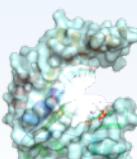




WELCOME



# LOGISTICS

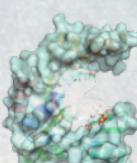
- doors
- water
- restrooms
- breaks and lunch
- theory / practice ratio
- computer setup
- wifi
- skill level

# WHY ARE WE HERE ?



# AGENDA

- 09:00 am - 09:30 am Registration
- **09:30 am - 11:20 am ML review**
- 11:20 am - 11:30 am Break
- **11:30 am - 01:00 pm Neural Nets**
- 01:00 pm - 02:00 pm Lunch
- **02:00 pm - 03:50 pm Gradient descent**
- 03:50 pm - 04:00 pm Break
- **04:00 pm - 05:15 pm Overfitting**
- 05:15 pm - 05:30 pm Wrap up



Catalit LLC

# GOALS

- Learn some cool DL
- Meet fellow students
- Have fun



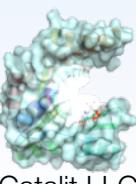
# WHAT TO EXPECT

- Accelerated learning
- Interactive
- Fun / Discovery



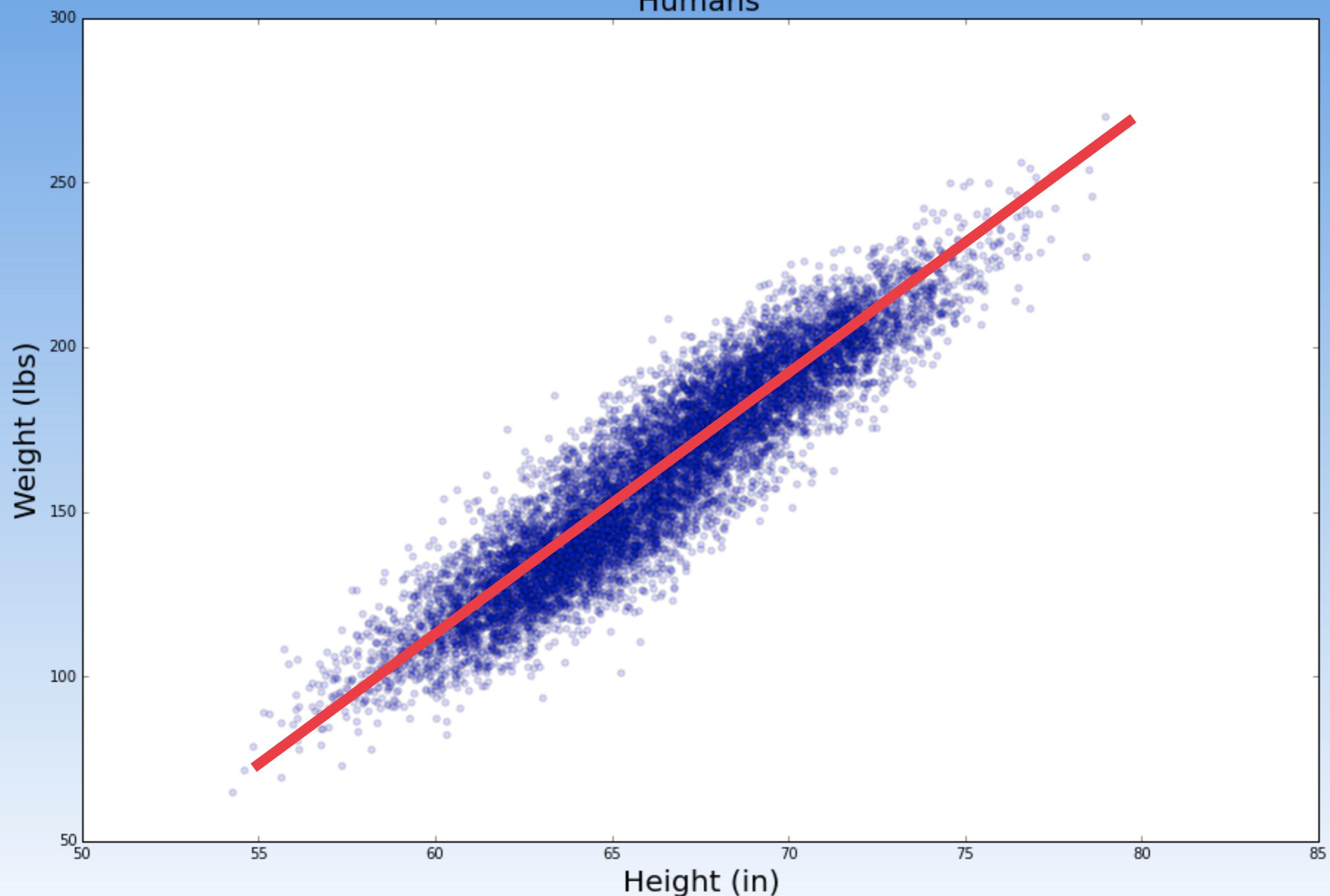
Catalit LLC

# MACHINE LEARNING RECAP

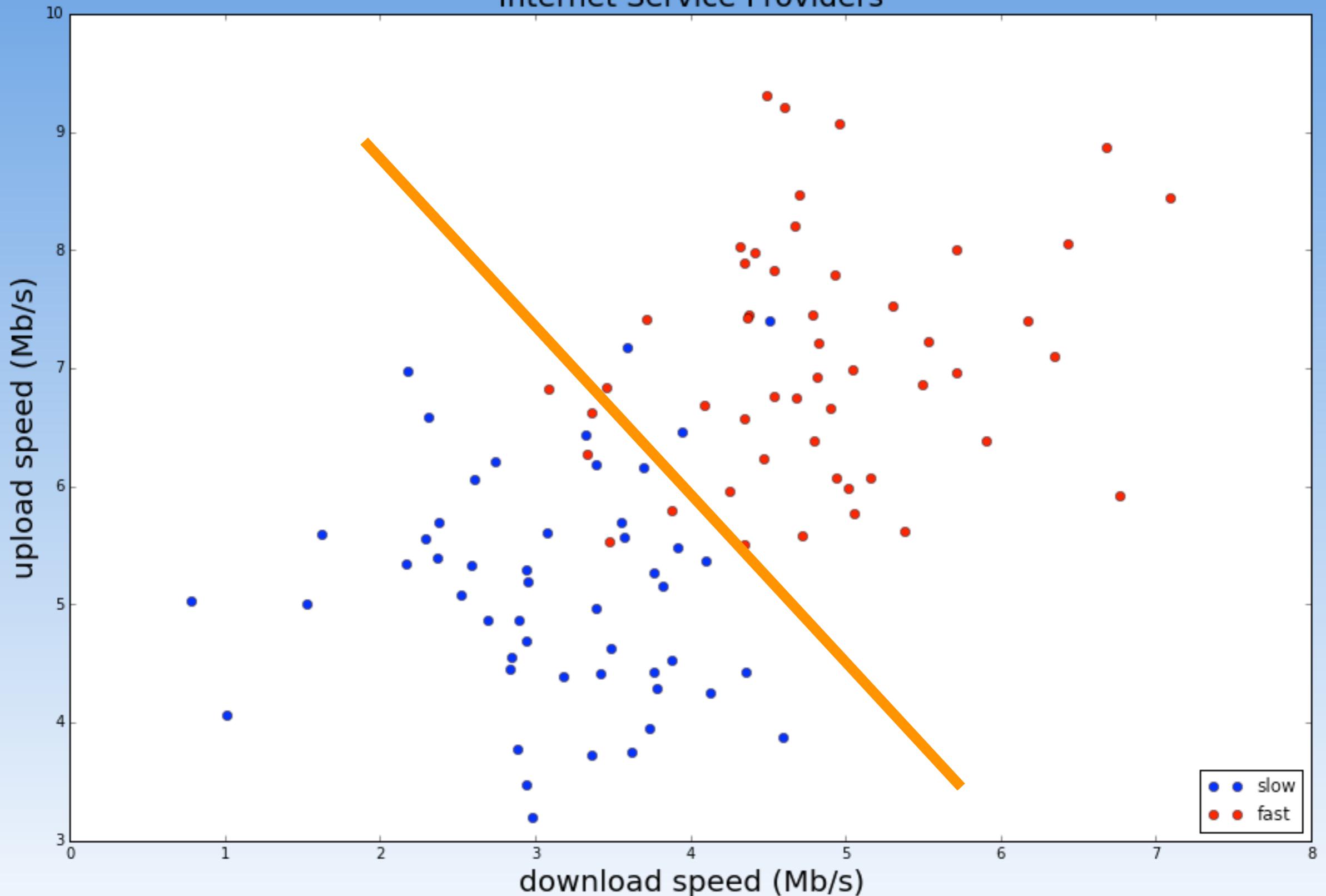


Catalit LLC

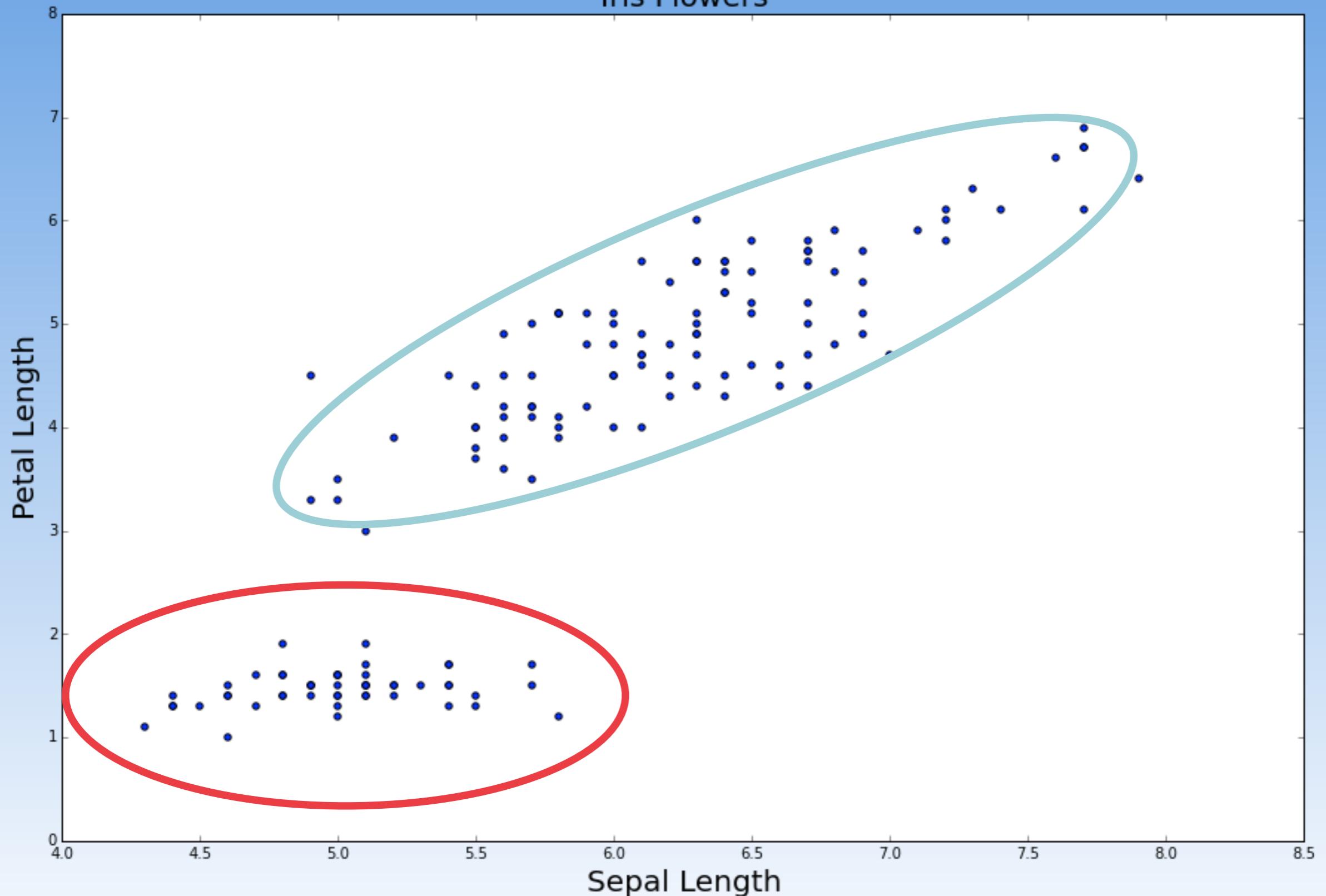
## Humans



## Internet Service Providers



## Iris Flowers



# FEATURES & LABELS

	Age	Gender	Annual Salary	Months in residence	Months in job	Current Debt	Paid off credit
	23	M	\$30,000	36	12	\$5,000	Yes
Client 2	30	F	\$45,000	12	12	\$1,000	Yes
Client 3	19	M	\$15,000	3	1	\$10,000	No
Client 4	25	M	\$25,000	12	27	\$15,000	?

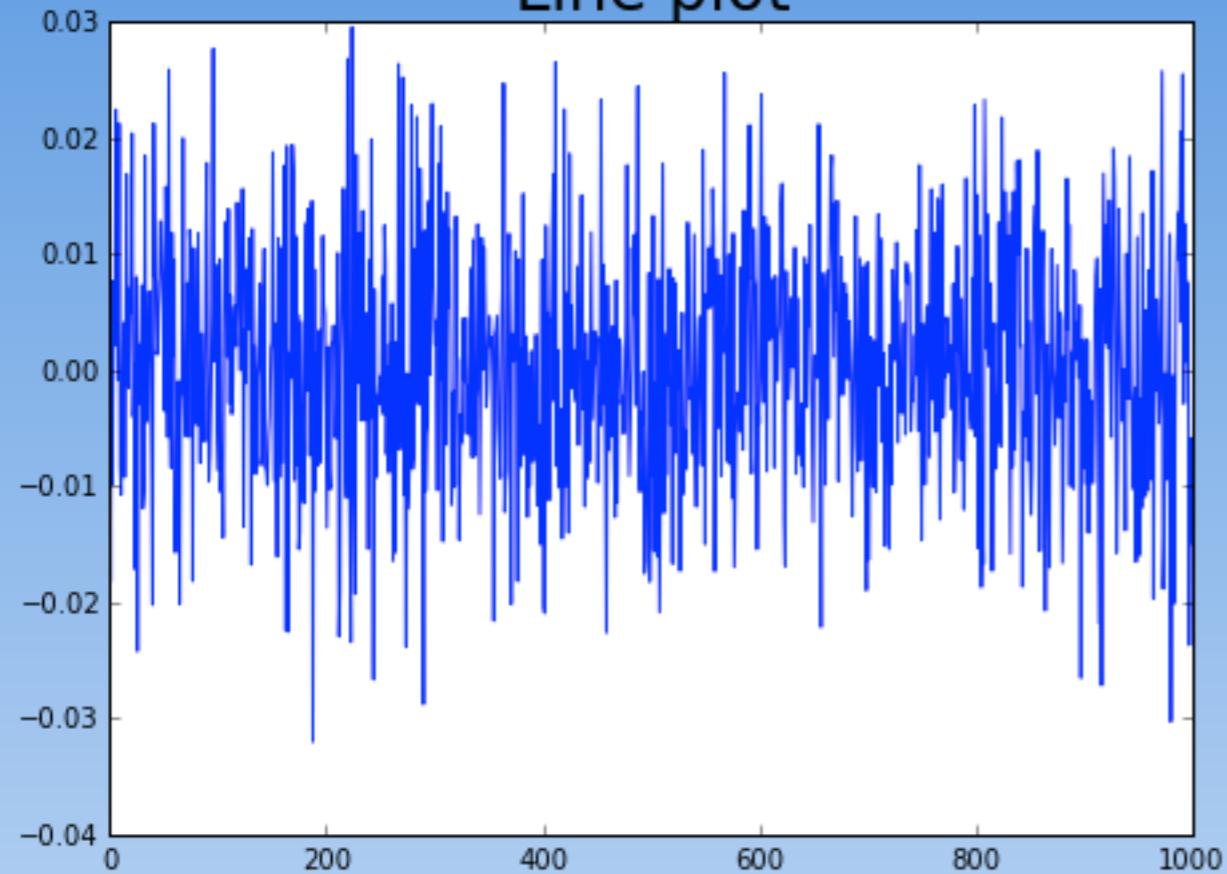
# SUPERVISED & UNSUPERVISED



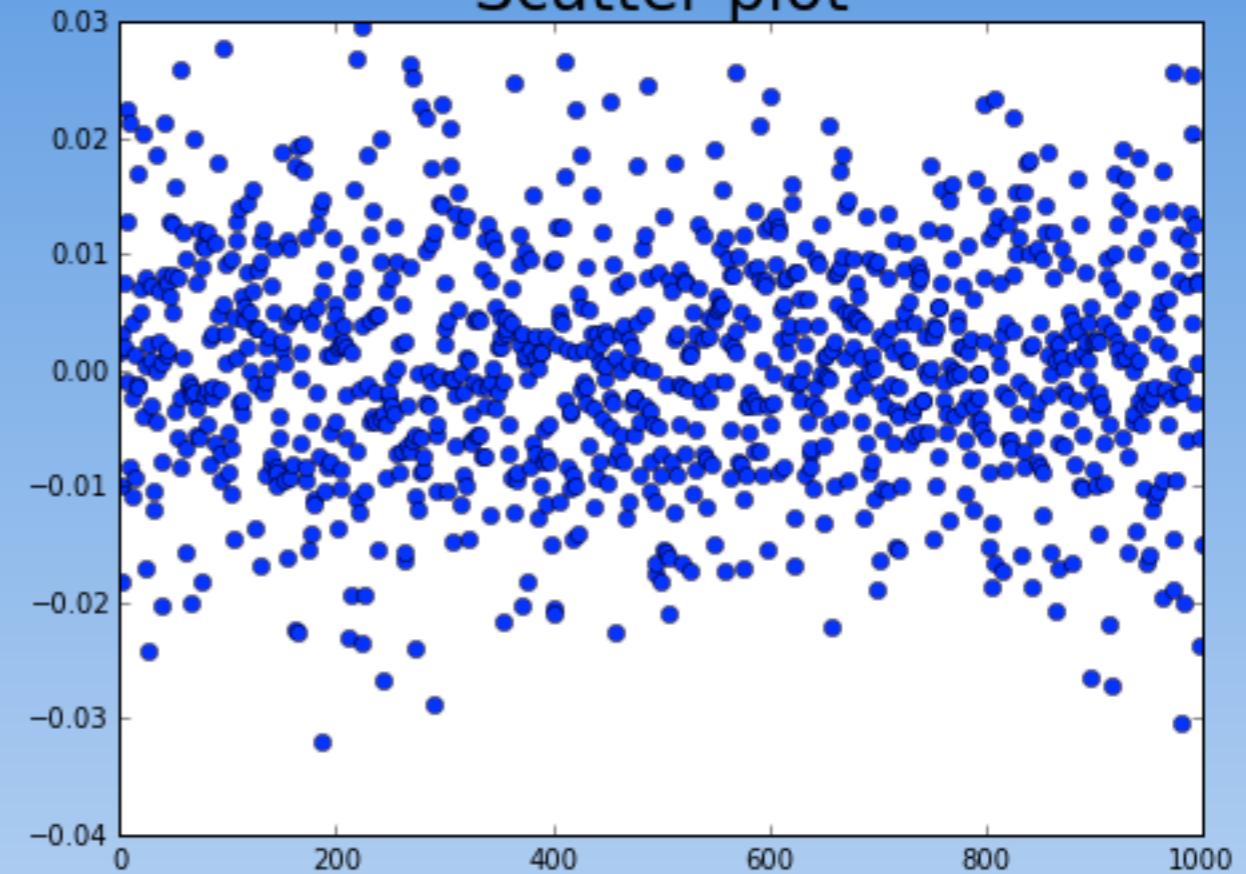
# COMBINED

	CONTINUOUS	CATEGORICAL
SUPERVISED	REGRESSION	CLASSIFICATION
UNSUPERVISED	DIMENSION REDUCTION	CLUSTERING

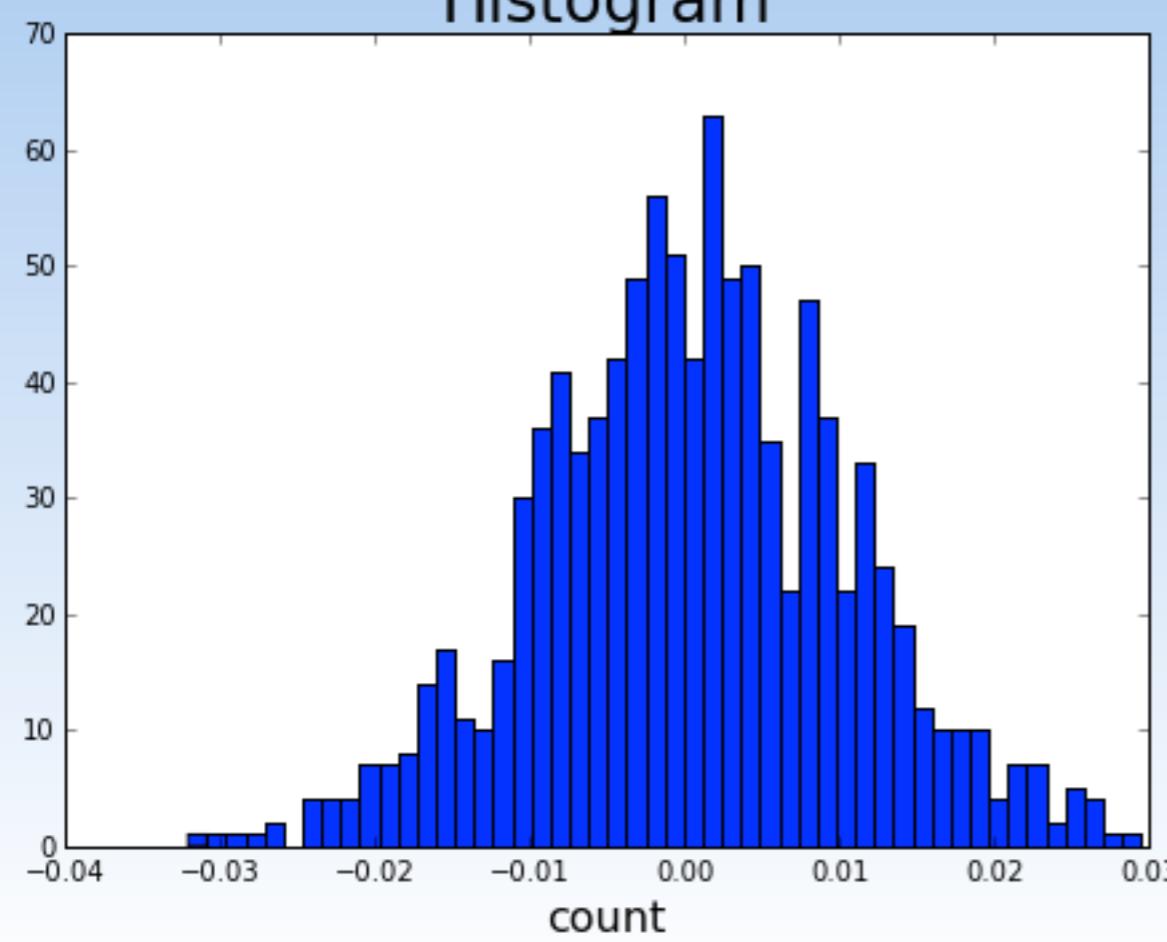
### Line plot



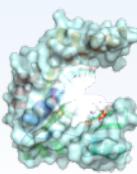
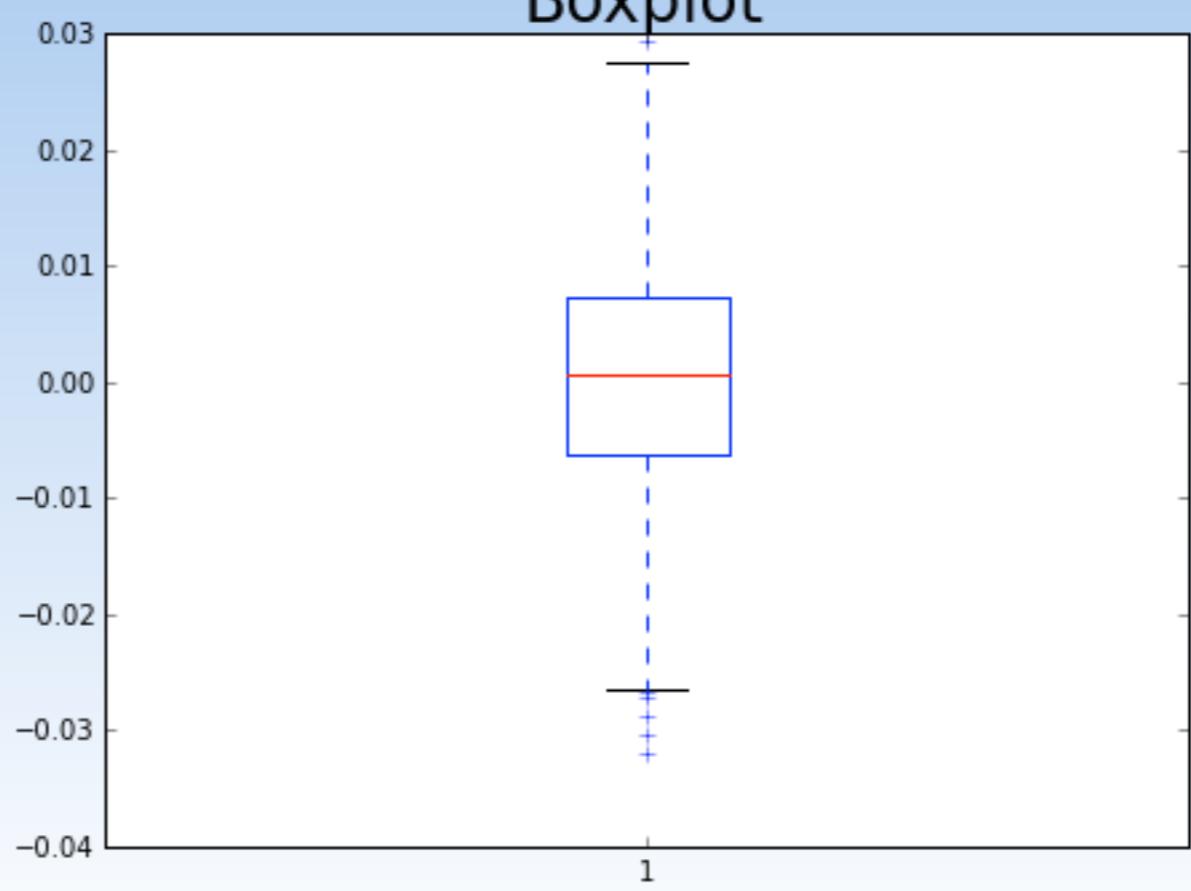
### Scatter plot



### Histogram

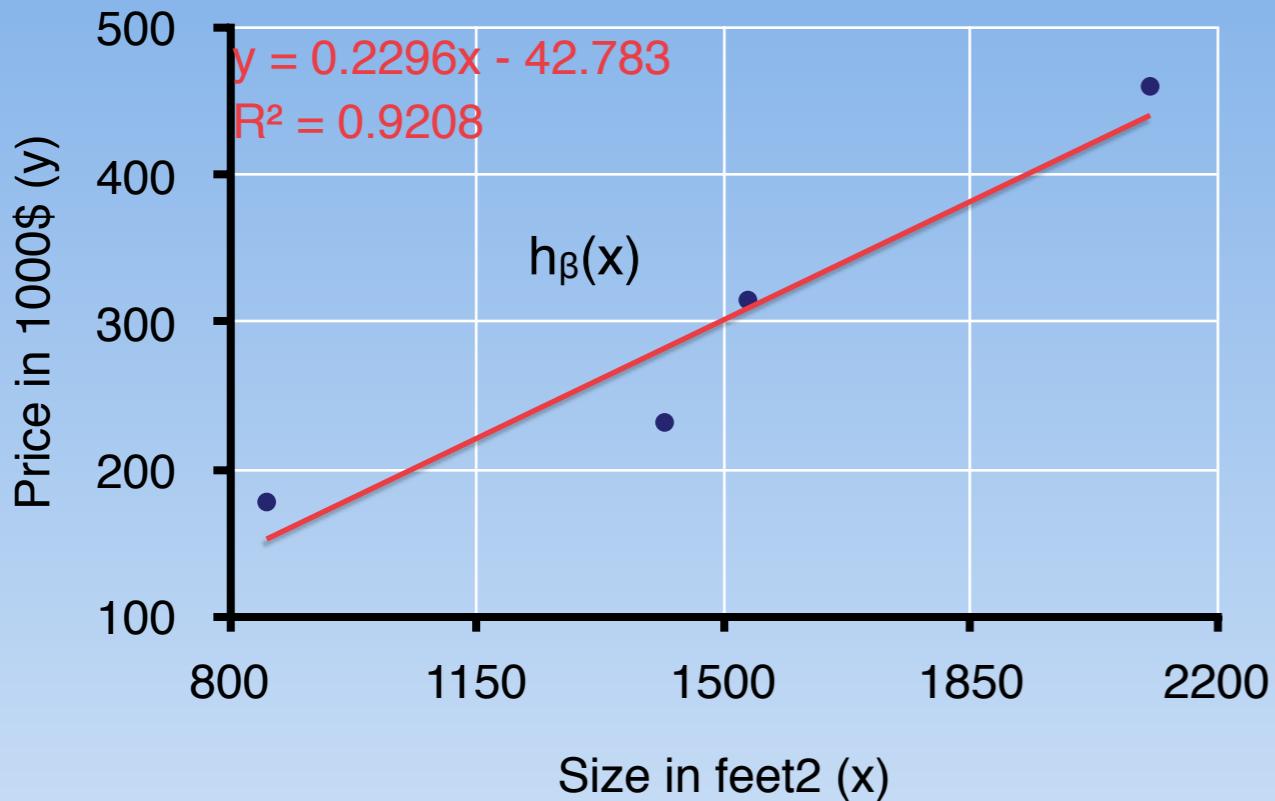


### Boxplot



Catalit LLC

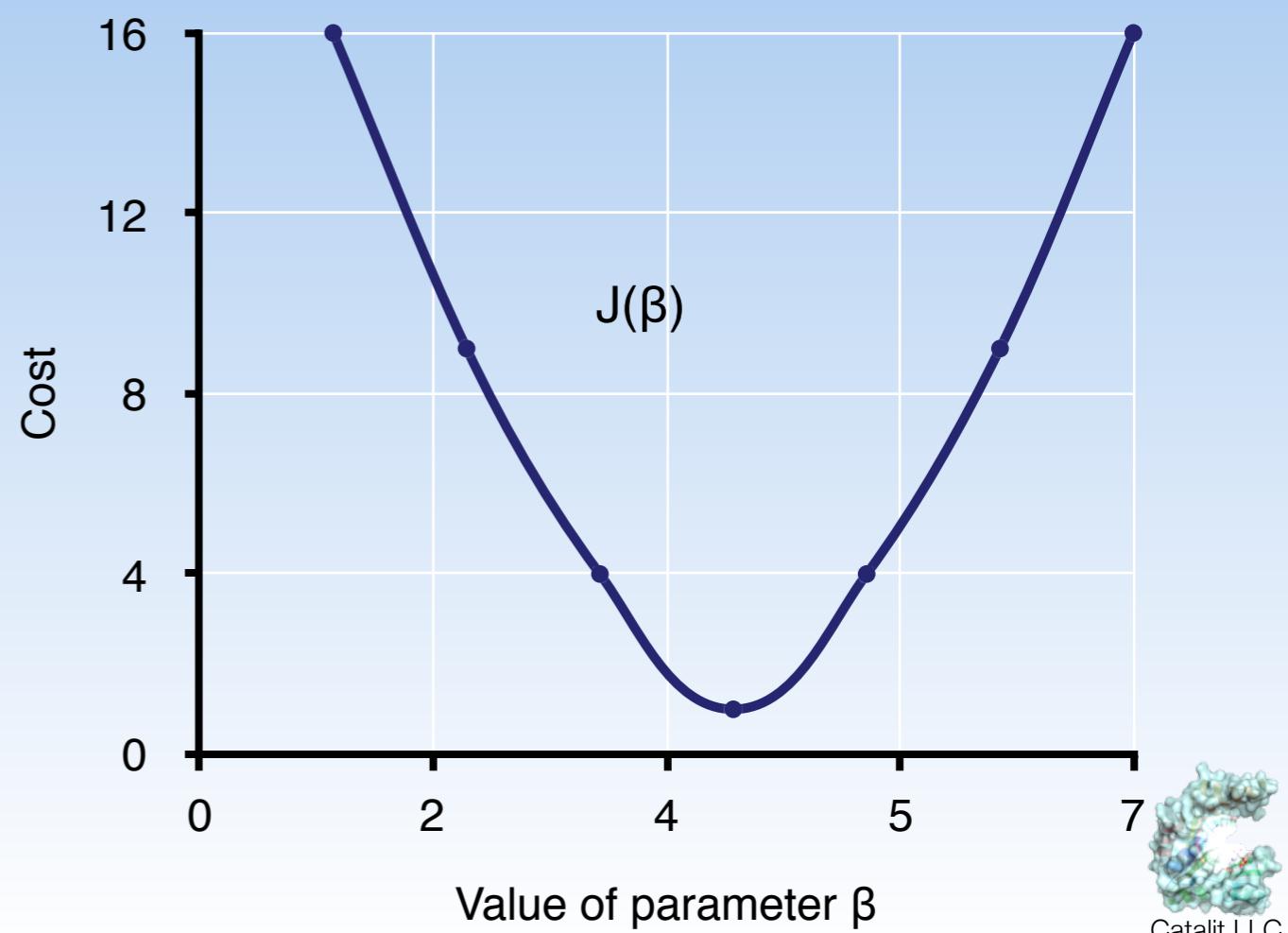
# COST FUNCTION



Minimize Cost

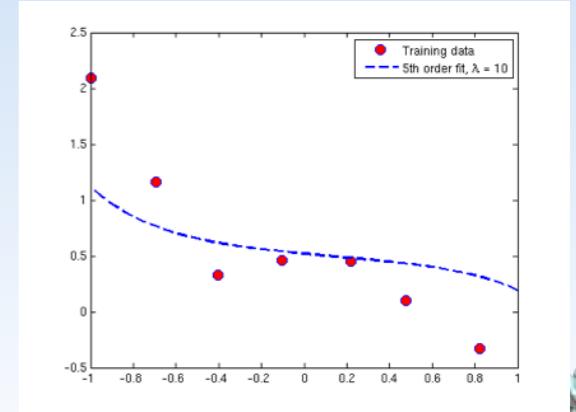
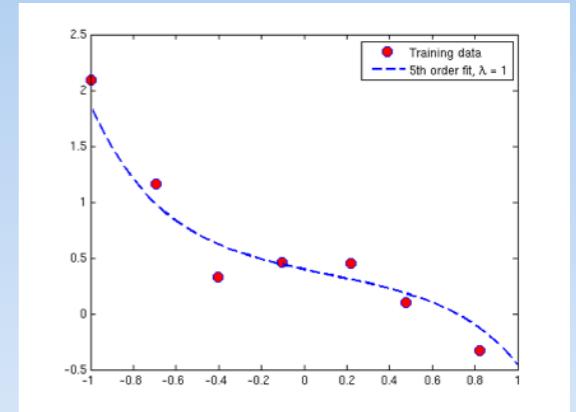
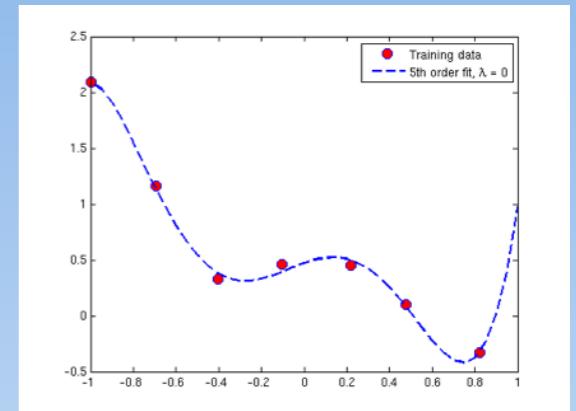
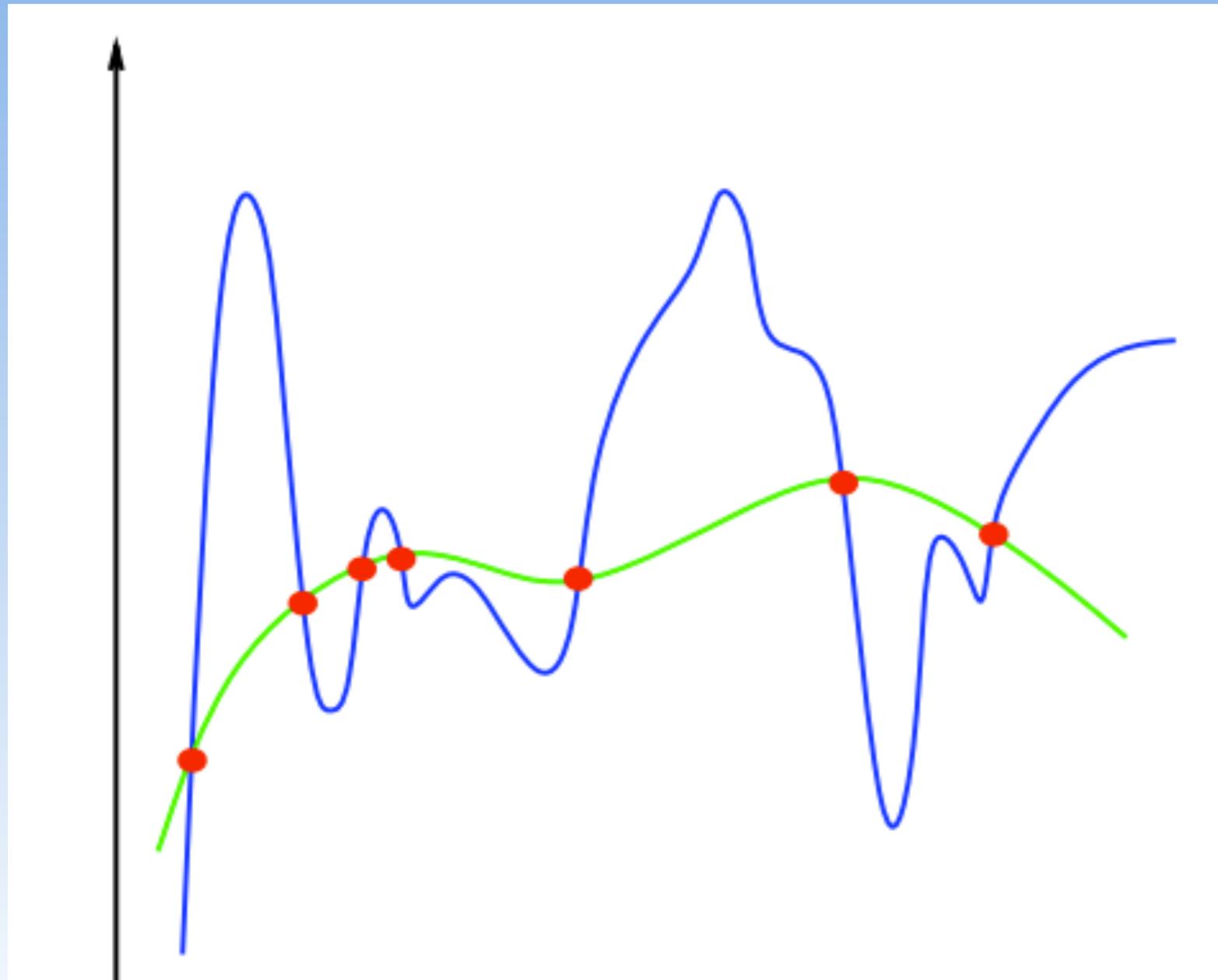


Define Hypothesis  
Define Cost



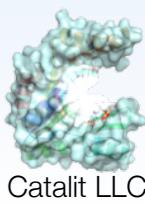
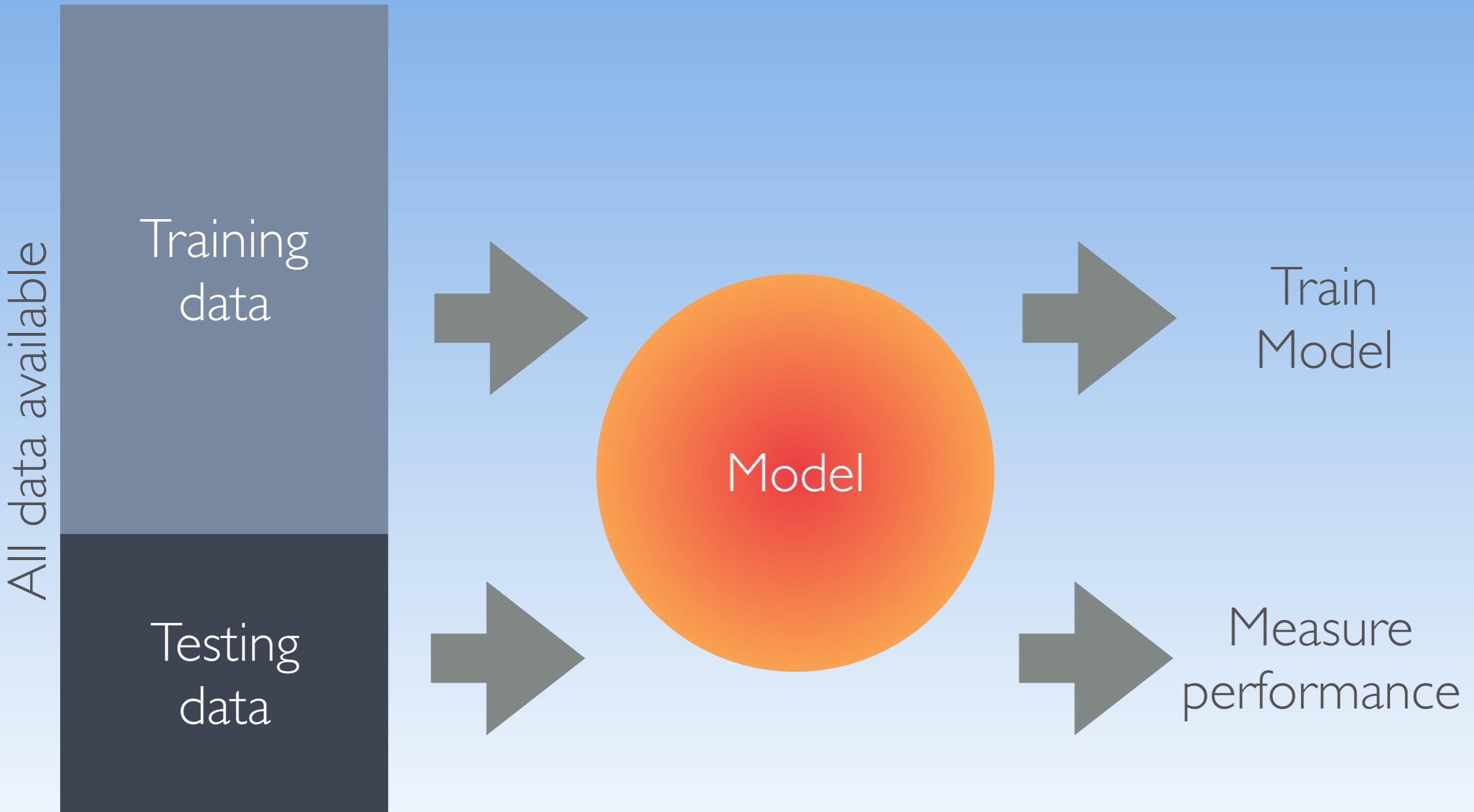
# OVERFITTING & REGULARIZATION

Regularization





# TRAIN - TEST SPLIT

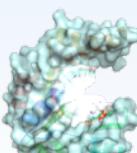


Catalit LLC

# PRECISION - RECALL & ACCURACY

- **Precision:** When test is positive, how often is prediction correct?
  - TP / test yes
- **Recall:** When actual value is positive, how often is prediction correct?
  - TP / actual yes
- **Accuracy:** Overall, how often is it correct?
  - $(TP + TN) / \text{total}$

	<i>Condition Positive</i>	<i>Condition Negative</i>
<i>Test Positive</i>	<b>TRUE POSITIVE</b>	<b>FALSE POSITIVE</b>
<i>Test Negative</i>	<b>FALSE NEGATIVE</b>	<b>TRUE NEGATIVE</b>



# CROSS-VALIDATION



Validation Set



Training Set

Round 1



Validation  
Accuracy:

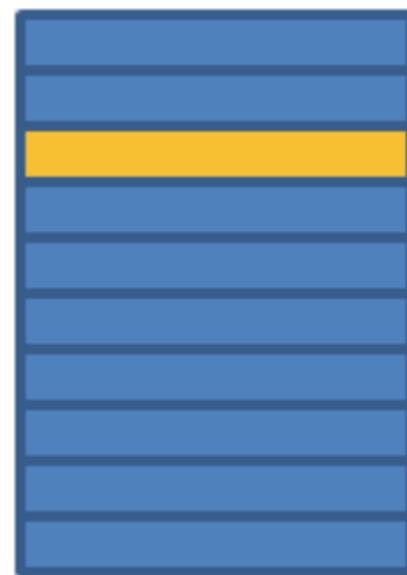
93%

Round 2



90%

Round 3



91%

Round 10

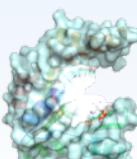
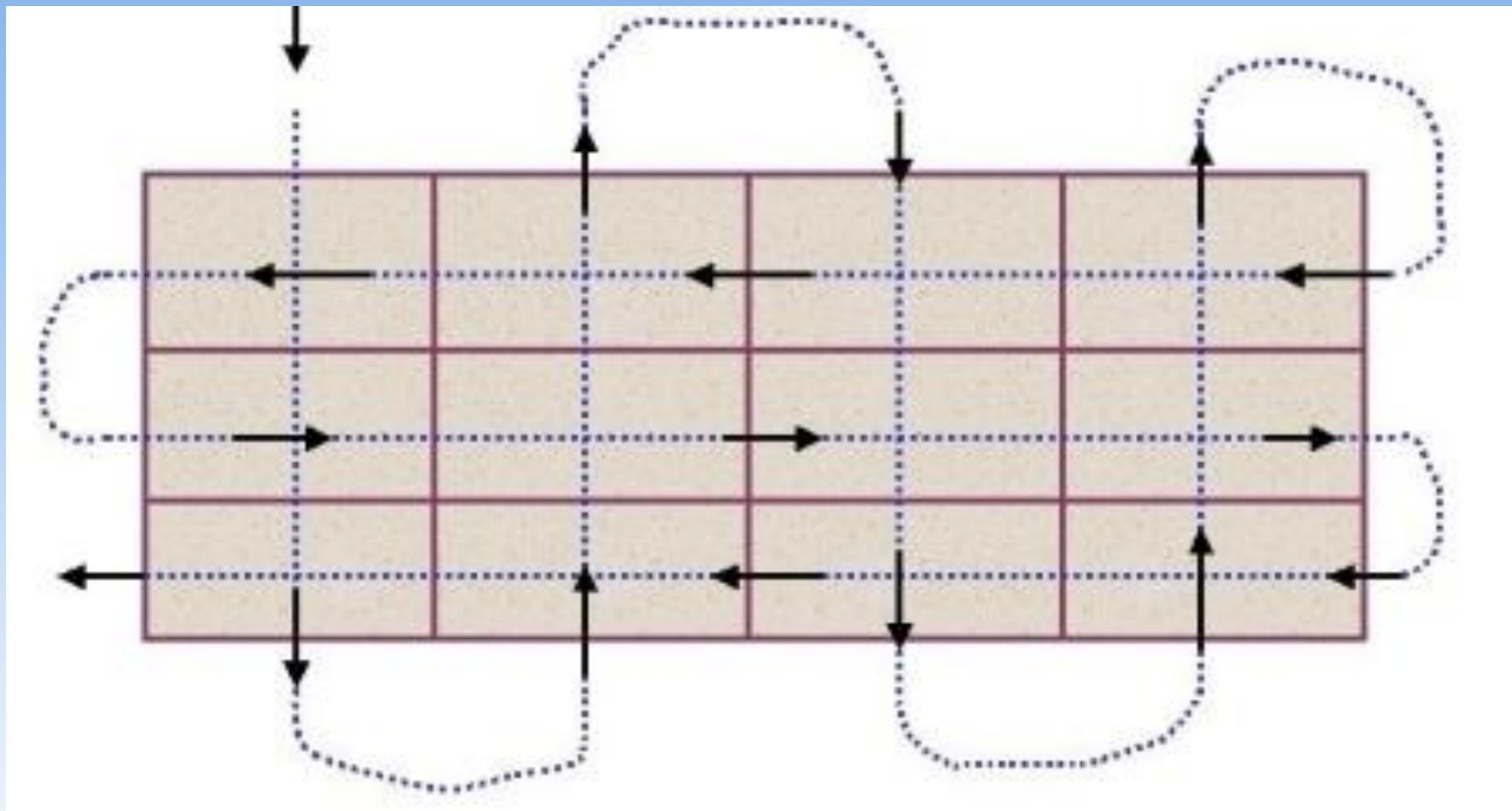


95%

...

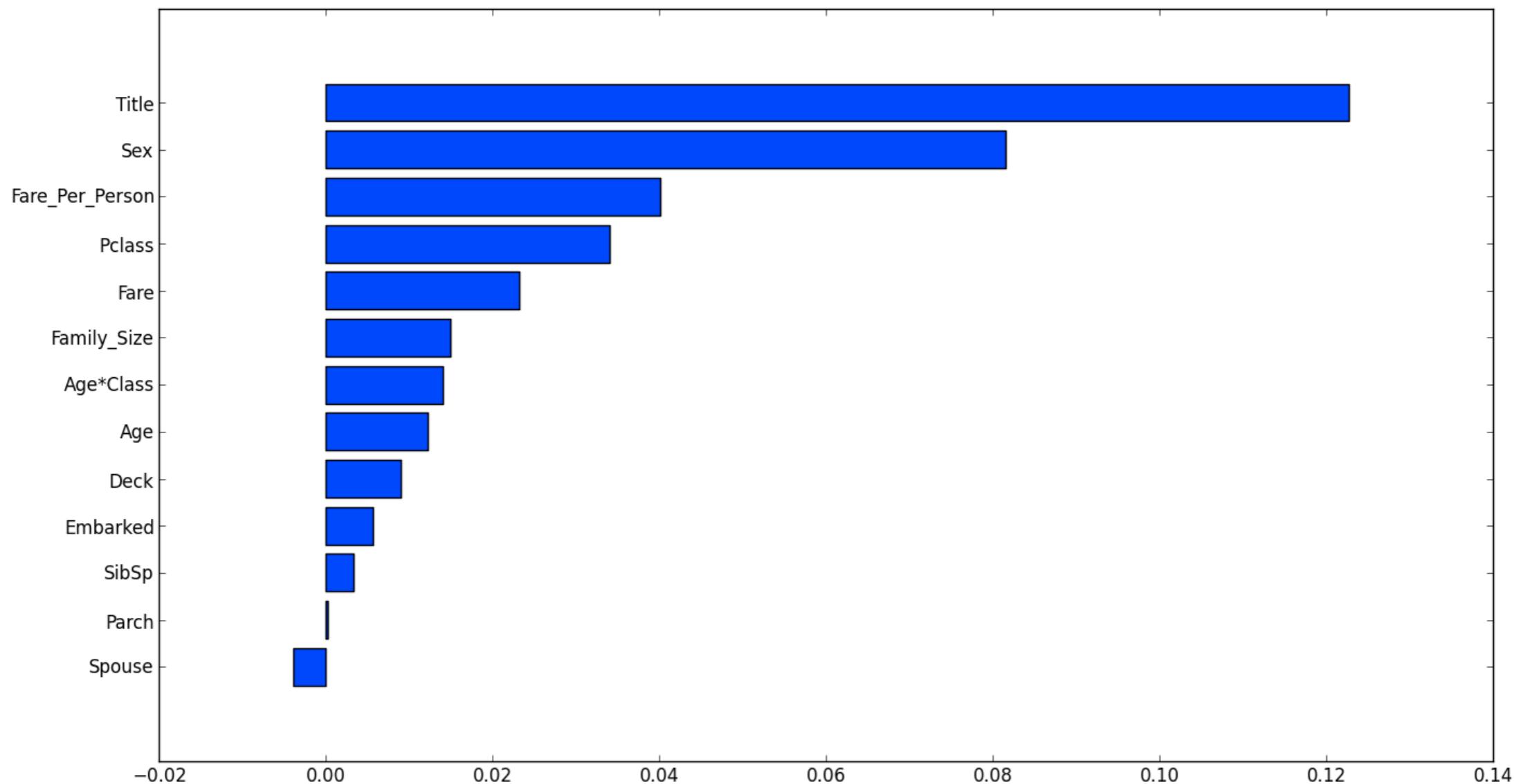
Final Accuracy = Average(Round 1, Round 2, ...)

# GRID SEARCH

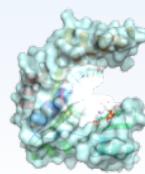
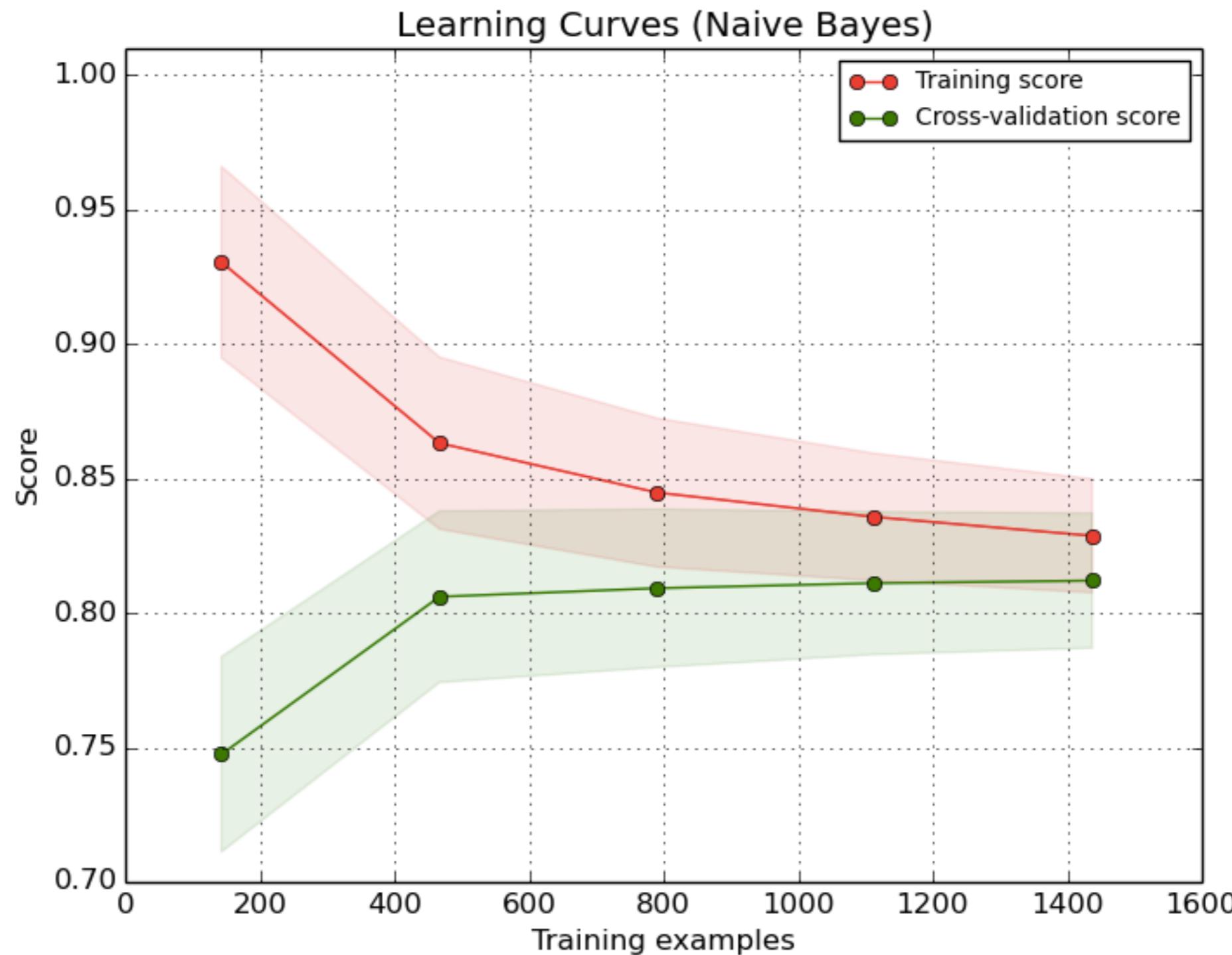


Catalit LLC

# FEATURE SELECTION



# LEARNING CURVES



# ENSEMBLE METHODS



Bagging  
Boosting  
Random Forest

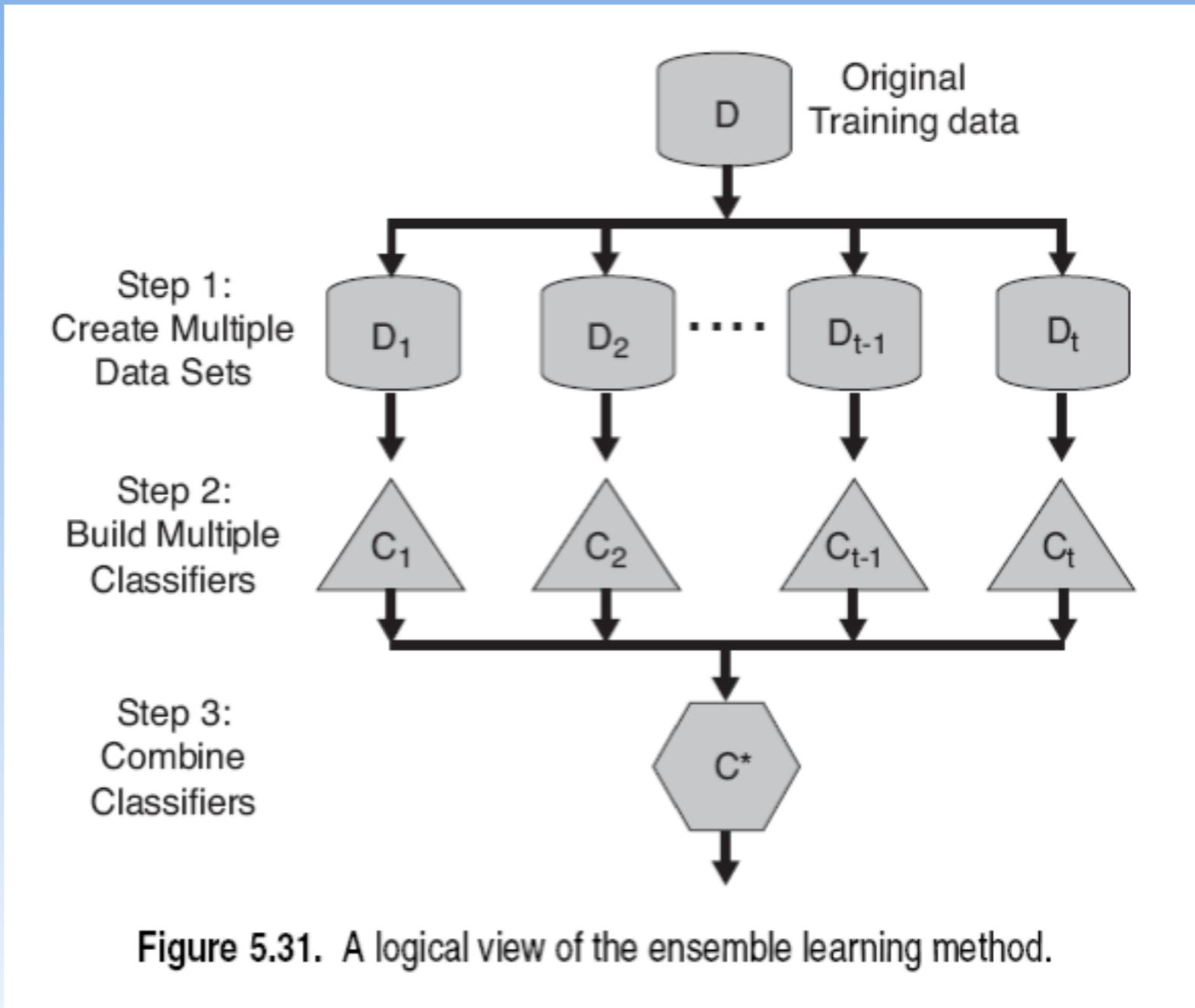
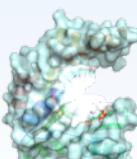


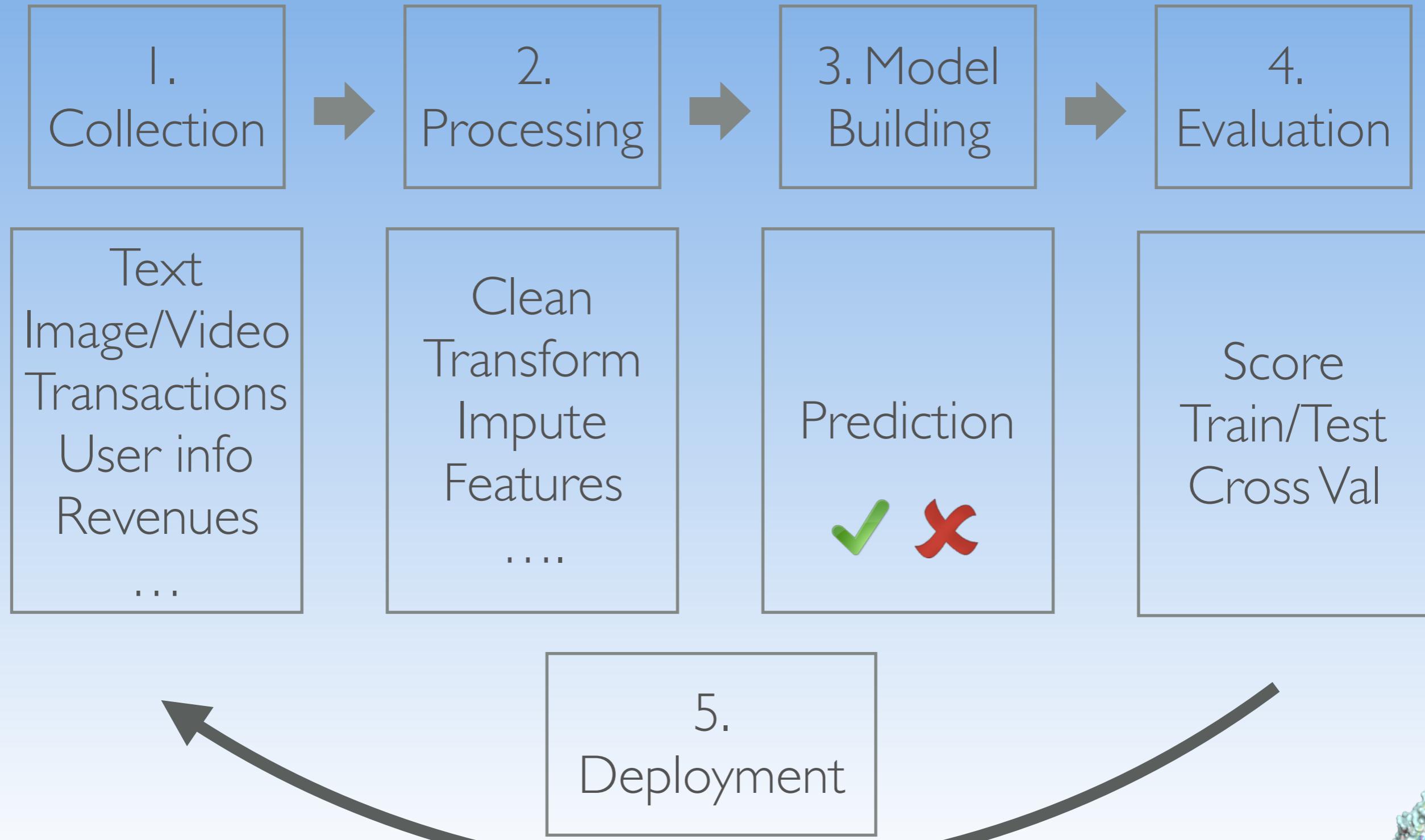
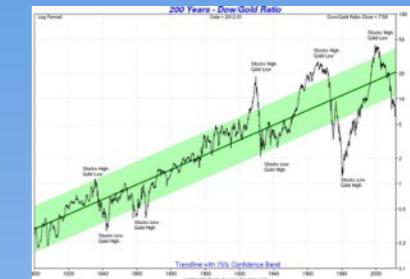
Figure 5.31. A logical view of the ensemble learning method.



Catalit LLC

## Exploration

# ML STEPS

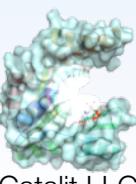


# MAIN LIMITATIONS OF ML



Catalit LLC

# DEEP LEARNING APPLICATIONS



Catalit LLC

# APPLICATIONS

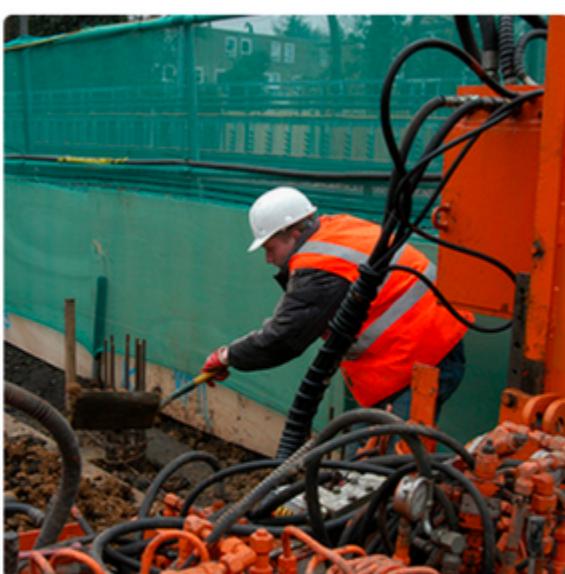
Object recognition



Caption generation



"man in black shirt is playing guitar."

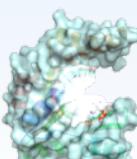
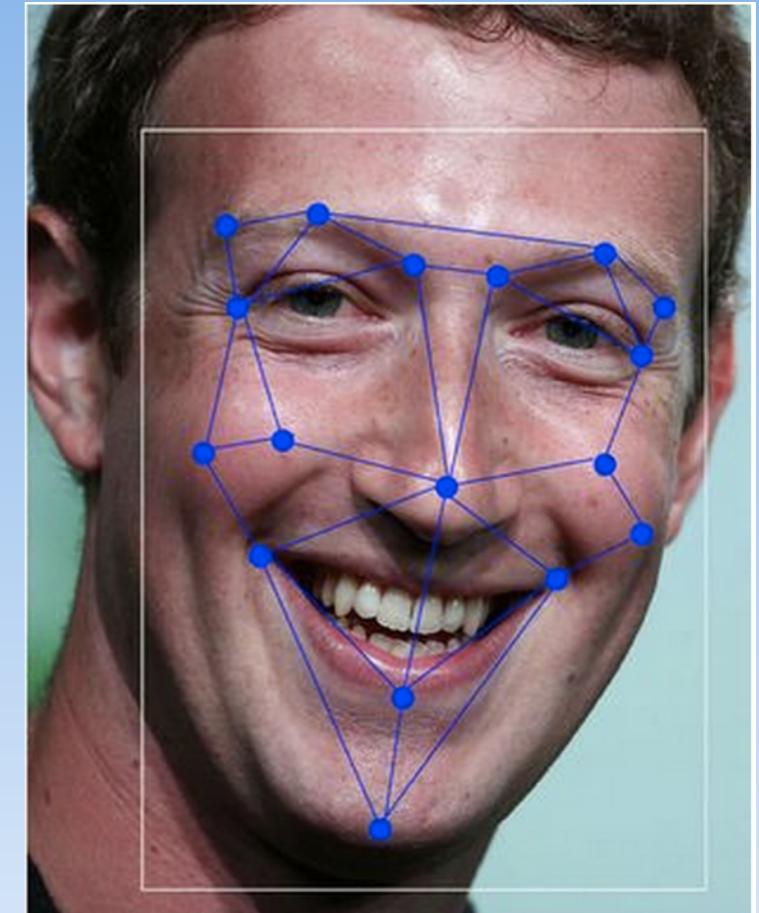
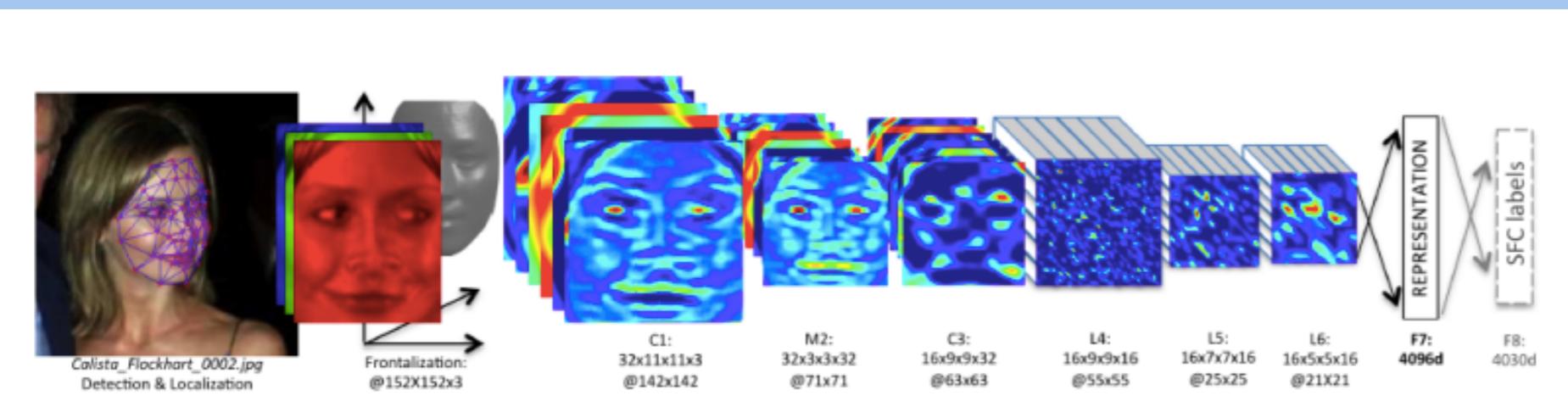


"construction worker in orange safety vest is working on road."

mite	container ship	motor scooter	leopard
black widow	lifeboat	go-kart	jaguar
cockroach	amphibian	moped	cheetah
tick	fireboat	bumper car	snow leopard
arfish	drilling platform	golfcart	Egyptian cat
mushroom	cherry	Madagascar cat	
vertible	dalmatian	squirrel monkey	
grille	agaric	spider monkey	
pickup	mushroom	titi	
wagon	jelly fungus	indri	
engine	gill fungus	howler monkey	
	dead-man's-fingers		

# APPLICATIONS

## Face recognition

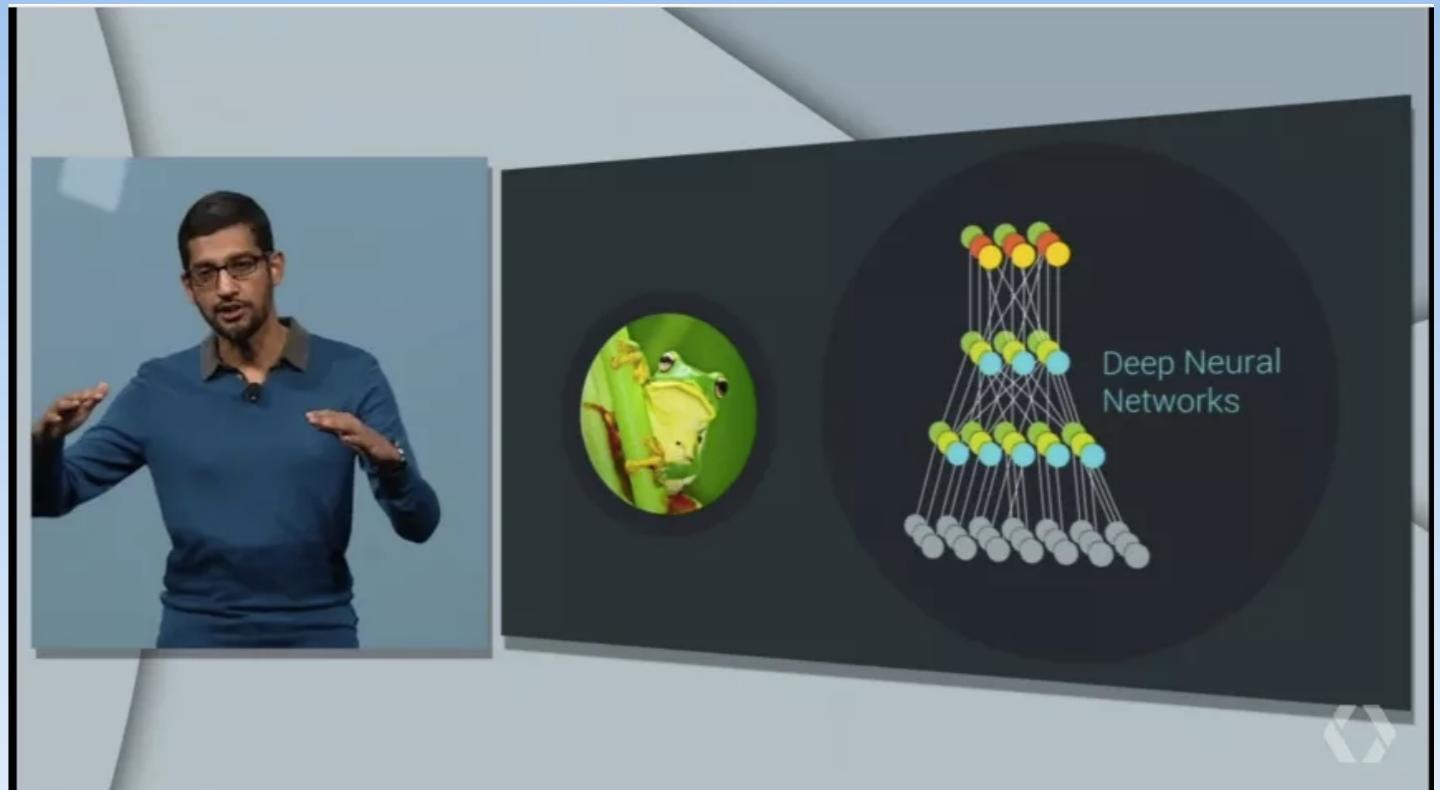


Catalit LLC

# APPLICATIONS

## Speech recognition and synthesis

*Google says its speech  
recognition technology  
now has only an 8%  
word error rate...*



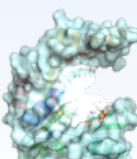
# APPLICATIONS

## Neural Machine Translation

Economic growth has slowed down in recent years .

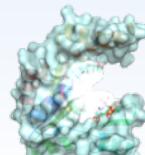


La croissance économique s'est ralentie ces dernières années .



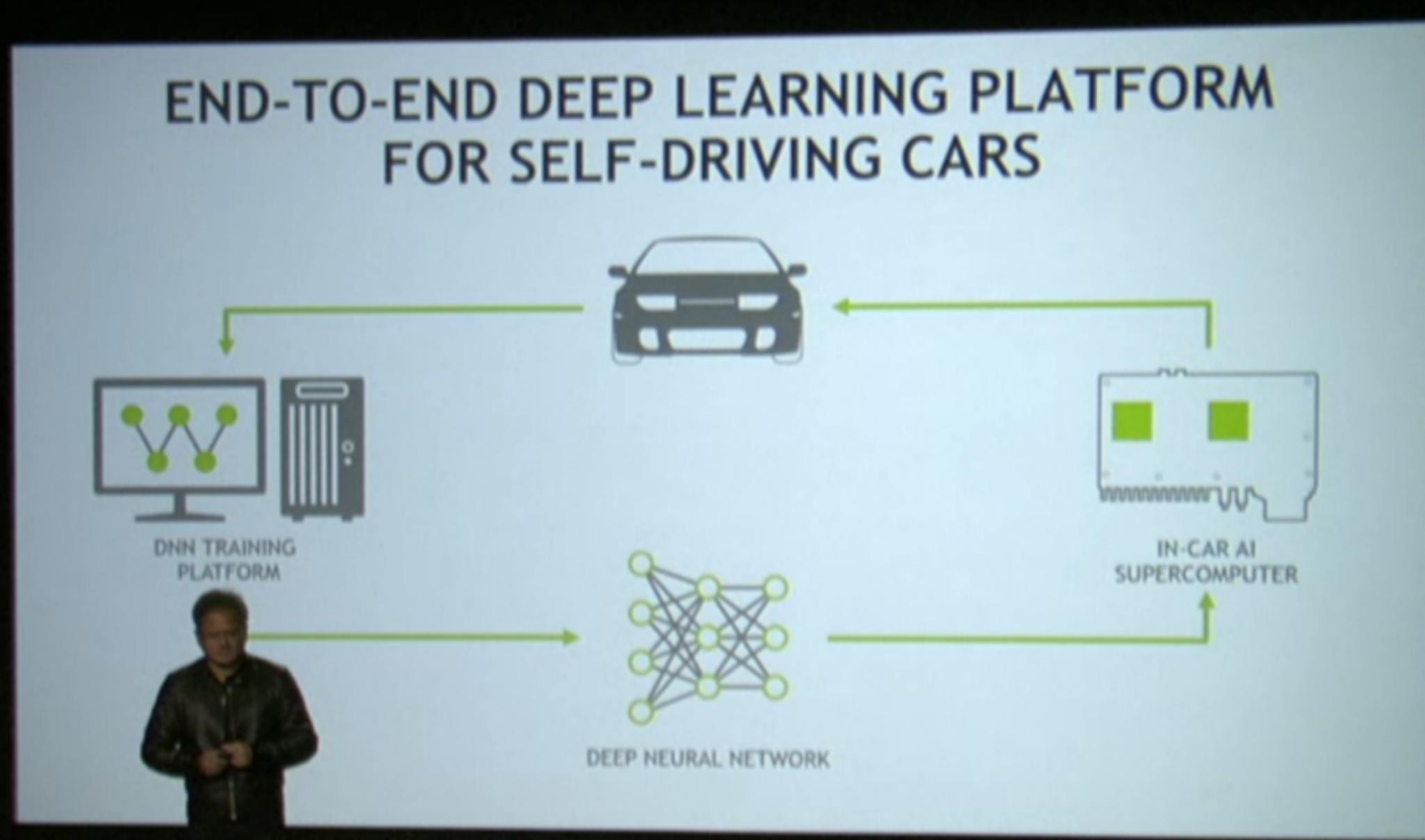
# APPLICATIONS

## Stock market prediction

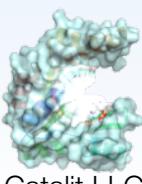


Catalit LLC

# APPLICATIONS

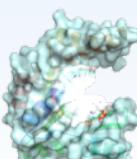


# DATA SCIENCE WITH PYTHON



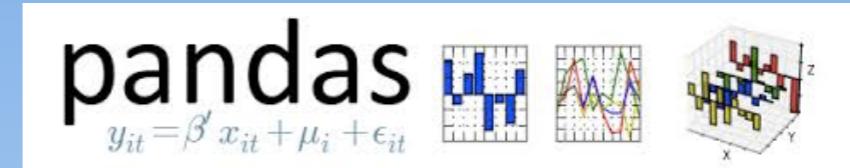
Catalit LLC

# OPEN SOURCE

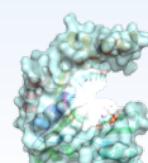


Catalit LLC

# PYDATA ECHOSYSTEM



Bokeh



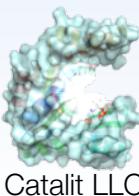
Catalit LLC

# goo.gl/oqj8LN

DLWorkshop-02-24-2017		
Name		Date Modified
0_Intro.ipynb	✓	Today, 8:52 PM
1a-Introduction_to_Keras.ipynb	✓	Today, 8:52 PM
1b-Separability_a...dden_Layers.ipynb	✓	Today, 8:52 PM
2-Learning_with_...ient_Descent.ipynb	✓	Today, 8:52 PM
3-Overfitting_and_Regularization.ipynb	✓	Today, 8:52 PM
4-MNIST.ipynb	✓	Today, 8:52 PM
5-Classifying_Images_with_CNNs.ipynb	✓	Today, 8:52 PM
6-Mapping_Representations.ipynb	✓	Today, 8:52 PM
7-Sequence-Learning.ipynb	✓	Today, 8:52 PM
8-Sequence-Generation.ipynb	✓	Today, 8:52 PM
Check Environment.ipynb	✓	Today, 8:52 PM
▶ data	✓	Today, 8:52 PM
▶ environment-cpu.yml	✓	Today, 8:52 PM
▶ extras	✓	Today, 8:52 PM
▶ further_resources.md	✓	Today, 8:52 PM

# INTRO LAB

- open terminal
- cd <folder>
- conda env create -f environment-cpu.yml
- source activate dataweekends-dlintro
- jupyter notebook



Catalit LLC

# INTRO LAB

- open the notebook 0\_intro
- go through the exercises