Getting started with Machine Learning

HELLO!

Thomas Paula

Software Engineer

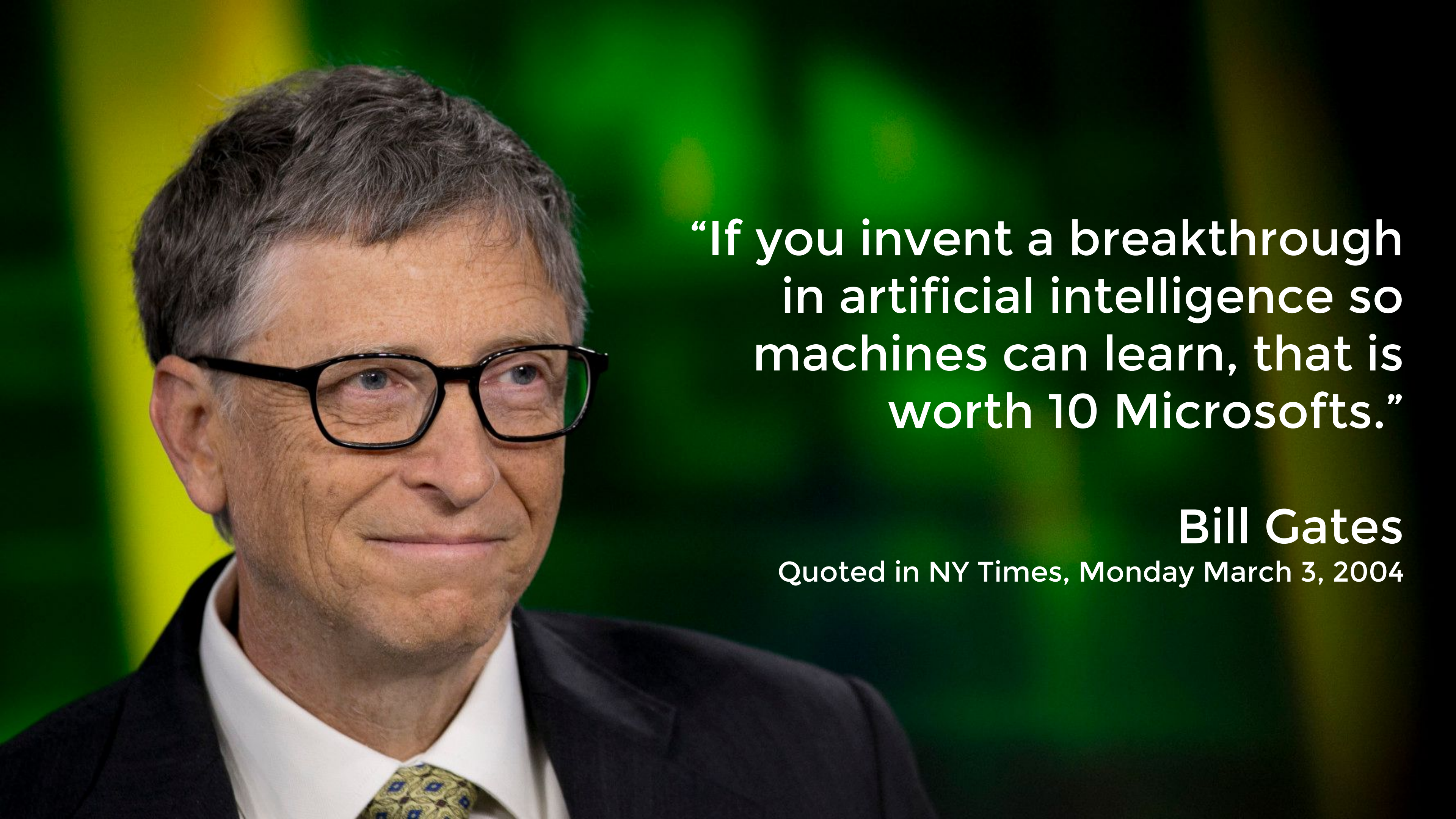
tsp.thomas@gmail.com

Humberto Marchezi

Software Engineer

hcmarchezi@gmail.com


How could we teach machines in a way that they could learn from experience?
1438


“If you invent a breakthrough in artificial intelligence so machines can learn, that is worth 10 Microsofts.”
Bill Gates
Quoted in NY Times, Monday March 3, 2004

1.

INTRODUCTION

Machine what?

“

- A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.

Tom M. Mitchell

“

- A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.

Tom M. Mitchell

“

- A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.

Tom M. Mitchell

“

- A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.

Tom M. Mitchell

Example

Task T: Classify an e-mail as spam or not spam.

Performance measure P: Percentage of e-mails correctly classified.

Training experience E: E-mails manually labeled by humans.

Supervised Learning

Unsupervised Learning

Learning Paradigms

Supervised Learning

Supervised Learning

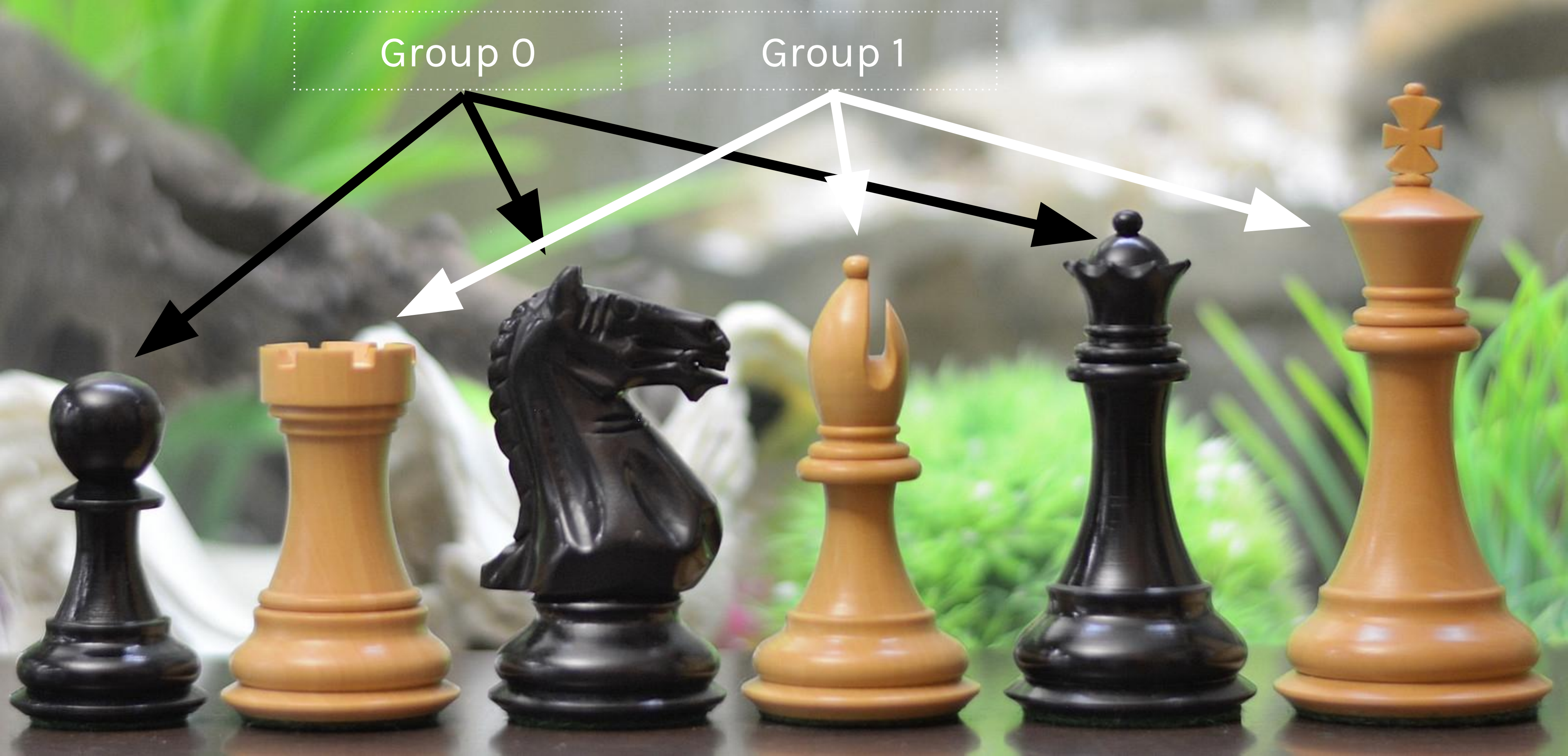
Supervised Learning

- Labeled data;
- Has a “supervisor”;
- Classifies unseen instances (classification) or predicts values (regression).

Unsupervised Learning


Unsupervised Learning
Smaller than others, with a ball on the top: group 0!
Not conic, nothing on the top: group 1!
Higher than others, with cross on the top: group 2!

Unsupervised Learning



Unsupervised Learning

- No labeled data;
- No “supervisor”;
- Finds the regularities in the input.

Classification

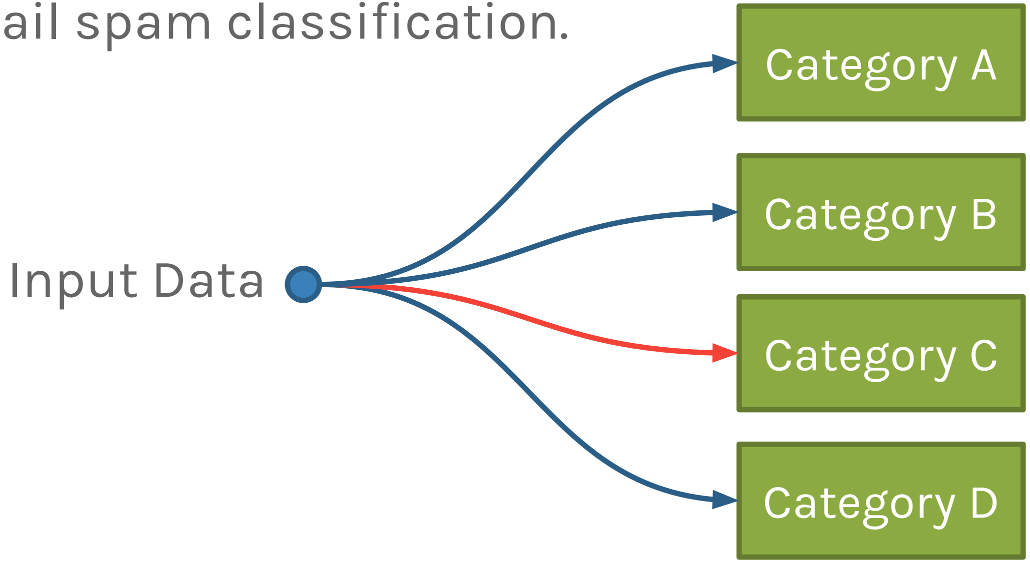
Regression

Clustering

Task Categories

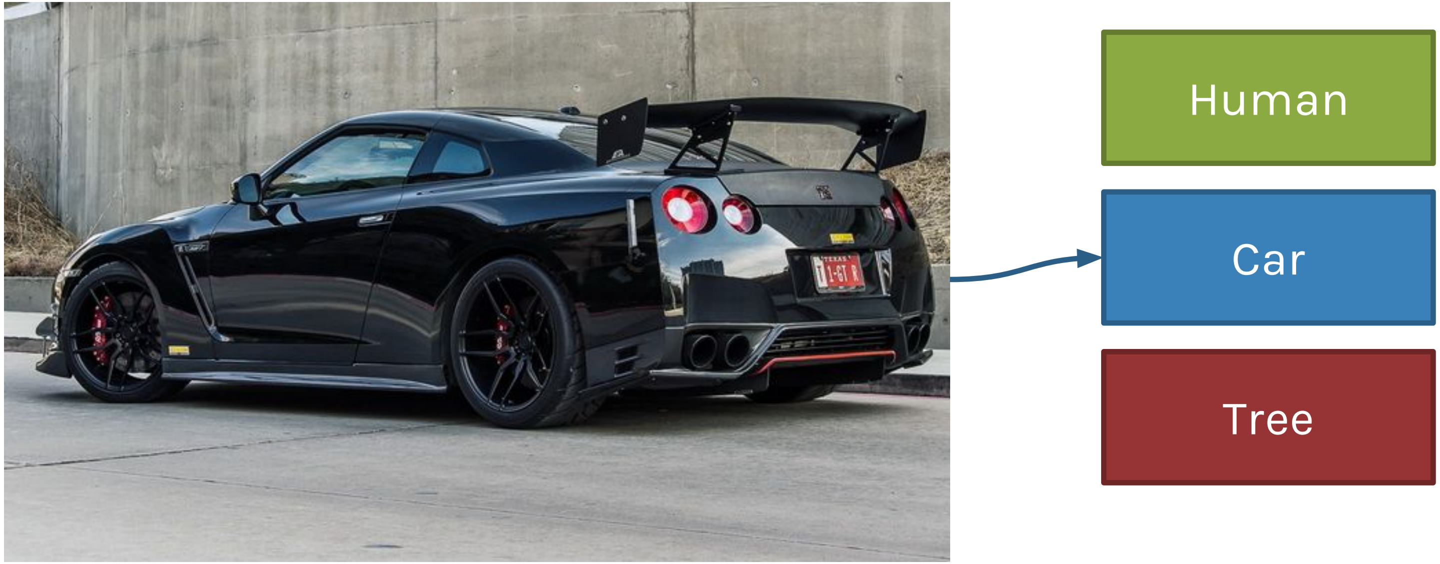
Classification

- Classifies a given element in a category;
- Output as discrete labels (categories);
- Email spam classification.



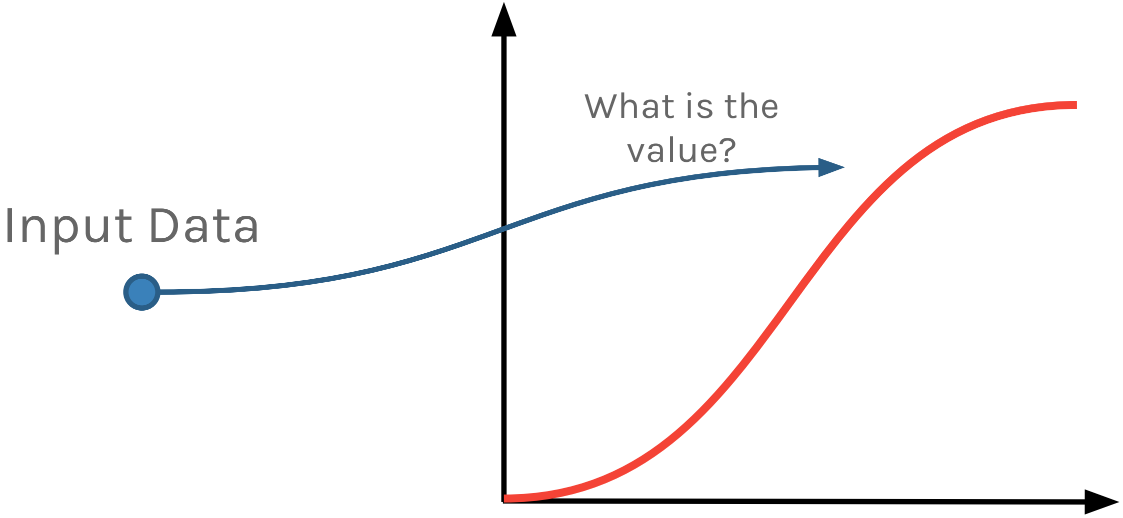
Classification

Classification



Regression

- Forecast/projection of given element
- Output as continuous values
- Predict house price given its features



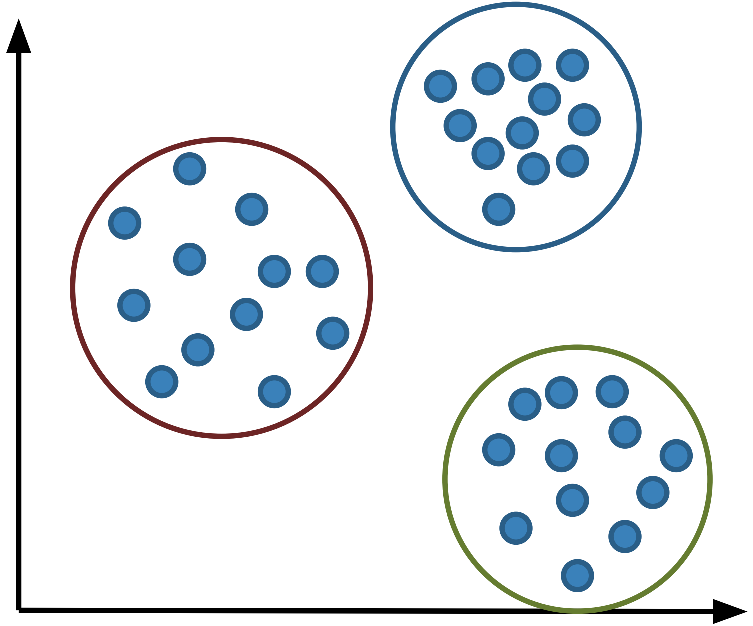
Regression

House features
Area crime rate: 2.5
Bedrooms: 5
Age: 45
Dist. to business district: 8

House value
USD 303,365.00

Clustering

- Find groups from a dataset of elements
- Related news in websites



Clustering

USA TODAY

See realtime coverage

Through adversity, Warriors show again why they're a force to be emulated

USA TODAY - 2 hours ago

OAKLAND - Every so often during this relentless Golden State Warriors run, there's a reminder that they haven't been the kings of the NBA mountain for nearly as long as it might seem.

Portland makes the shift, then Golden State changes the game ESPN (blog)
Warriors put off Curry questions with big comeback win over Blazers
Sports Illustrated

Related
Golden State Warriors »
Portland Trail Blazers »

Local Source: Was emotional loss to Golden State Warriors a death blow for the Trail Blazers?
OregonLive.com (blog)
Opinion: NBA: Thompson sparkles as Warriors overpower Blazers Inquirer.net
In Depth: Blazers take best shot they have in Game 2, but still can't faze Warriors CBSSports.com

Clustering


Through adversity, Warriors show again why they're a force to be emulated
USA TODAY - 2 hours ago
OAKLAND - Every so often during this relentless Golden State Warriors run, there's a reminder that they haven't been the kings of the NBA mountain for nearly as long as it might seem.
USA TODAY
See realtime coverage
Portland makes the shift, then Golden State changes the game ESPN (blog)
Warriors put off Curry questions with big comeback win over Blazers Sports Illustrated
Related
Golden State Warriors »
Portland Trail Blazers »
Local Source: Was emotional loss to Golden State Warriors a death blow for the Trail Blazers? OregonLive.com (blog)
Opinion: NBA: Thompson sparkles as Warriors overpower Blazers Inquirer.net
In Depth: Blazers take best shot they have in Game 2, but still can't faze Warriors CBSSports.com

2.

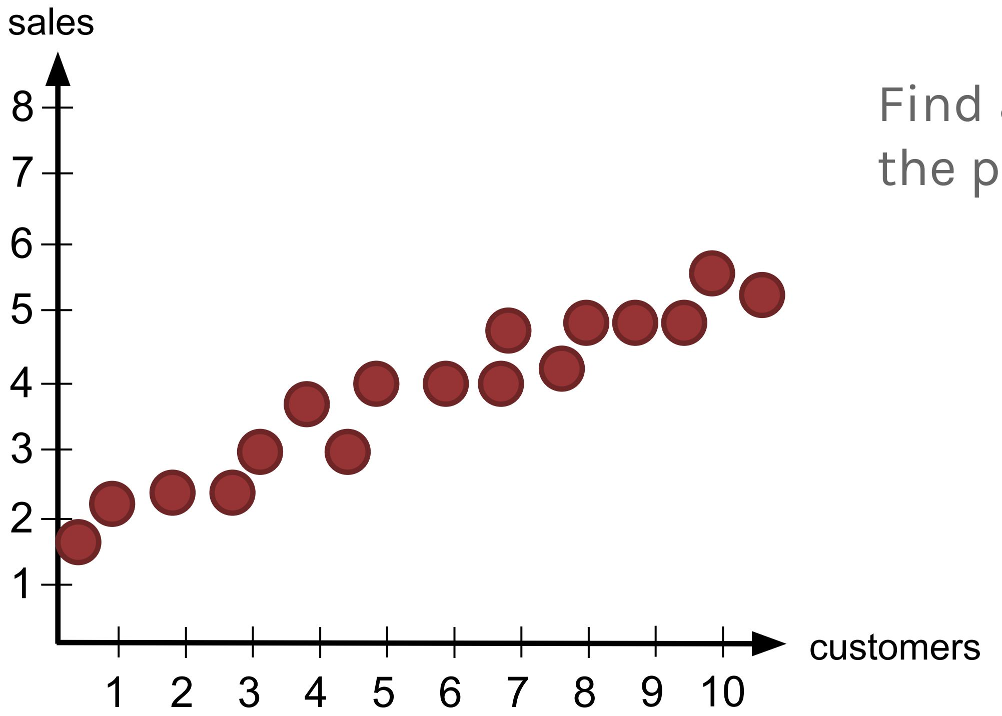
ALGORITHMS

The fun part ☺

Linear Regression

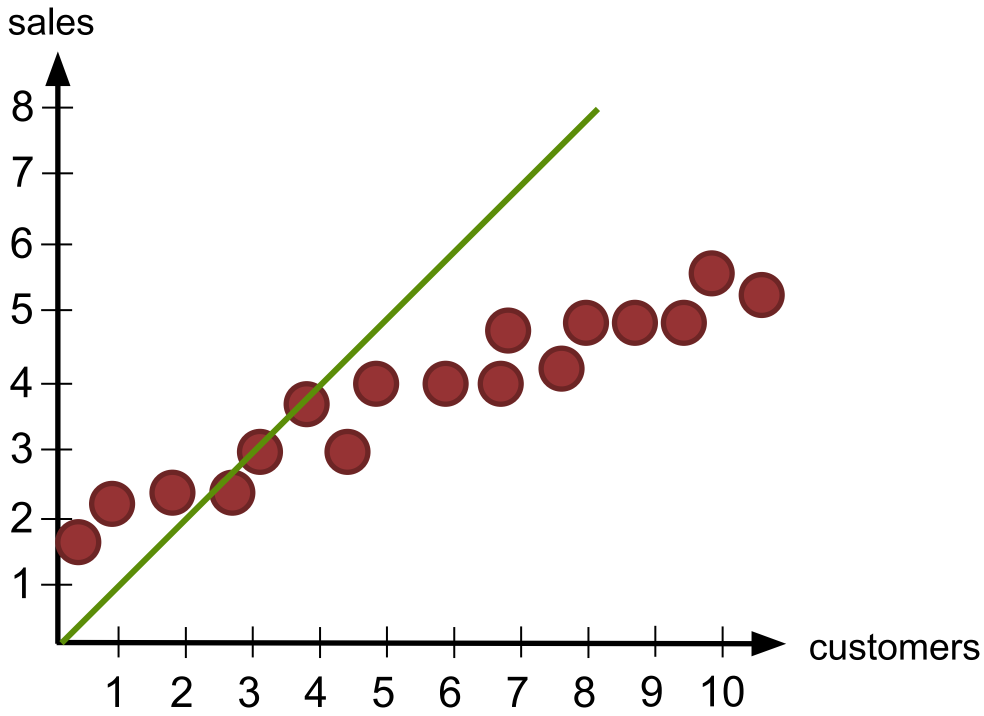
- Supervised learning;
- Regression.

Linear Regression

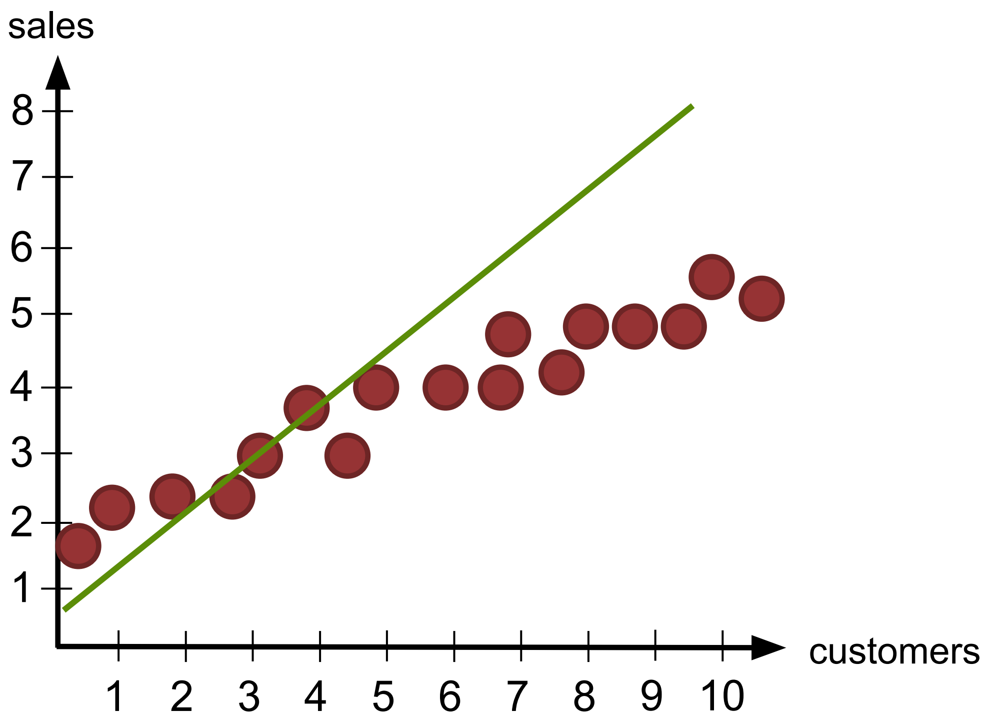


Find a line to fit the point cloud

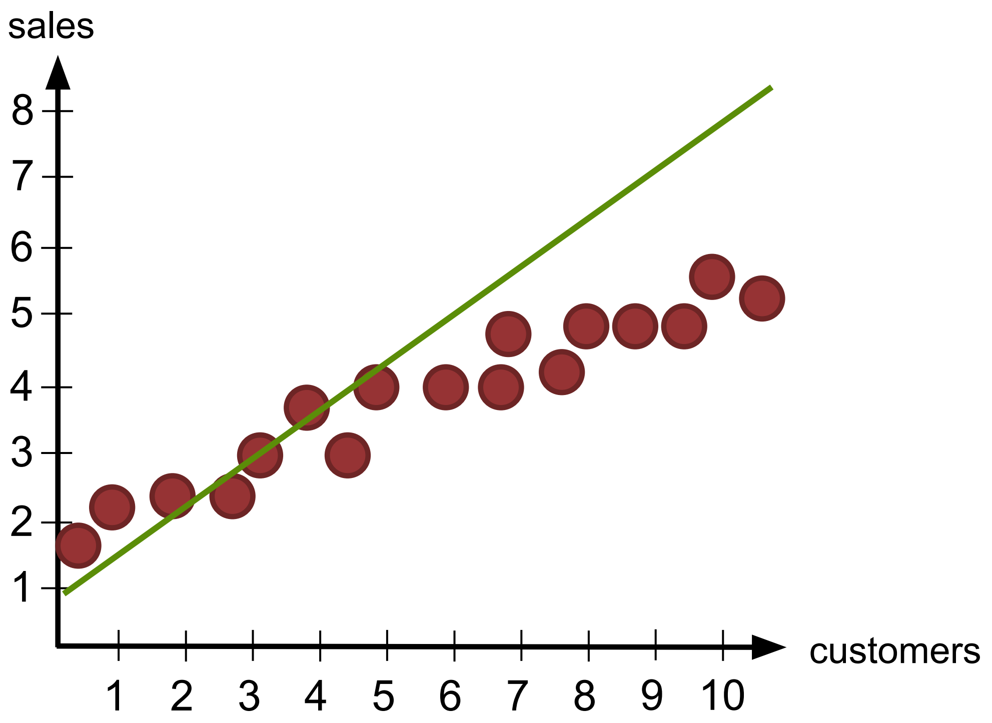
Linear Regression



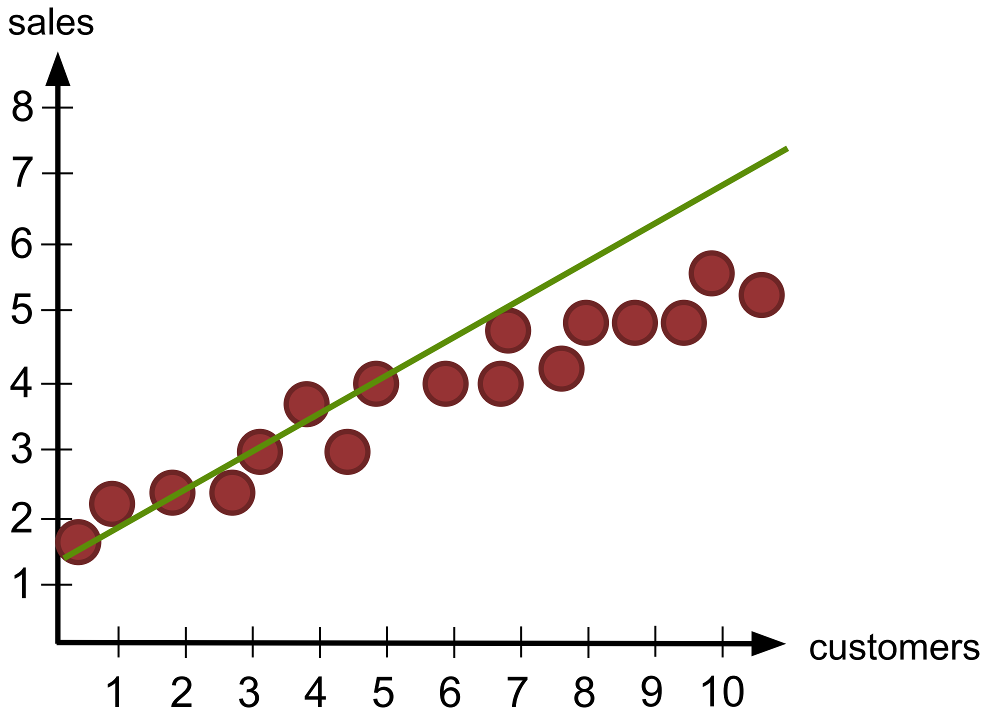
Linear Regression



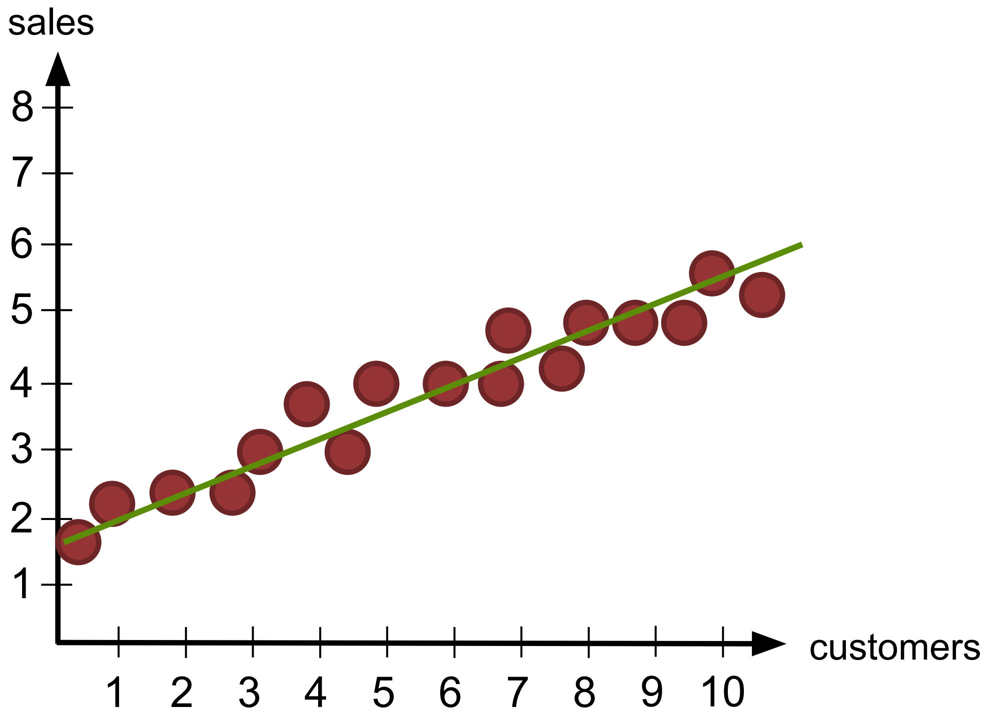
Linear Regression



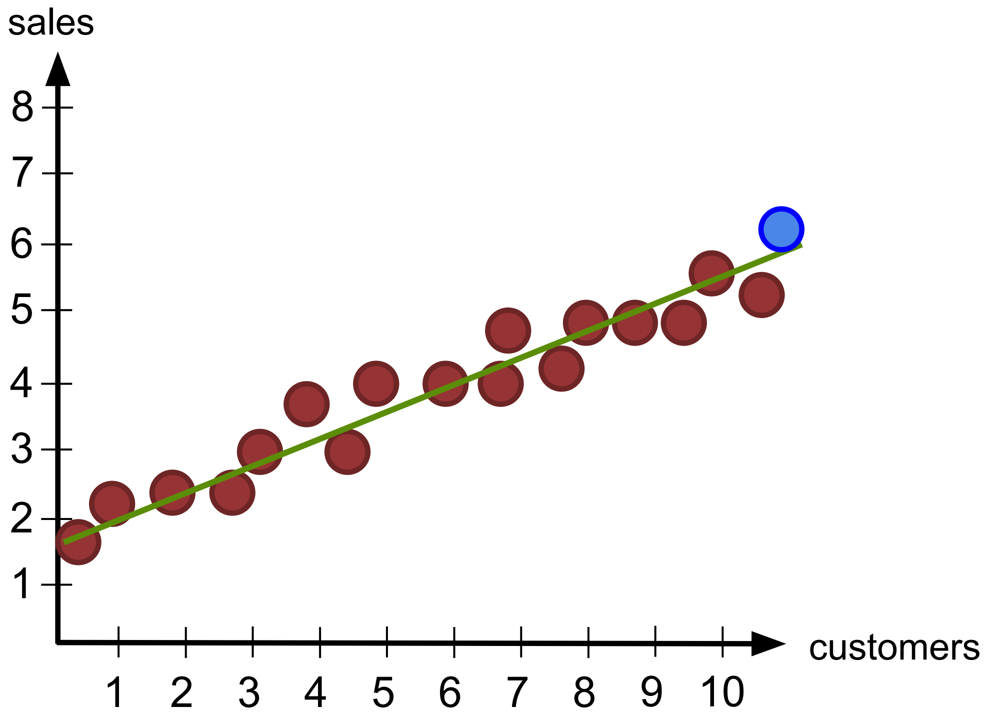
Linear Regression



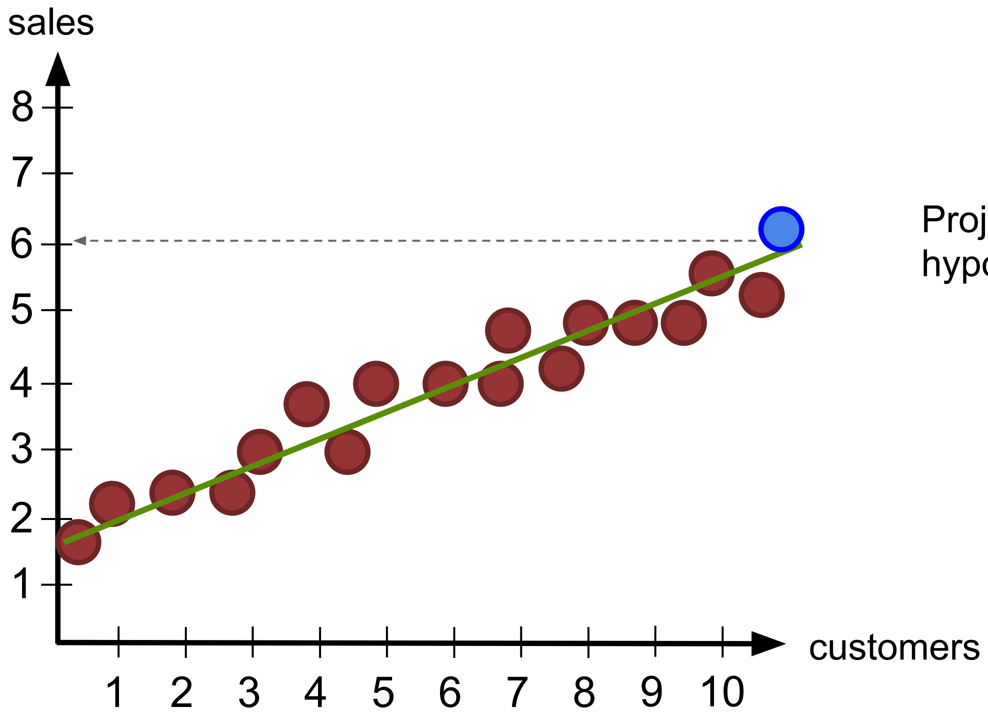
Linear Regression



Linear Regression

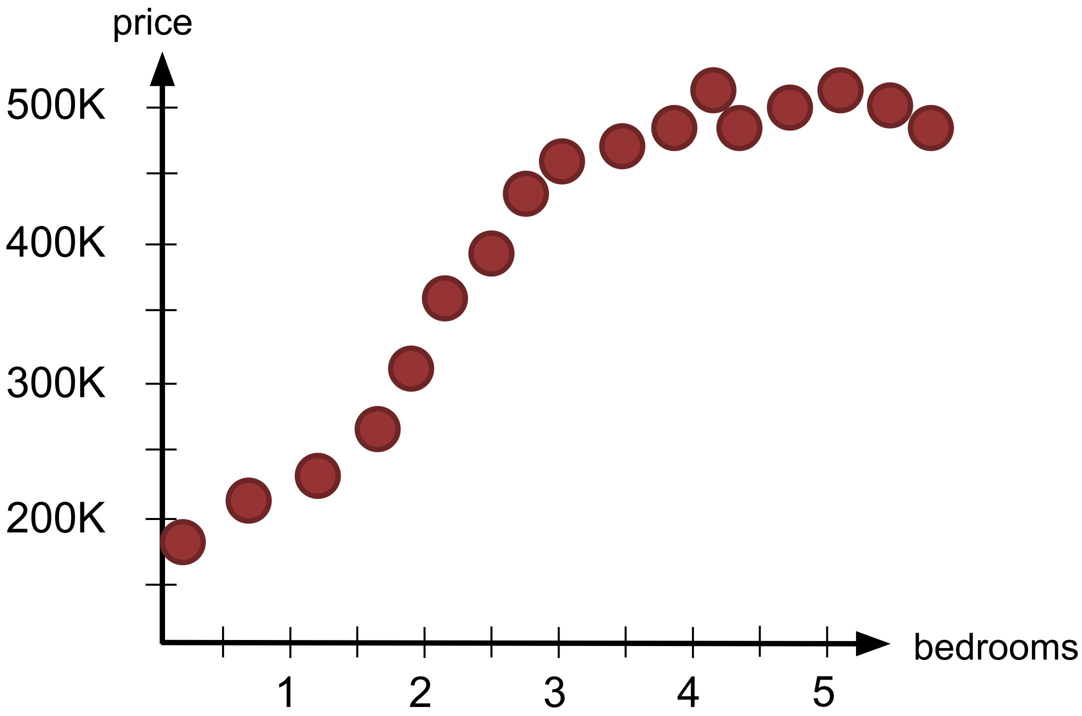


Linear Regression

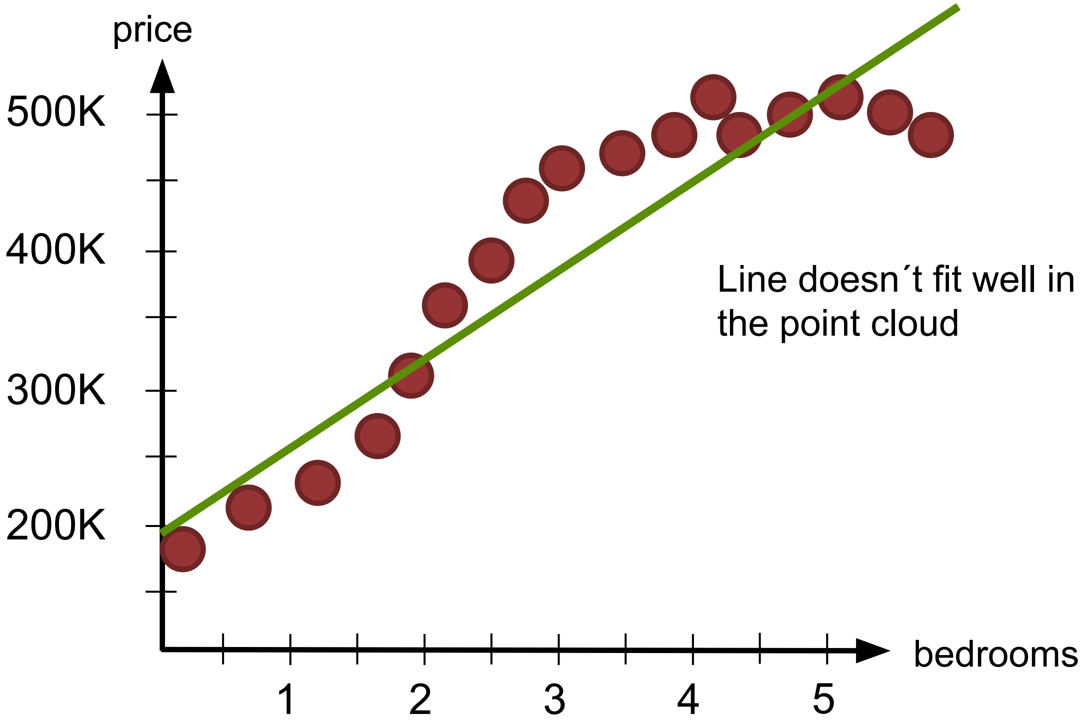


Project the value of hypothetical point.

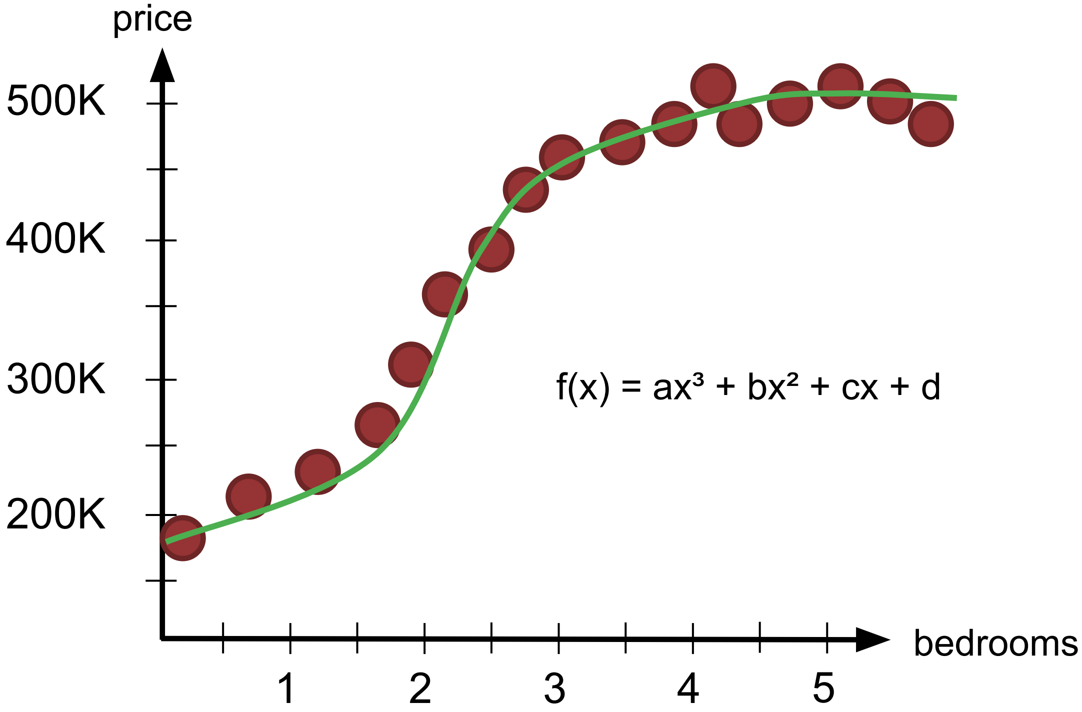
Polynomial Regression



Polynomial Regression



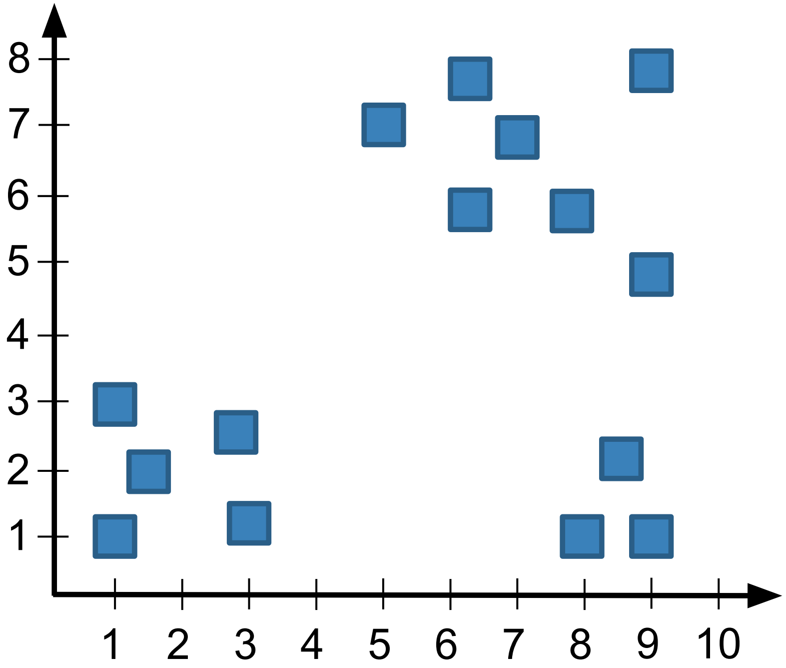
Polynomial Regression



K-means

- Unsupervised learning;
- Clustering;
- Uses K to specify the number of clusters.

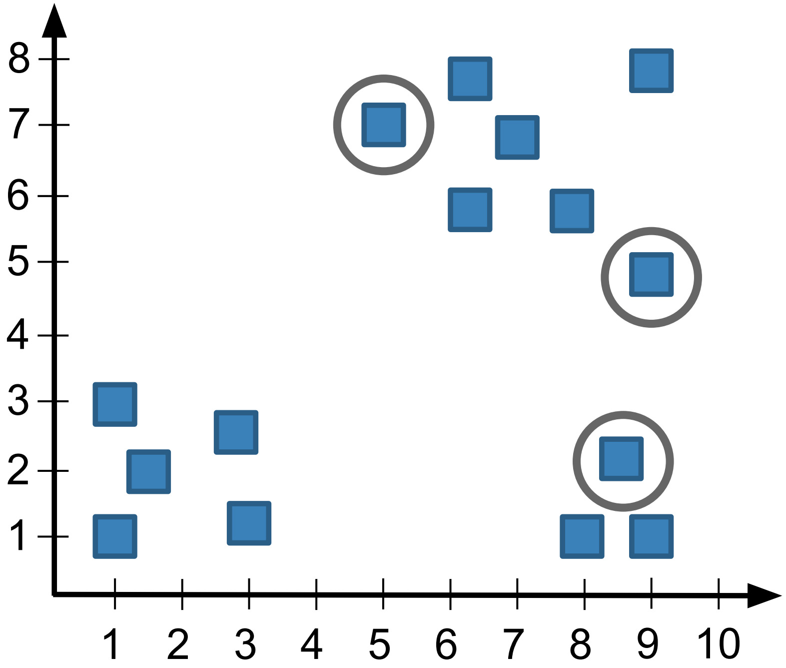
K-means



Find K clusters.

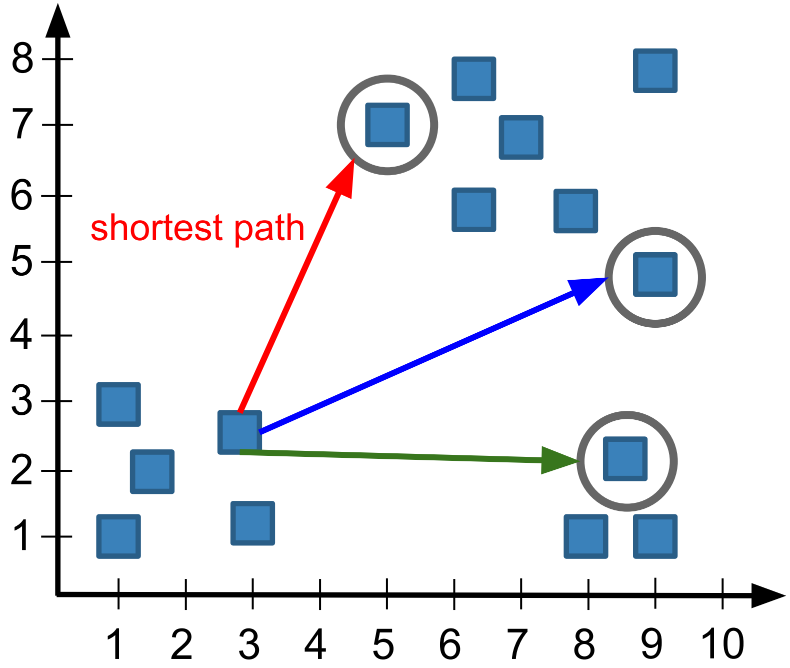
For K = 3

K-means



Step 1
Randomly choose K initial samples as cluster centroids

K-means

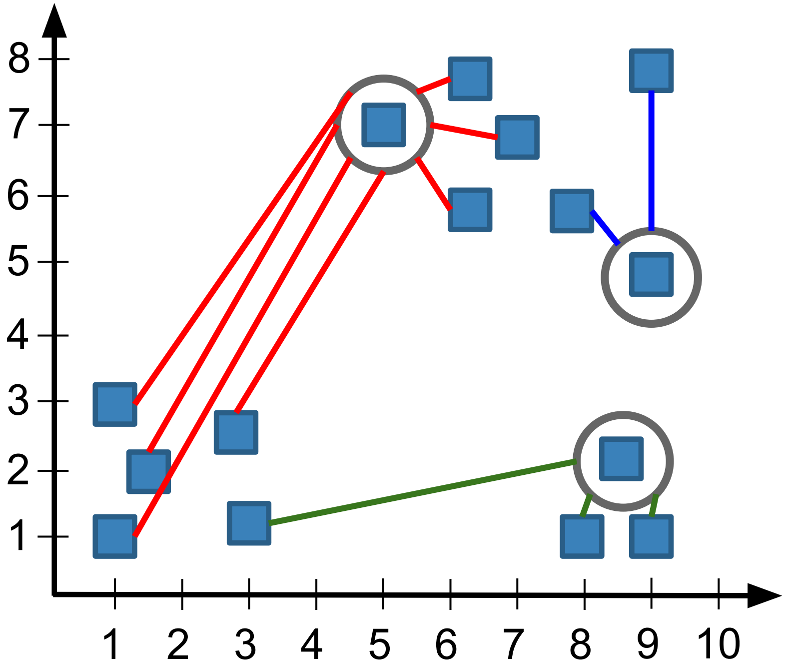


Step 2

For every sample,

- calculate distance to centroids
- pick cluster with closest centroid

K-means

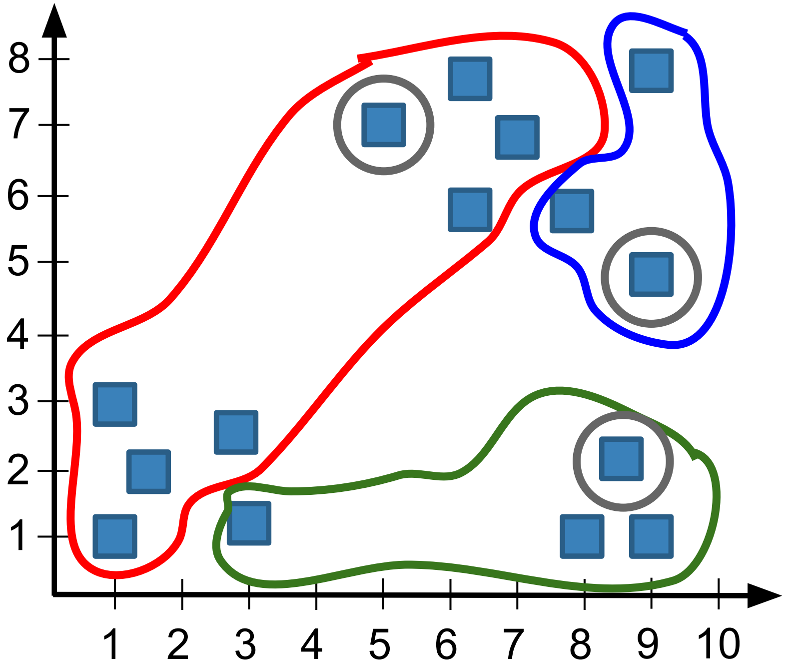


Step 2

For every sample,

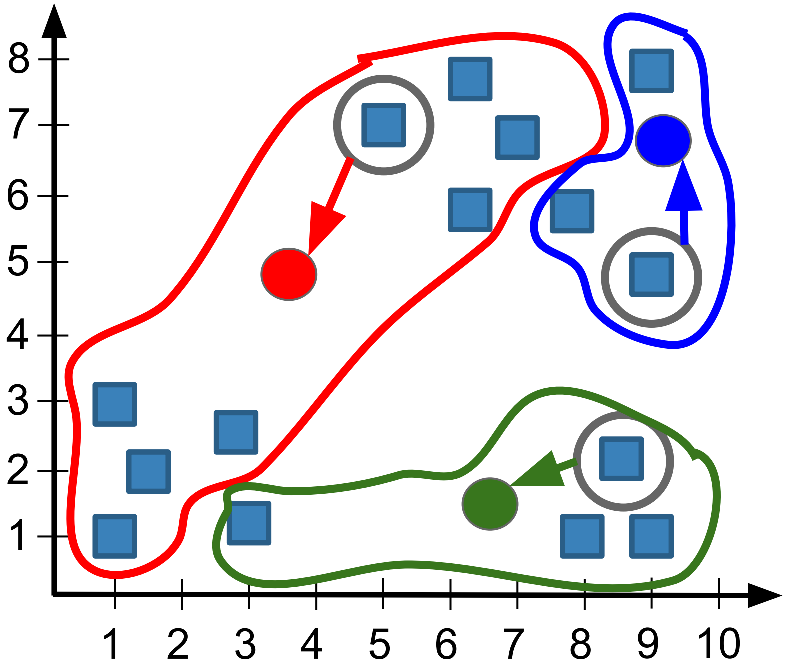
- calculate distance to centroids
- pick cluster with closest centroid

K-means



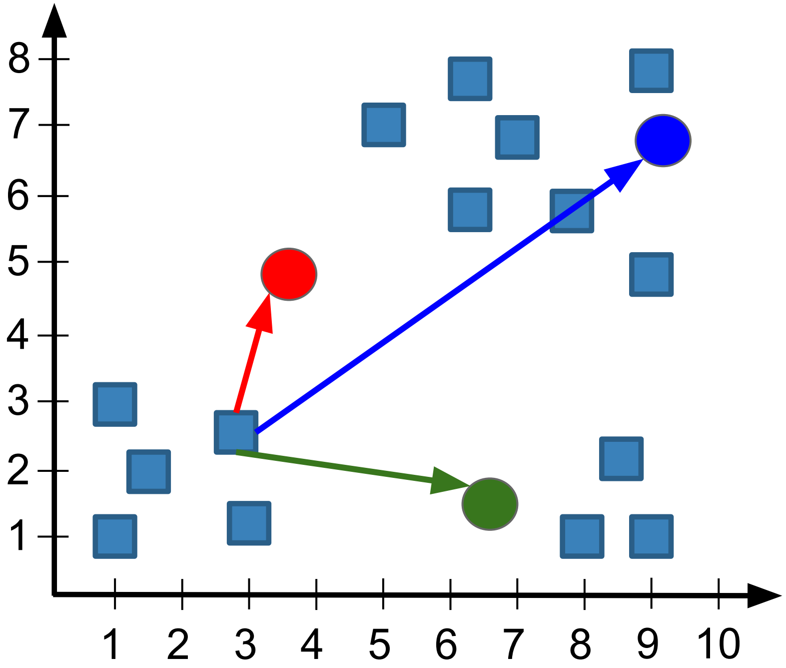
Resulting clusters

K-means



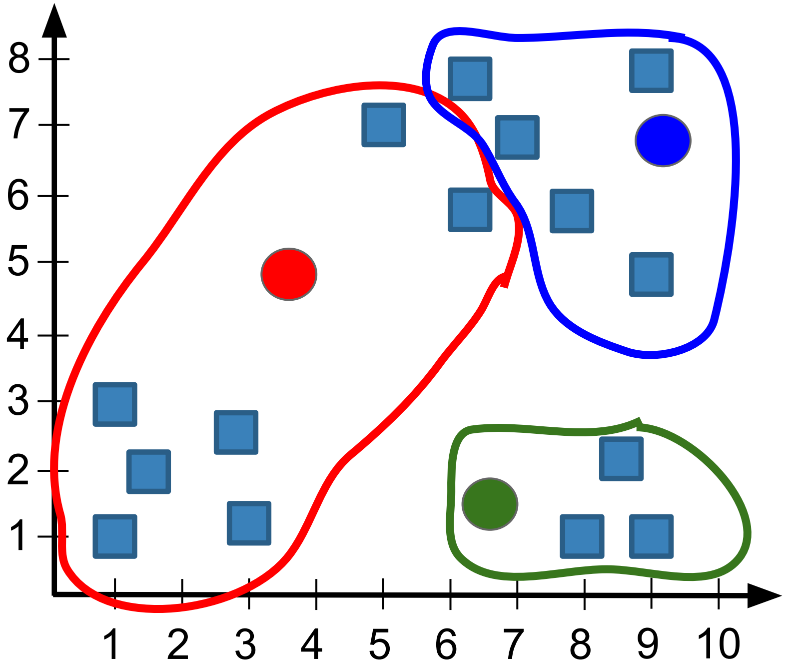
Step 3
Recalculate centroids from cluster members

K-means



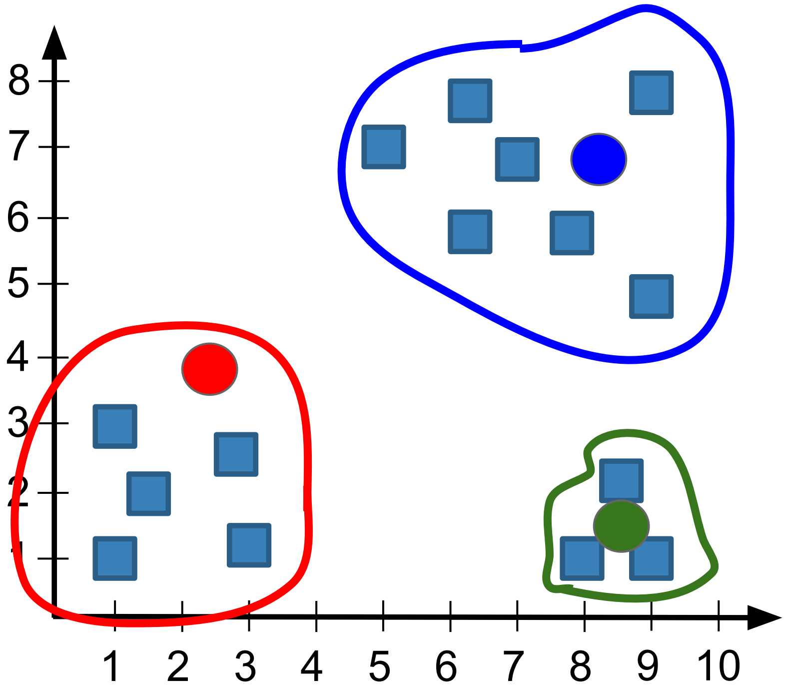
Step 4
If centroids move is bigger than distance D, go back to step 2 if new centroids

K-means



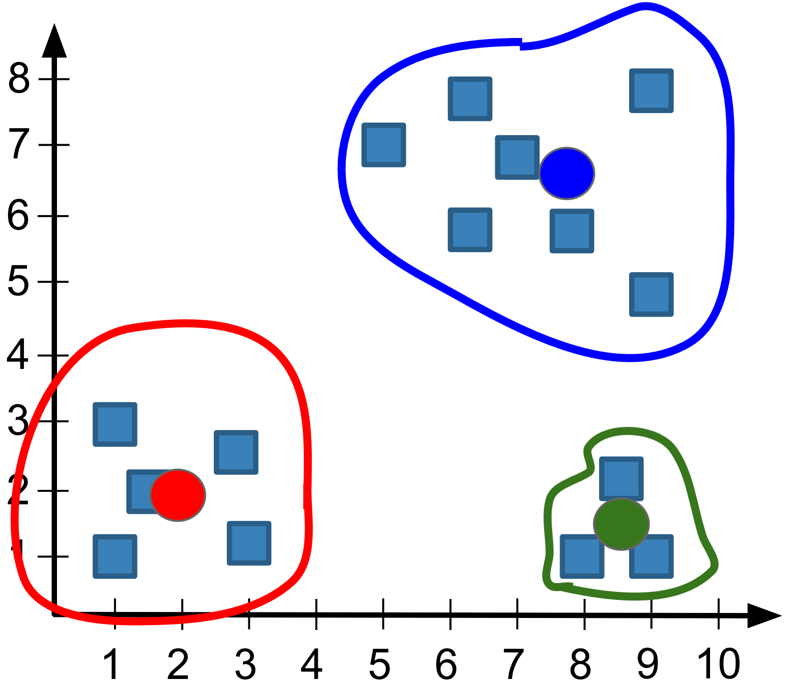
2nd iteration

K-means



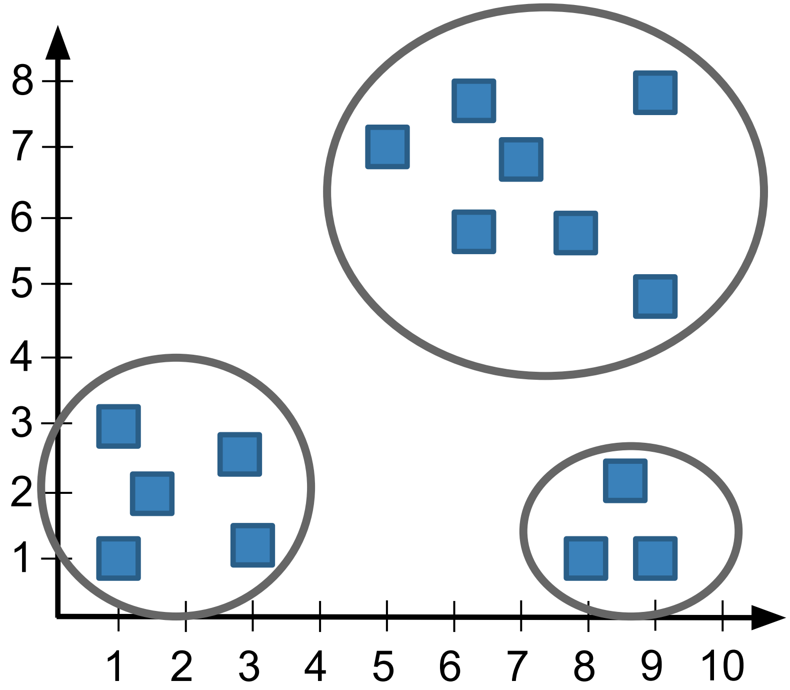
3rd iteration

K-means



4th iteration

K-means



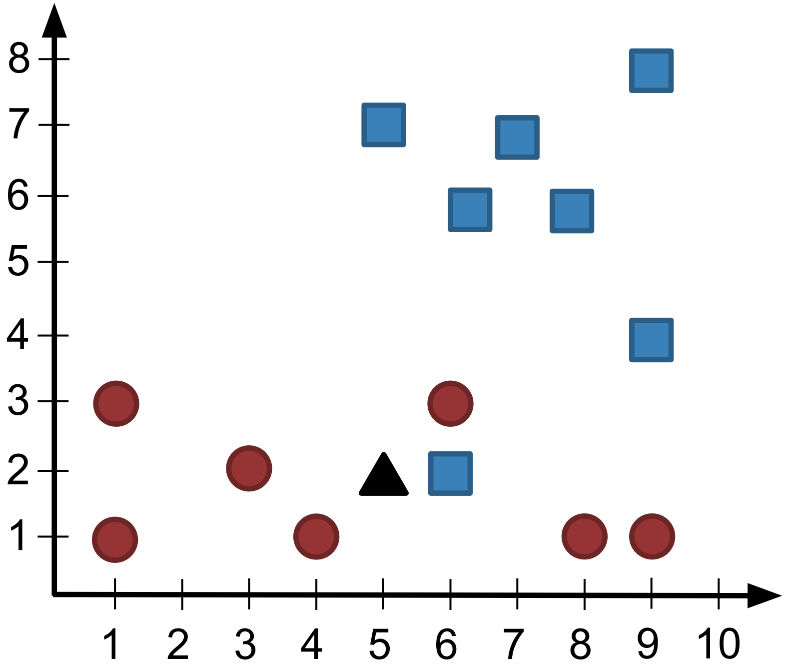
Find K clusters.
For K = 3

http://www.naftaliharris.com/blog/visualizing-k-means-clustering/

K-Nearest Neighbors (K-NN)

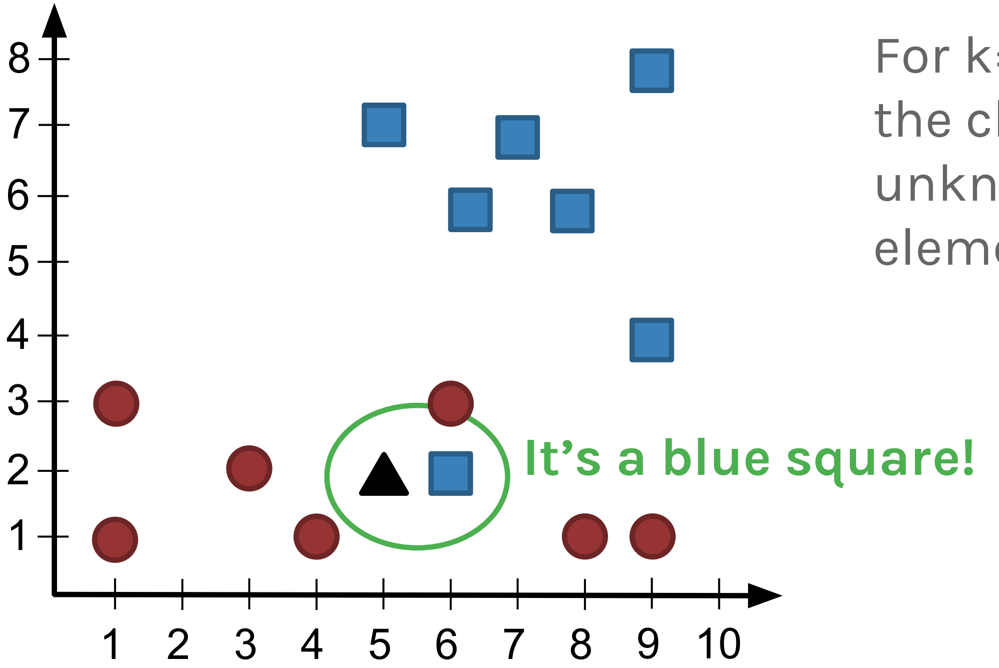
- Supervised learning;
- Classification and regression;
- Lazy learning;
- Uses the *k* closest instances to classify the unknown one;
- Easy to implement.

K-NN Example



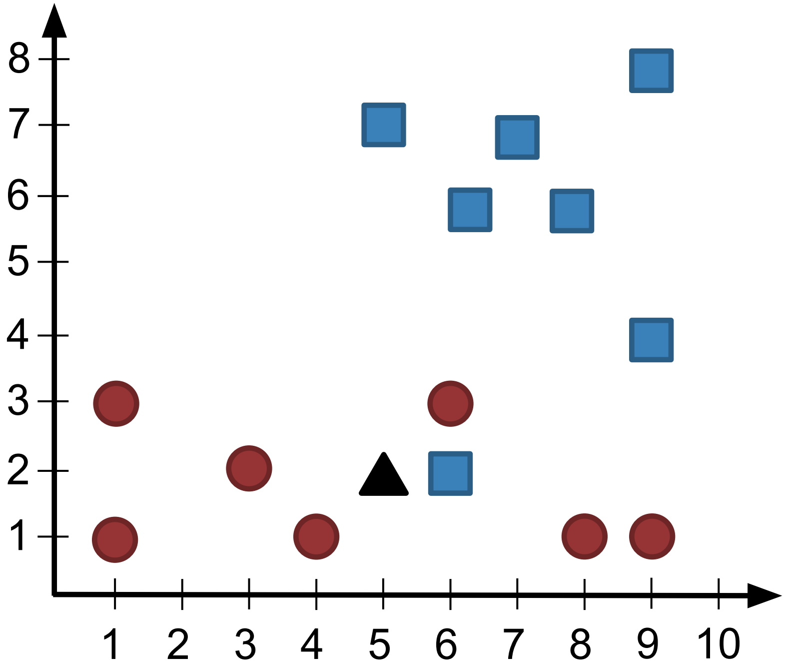
For k=1, what is the class of the unknown element (▲)?

K-NN Example



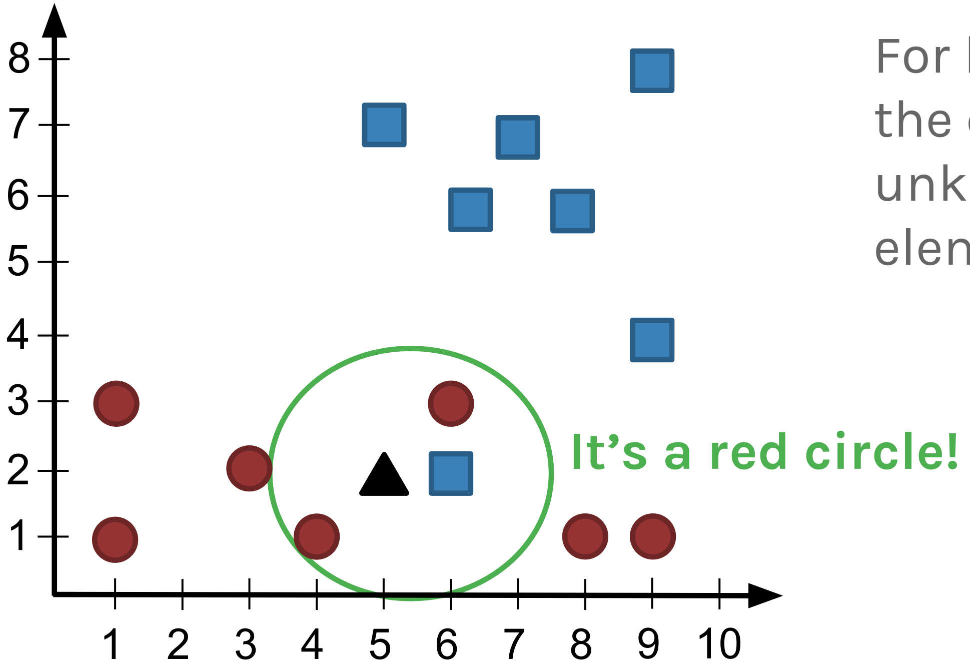
For k=1, what is the class of the unknown element (▲)?

K-NN Example

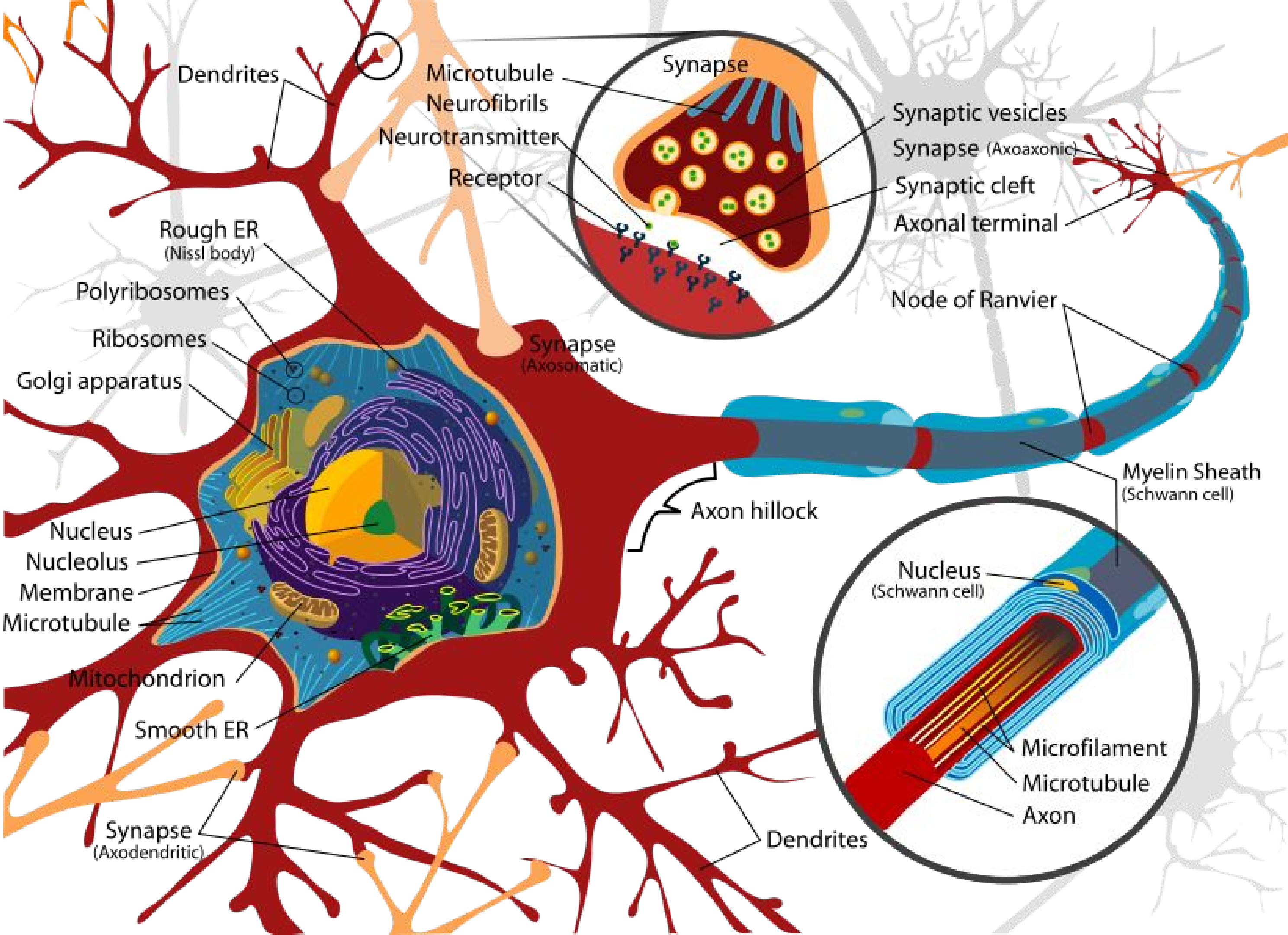


For k=3, what is the class of the unknown element (▲)?

K-NN Example

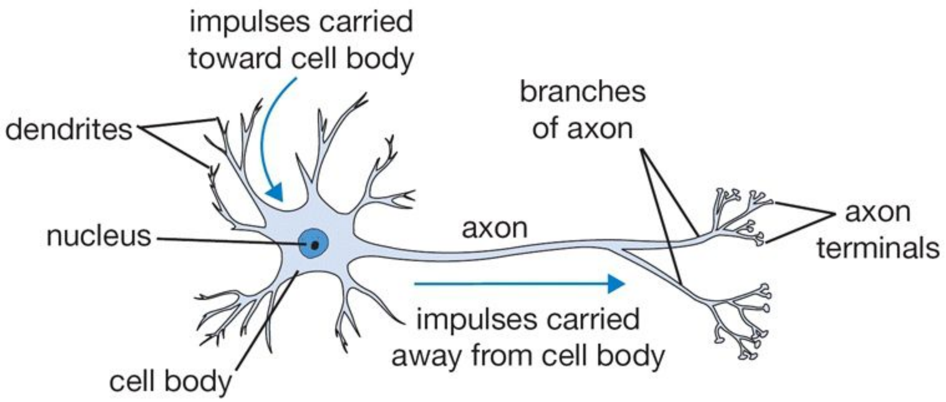


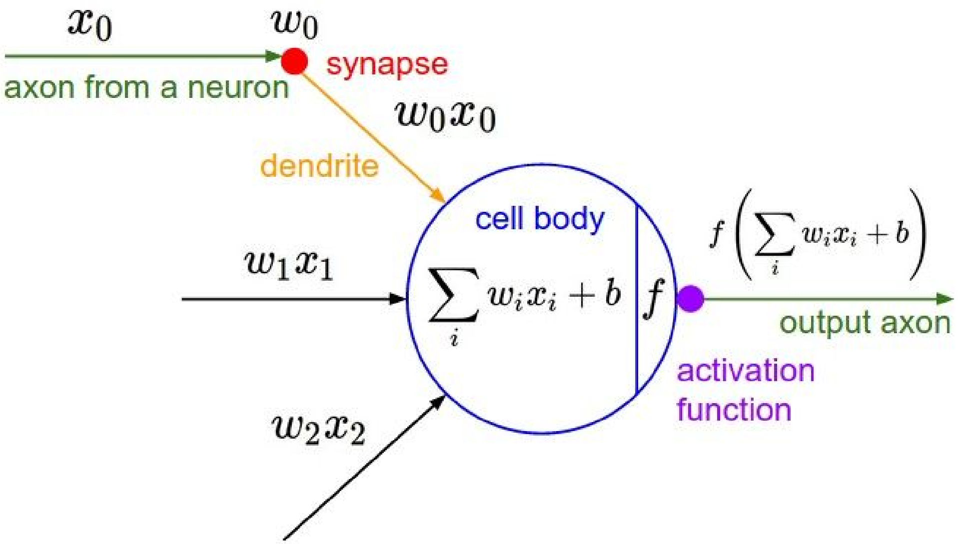
For k=3, what is the class of the unknown element (▲)?



Neural Network

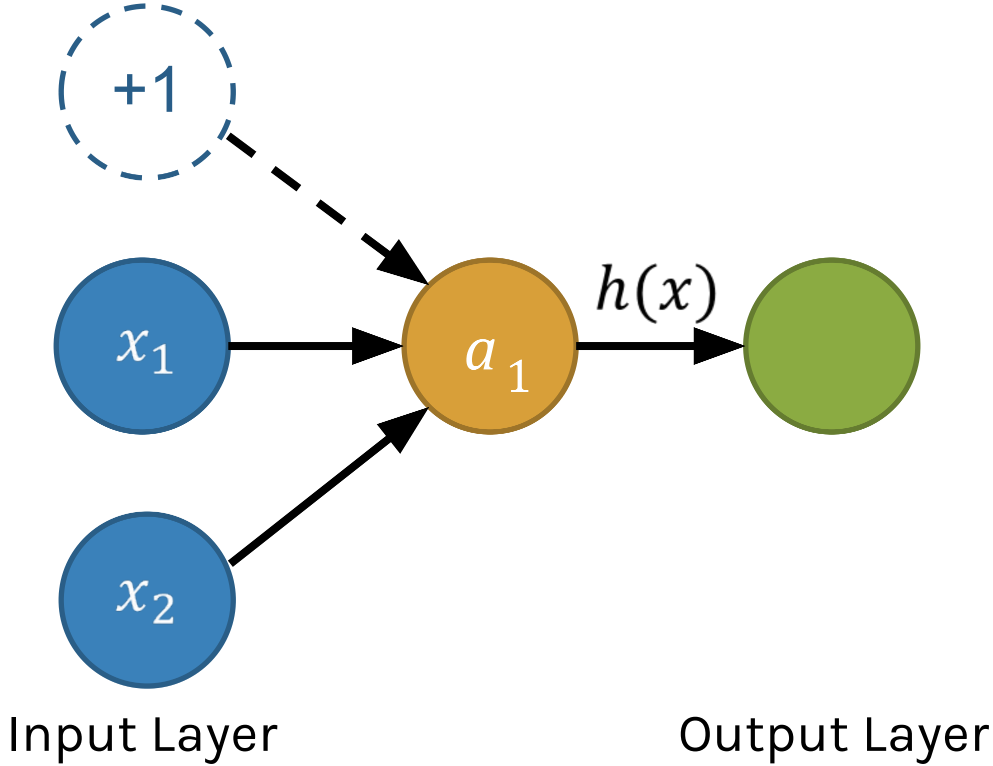
Neural Network



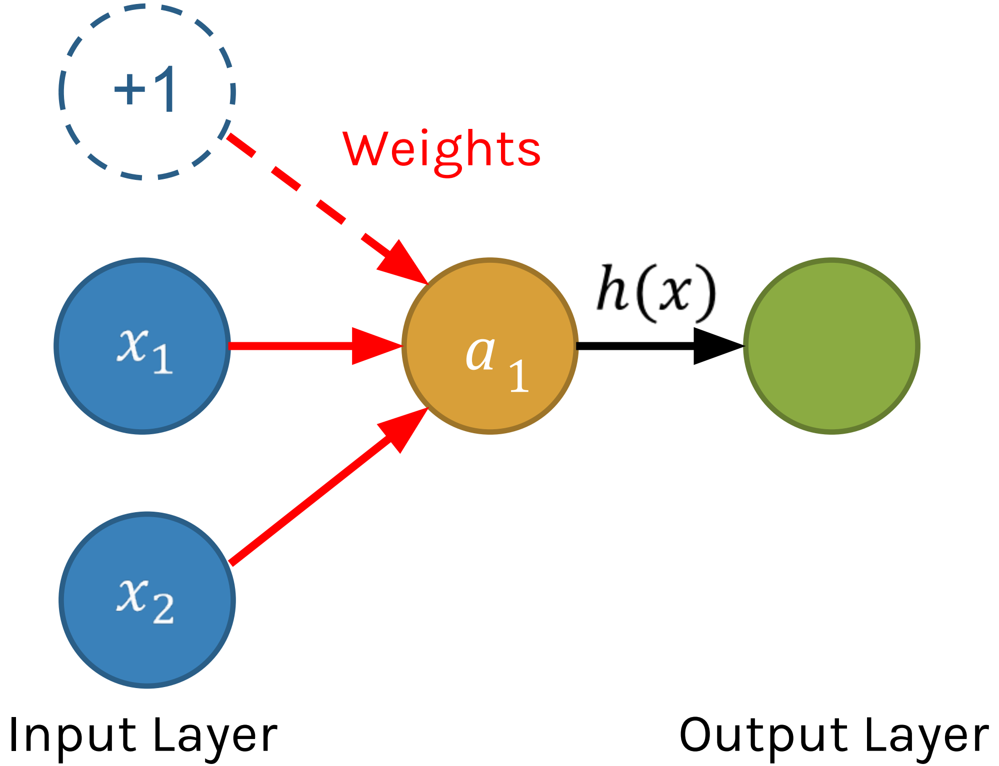


"Convolutional Neural Networks for Visual Recognition" (http://cs231n.github.io/convolutional-networks/)

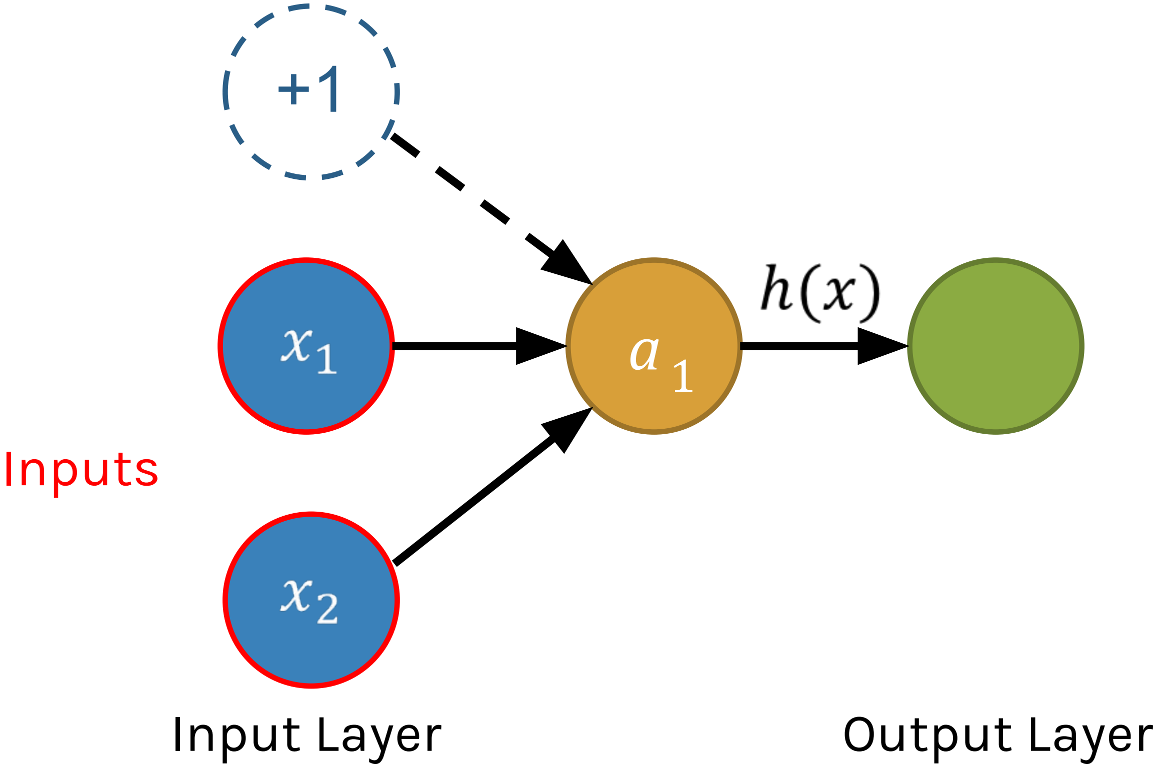
Neuron model



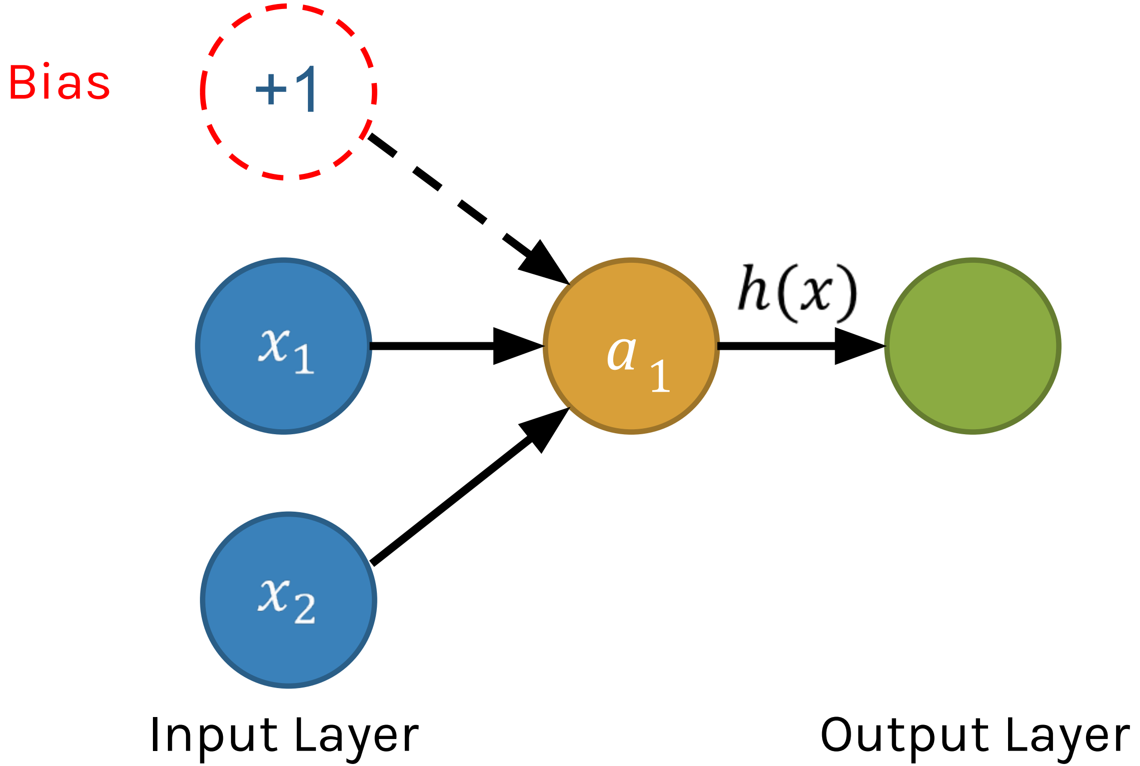
Neuron model



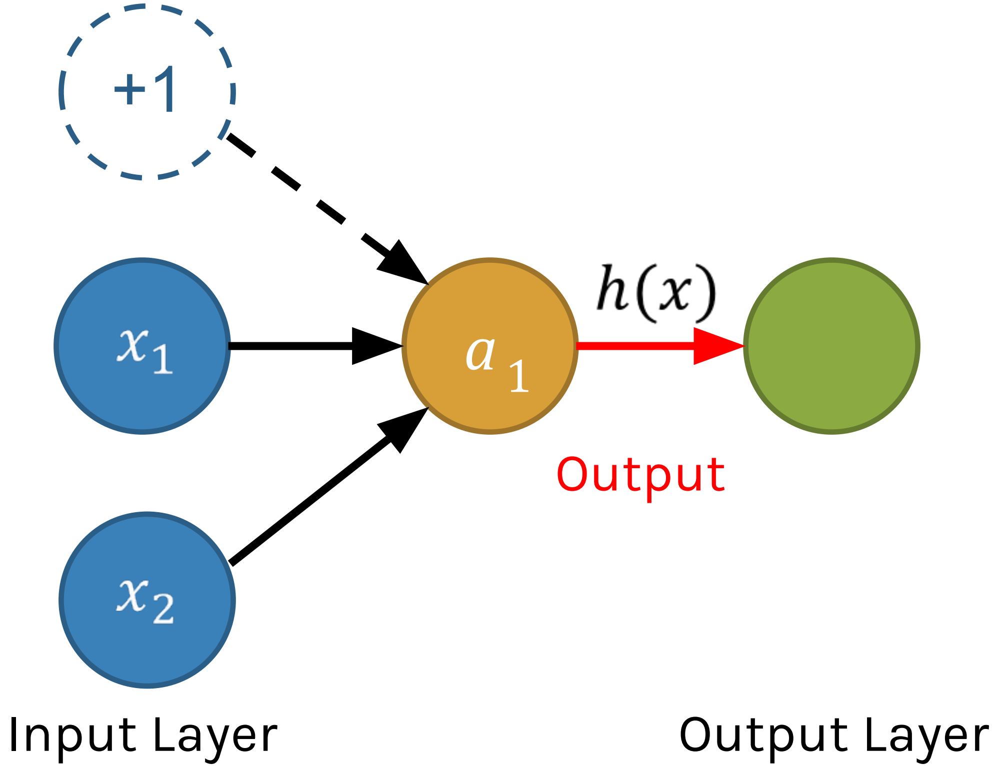
Neuron model



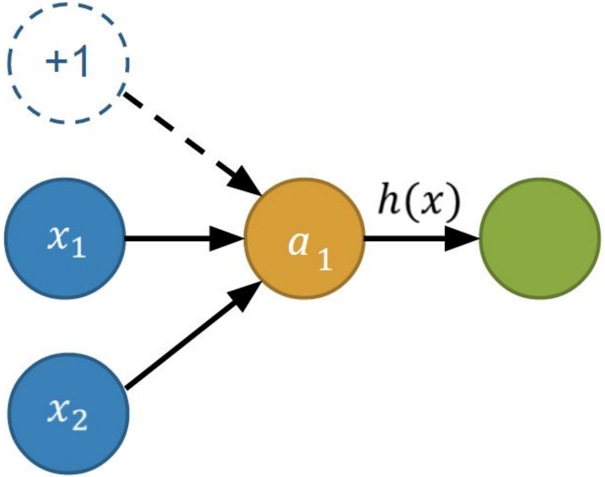
Neuron model



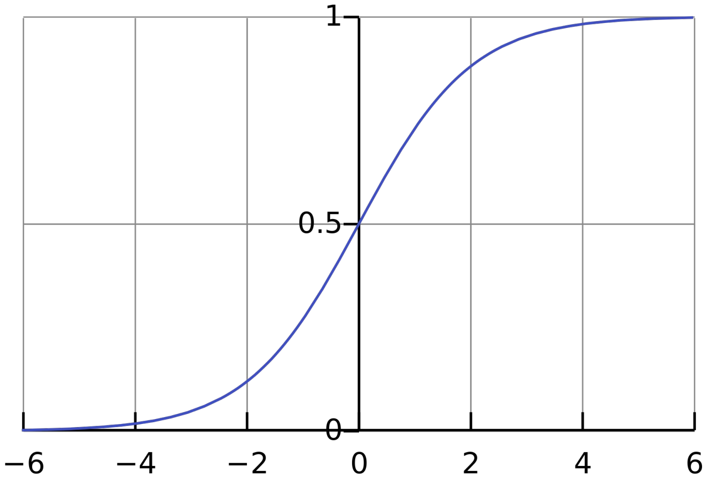
Neuron model



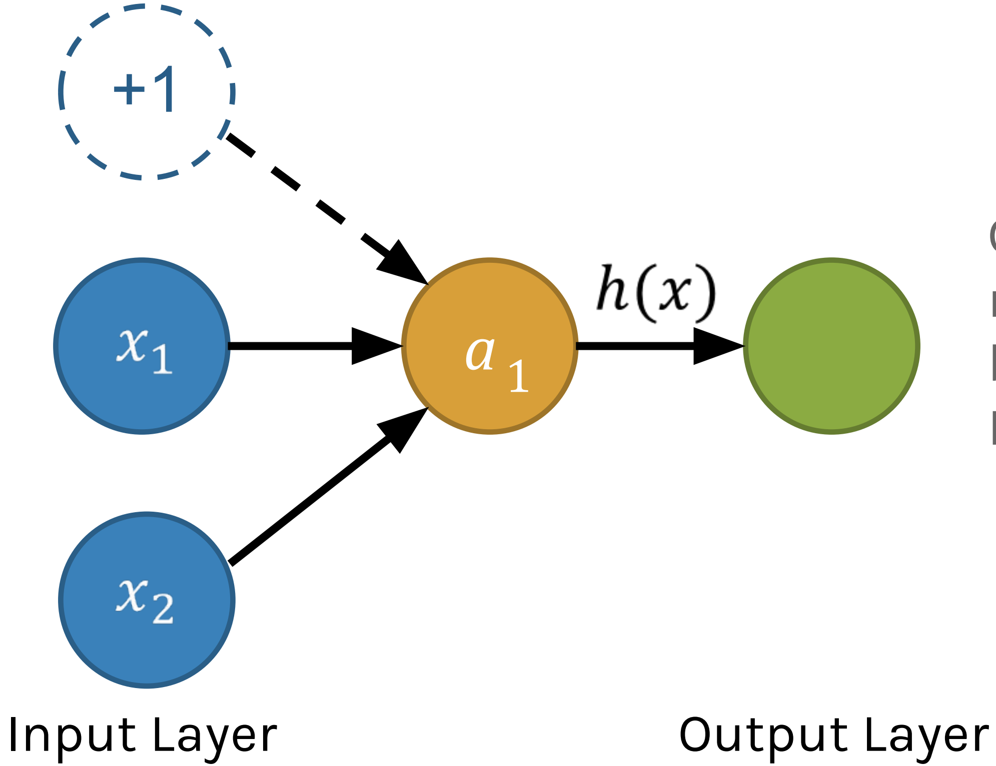
Sigmoid



$$h(x) = \frac{1}{1 + e^{-x}}$$

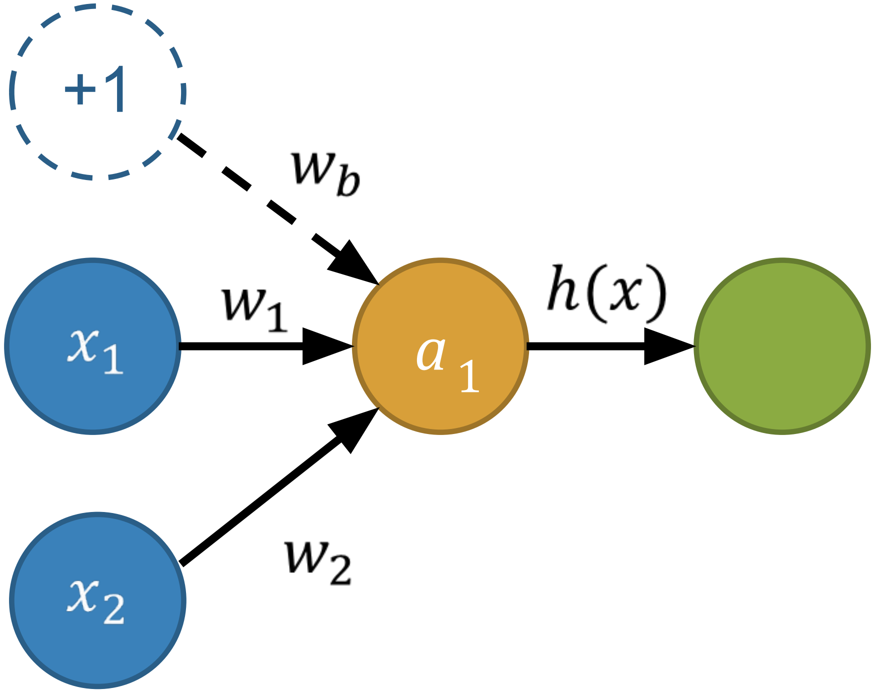


Neural Network Example



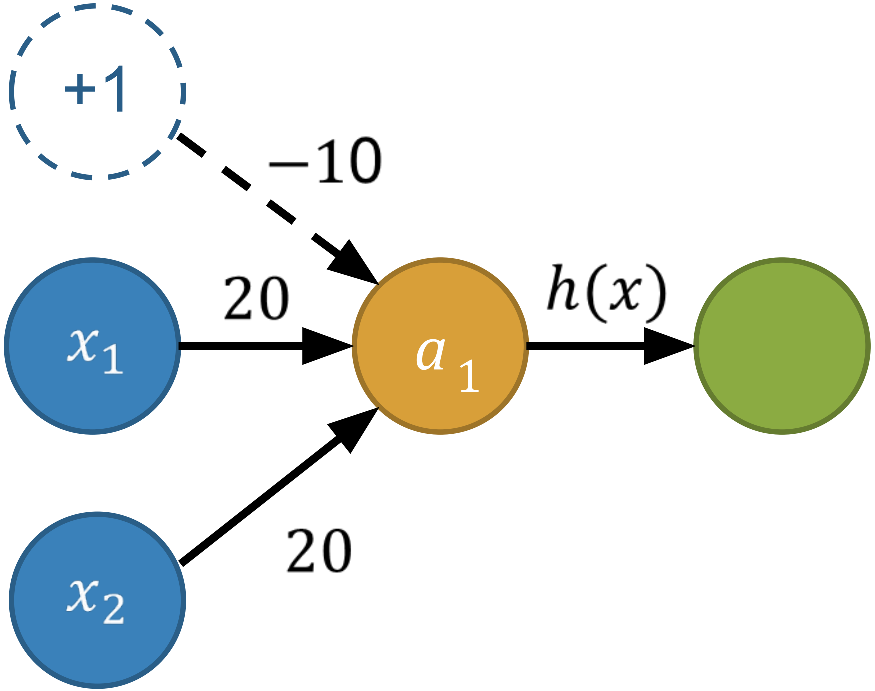
Create a neural network that behaves as an Logic OR.

Neural Network Example



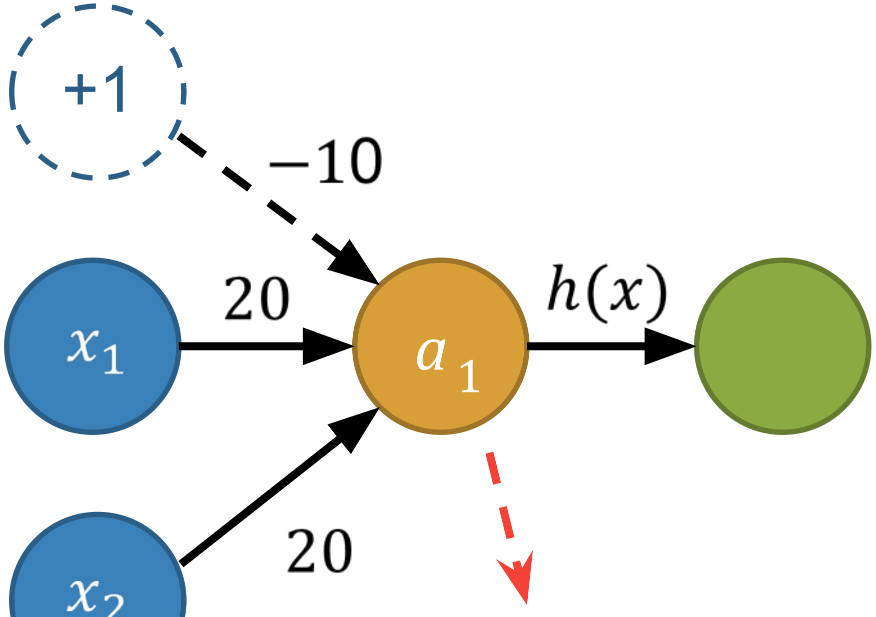
Input		Output
0	0	0
0	1	1
1	0	1
1	1	1

Neural Network Example



Input		Output
0	0	0
0	1	1
1	0	1
1	1	1

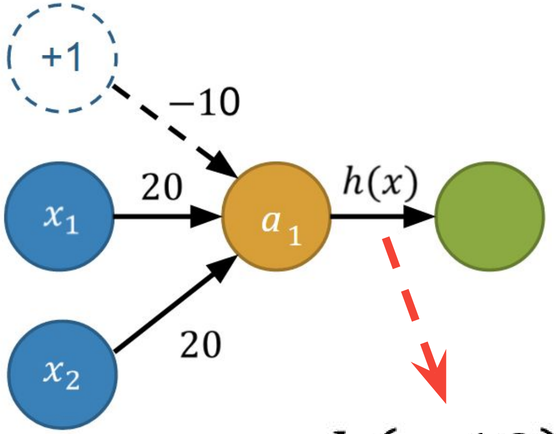
Neural Network Example



Input		Output
0	0	0
0	1	1
1	0	1
1	1	1

$$a_1 = -10*1 + 20*0 + 20*0 = -10$$

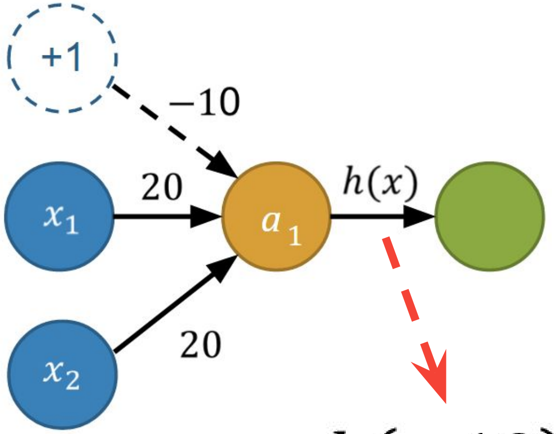
Neural Network Example



$$h(-10) = \frac{1}{1 + e^{-(-10)}} = 0.000045$$

Input		Output
0	0	0
0	1	1
1	0	1
1	1	1

Neural Network Example

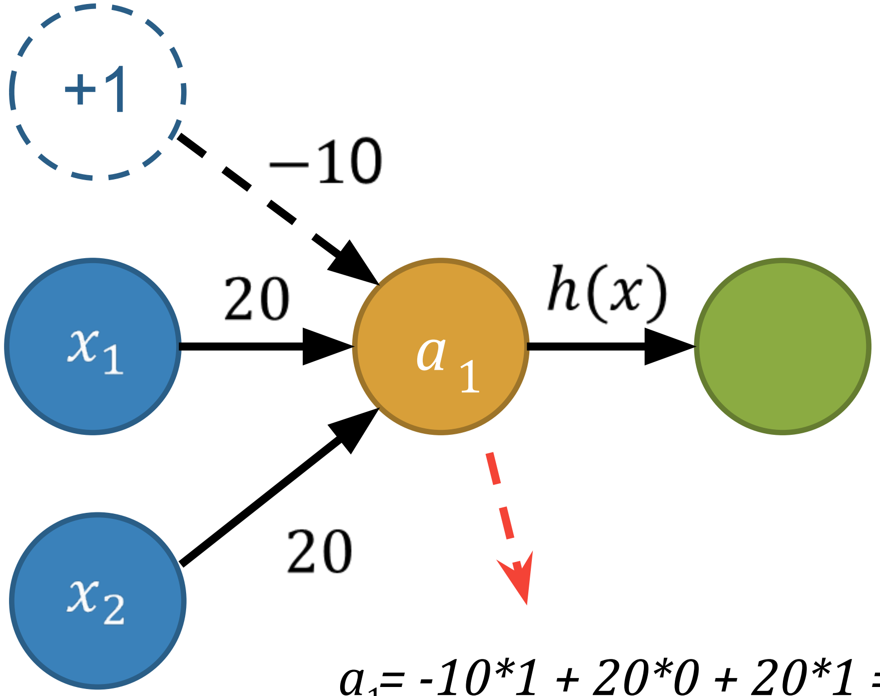


Input		Output
0	0	0
0	1	1
1	0	1
1	1	1

$$h(-10) = \frac{1}{1 + e^{-(-10)}} = 0.000045$$

$$0.000045 < 0.5 = 0$$

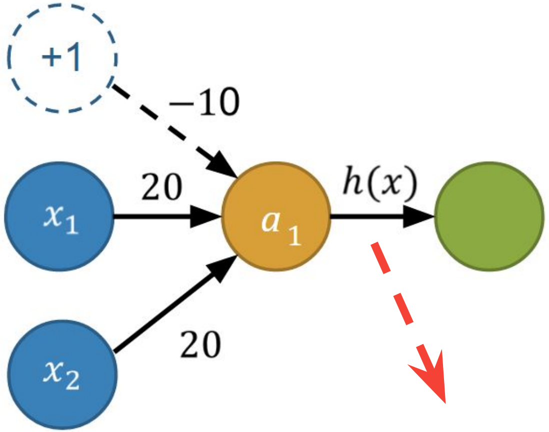
Neural Network Example 2



Input		Output
0	0	0
0	1	1
1	0	1
1	1	1

$a_1 = -10*1 + 20*0 + 20*1 = 10$

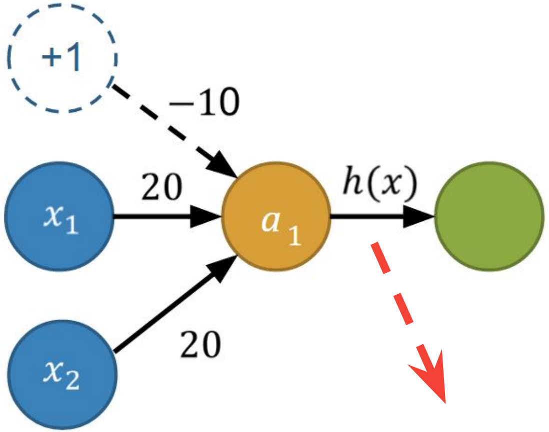
Neural Network Example 2



Input		Output
0	0	0
0	1	1
1	0	1
1	1	1

$$h(10) = \frac{1}{1 + e^{-(10)}} = 0.999955$$

Neural Network Example 2



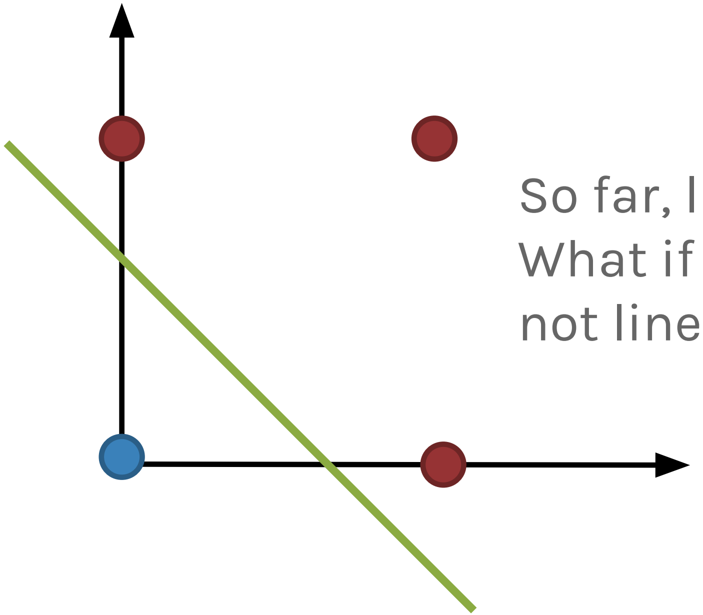
Input		Output
0	0	0
0	1	1
1	0	1
1	1	1

$$h(10) = \frac{1}{1 + e^{-(10)}} = 0.999955$$

$$0.99995 \geq 0.5 = 1$$

Neural Network Example (OR)

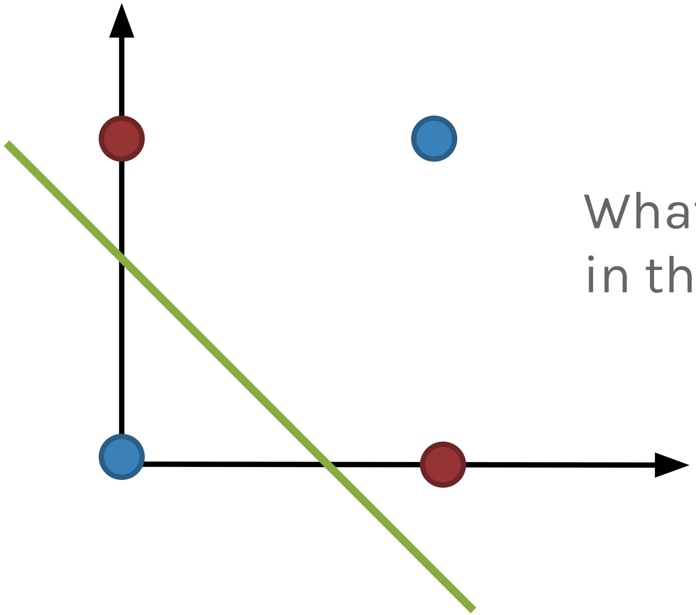
Input		Output
0	0	0
0	1	1
1	0	1
1	1	1



So far, linearly separable.
What if the problem is
not linearly separable?

Neural Network Example 3 (XOR)

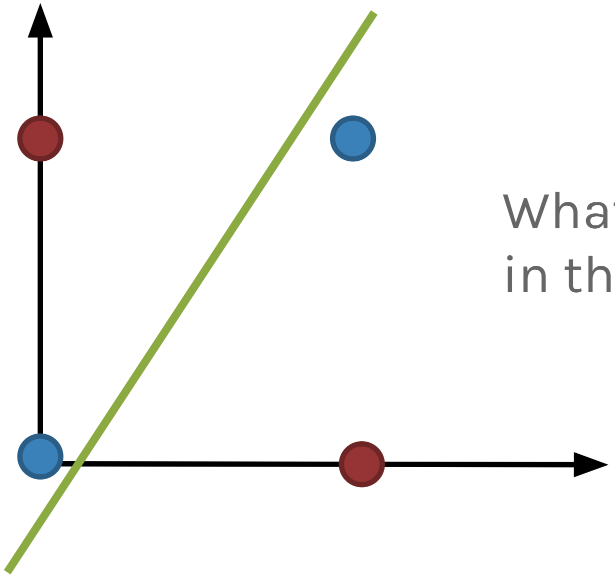
Input		Output
0	0	0
0	1	1
1	0	1
1	1	0



What could we do in this case?

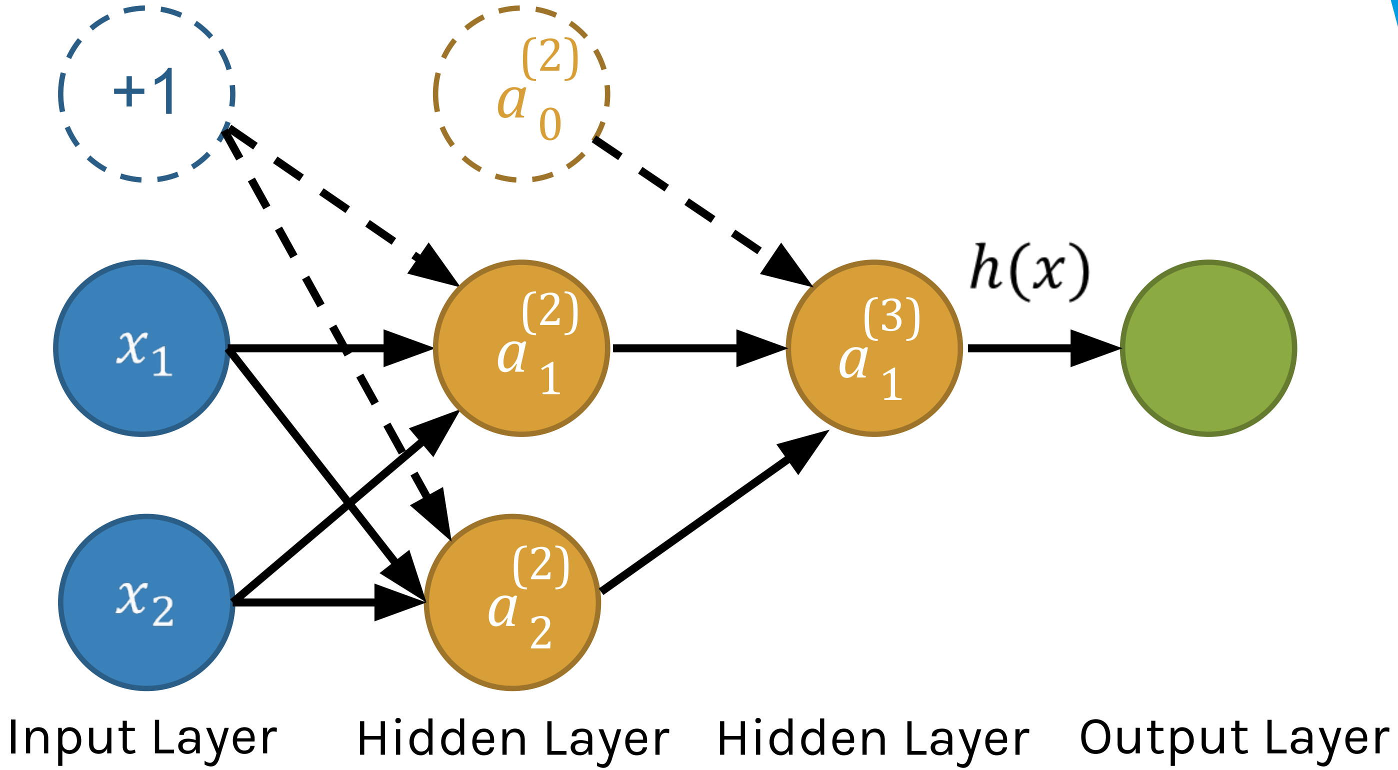
Neural Network Example 3 (XOR)

Input		Output
0	0	0
0	1	1
1	0	1
1	1	0

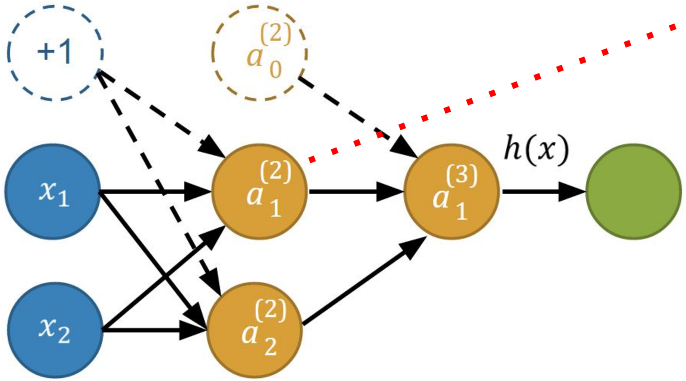


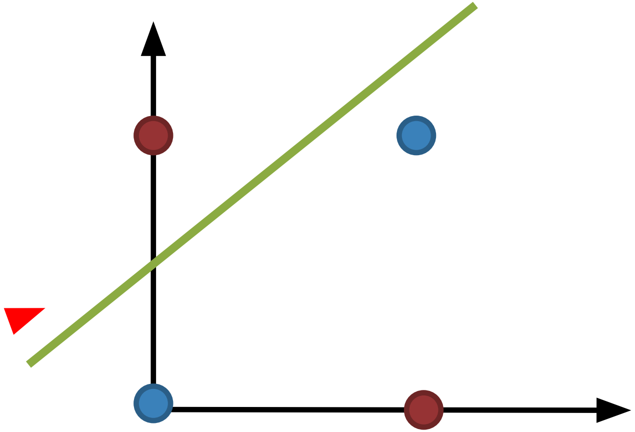
What could we do in this case?

Multi-Layer Perceptron

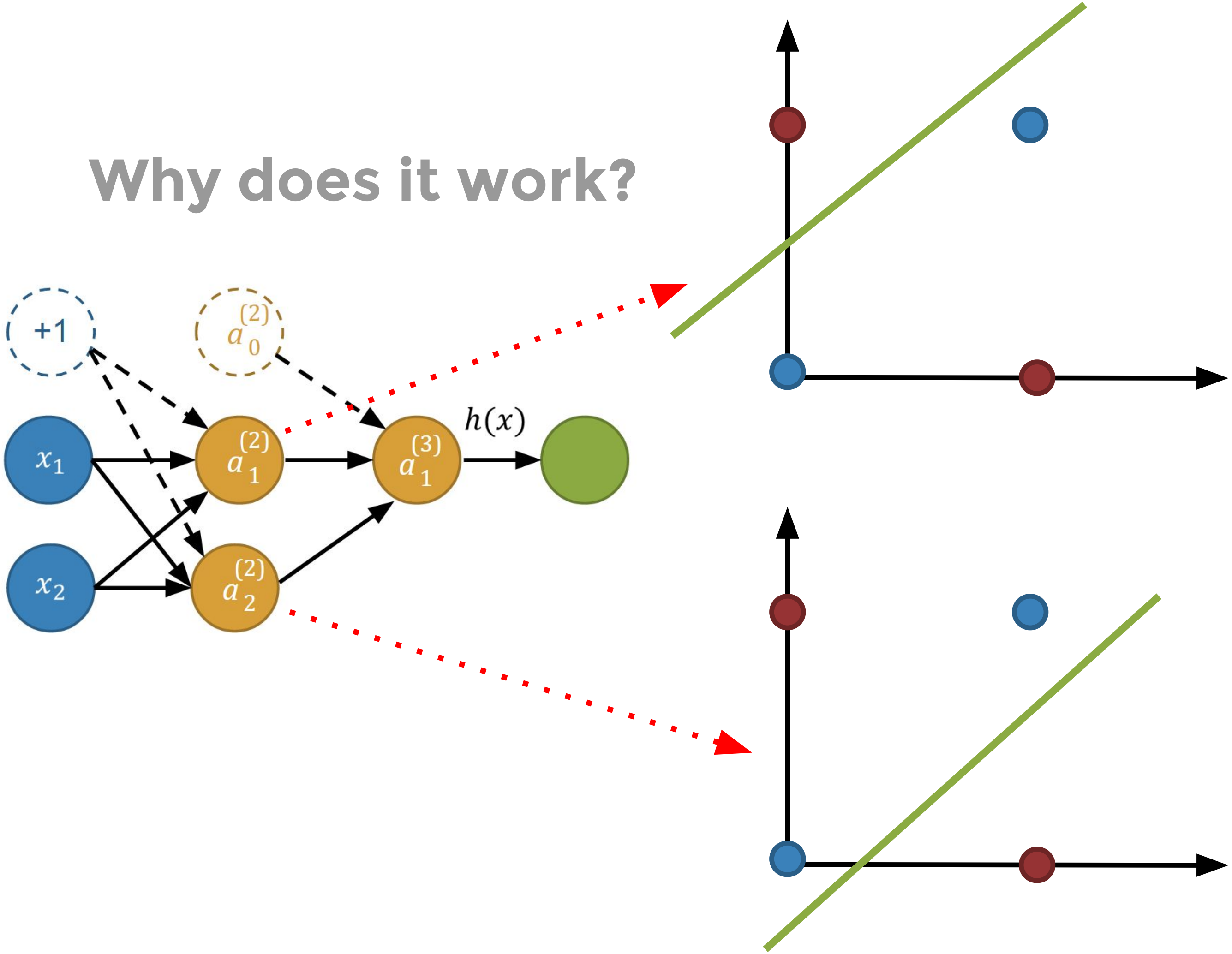


Why does it work?

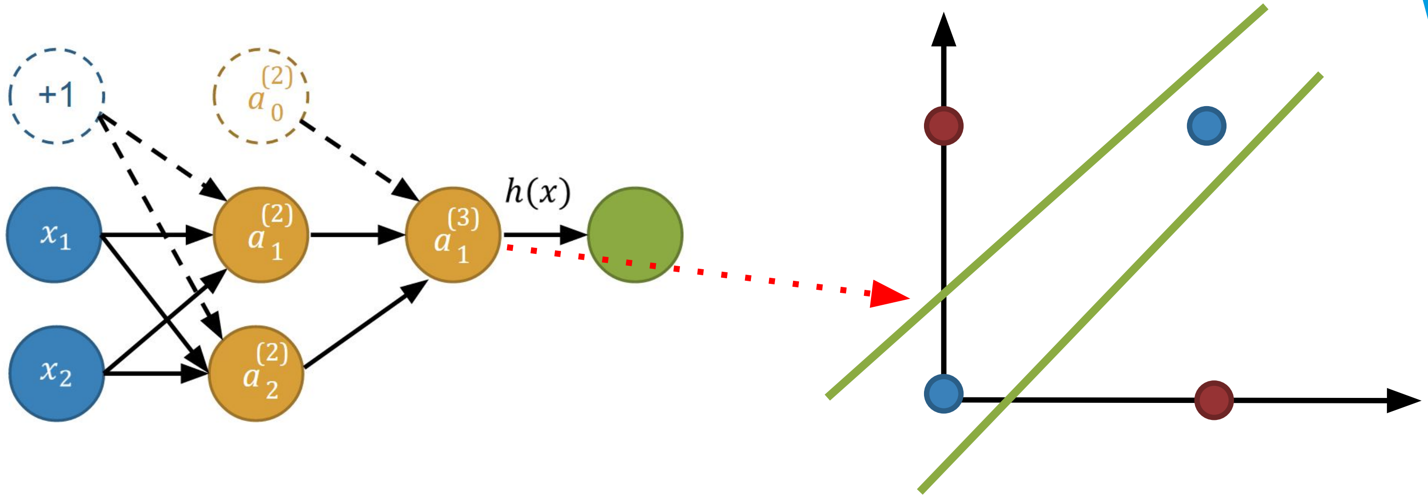




Why does it work?



Why does it work?



So what can a MLP solve?

“

- Can approximate any boolean function, with one hidden layer;
- Can approximate any function with two hidden layers.

Simon Haykin

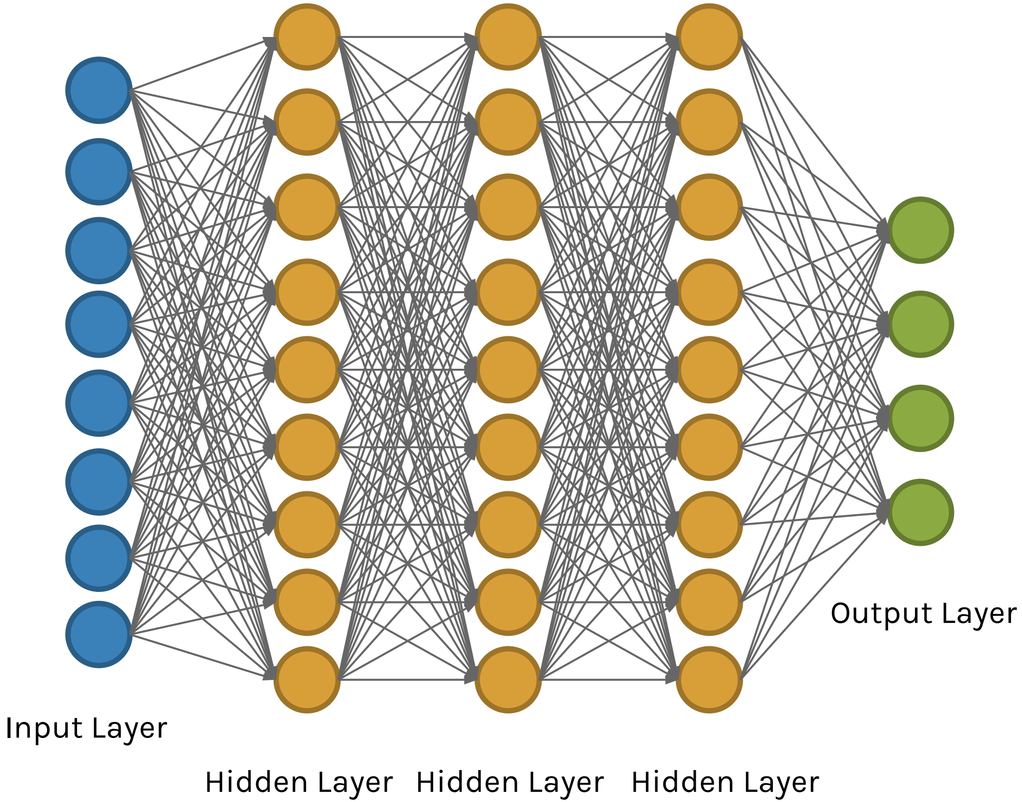
3.

DEEP LEARNING

“

- A machine learning approach that attempts to extract high-level and hierarchical features from raw data.

Deep Neural Network



Where is it used?

- Speech Recognition;
- Music Classification;
- Recommendation;
- Image Classification;
- ...

The “creators”

Yann LeCun

Geoffrey Hinton

Yoshua Bengio

AI Winter
and
the Canadian Mafia

So… why now?

- Better hardware
  - GPUs enable ~9x faster training compared to CPUs.
- High-quality datasets;
- New activation functions;
- Regularization methods.

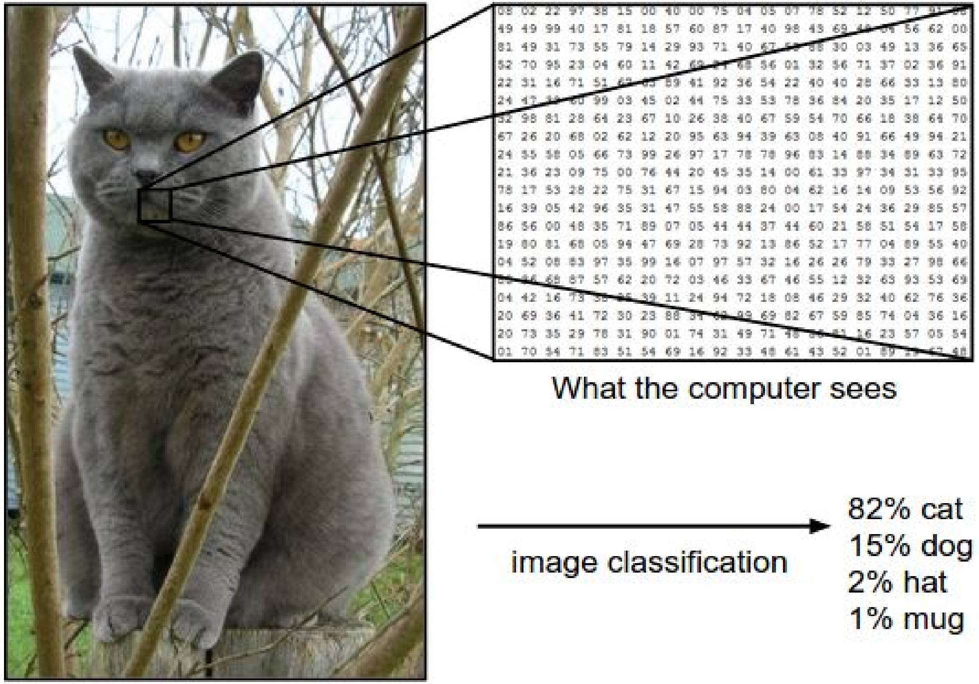
Different Implementations

- Convolutional Neural Networks (CNN);
- Deep Belief Networks;
- Autoencoders;
- Long Short Term Memory (LSTM)
- …

3.1

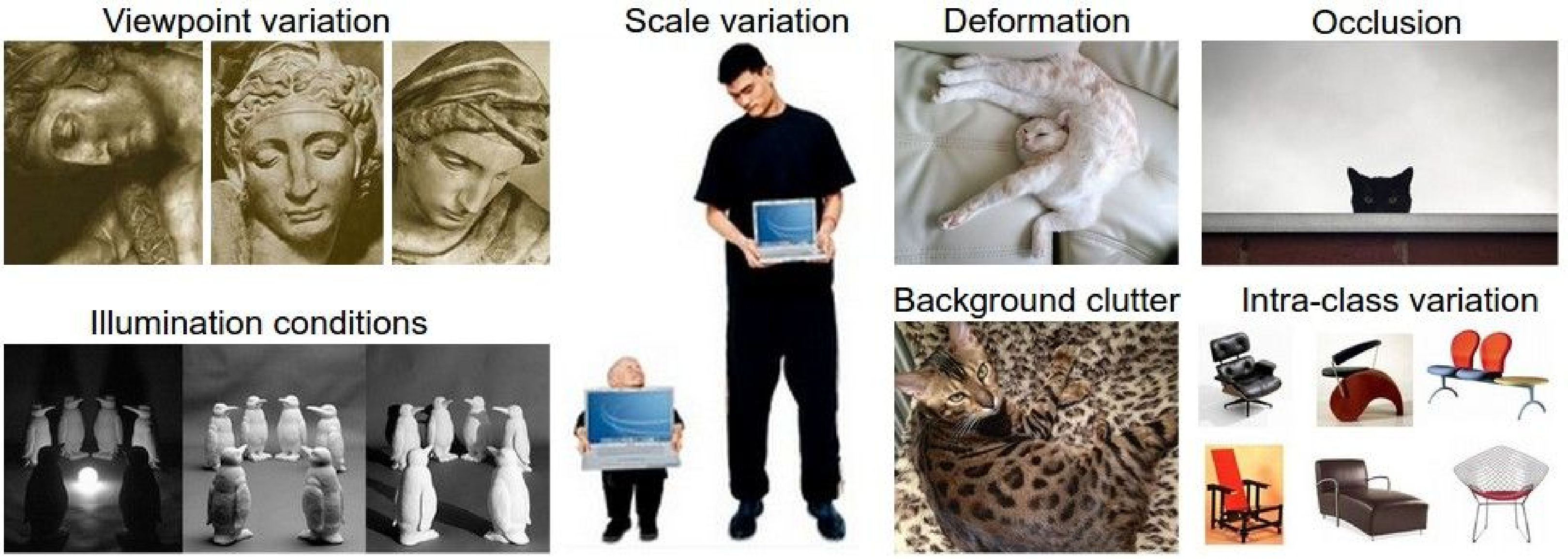
DEEP LEARNING EXAMPLE

Image Classification Problem

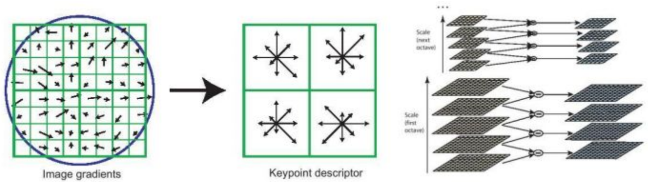


"Convolutional Neural Networks for Visual Recognition" (http://cs231n.github.io/convolutional-networks/)

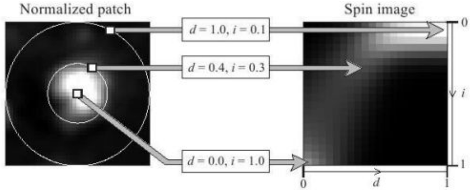
Image Classification - Challenges



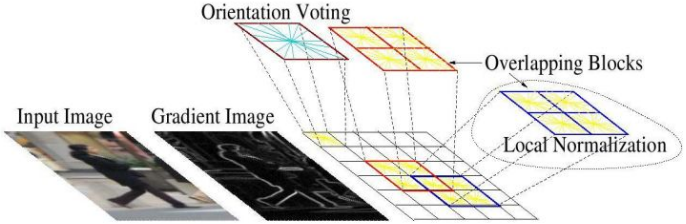
Traditional way



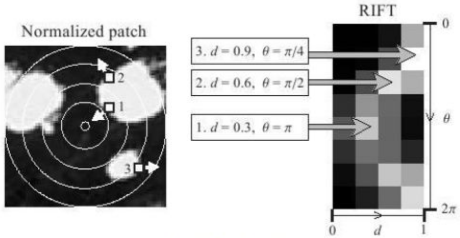
SIFT



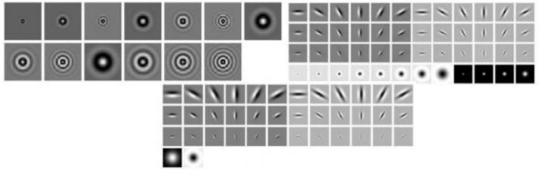
Spin image

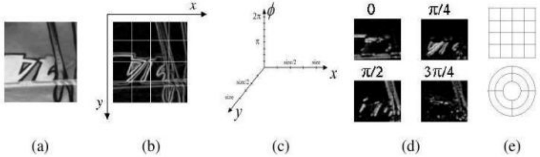


HoG



RIFT


Textons



GLOH

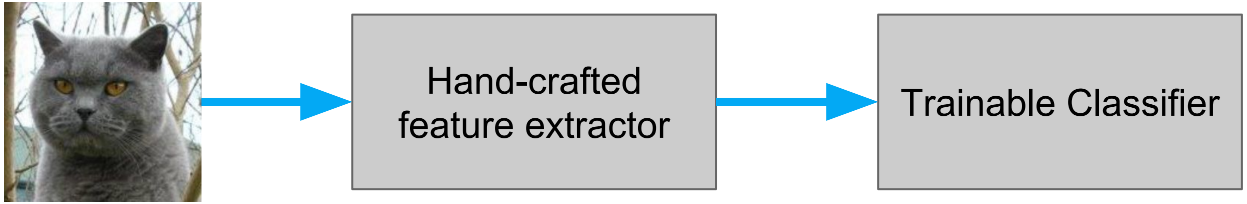
"Deep Learning Methods for Vision" (Honglak Lee, 2012)

Hand-crafted features

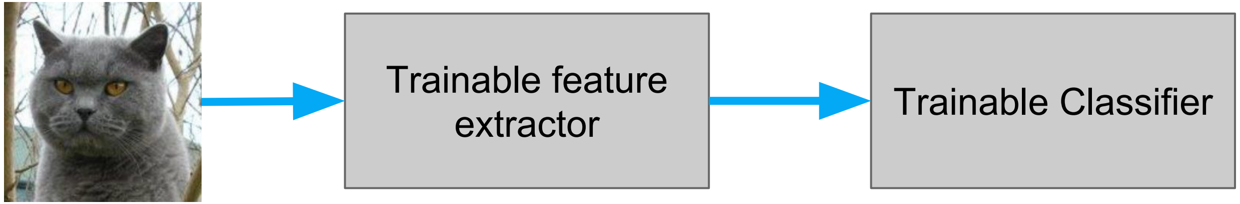
- Time consuming;
- Demand expert knowledge;
- Sometimes meaningful for a very specific scenario.

But what if we could learn feature extractors instead?

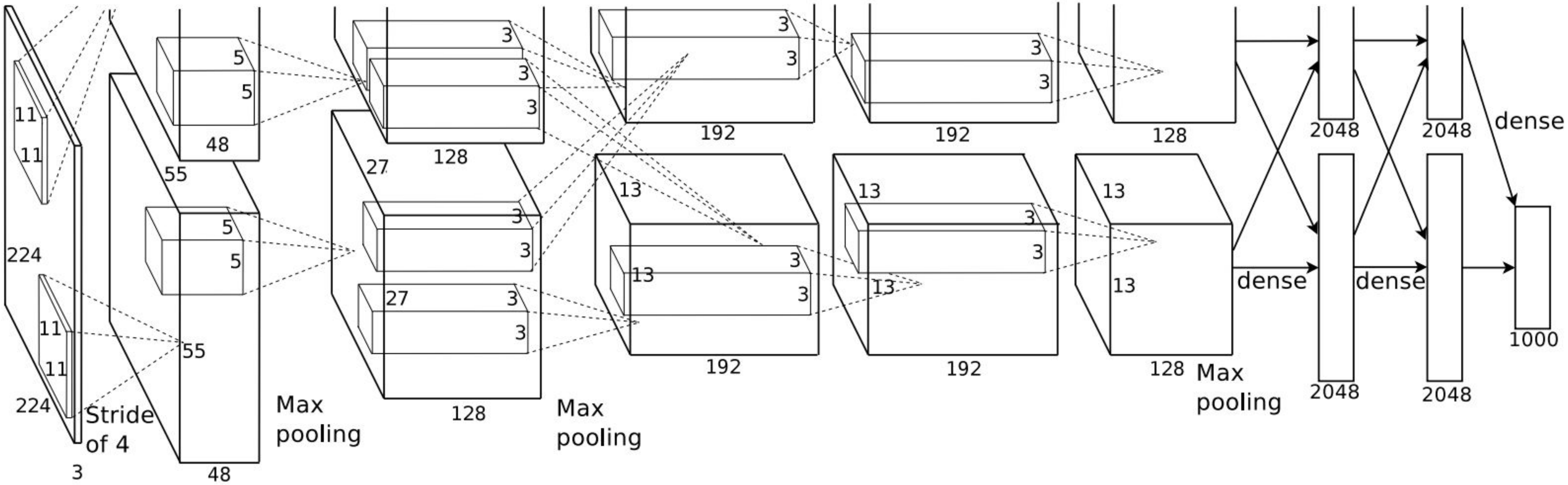
Traditional



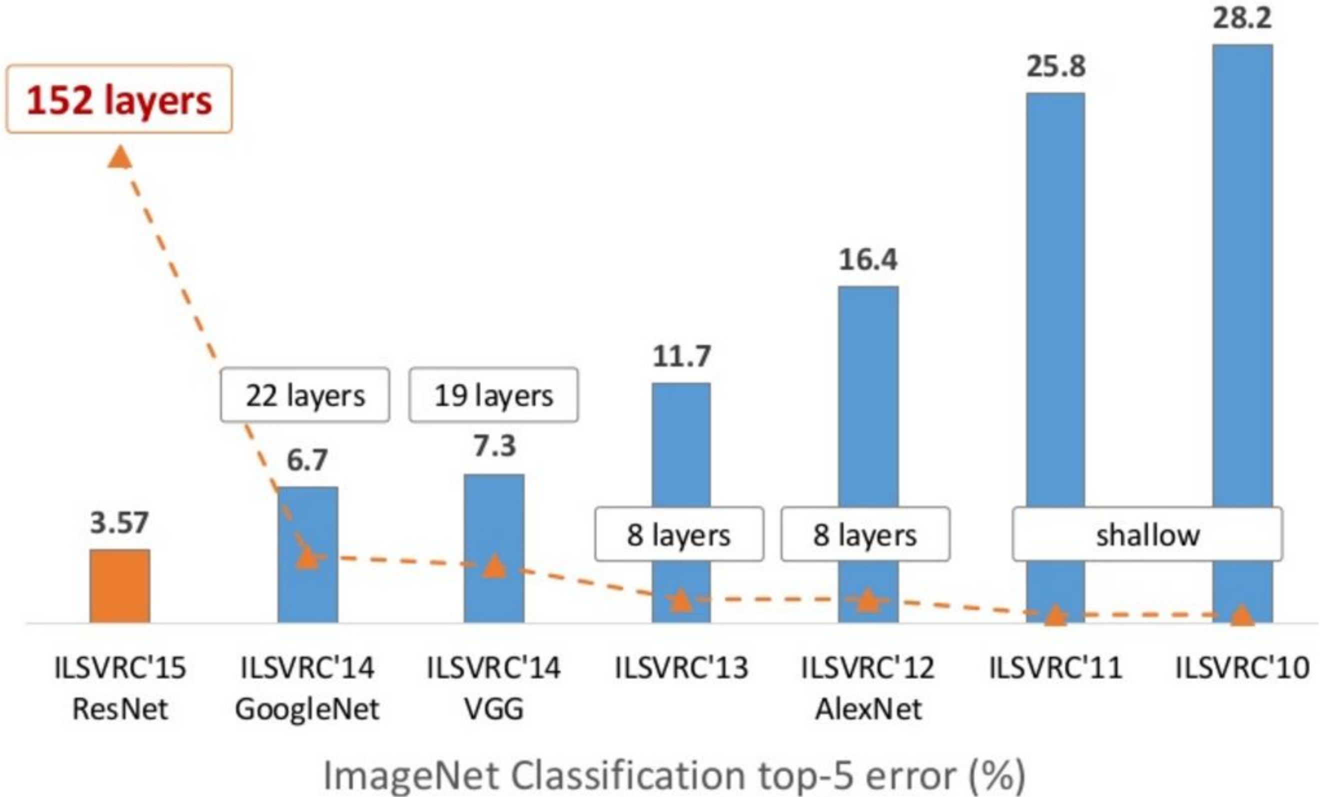
Deep Learning



Architecture Example



Architectures Overview



4.

HOW COULD WE APPLY IT AT OUR WORK ?

But my work
is not about
Machine Learning...

Steps

- Start with the problem
- Understand your data
- Try simple hypotheses first
- Evaluating and presenting results

Start with the problem

Talk to your business contact

Get **insights** about your problem domain

Ask about hypothetical **features**

Be clear about your **intentions**

Do not commit with results!

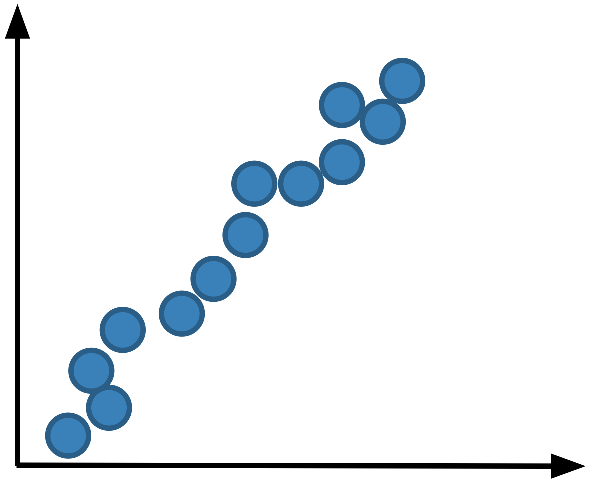
Start with the problem

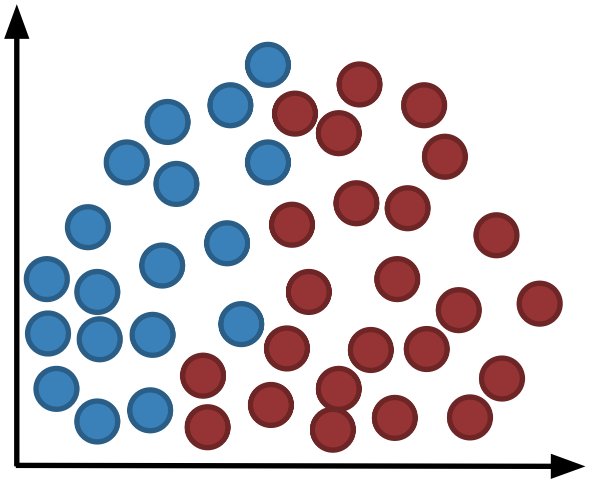
Where is the data ?
- Application logs
- User inputs

Search things like:
- Repetitive tasks by user
- Customer habits
- Patterns of system usage by the user

Understand your data

- Explore your data
- Plot your data
- Start with 2 or 3 attributes
- Try to spot some pattern





Understand your data

- Explore your data
- Plot your data
- Start with 2 or 3 attributes
- Try to spot some pattern

1 tablespoon extra-virgin olive oil
1 tablespoon chopped garlic
1/2 teaspoon kosher salt

Heat 1 tablespoon butter and olive oil together **in a large skillet over medium heat;**
cook and stir zucchini noodles (zoodles),

Try simple hypotheses first

- Start with simple hypotheses
  - Naive Bayes
  - Linear regression
  - Logistic regression
  - K-means
  - KNN
- Try more complex hypotheses later
  - SVM
  - Random Forest
  - Neural Networks

Evaluating and presenting results

- Define a metric based on the business goal
    - Error rate
    - F1 score
- Demonstrate results to business unit
- Use functional prototypes when possible

Market Analysis Case

Examples

Market Analysis Case

- Optimize subscription plans
- Find customer groups
- **Use and analyze the application log**

application logs:
```
[2015-01-05 09:12:34] user id=543 clicked buy button for app id=12
[2015-01-06 08:10:31] user id=143 clicked buy button for book id=786
[2015-01-06 09:22:06] user id=563 clicked buy button for music id=900
[2015-01-06 13:12:34] user id=543 clicked buy button for music id=34
[2015-01-07 15:45:22] user id=890 clicked more details button for movie id=189
[2015-01-07 17:21:14] user id=897 clicked more details button for app id=231
[2015-01-08 08:22:02] user id=239 clicked buy button for book id=786
[2015-01-09 09:00:11] user id=253 clicked buy button for music id=373
[2015-01-09 17:12:56] user id=389 clicked more details button for app id=262
[2015-01-09 20:12:04] user id=141 clicked buy button for app id=145
```

Market Analysis Case

- Optimize subscription plans
- Find customer groups
- **Use and analyze the application log**

application logs:
[2015-01-05 09:12:34] **user id=543** clicked buy button for **app id=12**
[2015-01-06 08:10:31] **user id=143** clicked buy button for **book id=786**
[2015-01-06 09:22:06] **user id=563** clicked buy button for **music id=900**
[2015-01-06 13:12:34] **user id=543** clicked buy button for **music id=34**
[2015-01-07 15:45:22] user id=890 clicked more details button for movie id=189
[2015-01-07 17:21:14] user id=897 clicked more details button for app id=231
[2015-01-08 08:22:02] **user id=239** clicked buy button for **book id=786**
[2015-01-09 09:00:11] **user id=253** clicked buy button for **music id=373**
[2015-01-09 17:12:56] user id=389 clicked more details button for app id=262
[2015-01-09 20:12:04] **user id=141** clicked buy button for **app id=145**

Market Analysis Case

- Clustering algorithm can be used
- Present results back to business unit

Group 1: App Buyers
+- 75% Applications
+- 15% Movies
+- 5% Music
+- 5% Books

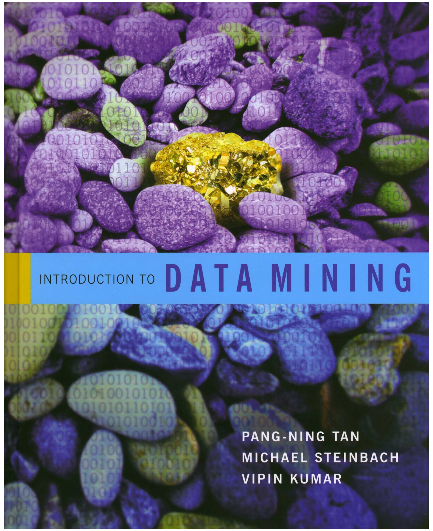
Group 2: Media Users
+- 10% Applications
+- 40% Movies
+- 40% Music
+- 10% Books

Group 3: Book Readers
+- 5% Applications
+- 2.5% Movies
+- 7.5% Music
+- 90% Books

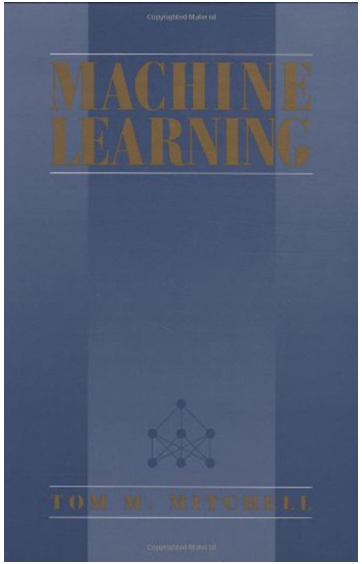
5.

BOOKS & COURSES

Books



Introduction to Data Mining
Pang-Ning Tan



Machine Learning
Tom Mitchell

Machine Learning

About this course: All learners in Stanford University's Machine Learning course on Coursera will now have the option to purchase a Course Certificate for Machine Learning until May 31, 2016.

May 31st is the last date for purchasing a certificate, after which the course will be offered without a

More

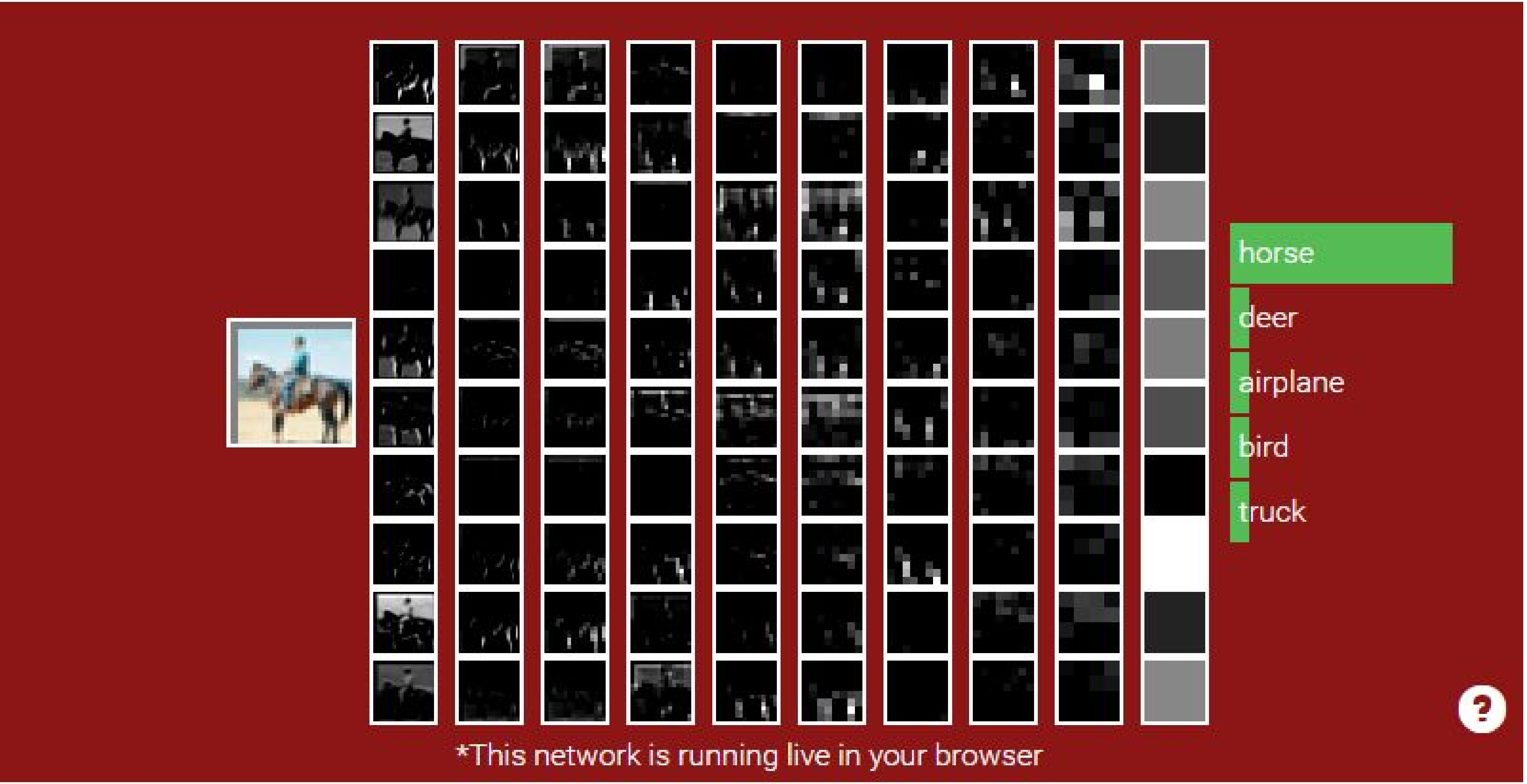
Created By:

Stanford

Taught By: Andrew Ng, Associate Professor, Stanford University; Chief Scientist, Baidu; Chairman and Co-founder, Coursera

MOOCs

CS231n: Convolutional Neural Networks for Visual Recognition



Other

THANK YOU!

Any questions?