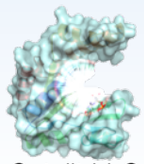
DataWknds.

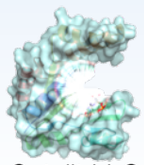
CLASSIFICATION



Catalit LLC

CLASSIFICATION

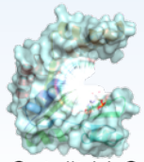
	CONTINUOUS	CATEGORICAL
SUPERVISED	?	?
UNSUPERVISED	?	?



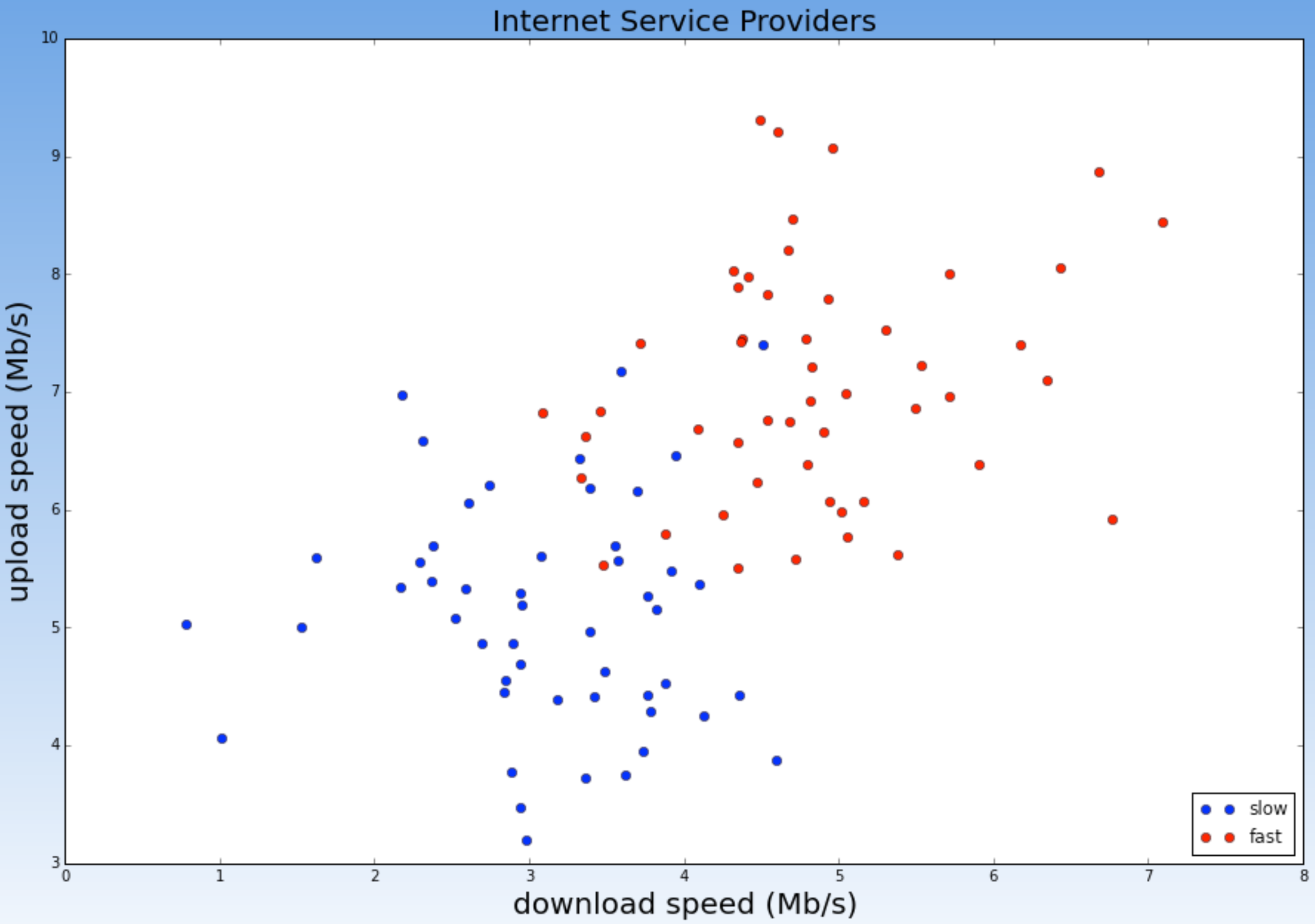
Catalit LLC

CLASSIFICATION

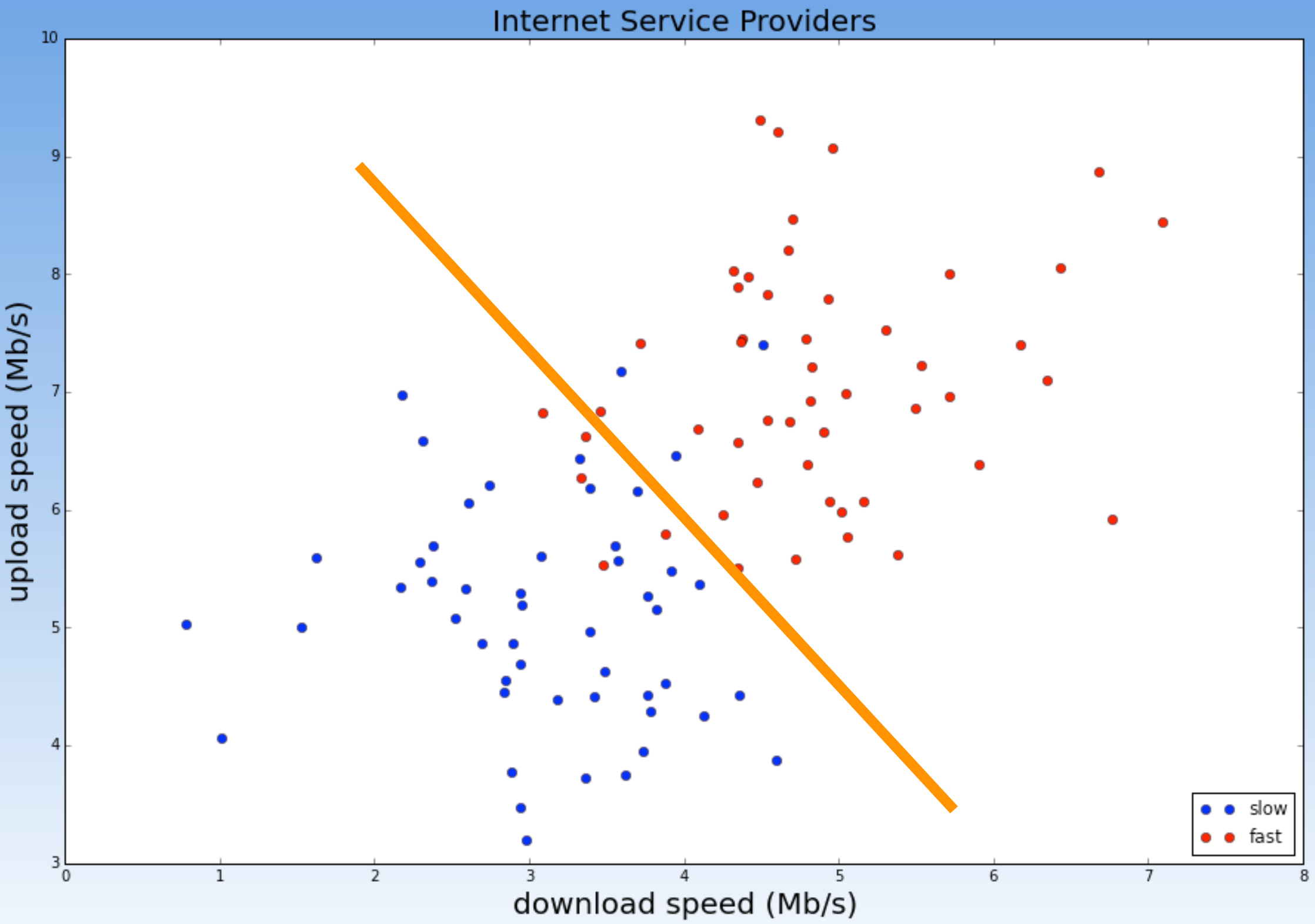
	CONTINUOUS	CATEGORICAL
SUPERVISED	REGRESSION	CLASSIFICATION
UNSUPERVISED	DIMENSION REDUCTION	CLUSTERING



Catalit LLC


Internet Service Providers
upload speed (Mb/s)
download speed (Mb/s)
slow
fast
10
9
8
7
6
5
4
3
0
1
2
3
4
5
6
7
8

Catalit LLC


Internet Service Providers
upload speed (Mb/s)
download speed (Mb/s)
slow
fast
10
9
8
7
6
5
4
3
0
1
2
3
4
5
6
7
8

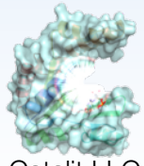
Catalit LLC

BINARY CLASSIFICATION

Features

Labels

	Age	Gender	Annual Salary	Months in residence	Months in job	Current Debt	Paid off credit
Data Point → Client 1	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

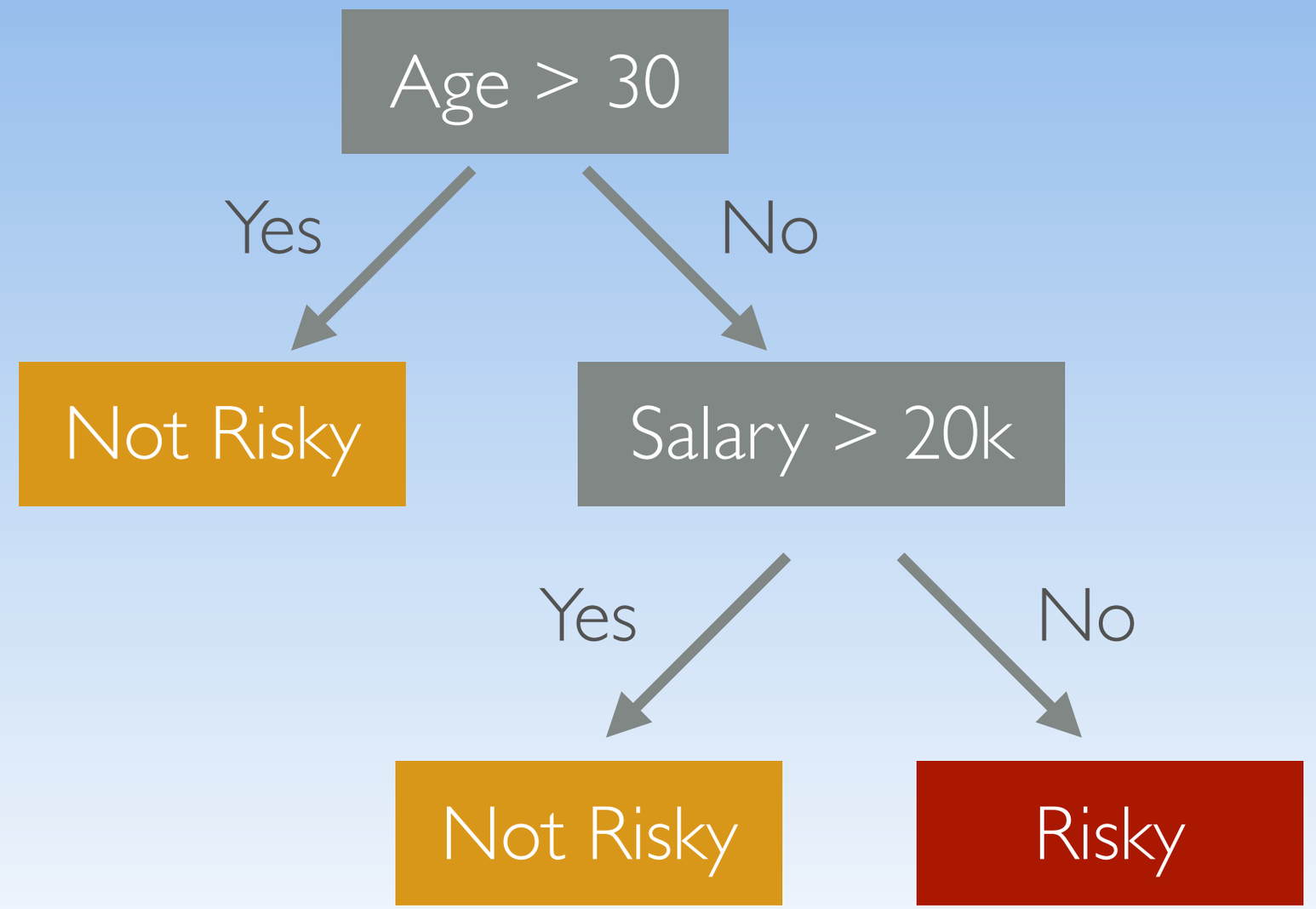


Catalit LLC

DECISION TREE

- GOAL: make optimal decision at each node

- Advantages:
  - easy to interpret
  - fast prediction
  - rules based



DECISION TREE

Binary splits

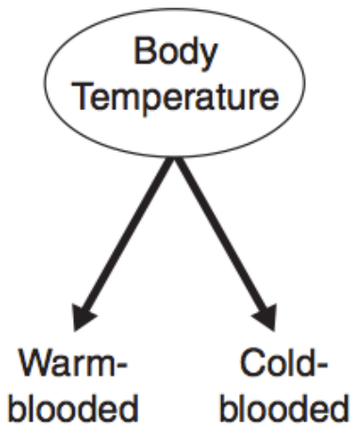
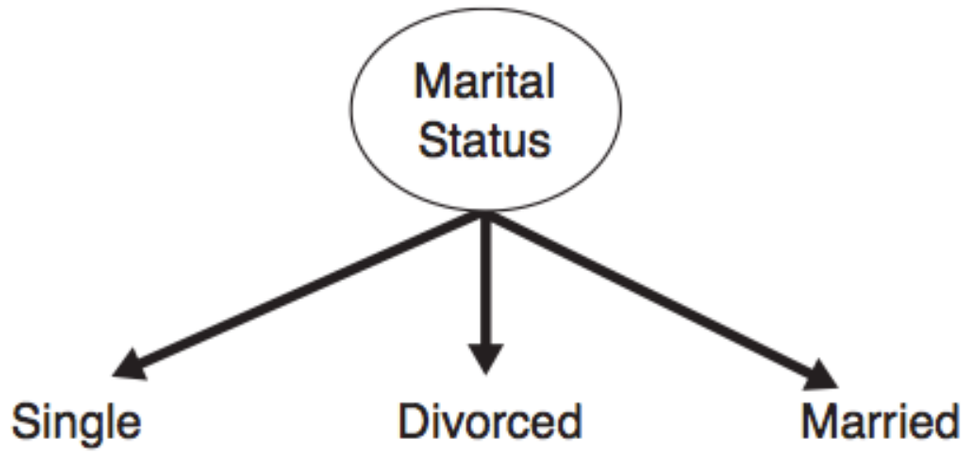


Figure 4.8. Test condition for binary attributes.

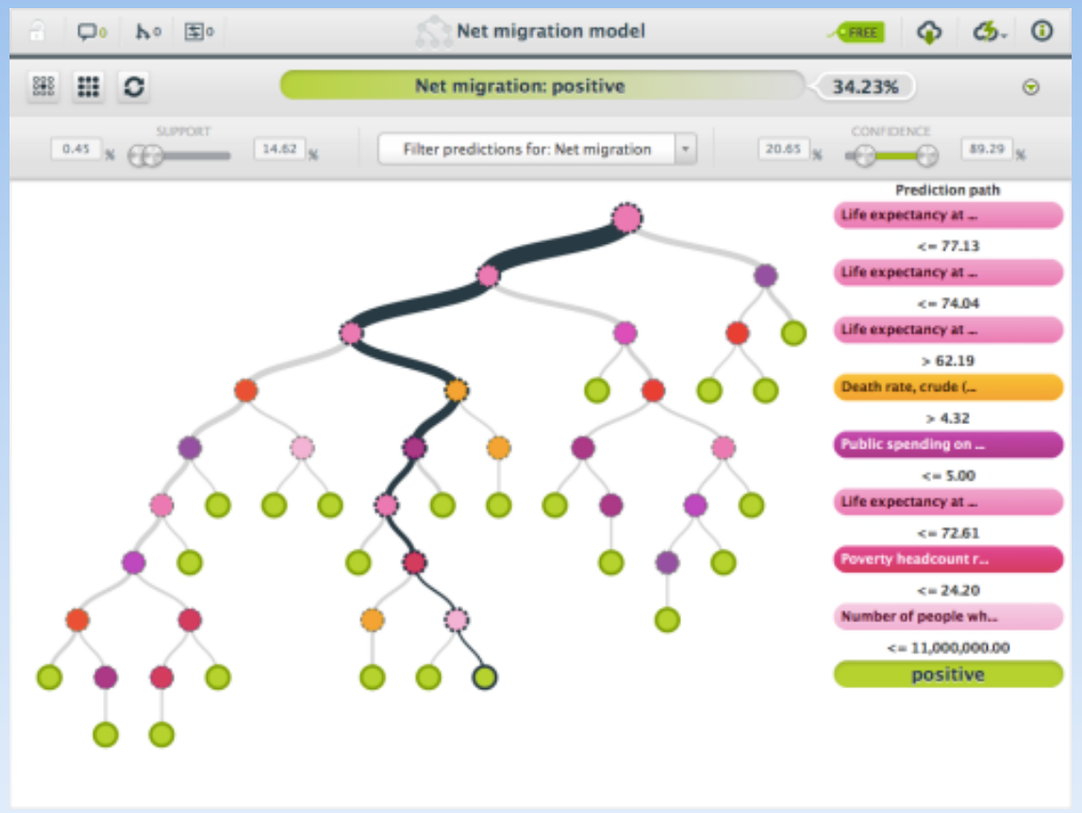
Multiway splits

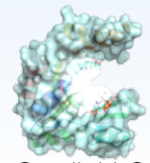


(a) Multiway split

APPLICATIONS

- Lead scoring
- Customer churn
- User segmentation

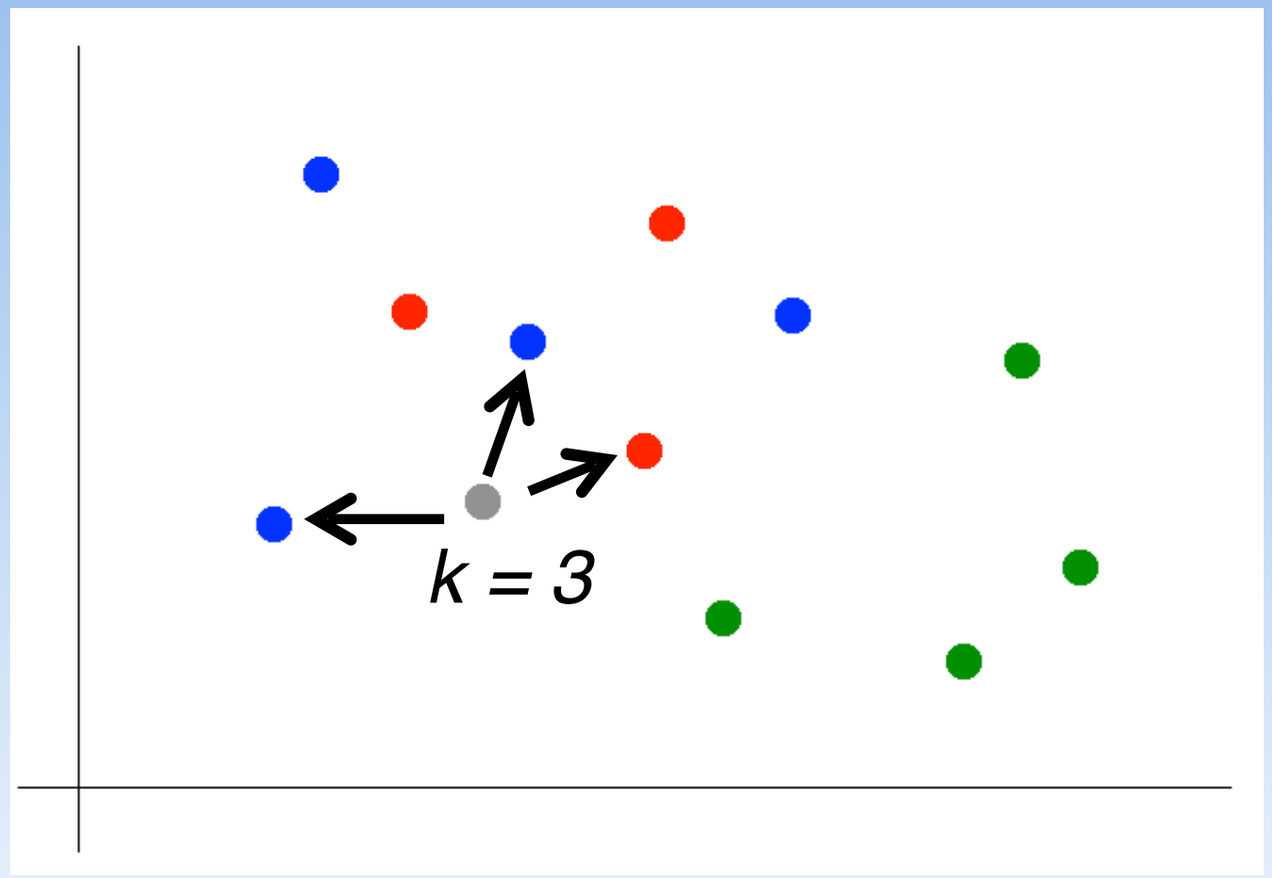


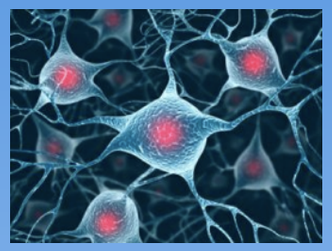


Catalit LLC

K-NEAREST NEIGHBORS

- GOAL: majority vote of nearby points
- Advantages:
  - easy to understand
  - any boundary

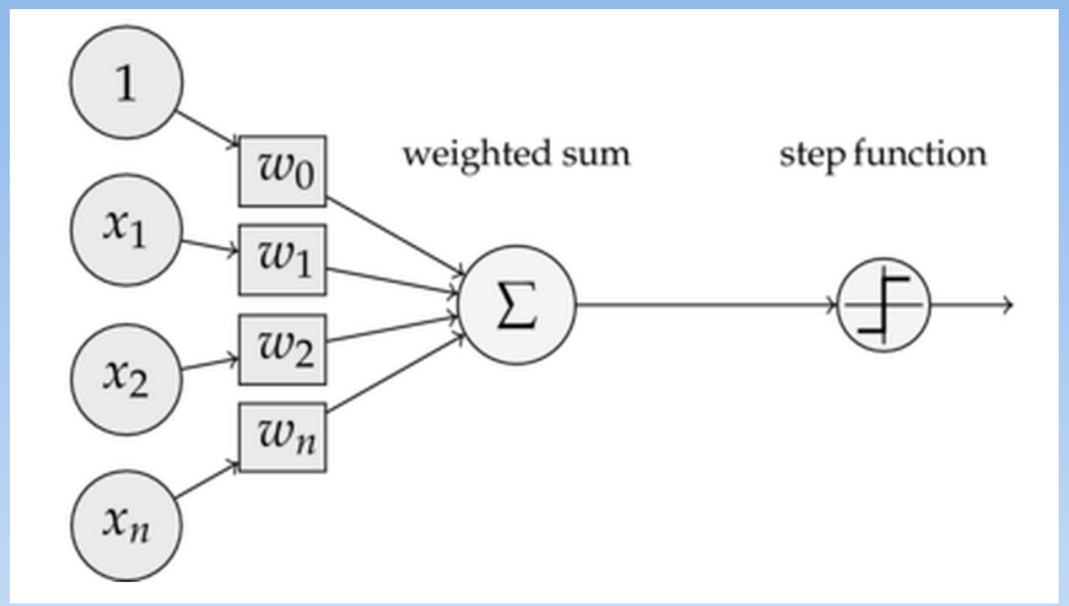




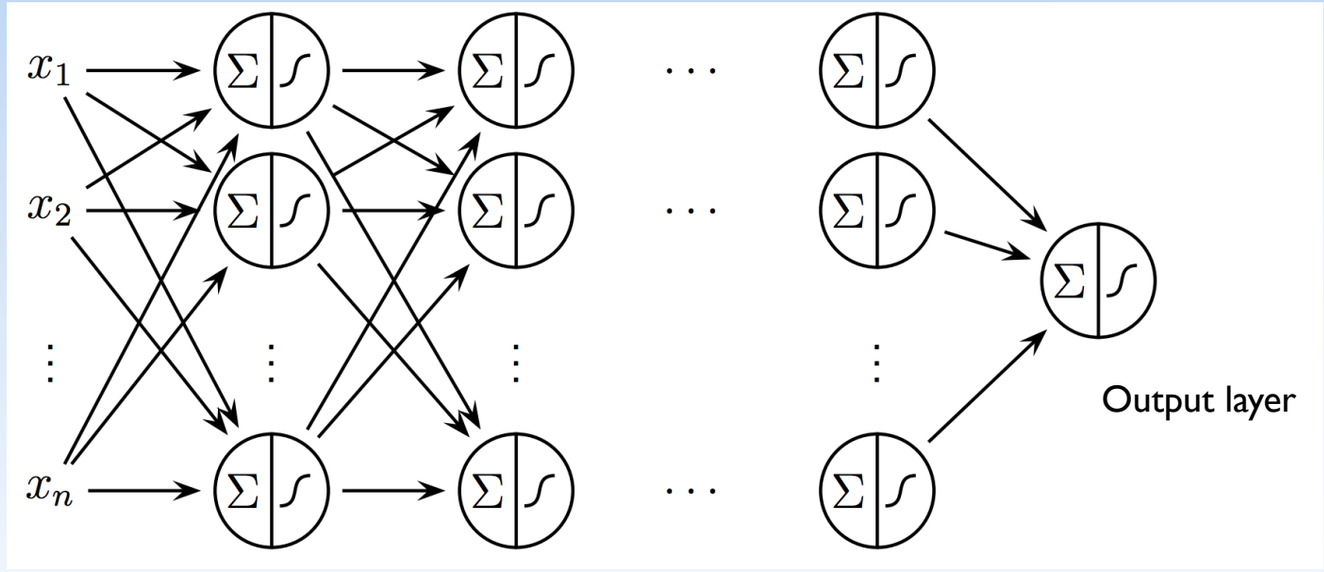
NEURAL NETWORKS

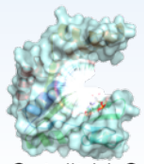
- GOAL: adjust weights to minimize output error

- Perceptron



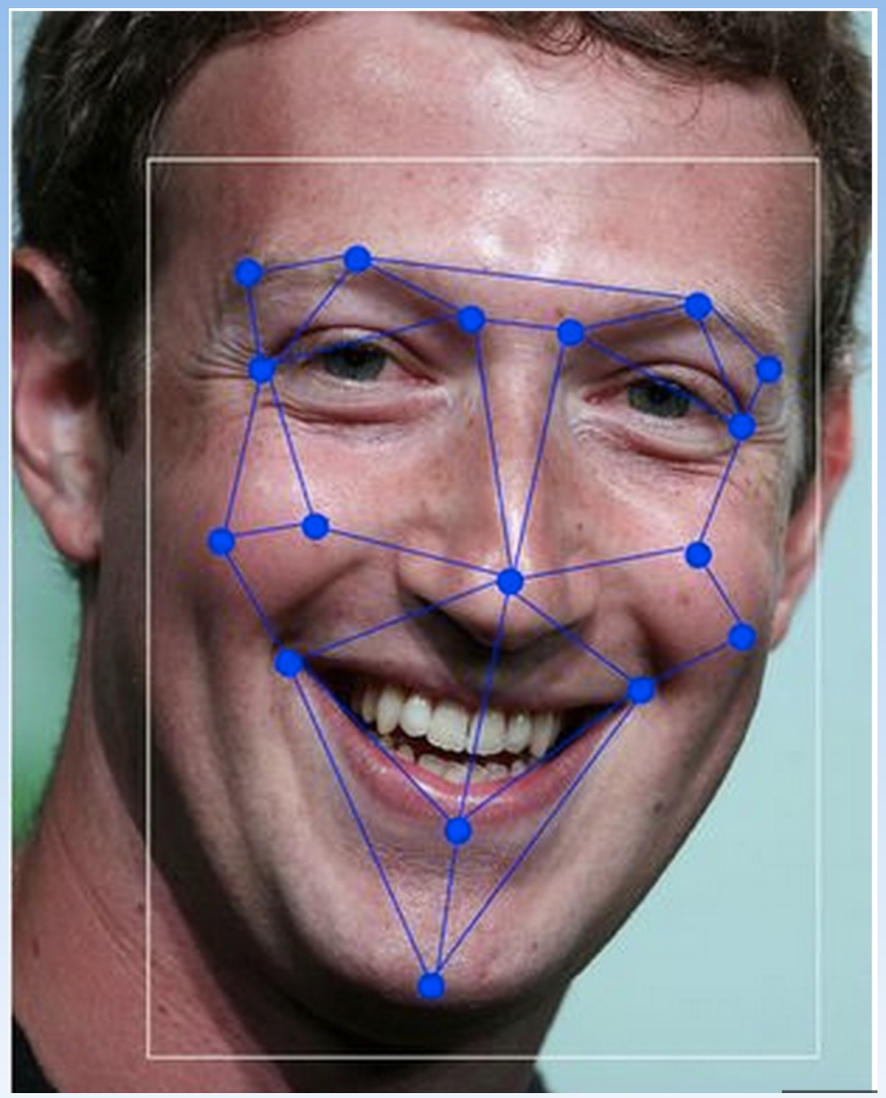
- Multi-layer

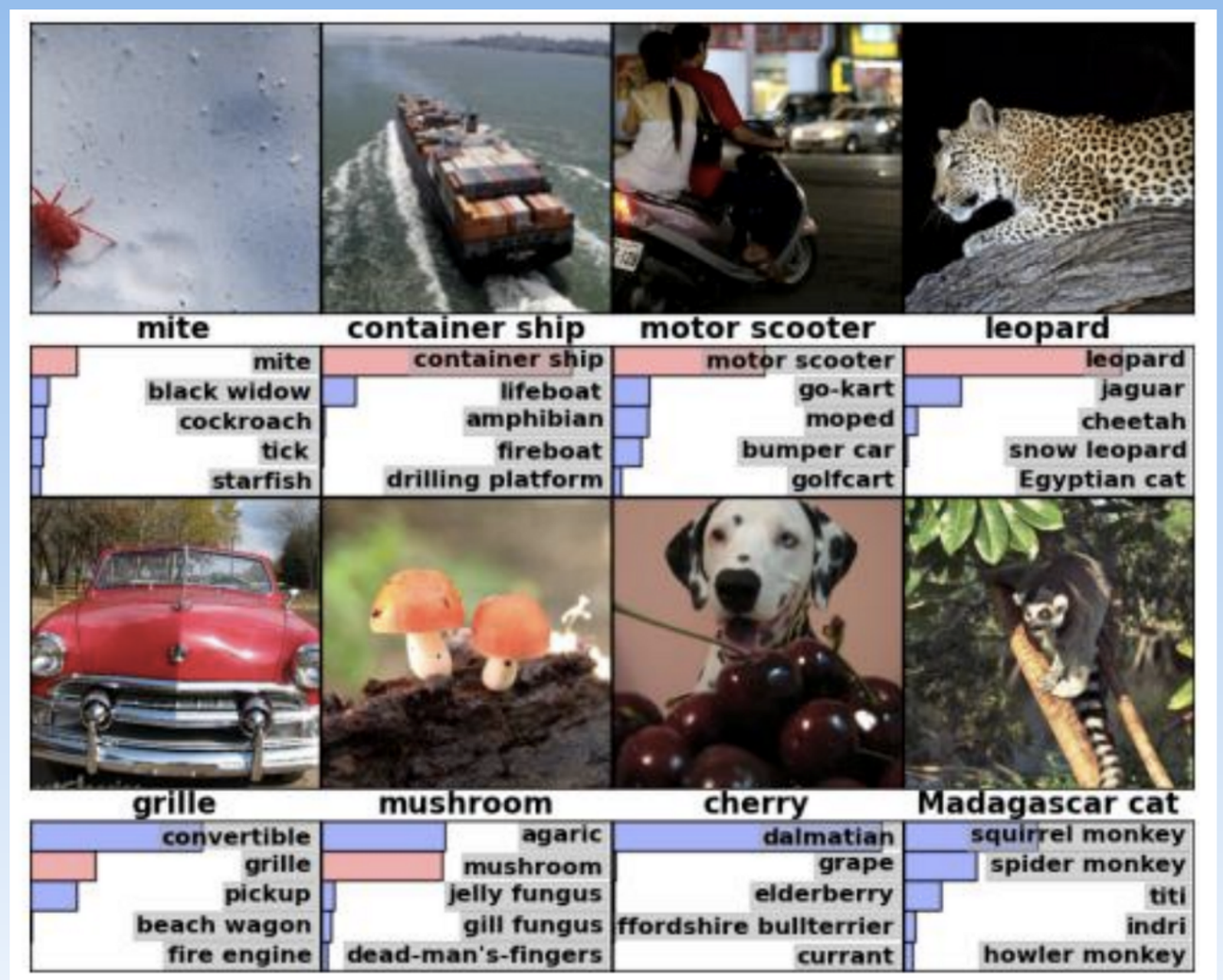




Catalit LLC

APPLICATIONS OF NEURAL NETWORKS



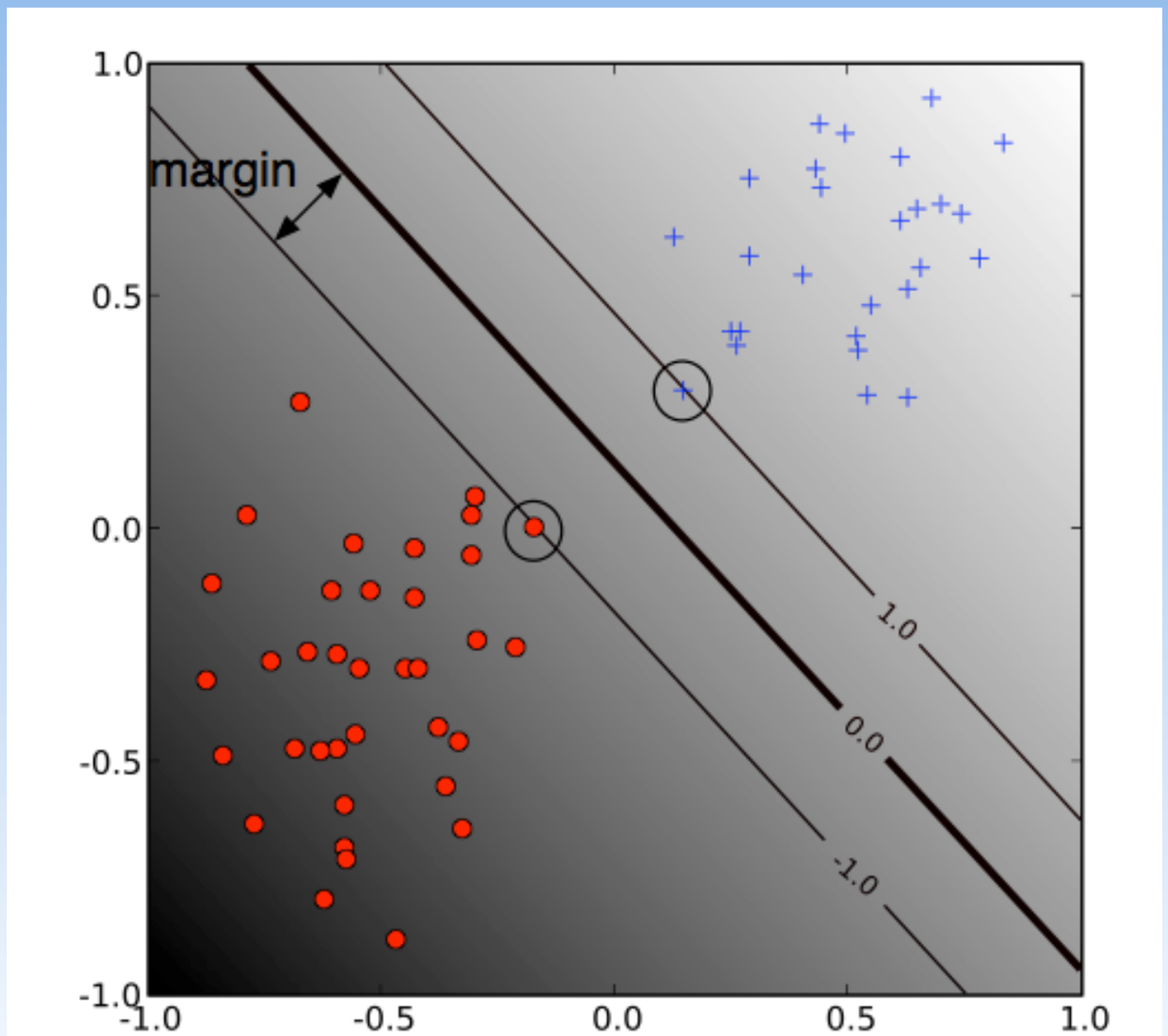


https://research.facebook.com/publications/480567225376225/deepface-closing-the-gap-to-human-level-performance-in-face-verification/

http://www.image-net.org/challenges/LSVRC/

SUPPORT VECTOR MACHINE

- GOAL: Search for a separating hyperplane that maximizes margin



- Advantages:
  - powerful in non-linear boundaries
  - transform data to higher dimension

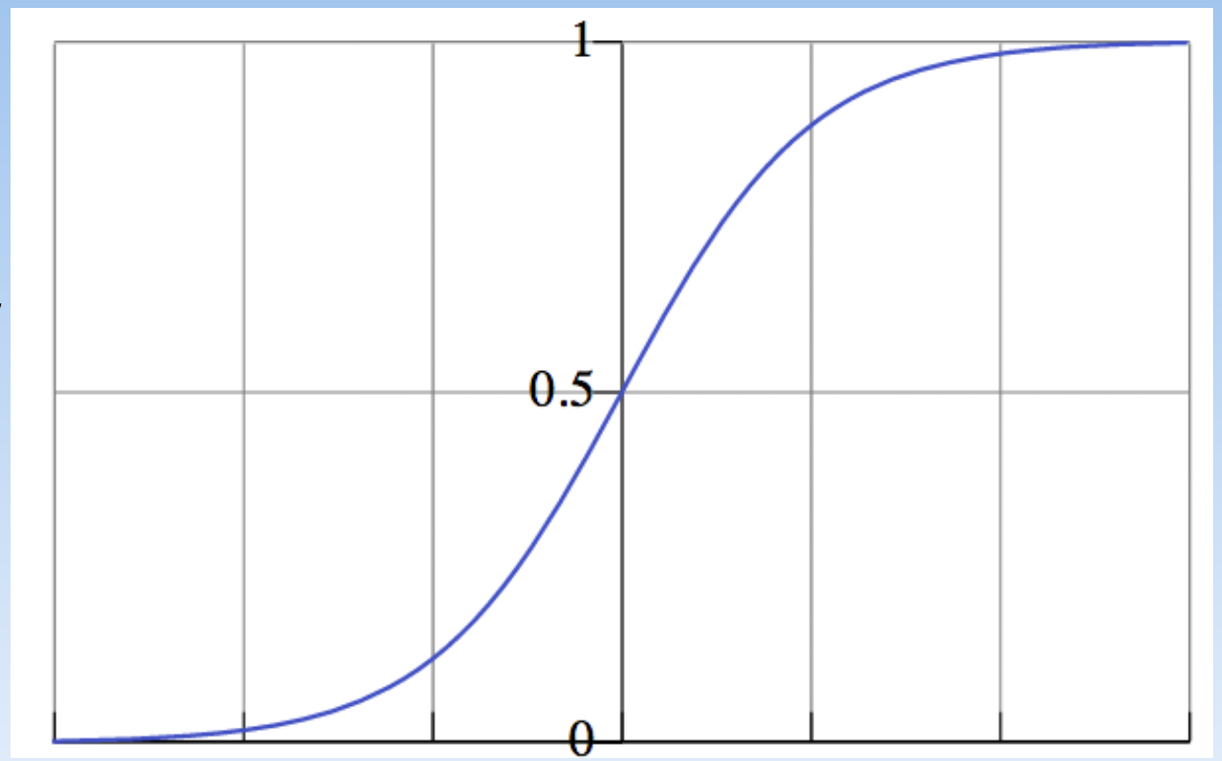
APPLICATIONS OF SVM

- Transmission
  - IP address —167.12.24.555
  - Sender URL -- one-spam.com
- Email header
  - From --“admin@one-spam.cpm”
  - To --“undisclosed”
  - cc
- Email Body
  - # of paragraphs
  - # words
- Email structure
  - # of attachments
  - # of links

LOGISTIC REGRESSION

- GOAL: minimize squared error for probability

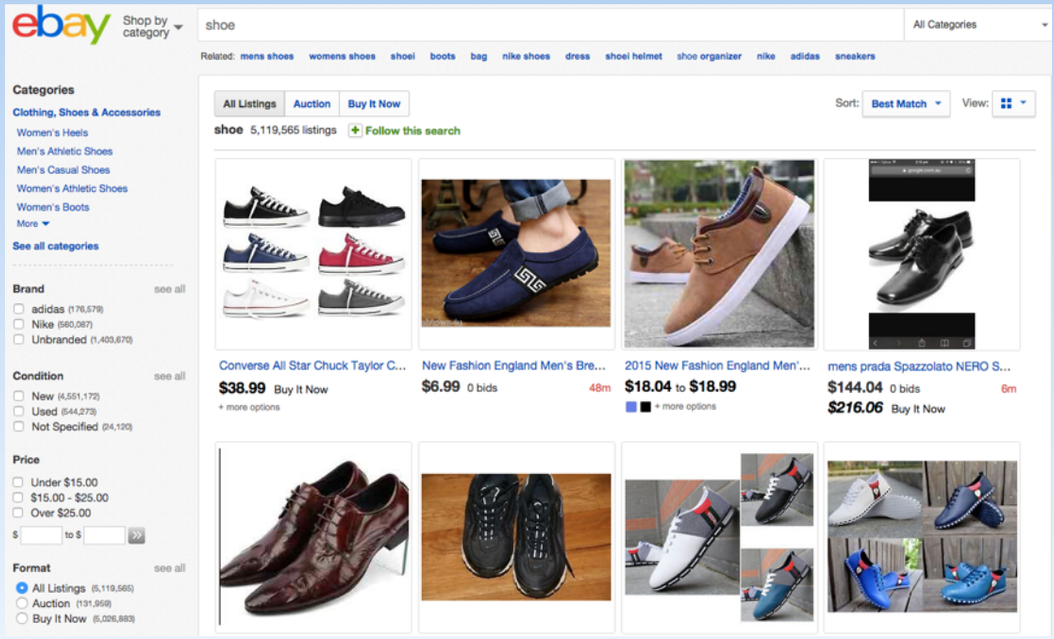
- Advantages:
  - simple function
  - can be parallelized
  - large scale

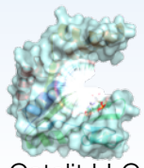


APPLICATIONS

- Click prediction
  - Search ranking (web pages, products)
  - Online advertising
  - Recommendation

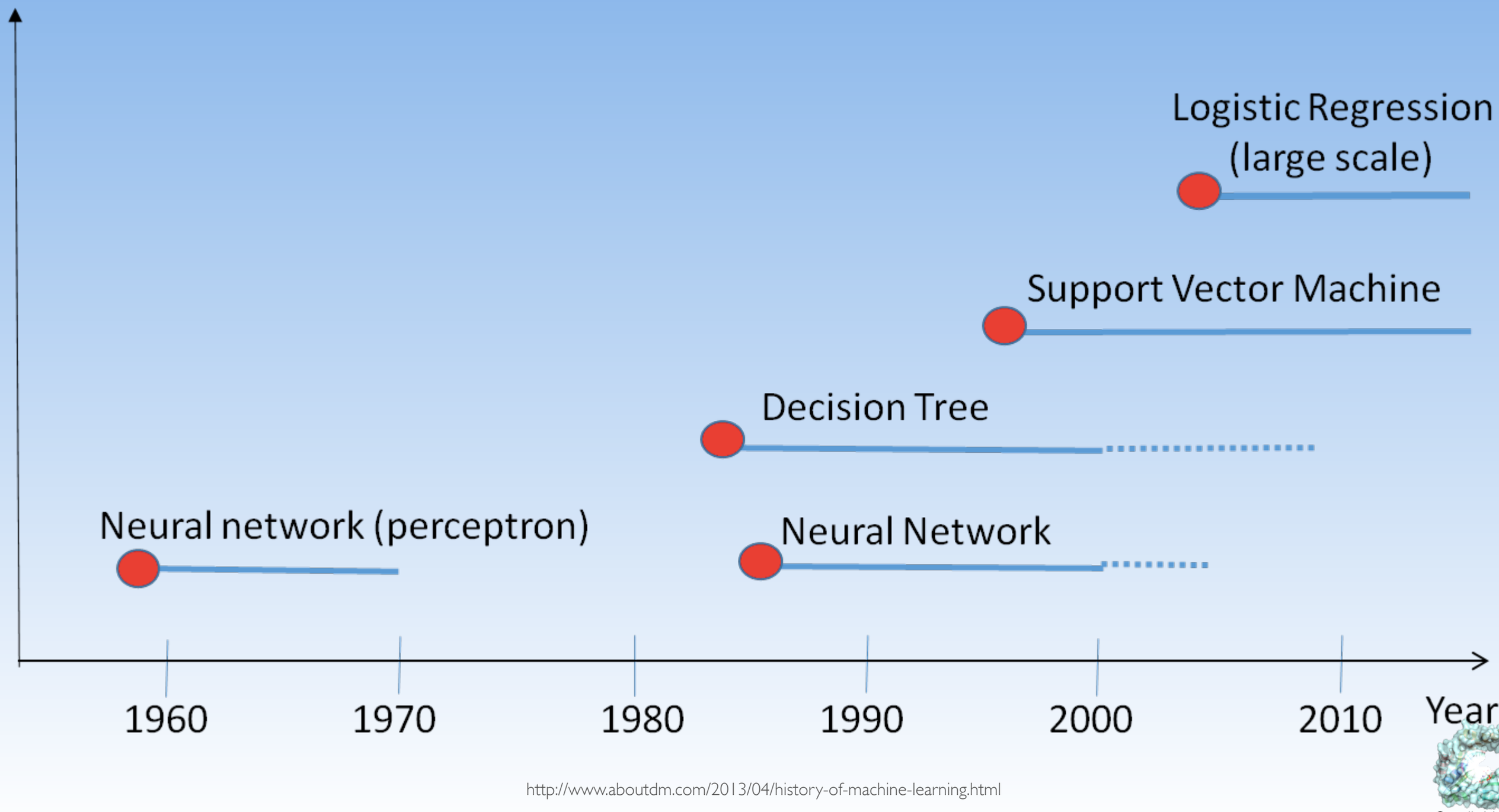
- The model
  - Output: Click/no click
  - Input features:
    - page content
    - search keyword
    - User information





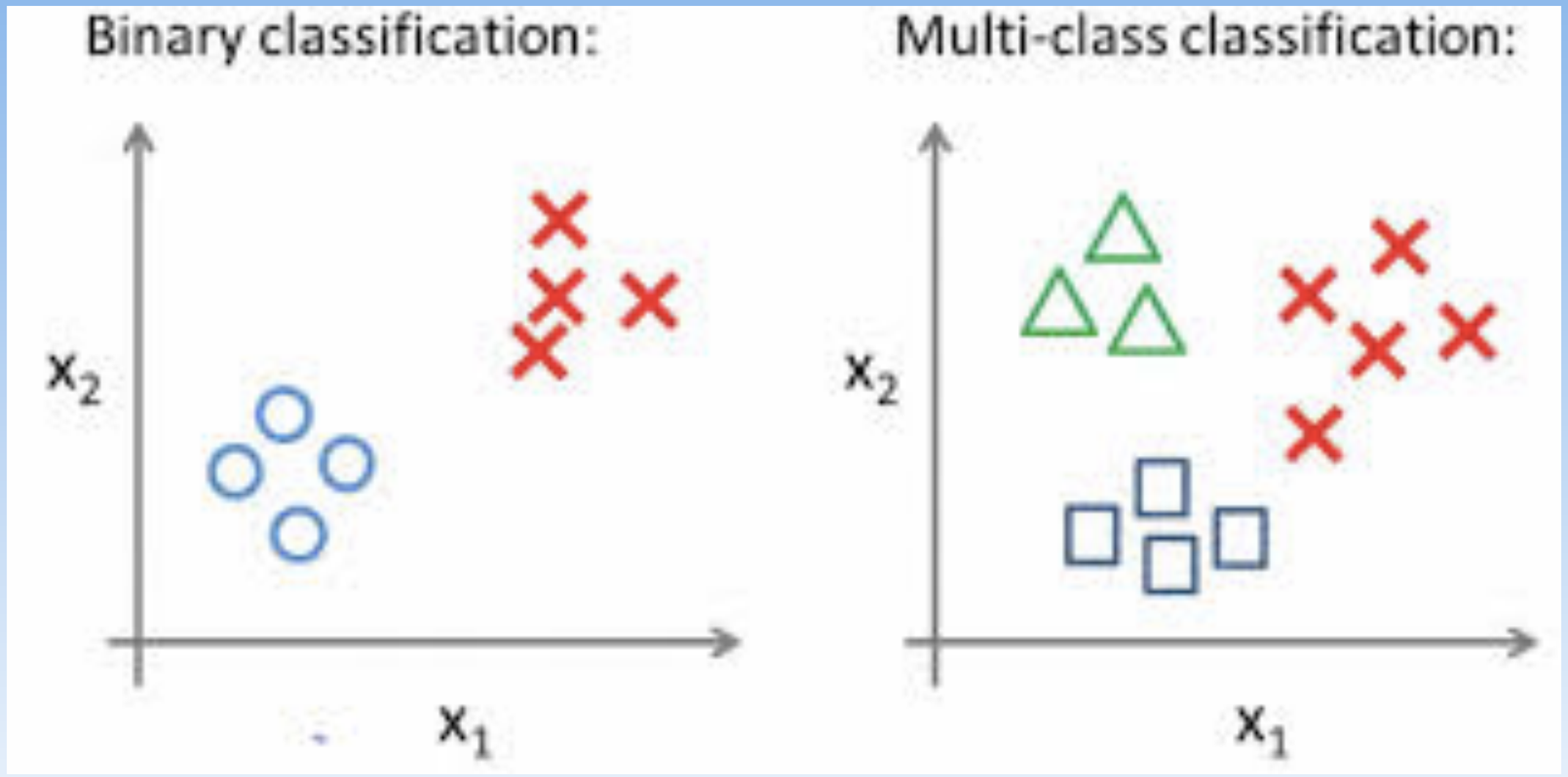
Catalit LLC

HISTORY OF SUPERVISED LEARNING



http://www.aboutdm.com/2013/04/history-of-machine-learning.html

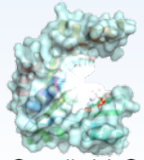
BINARY / MULTI CLASS



Catalit LLC

TRAIN - TEST SPLIT

http://todaysabundantliving.com/wp-content/uploads/2013/08/IMG_2221.jpg



Catalit LLC

TRAIN - TEST SPLIT

All data available

Training data

Testing data

Model

Train Model

Measure performance

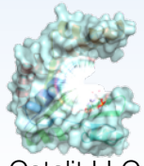
CONFUSION MATRIX

	Test Negative	***Test Positive***
Condition Negative	***TRUE NEGATIVE***	***FALSE POSITIVE (Type I error)***
Condition Positive	***FALSE NEGATIVE (Type II error)***	***TRUE POSITIVE***

Catalit LLC

CONFUSION MATRIX

	Class A	***Class B***	***Class C***	***Class D***
Prediction A	***84***			
Prediction B		***69***	***3***	
Prediction C	***2***		***78***	
Prediction D				***91***

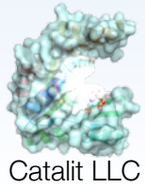


Catalit LLC

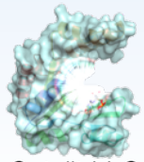
PRECISION - RECALL & ACCURACY

- **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) / total

	Test Negative	***Test Positive***
Condition Negative	***TRUE NEGATIVE***	***FALSE POSITIVE***
Condition Positive	***FALSE NEGATIVE***	***TRUE POSITIVE***



LAB CLASSIFICATION



Catalit LLC