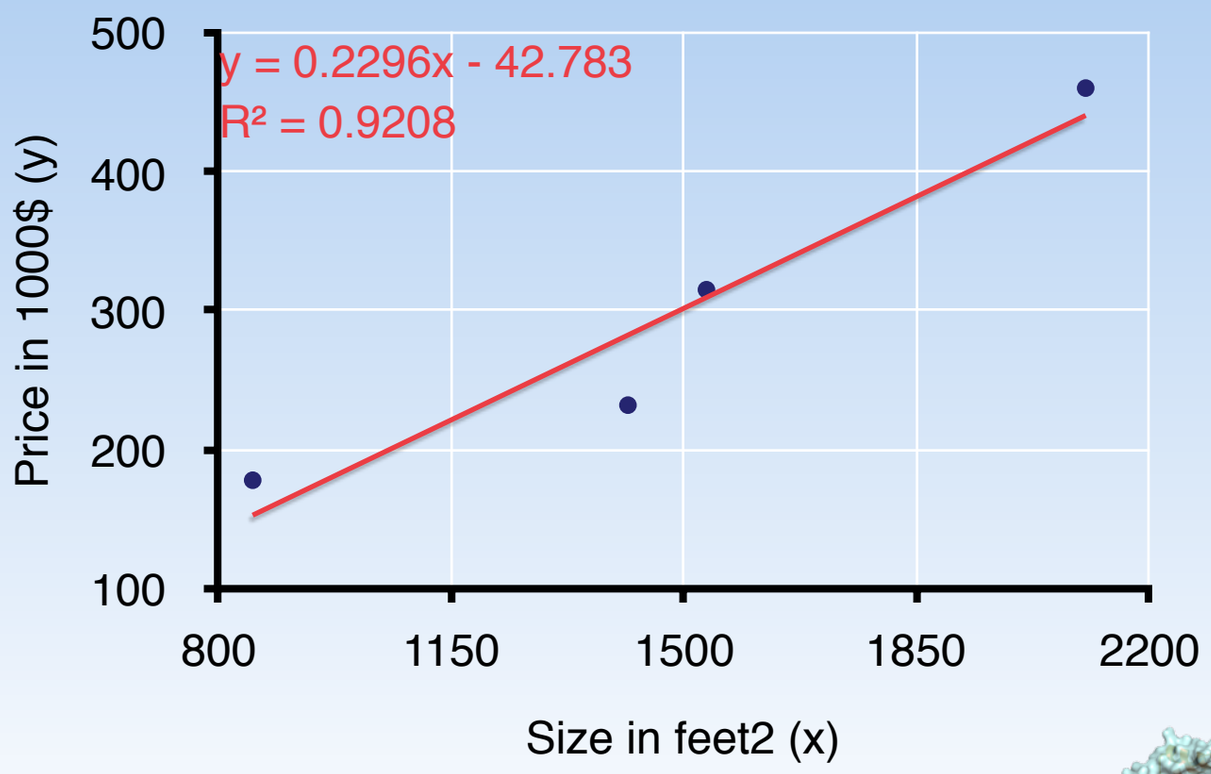
Hypothesis

$$y = h(x) = \beta_0 + \beta_1 x$$

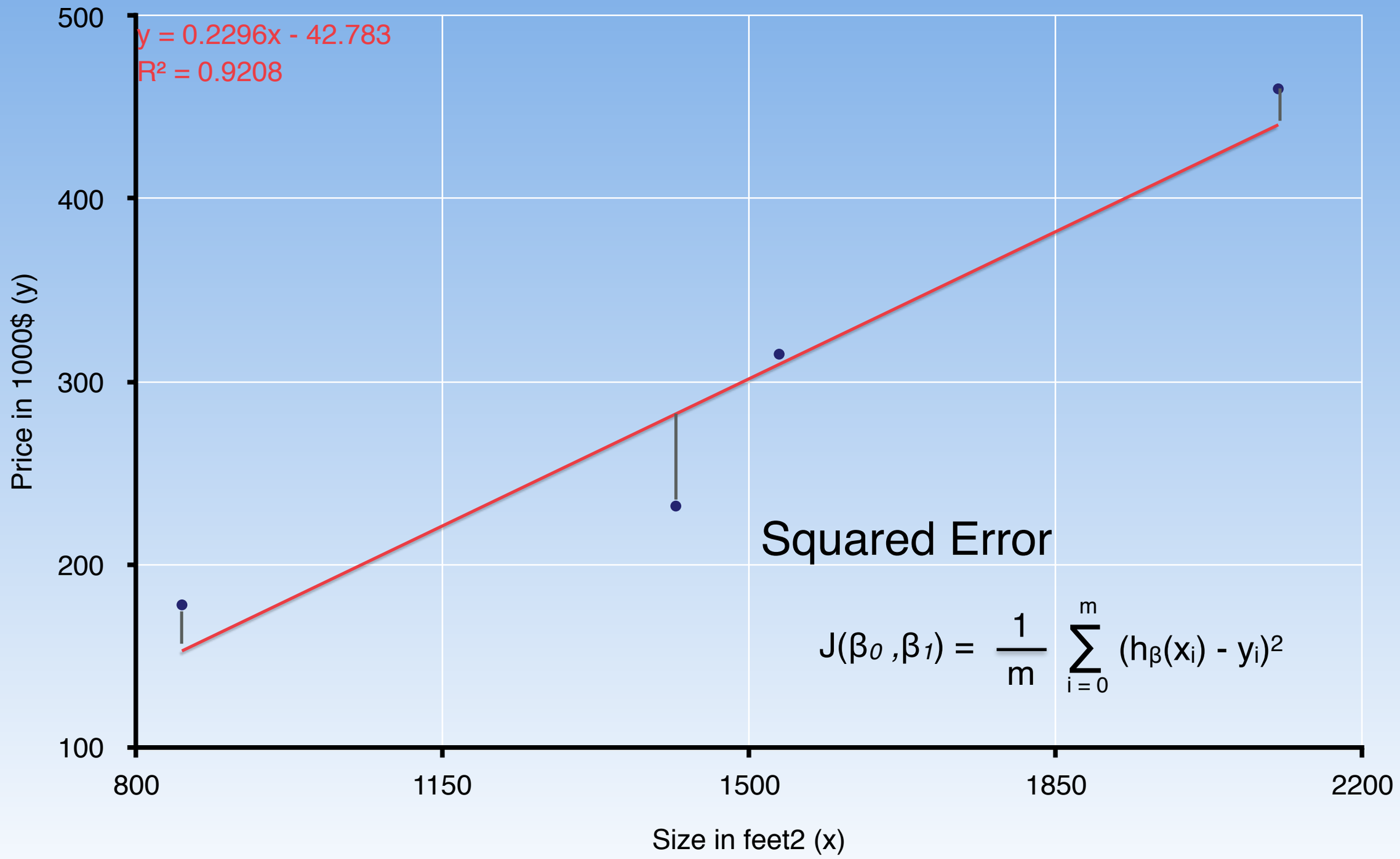
How do I determine the βs?

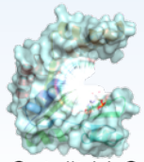
IDEA: choose the βs to minimize distance of h(x) from training data (x, y)

Size in feet² (x)	*Price in 1000$ (y)*
2104	*460*
1416	*232*
1534	*315*
852	*178*
…	*…*



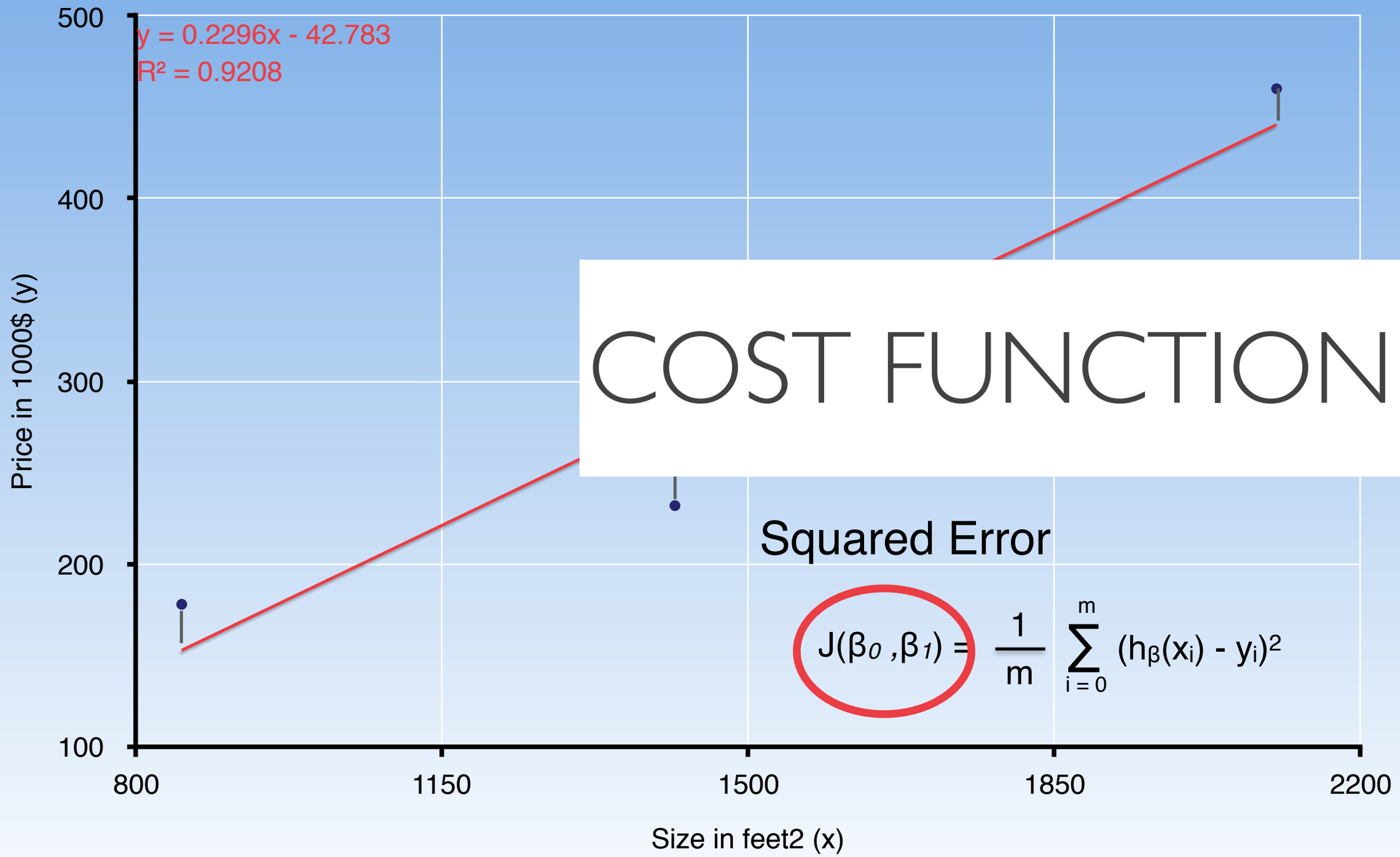
DISTANCE





Catalit LLC

DISTANCE

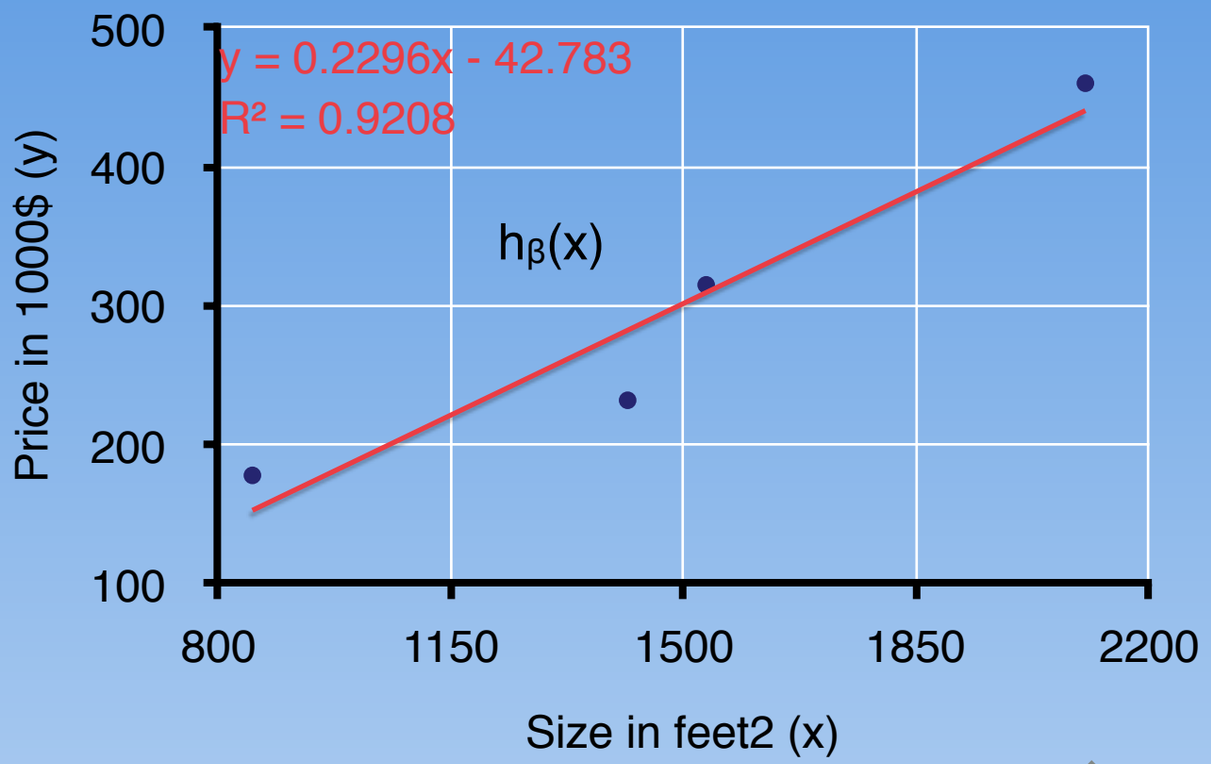


COST FUNCTION

Squared Error

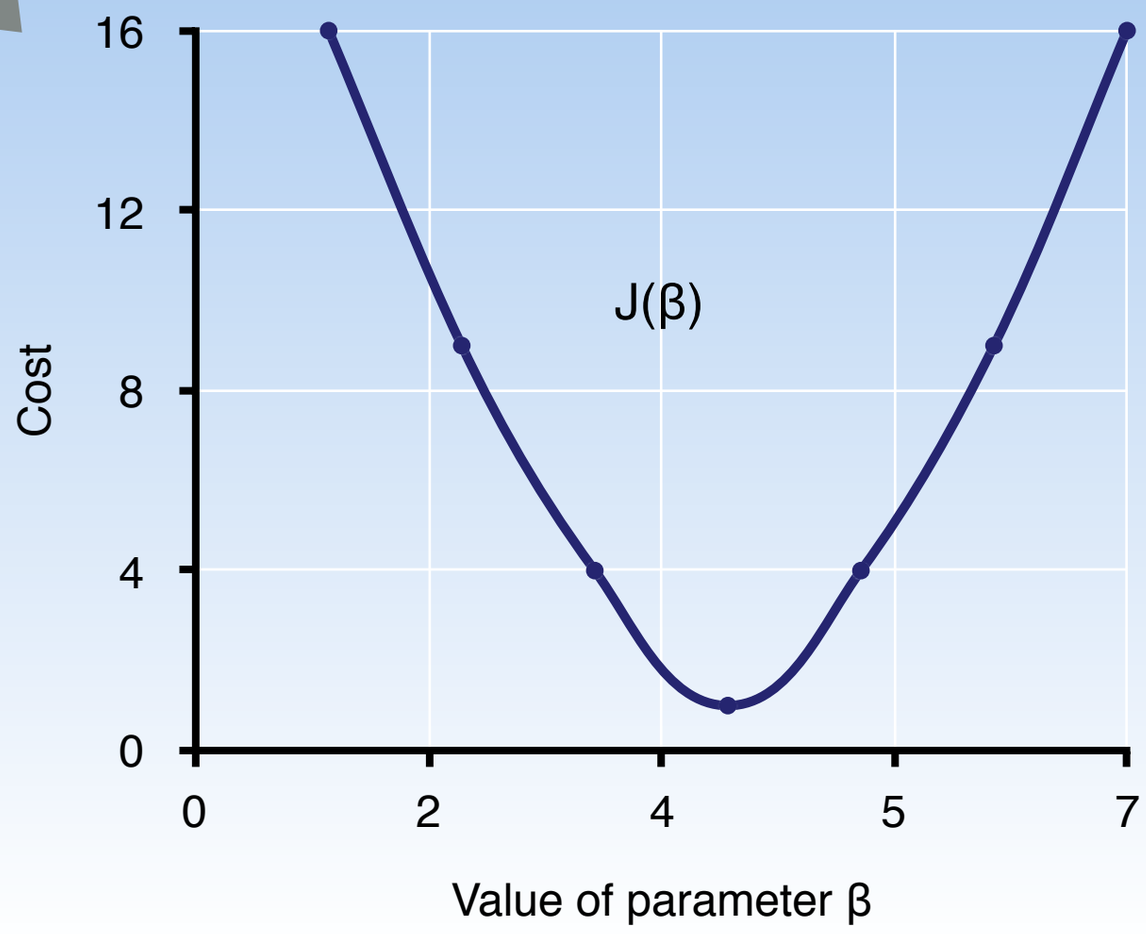
$$J(\beta_0, \beta_1) = \frac{1}{m} \sum_{i=0}^{m} (h_\beta(x_i) - y_i)^2$$

Catalit LLC

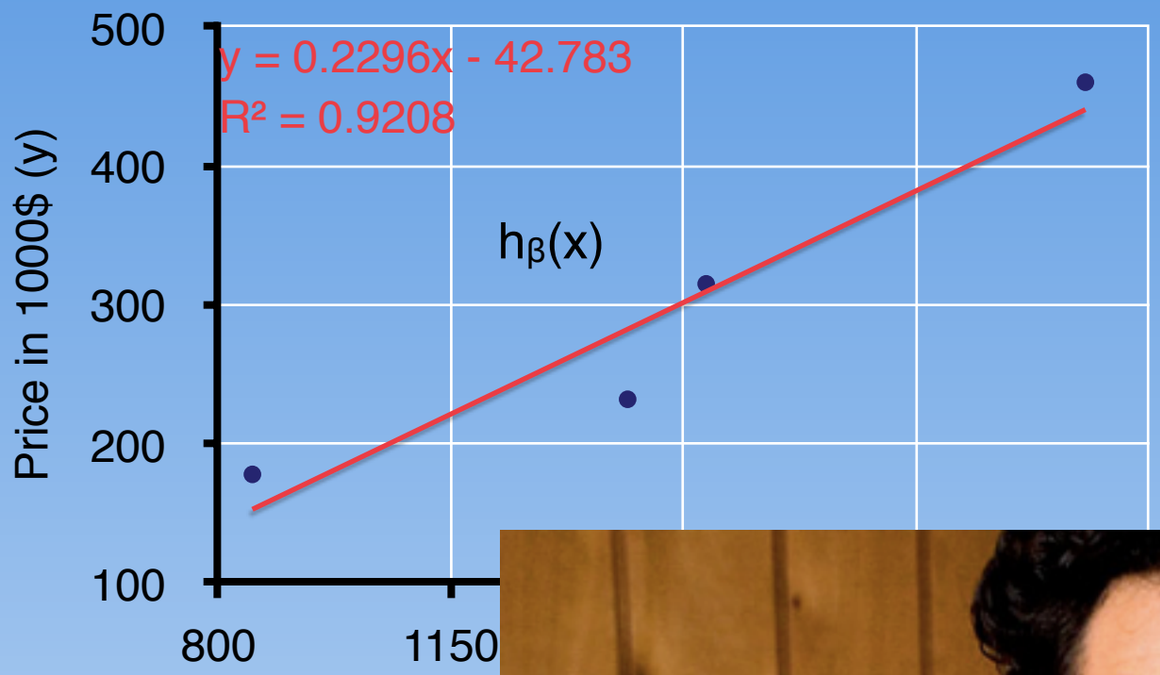


Define Hypothesis
Define Cost

Minimize Cost

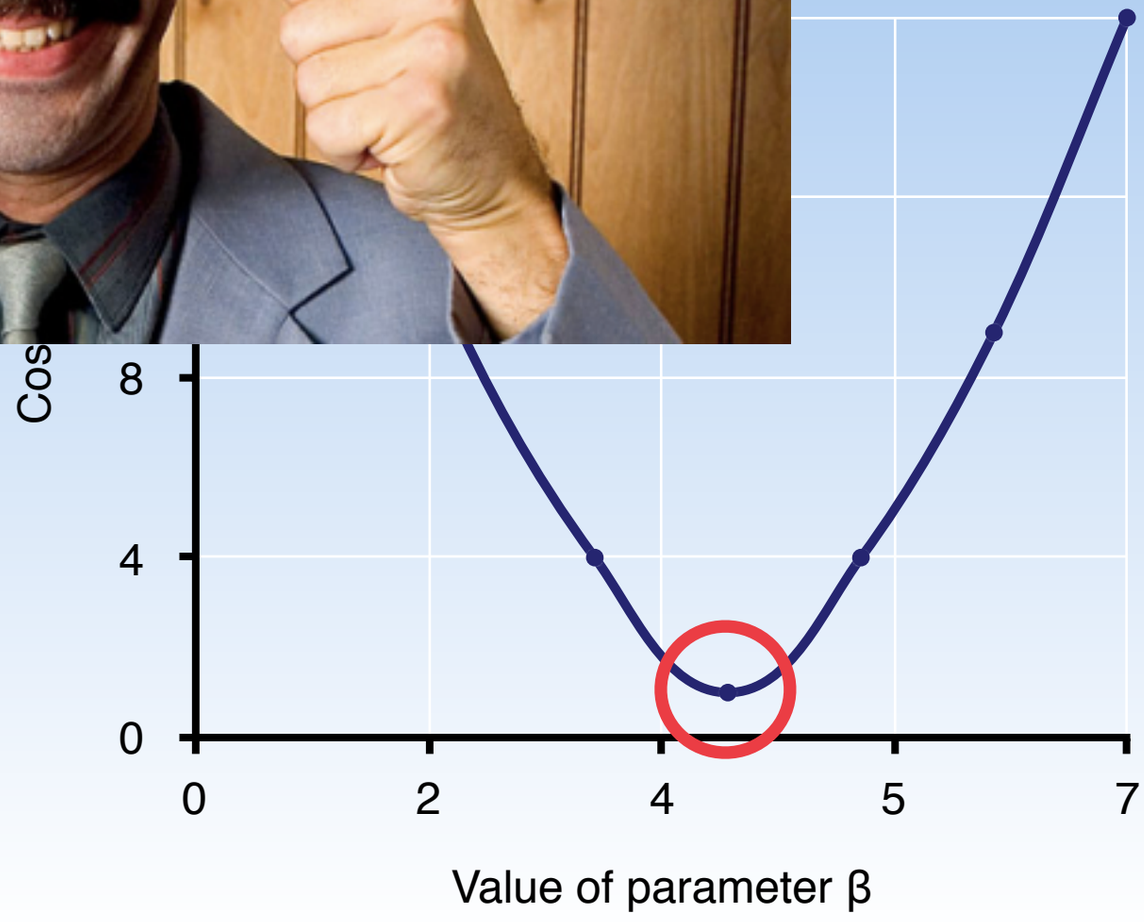


Catalit LLC

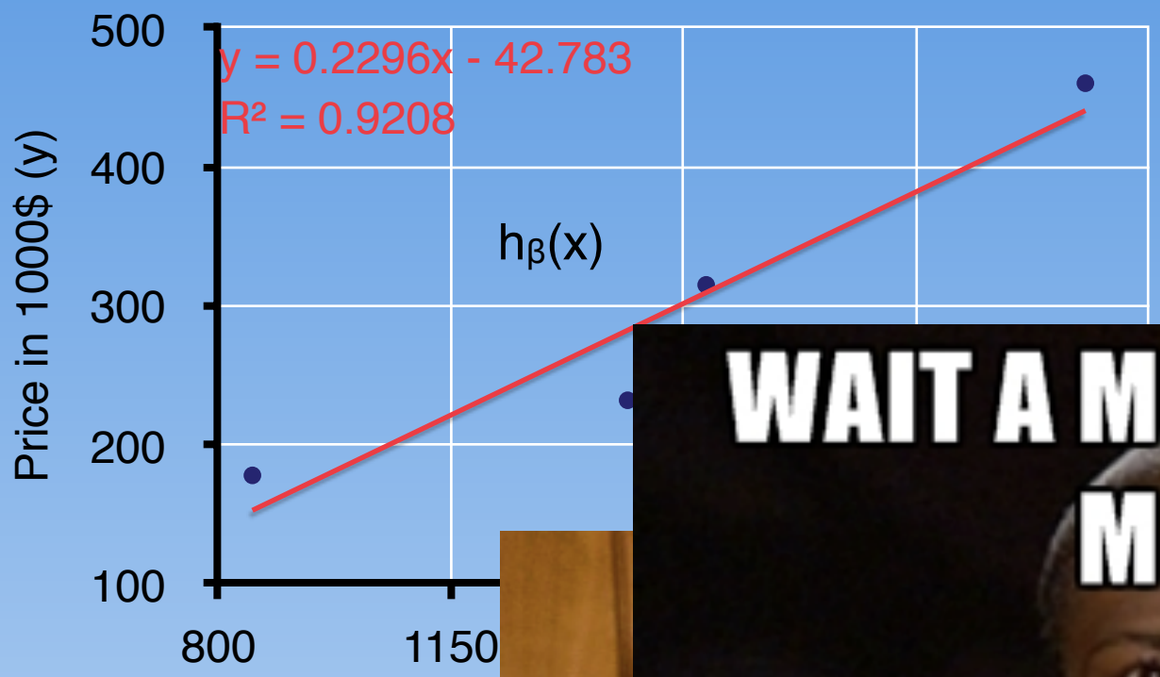


Define Hypothesis
Define Cost

Minimize Cost

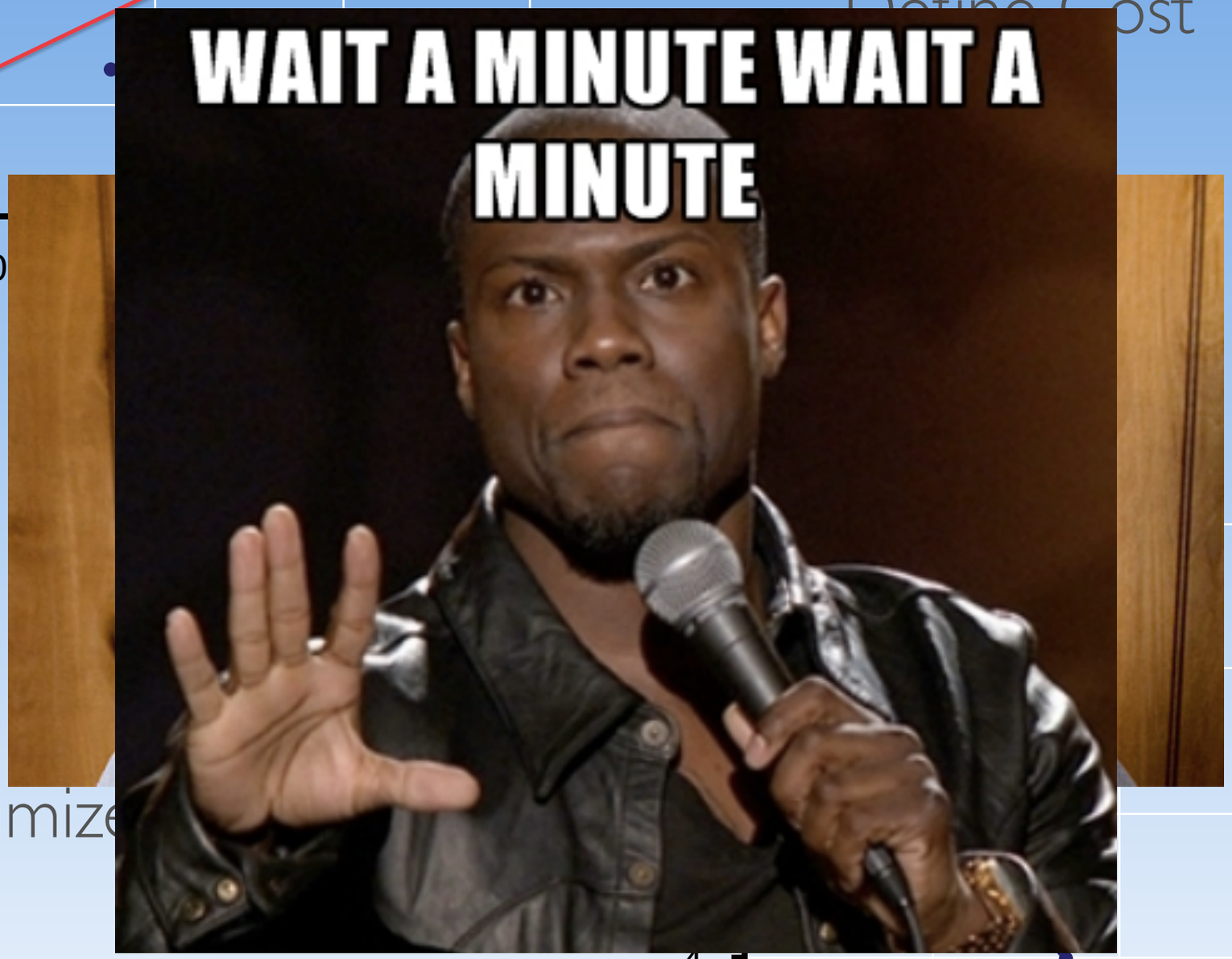


Catalit LLC

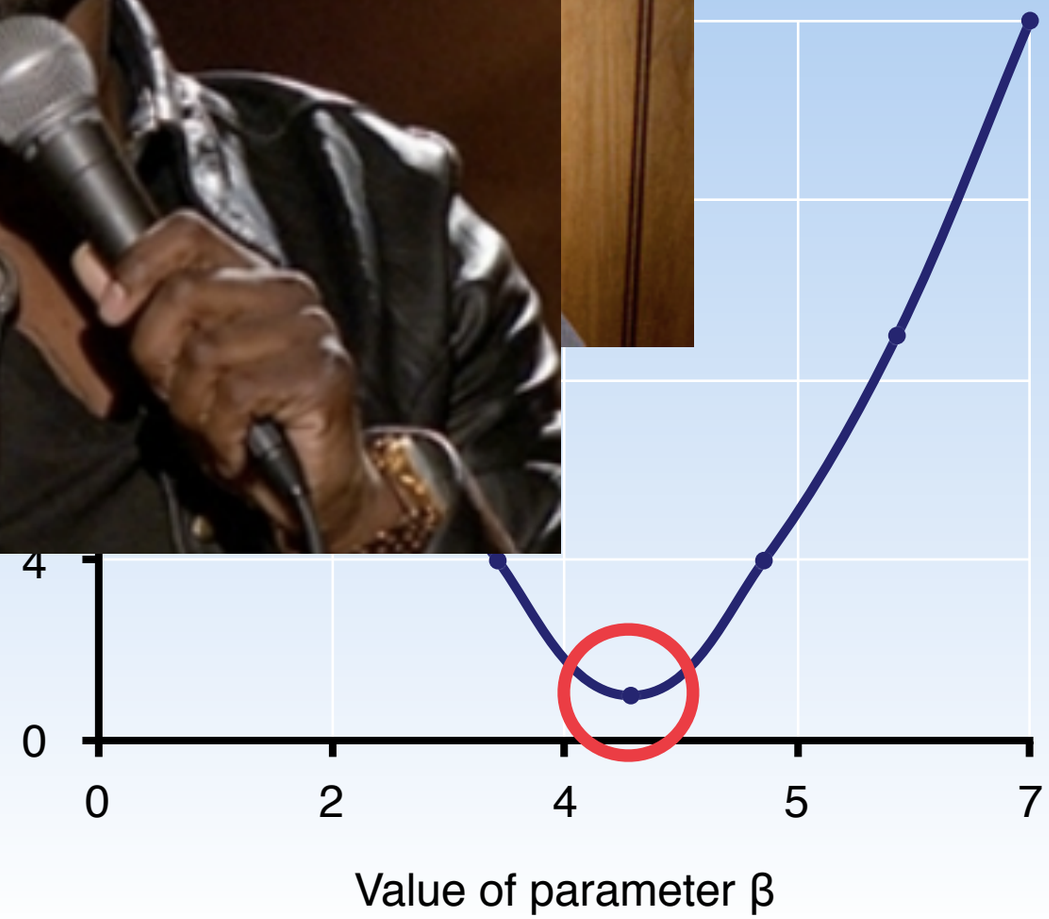


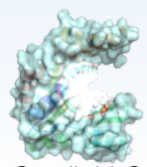
Define Hypothesis

Define Cost

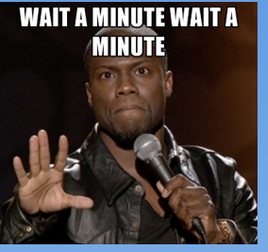


Minimize



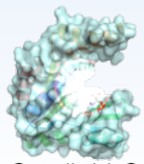


Catalit LLC

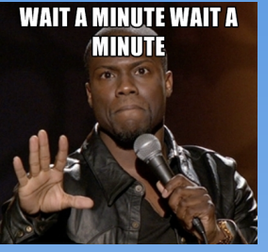


MANY FEATURES

Size in feet² (x1)	*Age (x2)*	*Number of rooms (x3)*	*…*	*Price in 1000$ (y)*
2104	*12*	*3*	*…*	*460*
1416	*4*	*2*	*…*	*232*
1534	*23*	*3*	*…*	*315*
852	*7*	*1*	*…*	*178*
…	*…*	*…*	*…*	*…*



Catalit LLC

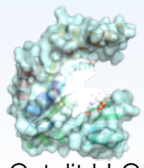


MANY FEATURES

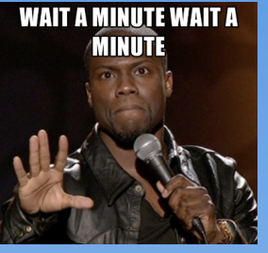
Size in feet² (x1)	*Age (x2)*	*Number of rooms (x3)*	*…*	*Price in 1000$ (y)*
2104	*12*	*3*	*…*	*460*
1416	*4*	*2*	*…*	*232*
1534	*23*	*3*	*…*	*315*
852	*7*	*1*	*…*	*178*
…	*…*	*…*	*…*	*…*

No problem!

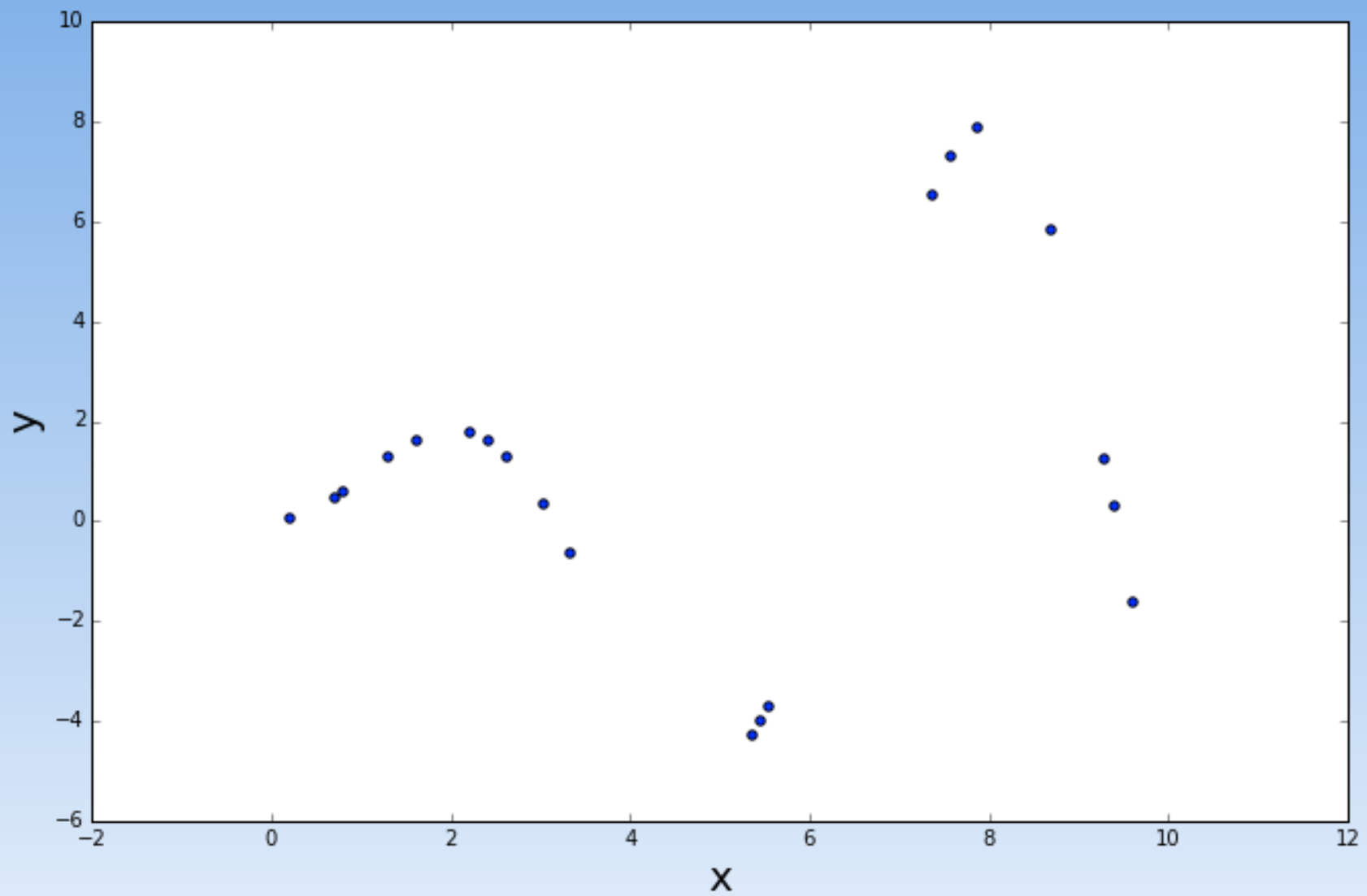
$$y = h(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$



Catalit LLC

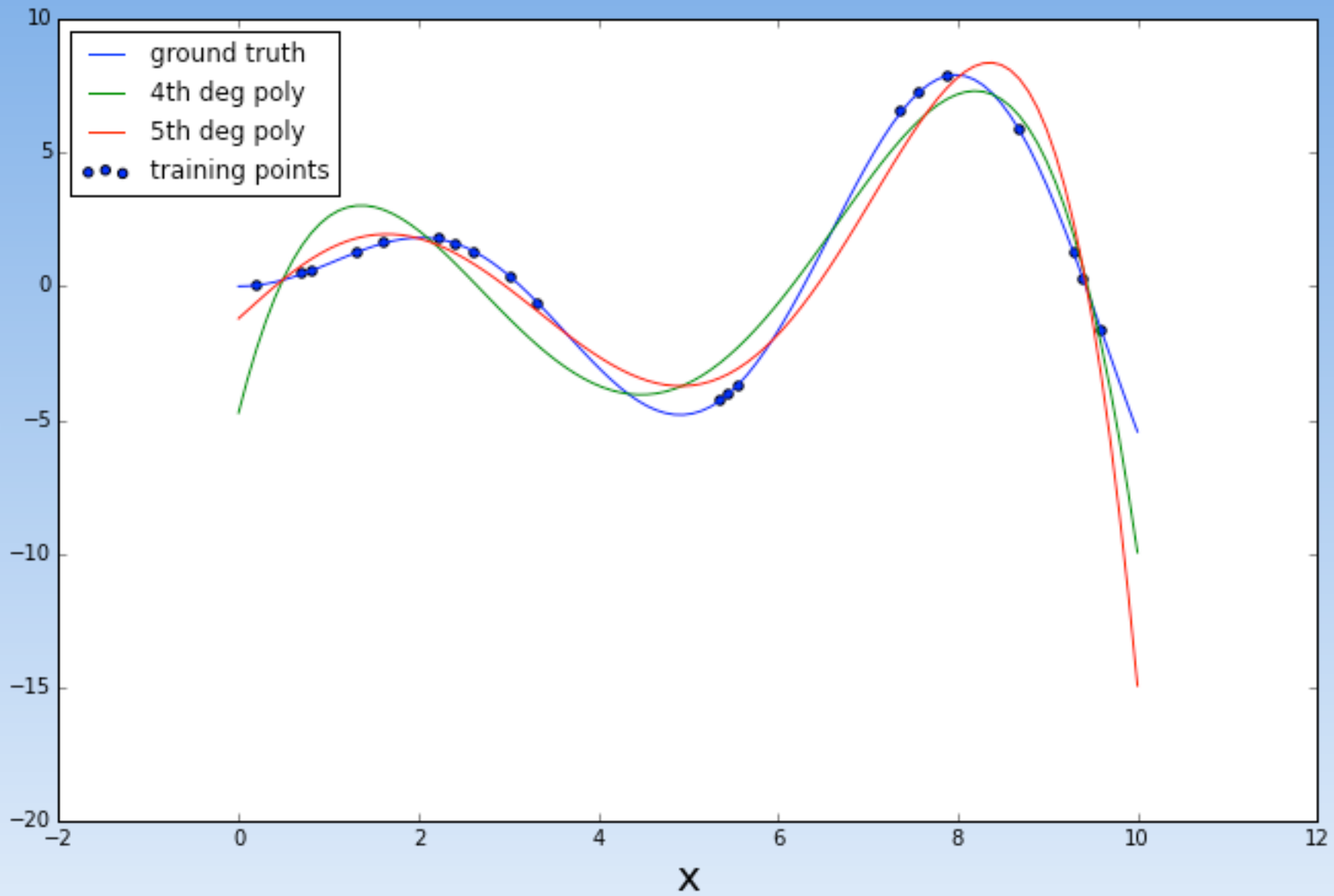


NON LINEAR RELATION



Catalit LLC

NON LINEAR RELATION



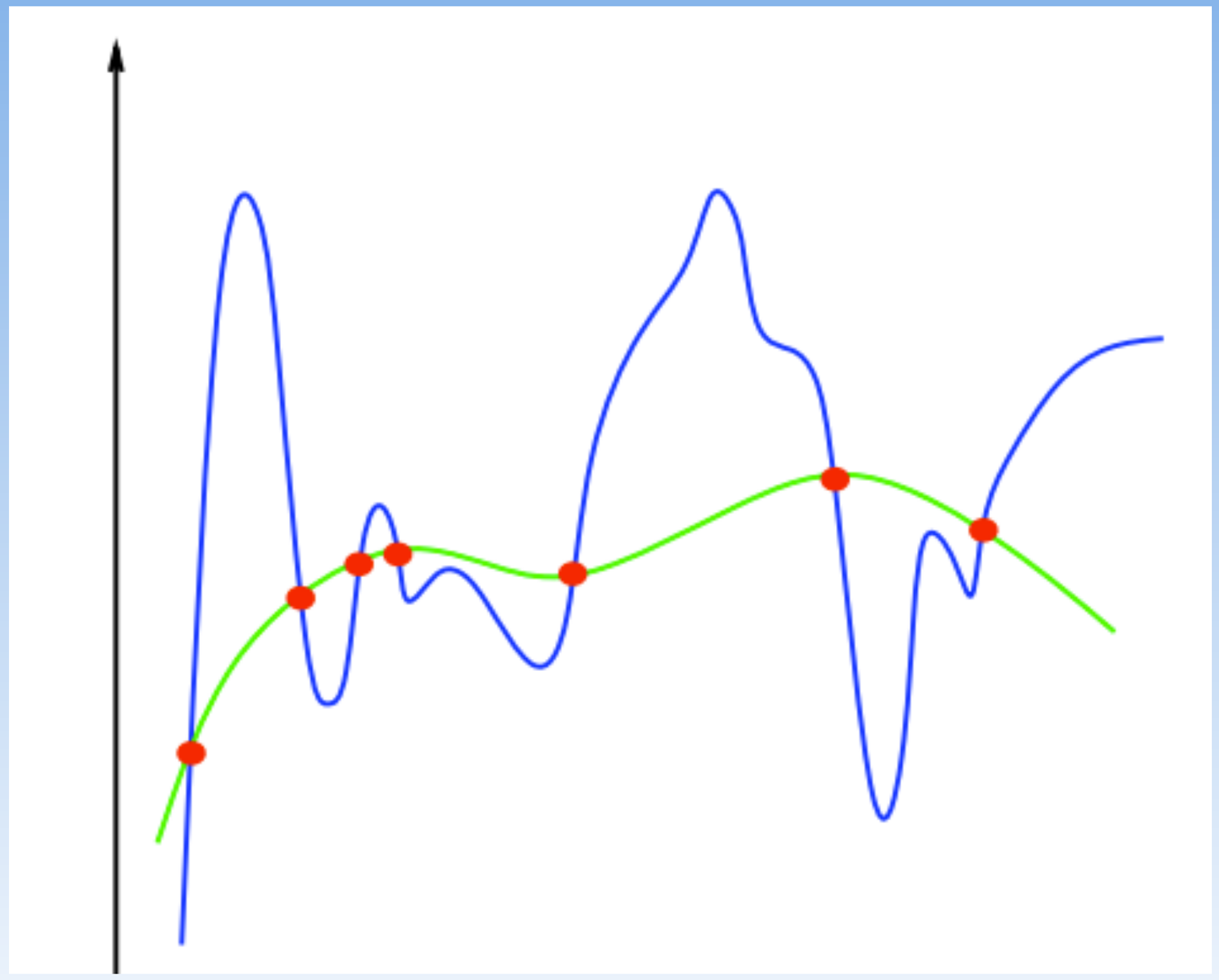
No problem! => polynomial features

$$y = h(x) = \beta_0 + \beta_1 x + \beta_2 f(x^2) + \ldots + \beta_n f(x^n)$$

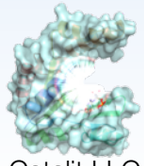
Catalit LLC

APPLICATIONS

OVERFITTING

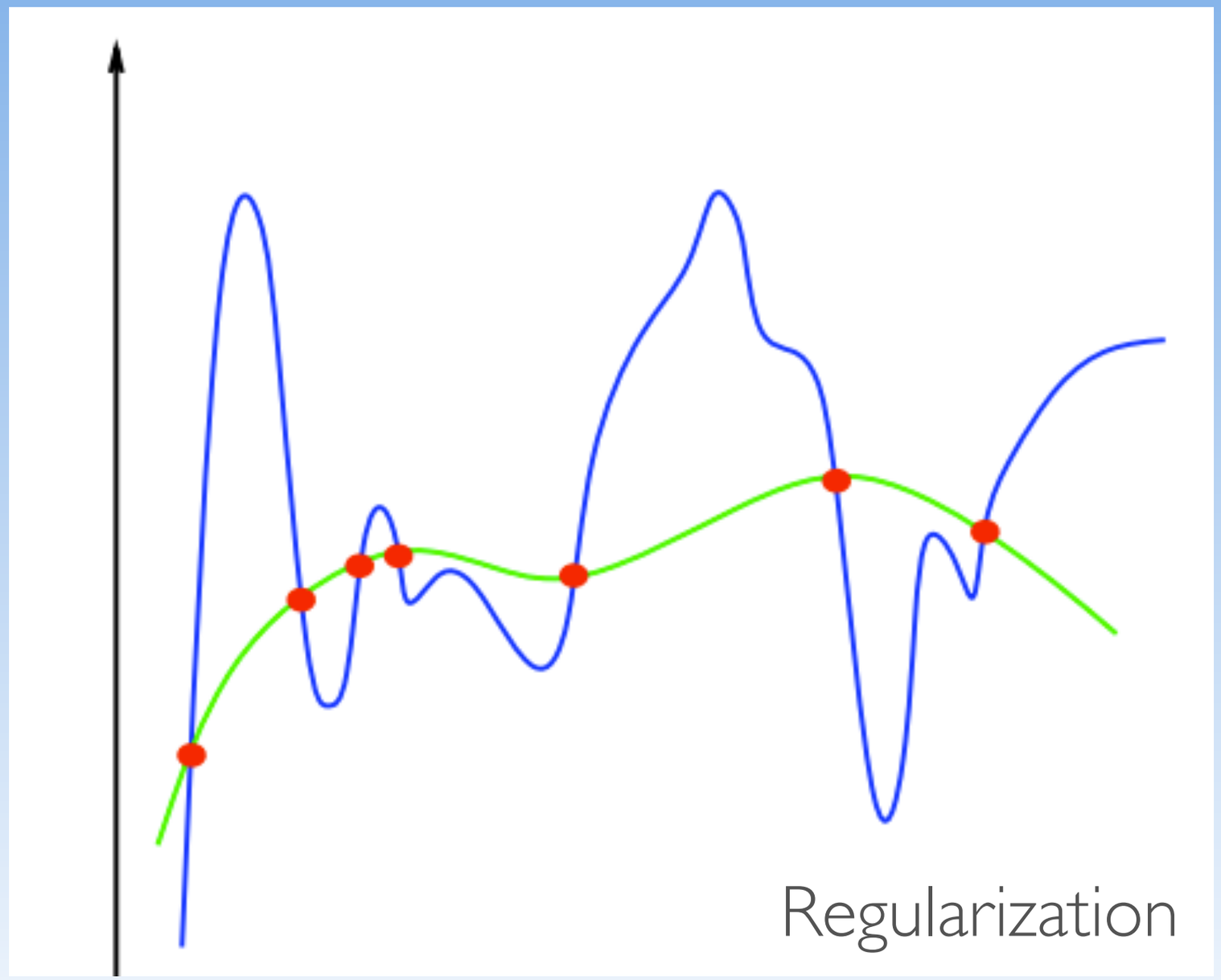


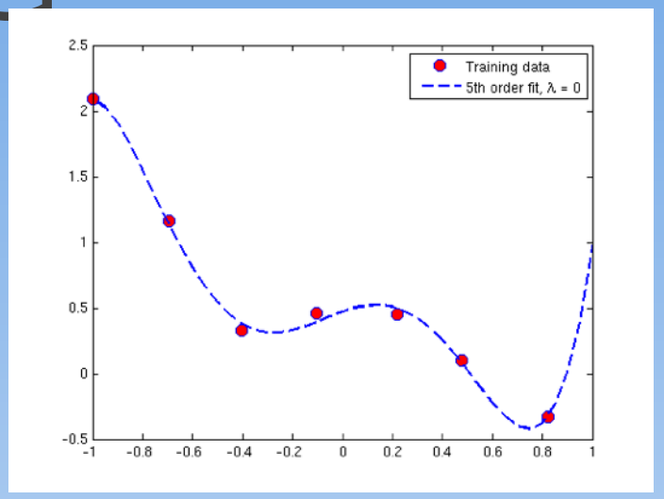
http://www.mit.edu/~9.520/spring12/slides/class02/class02.pdf

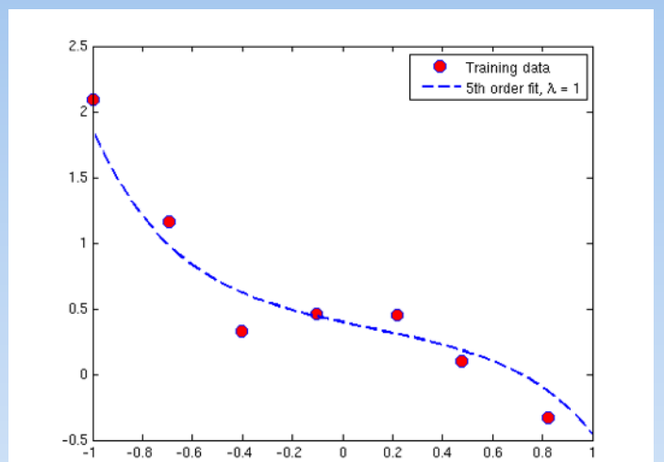


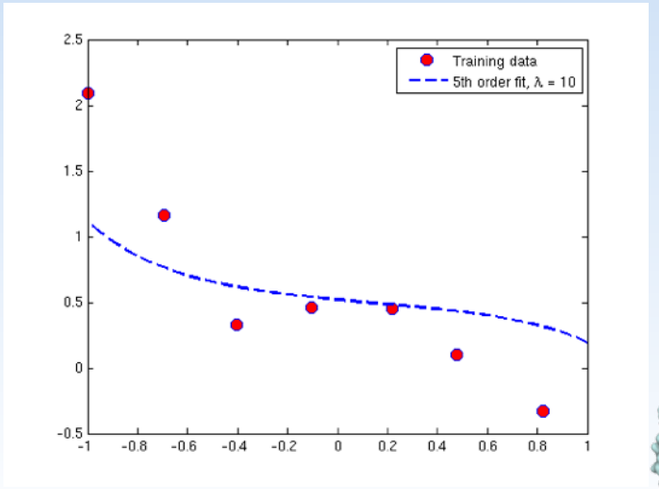
Catalit LLC

OVERFITTING

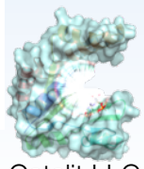








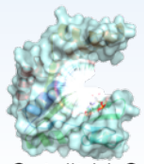
http://www.mit.edu/~9.520/spring12/slides/class02/class02.pdf



Catalit LLC

REGRESSION LAB

- python 02_regression.py
- Load a dataset
- Fit a linear model
- Discuss result



Catalit LLC