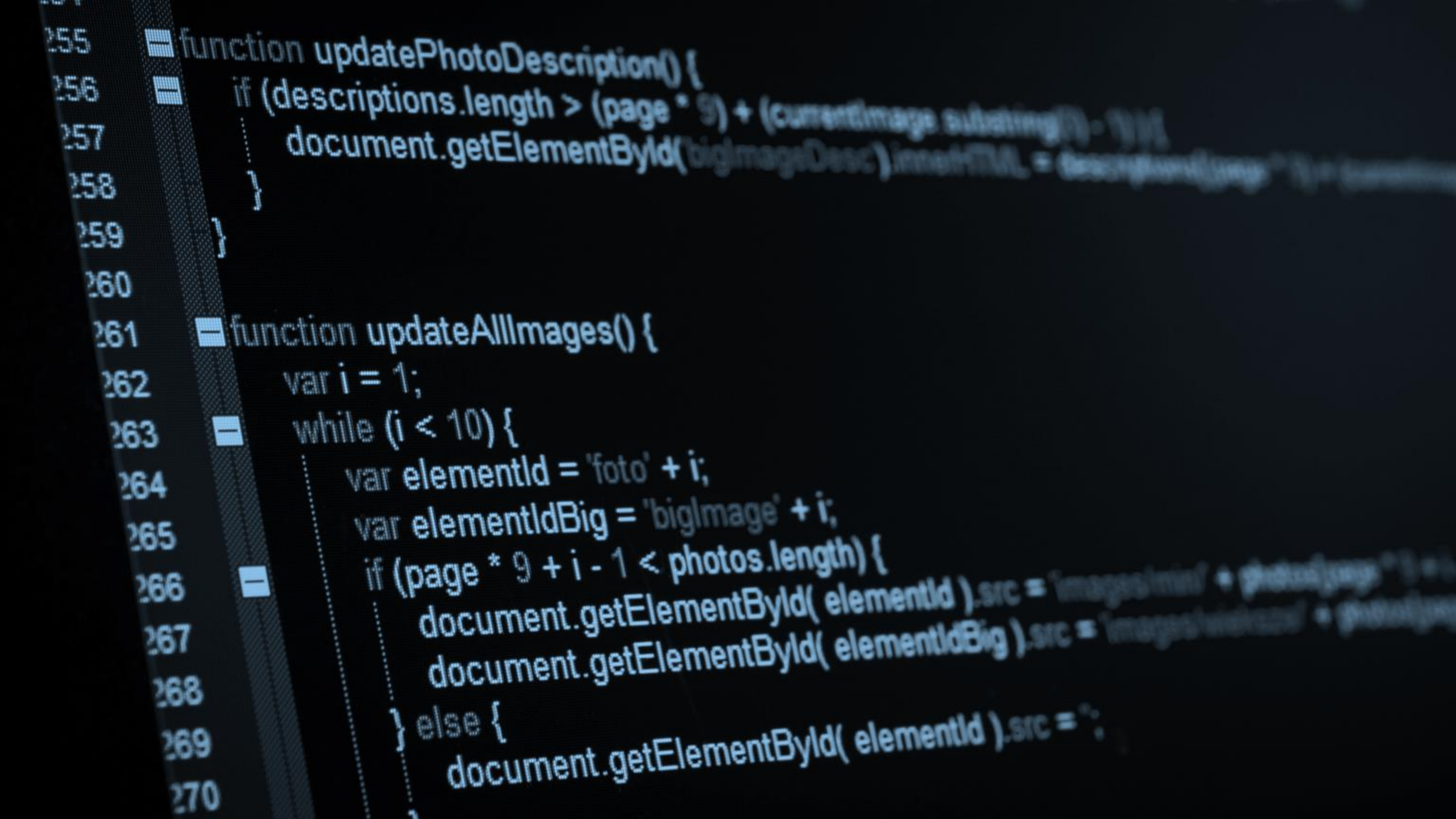
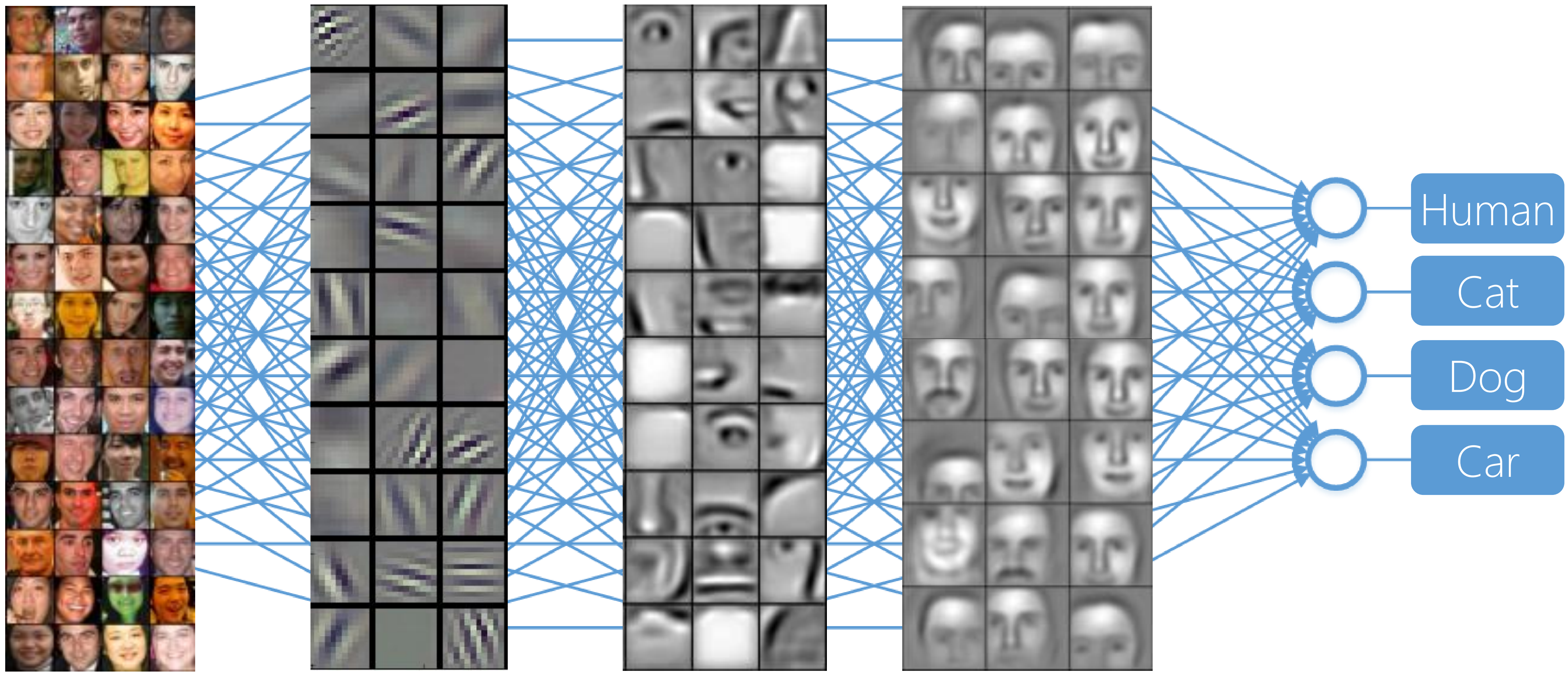
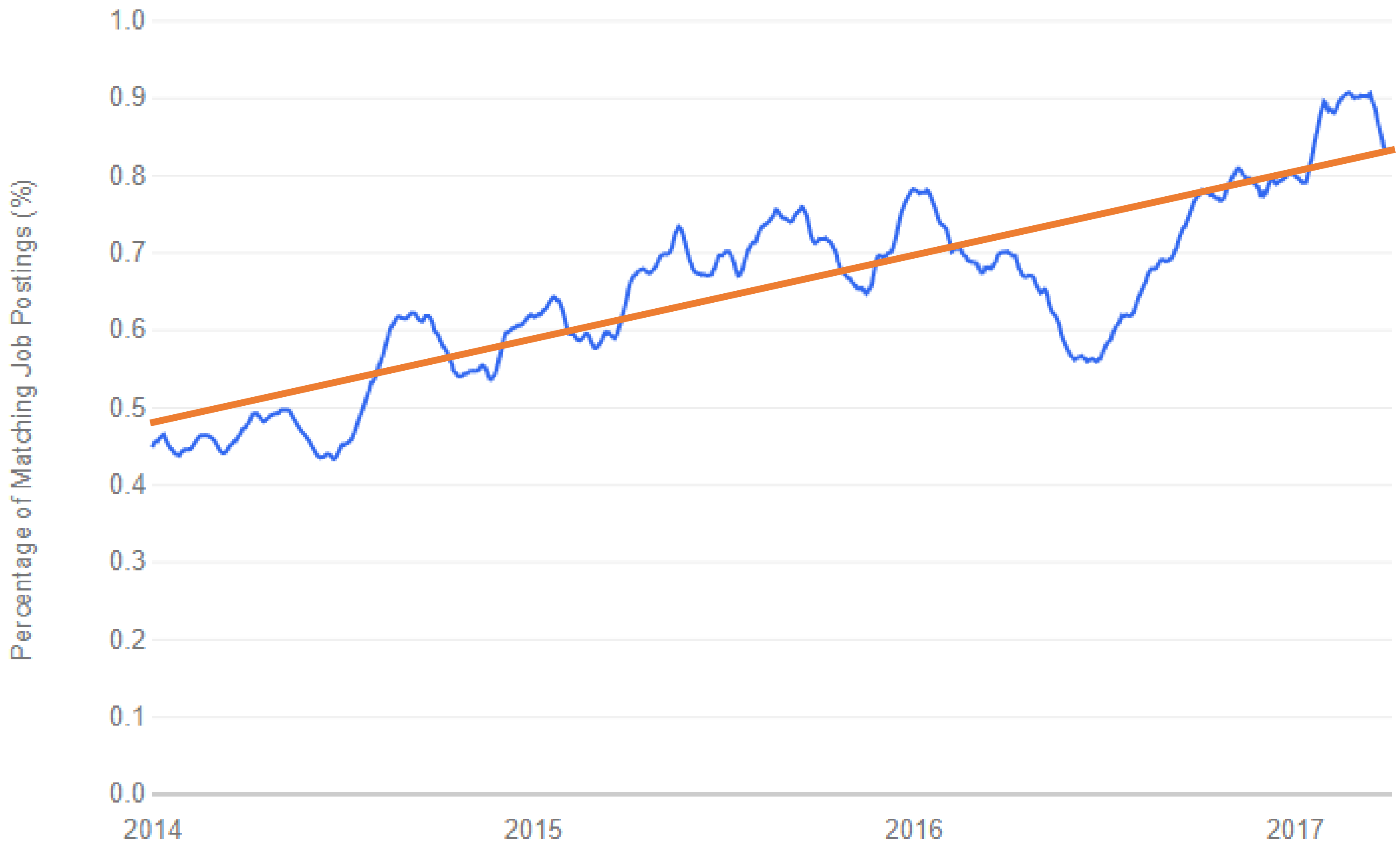
Practical Machine Learning with R

@MatthewRenze


function updatePhotoDescription() {
if (descriptions.length > (page * 9) + (currentImage.substring(
document.getElementById(
}
}
function updateAllImages() {
var i = 1;
while (i < 10) {
var elementId = 'foto' + i;
var elementIdBig = 'bigImage' + i;
if (page * 9 + i - 1 < photos.length) {
document.getElementById(elementId).src =
document.getElementById(elementIdBig).src =
} else {
document.getElementById(elementId).src = '';

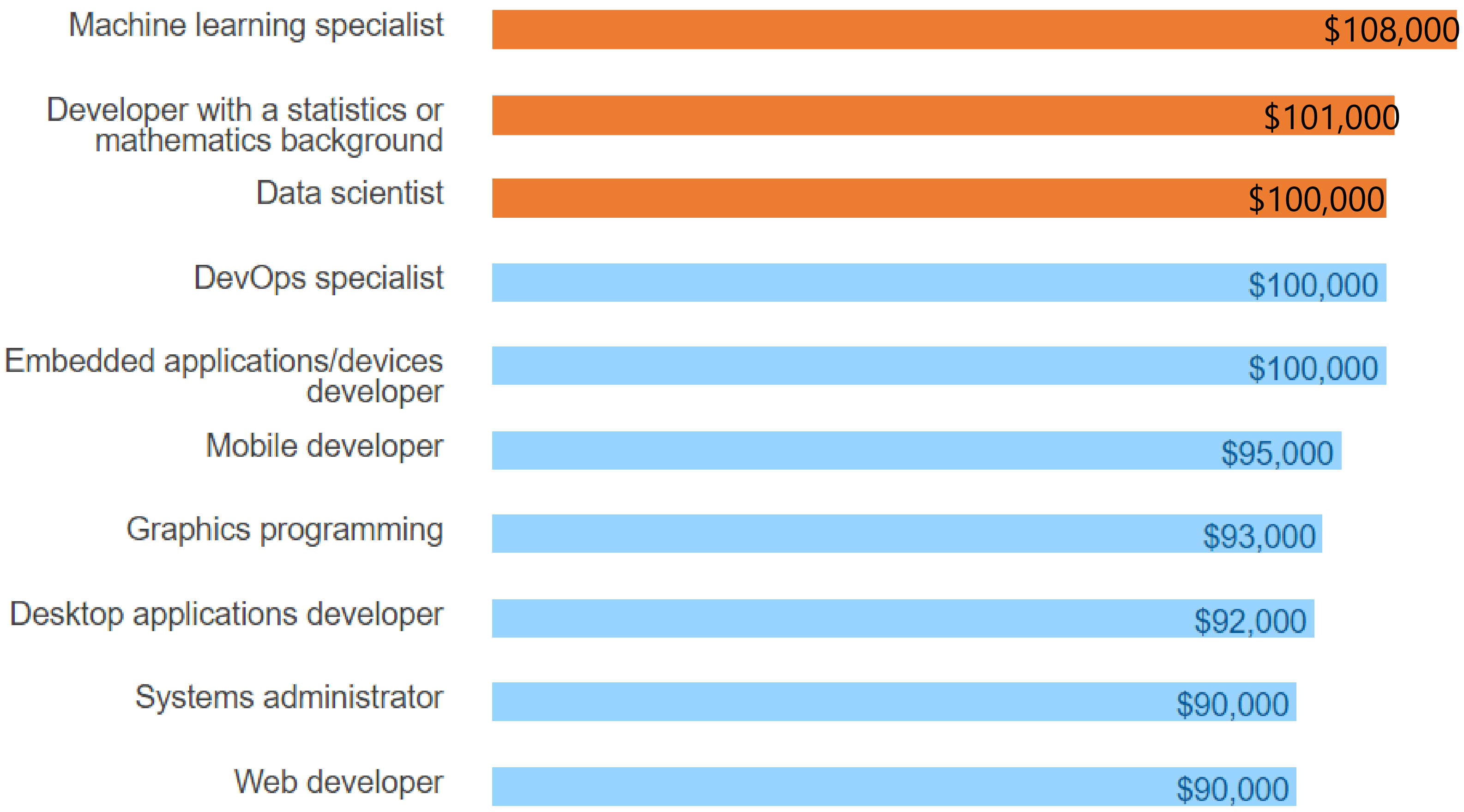

Human
Cat
Dog
Car

Job Postings for Machine Learning

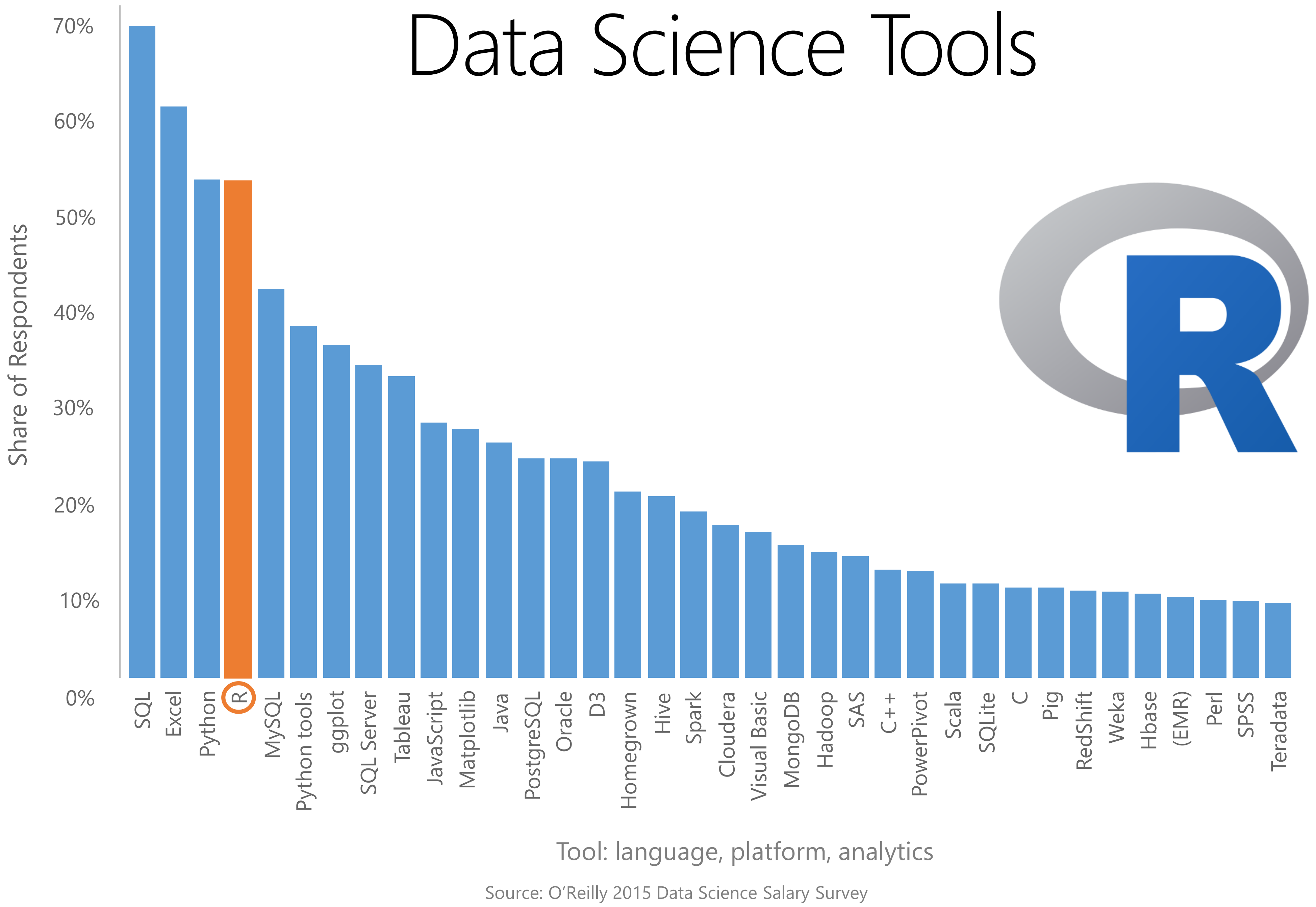


Source: Indeed.com

Average Salary by Job Type (USA)



Source: Stack Overflow 2017


Data Science Tools
Share of Respondents
70%
60%
50%
40%
30%
20%
10%
0%
SQL
Excel
Python
R
MySQL
Python tools
ggplot
SQL Server
Tableau
JavaScript
Matplotlib
Java
PostgreSQL
Oracle
D3
Homegrown
Hive
Spark
Cloudera
Visual Basic
MongoDB
Hadoop
SAS
C++
PowerPivot
Scala
SQLite
C
Pig
RedShift
Weka
Hbase
(EMR)
Perl
SPSS
Teradata
Tool: language, platform, analytics
Source: O'Reilly 2015 Data Science Salary Survey


R

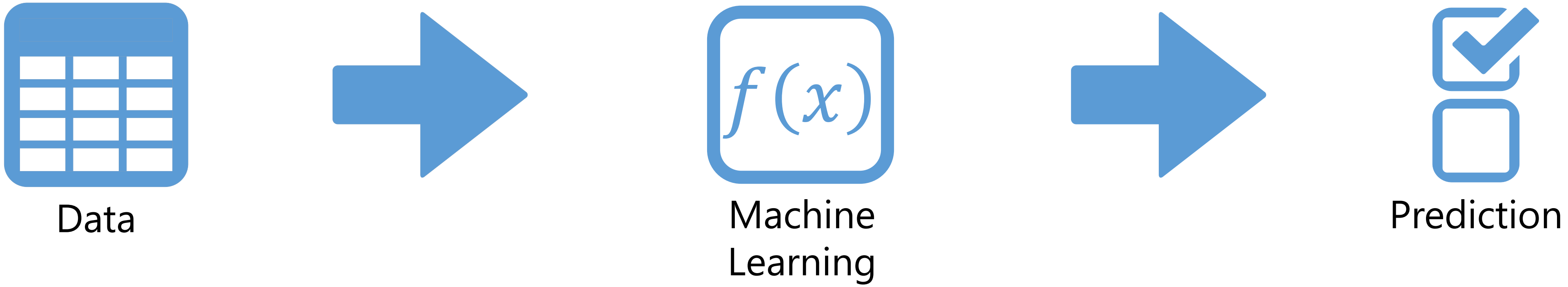
Overview

1. Intro to ML and R
2. Classification
3. Regression
4. Clustering
5. Ensemble Learning
6. Deep Learning
7. ML in Practice
8. ML in Production

How Does This Apply to Me?

- ☑ Make decisions using data
- ☑ Make predictions using data
- ☑ Make recommendations using data
- ☑ Automate these with code

Conceptual Model




About Me

Data Science Consultant

Education

- B.S. in Computer Science
- B.A. in Philosophy
- Data Science specializations

Community

- Public speaker
- Pluralsight author
- Microsoft MVP
- Open source

Schedule

Lectures (10 min)

Demos (10 min)

Labs (20 min)

Breaks (5 min)

Logistics

Pairing for labs is optional

Ask questions if needed

Come and go as needed

Feedback at the end

Labs

Labs

A

(Easy)

Labs

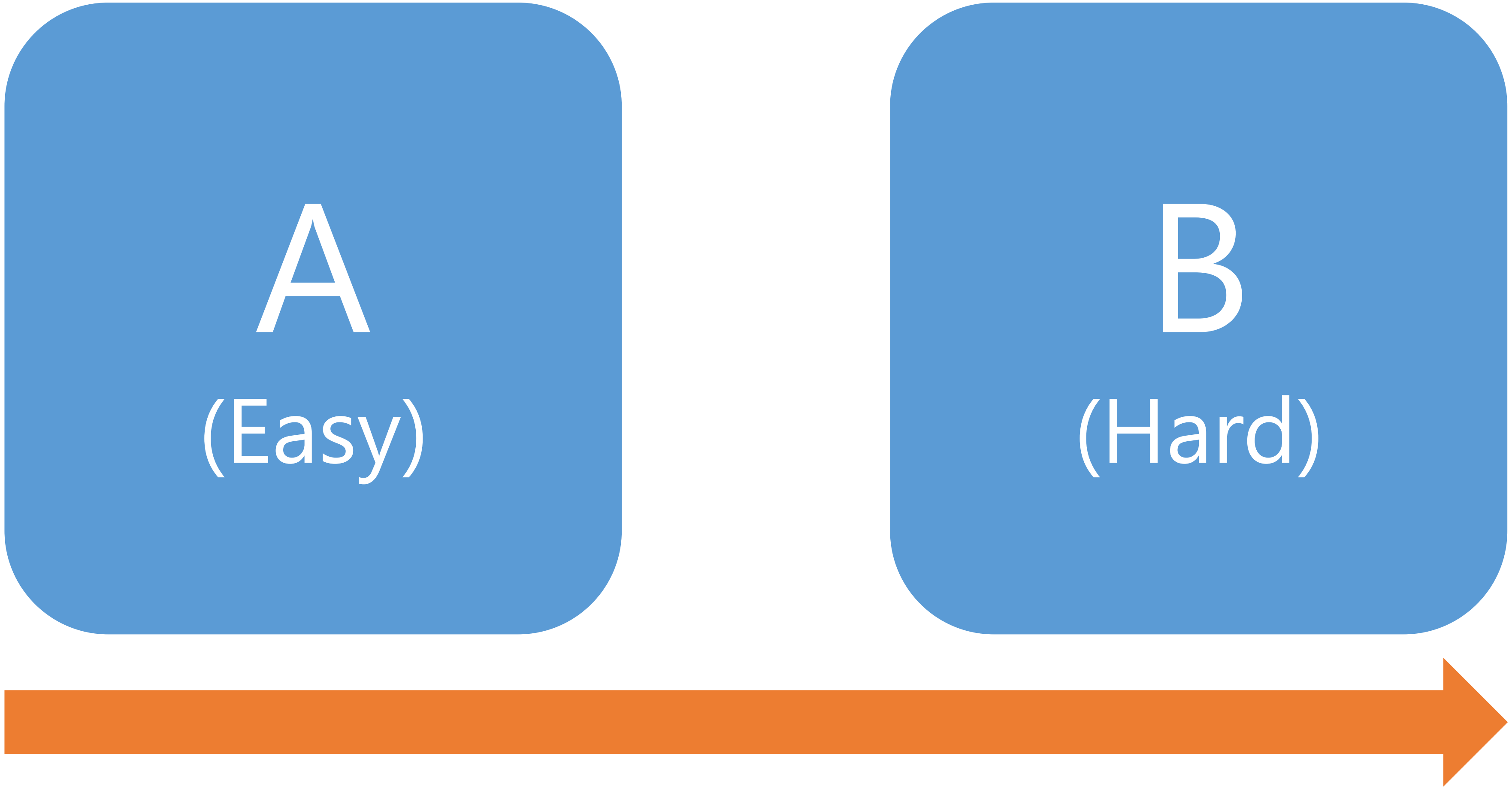
A

(Easy)

B

(Hard)

Labs

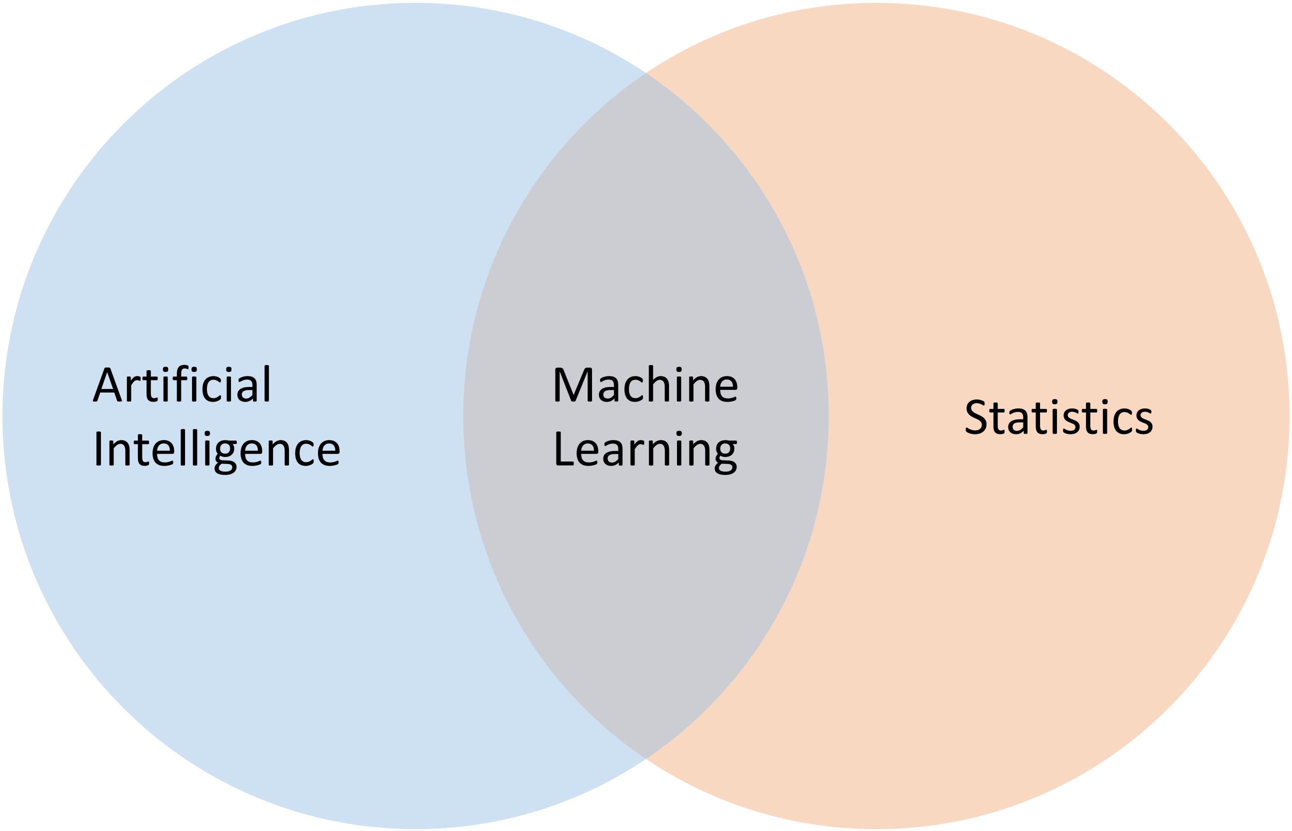


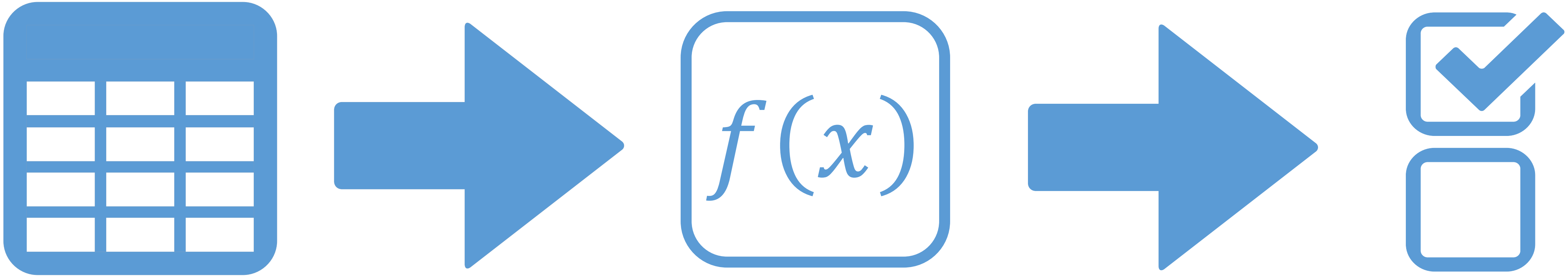
Workshop URL

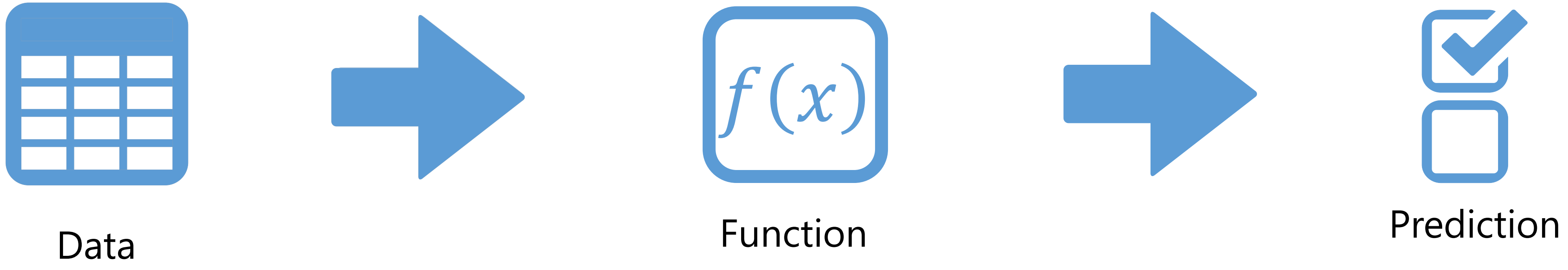
http://www.matthewrenze.com/workshops/practical-machine-learning-with-r/

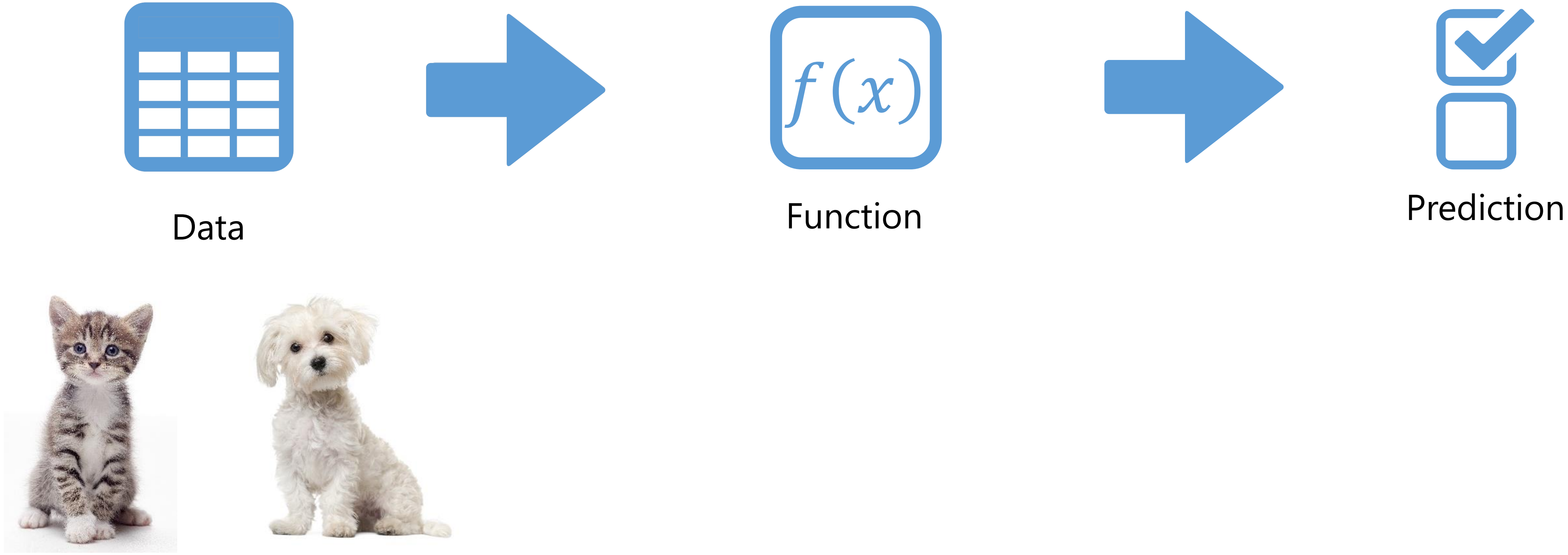
Introduction to Machine Learning

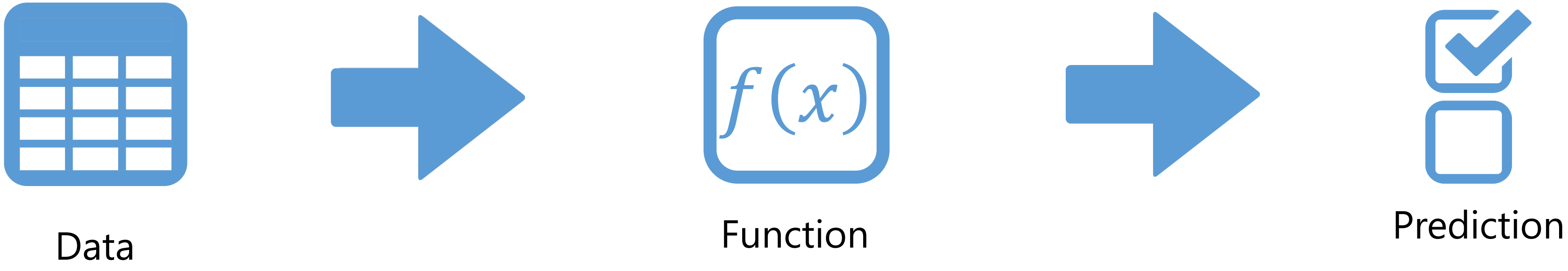
What is machine learning?


Artificial
Intelligence
Machine
Learning
Statistics


f(x)

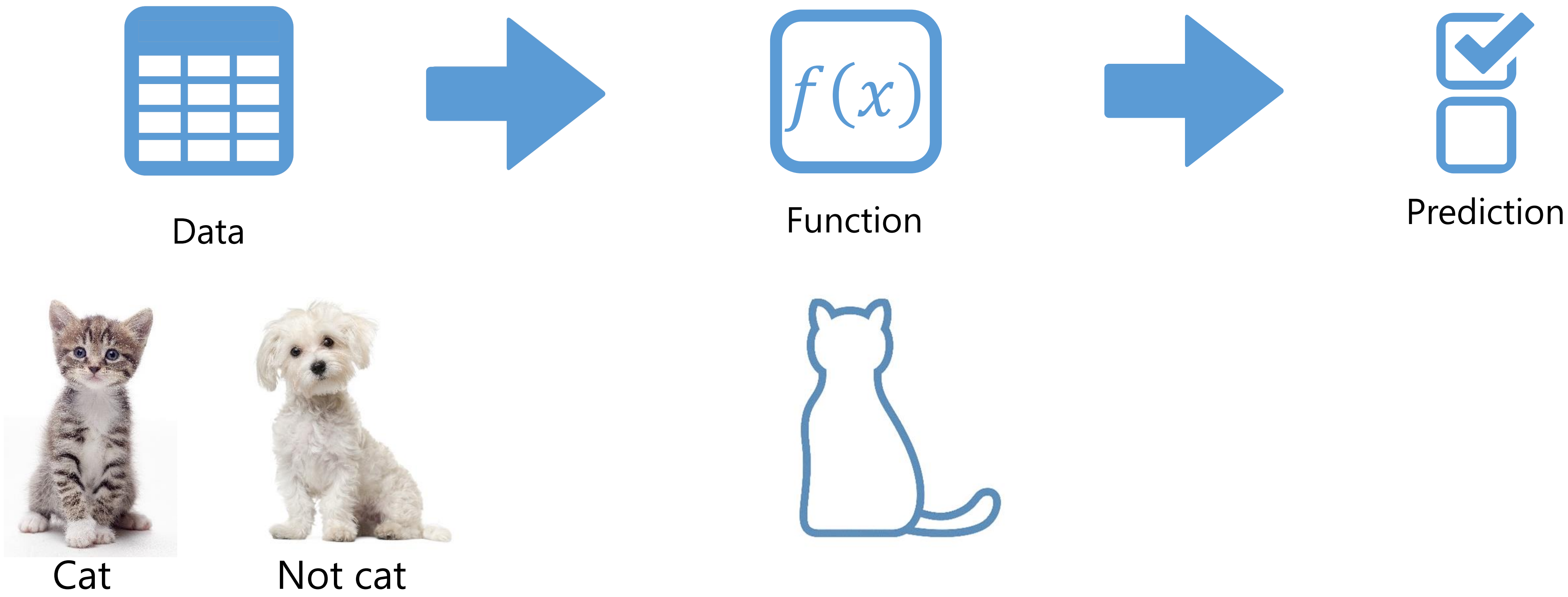

f(x)
Data
Function
Prediction

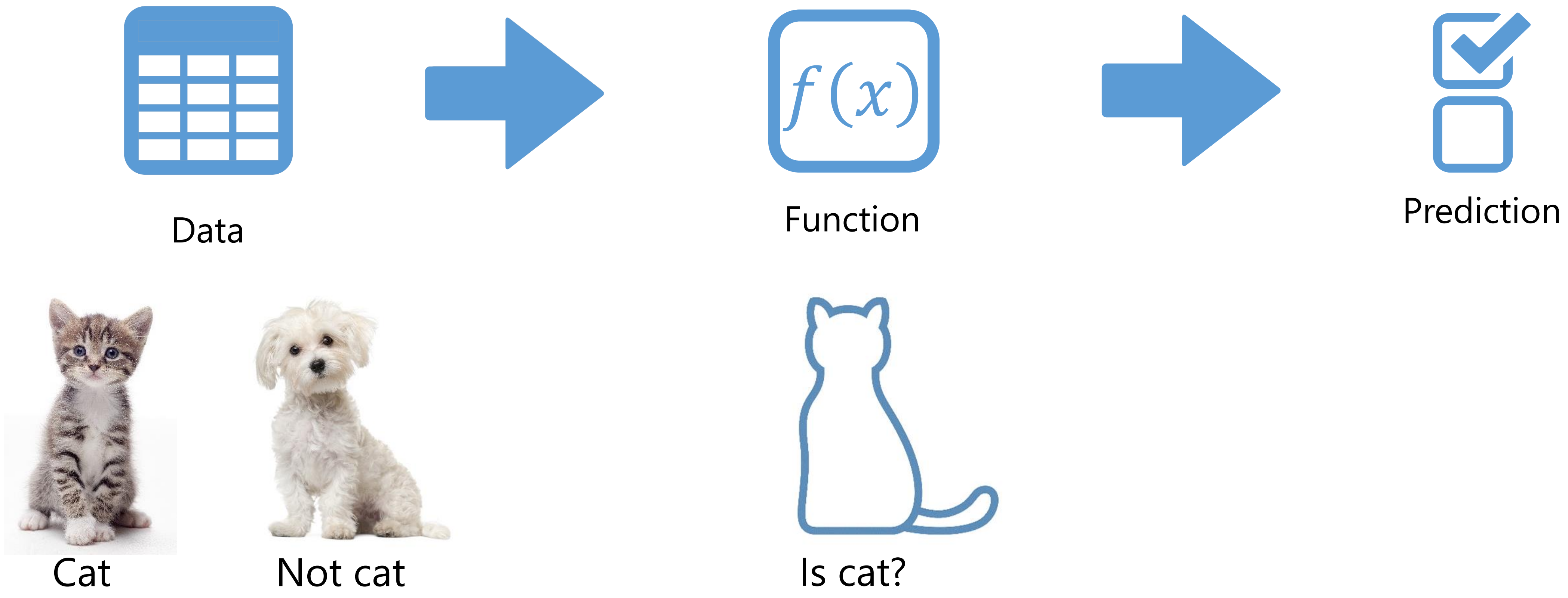

Data
f(x)
Function
Prediction

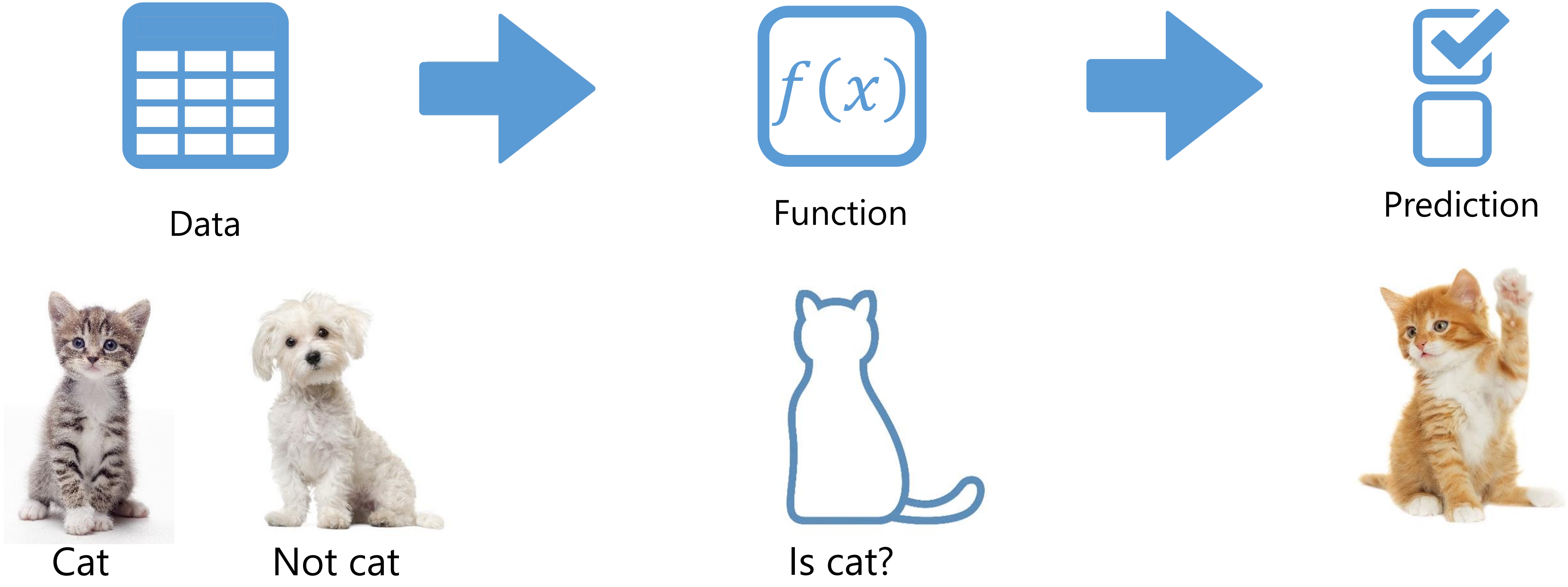


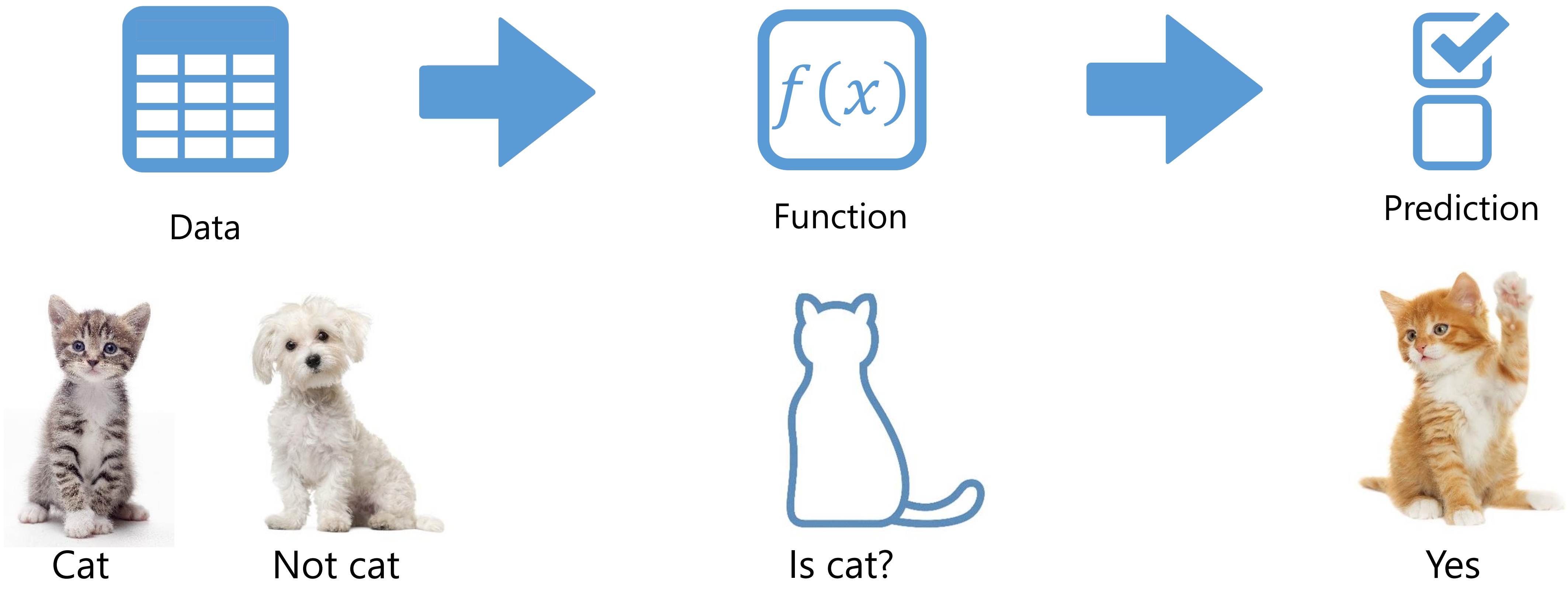

Cat


Not cat


Data
$f(x)$
Function
Prediction
Cat
Not cat

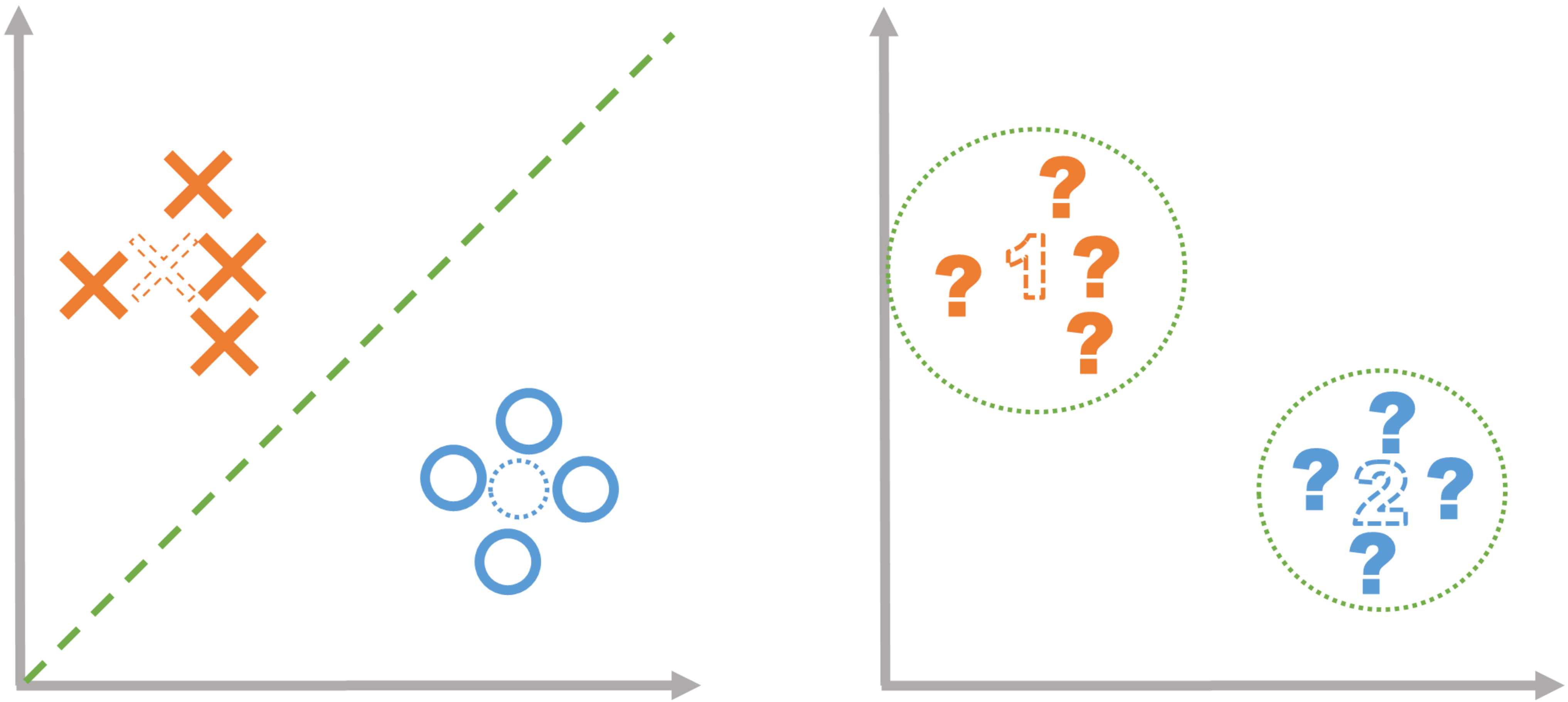

Data
Function
Prediction
Cat
Not cat
Is cat?


Data
Function
Prediction
Cat
Not cat
Is cat?


$f(x)$
Data
Function
Prediction
Cat
Not cat
Is cat?
Yes

What types of machine learning exist?

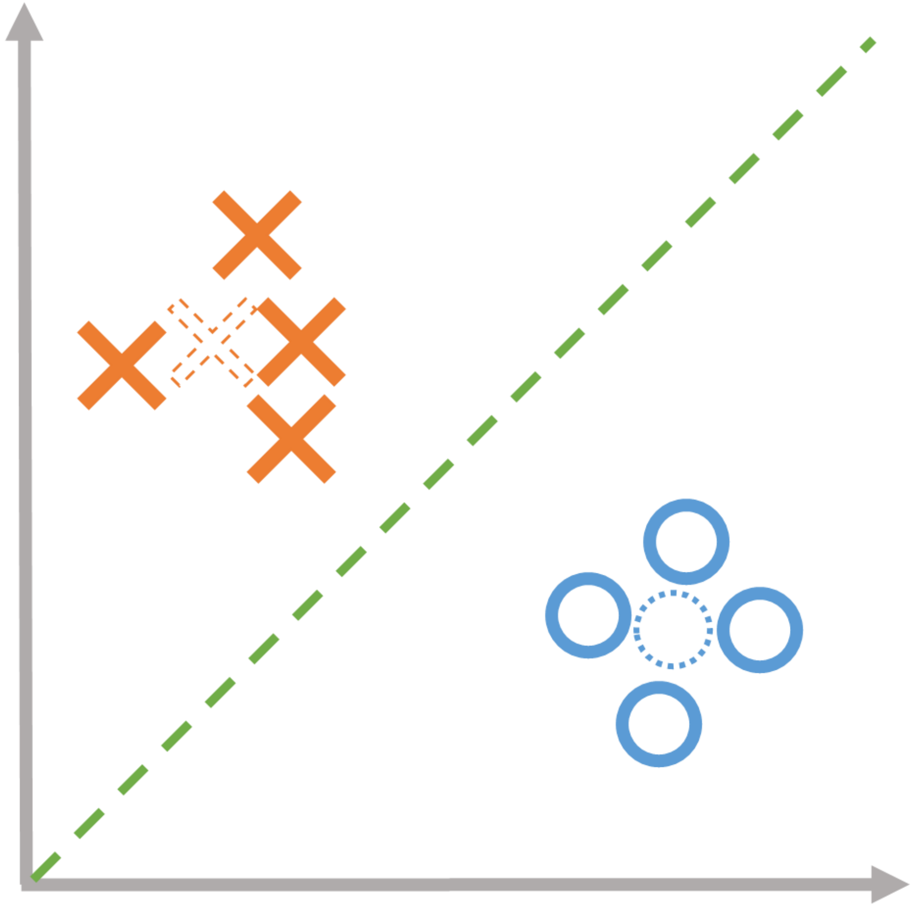
Types of Machine Learning



Supervised Learning

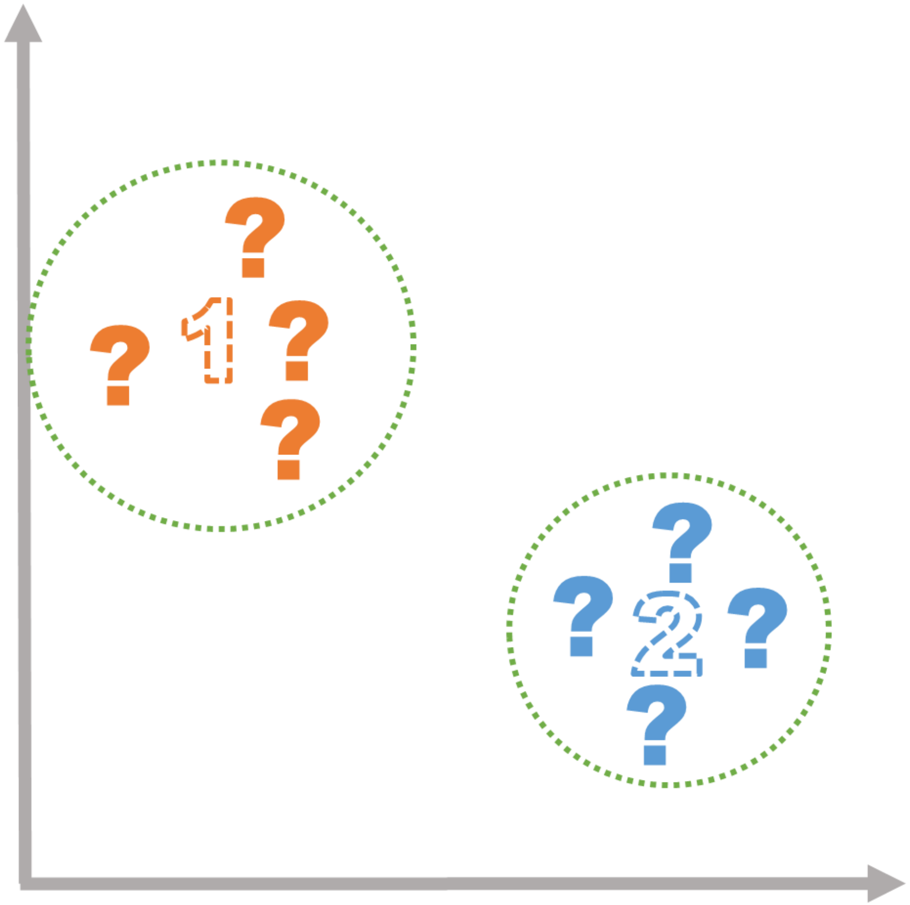
Unsupervised Learning

Types of Machine Learning



Supervised Learning

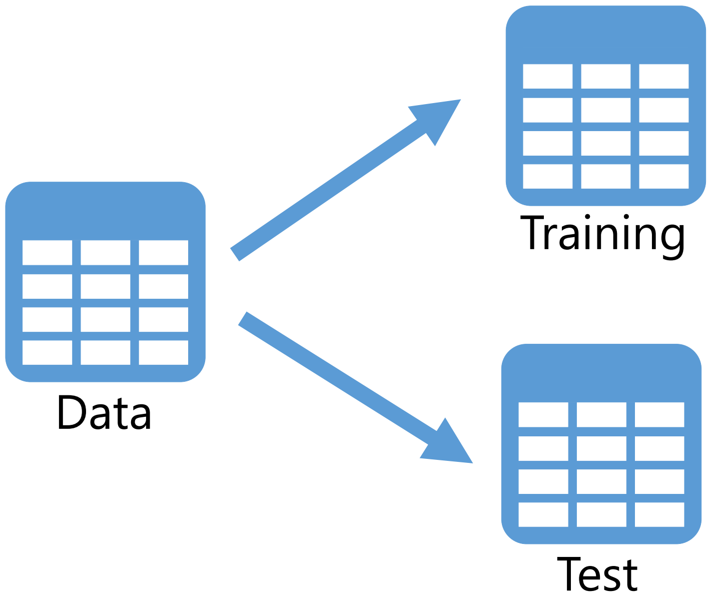
Types of Machine Learning

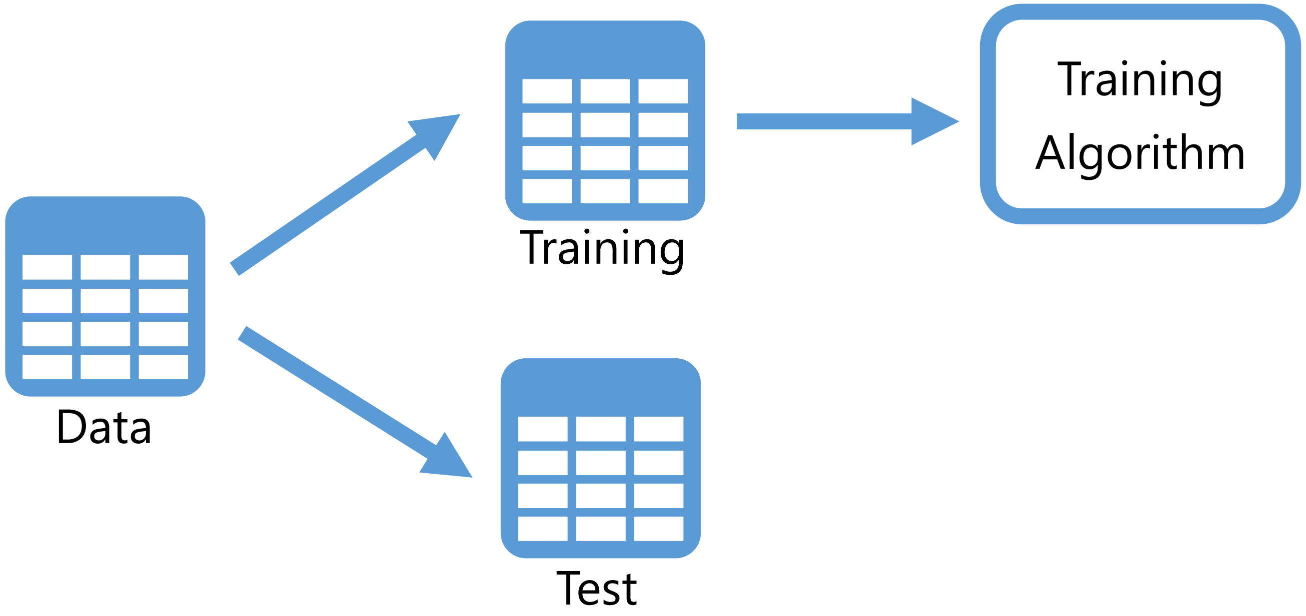


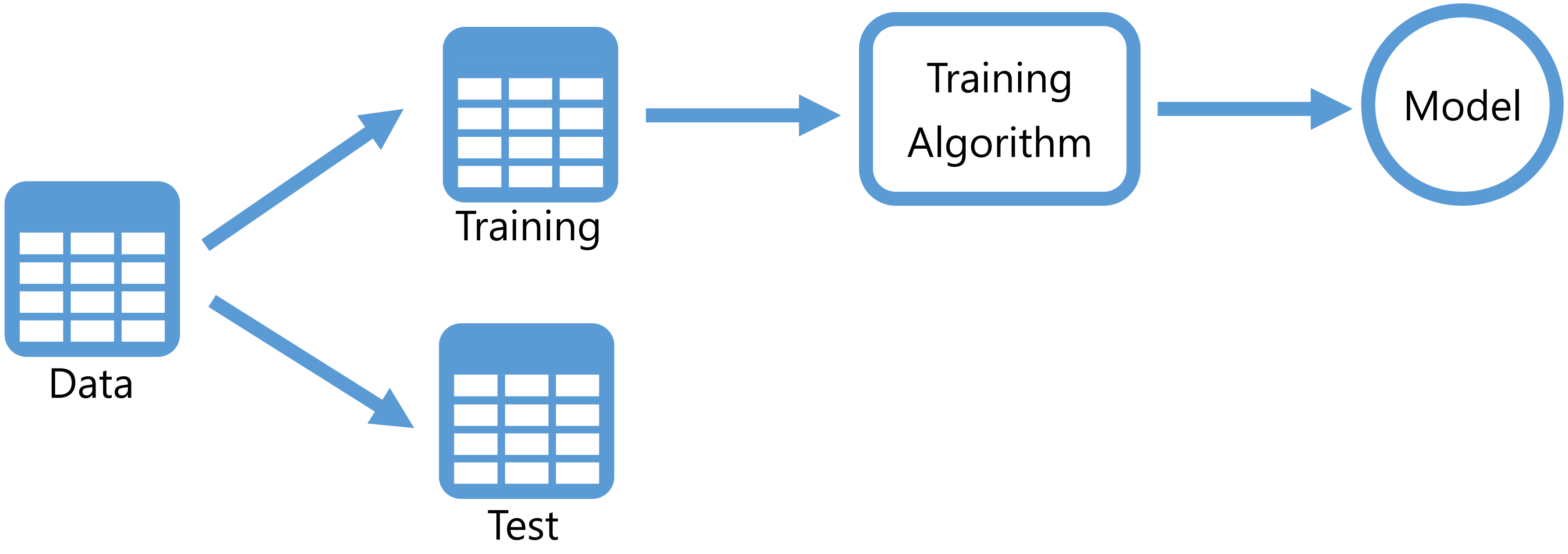
Unsupervised Learning

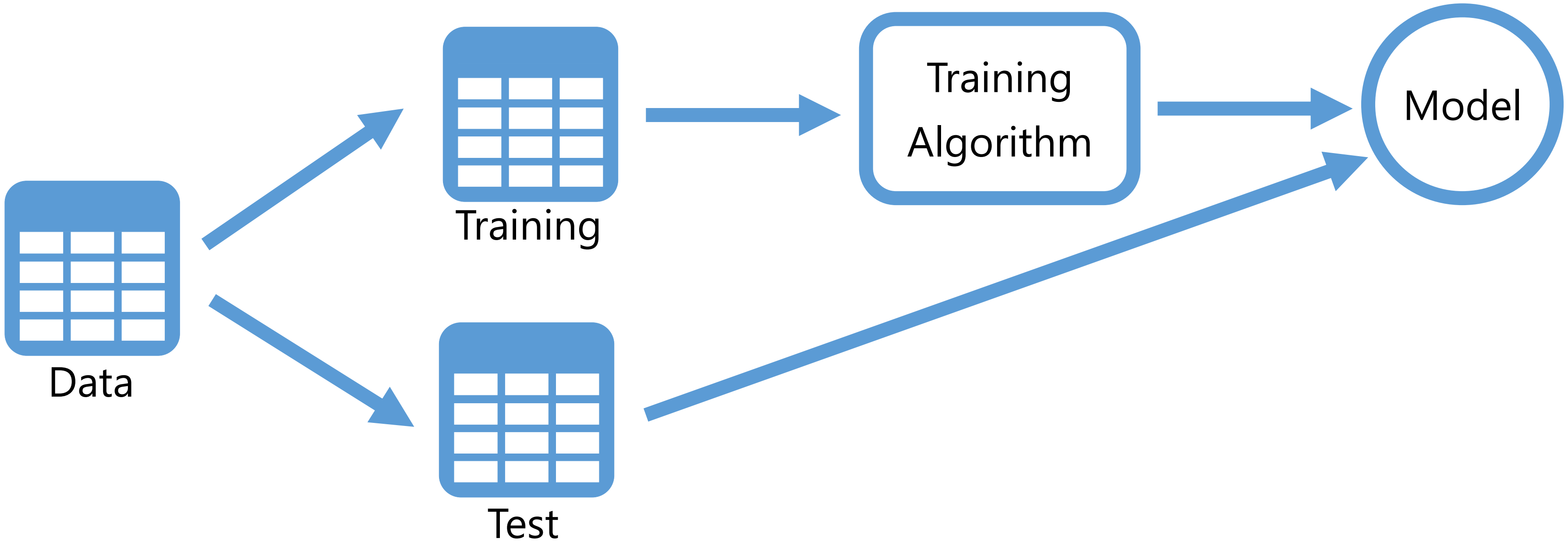
How does machine learning work?

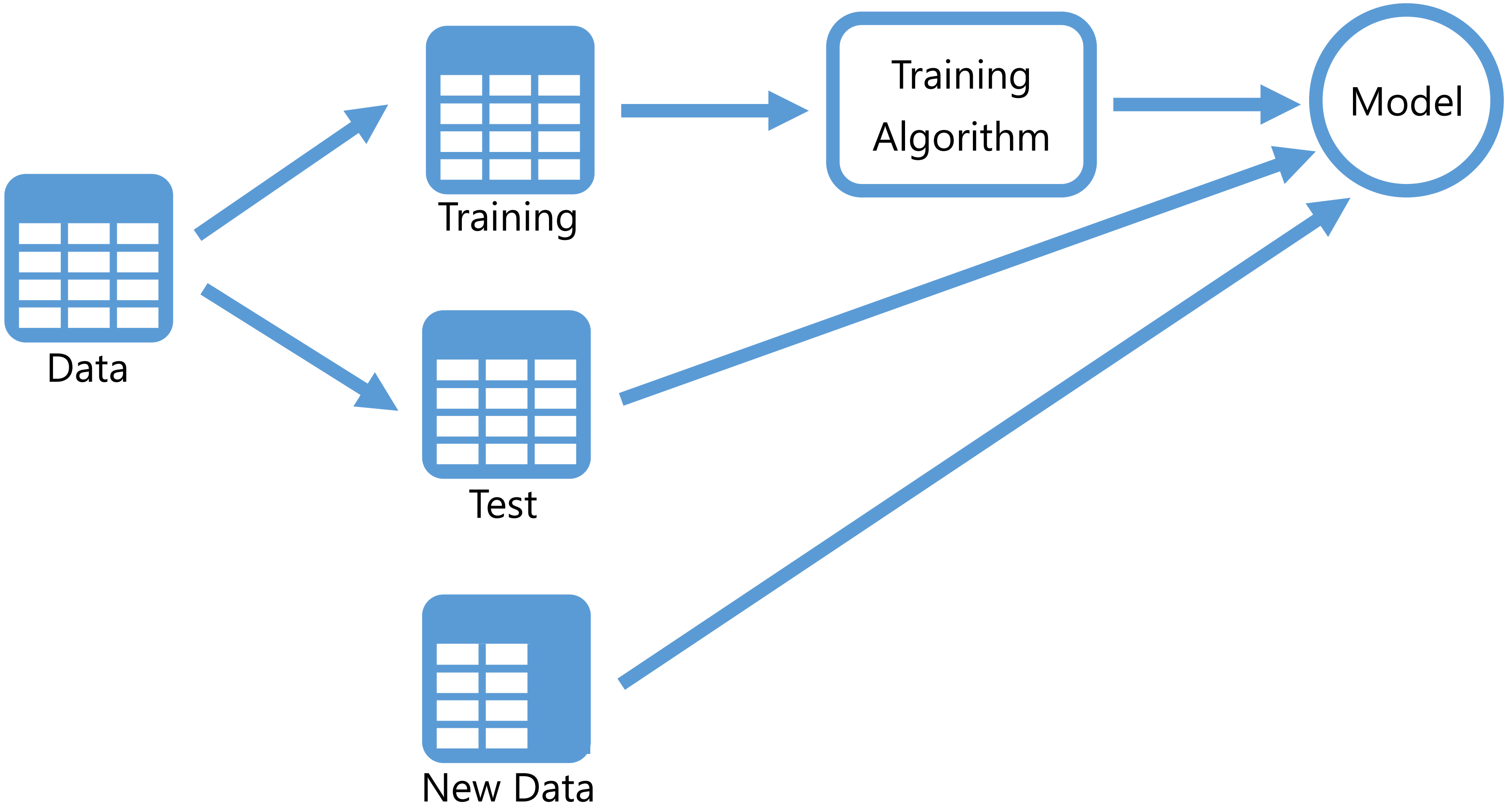
Data

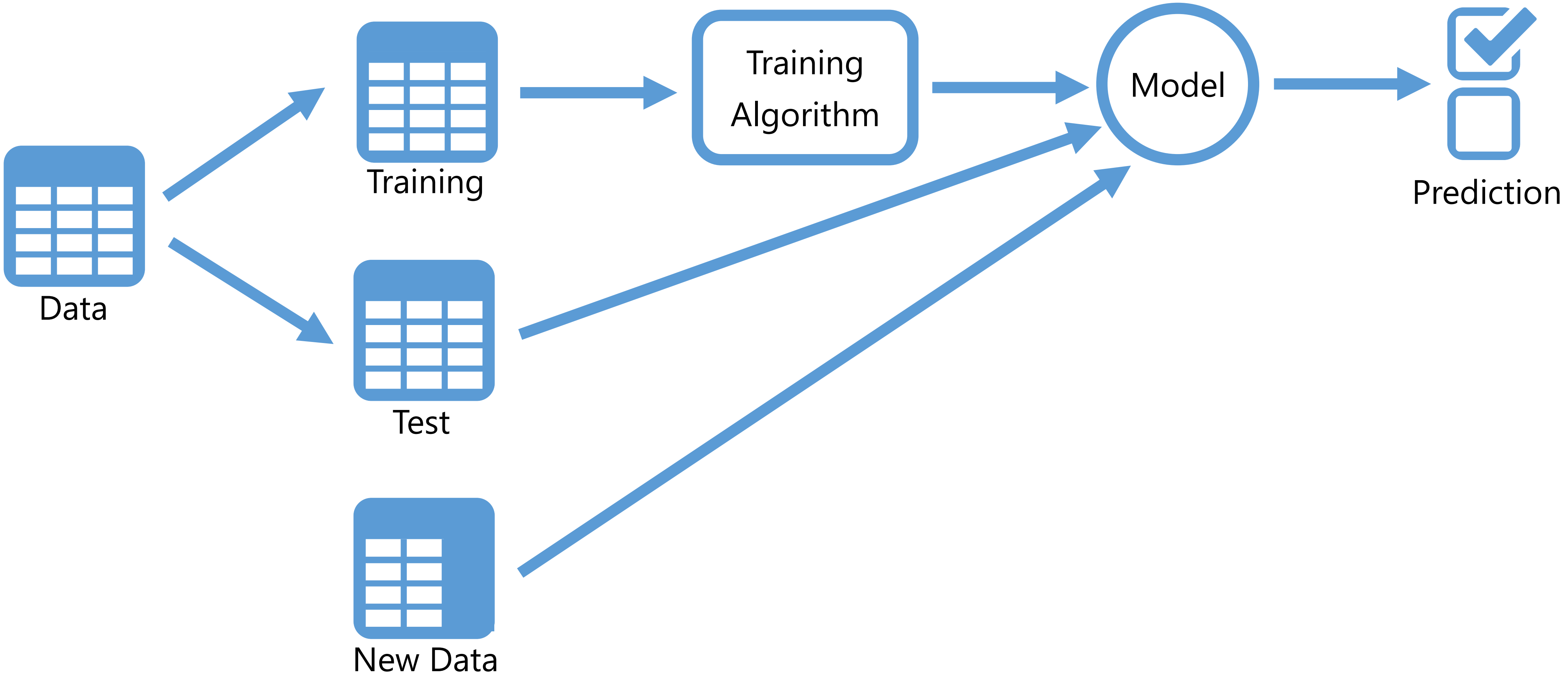

Training
Data
Test

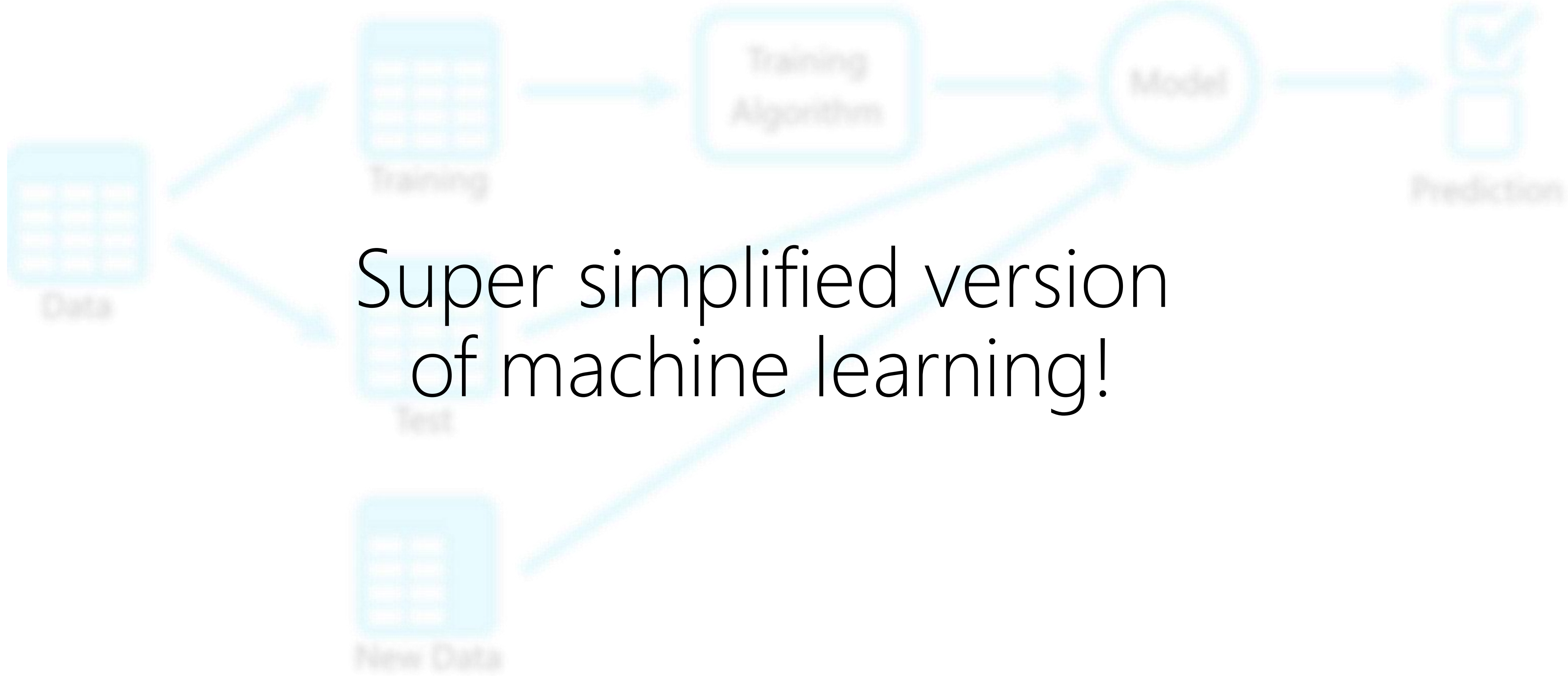

Data
Training
Test
Training Algorithm


Training
Algorithm
Model
Training
Data
Test


Data
Training
Test
Training
Algorithm
Model

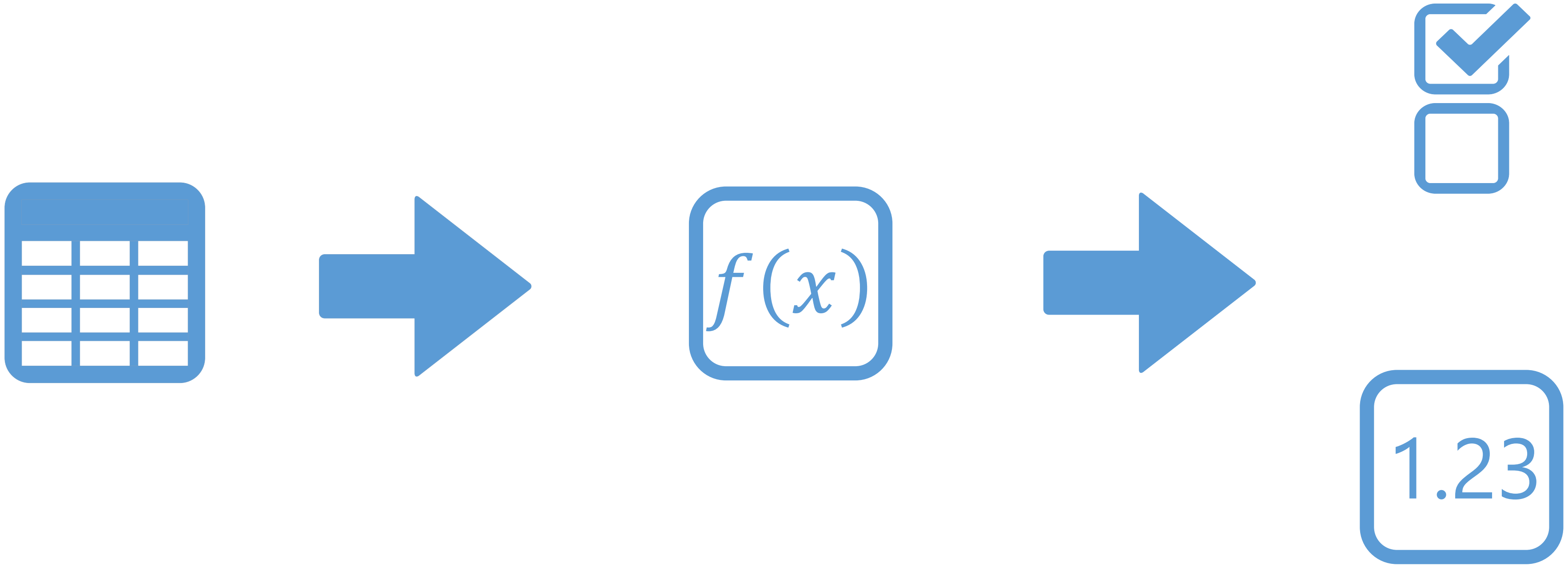

Training
Algorithm
Model
Training
Data
Test
New Data


Data
Training
Test
New Data
Training
Algorithm
Model
Prediction

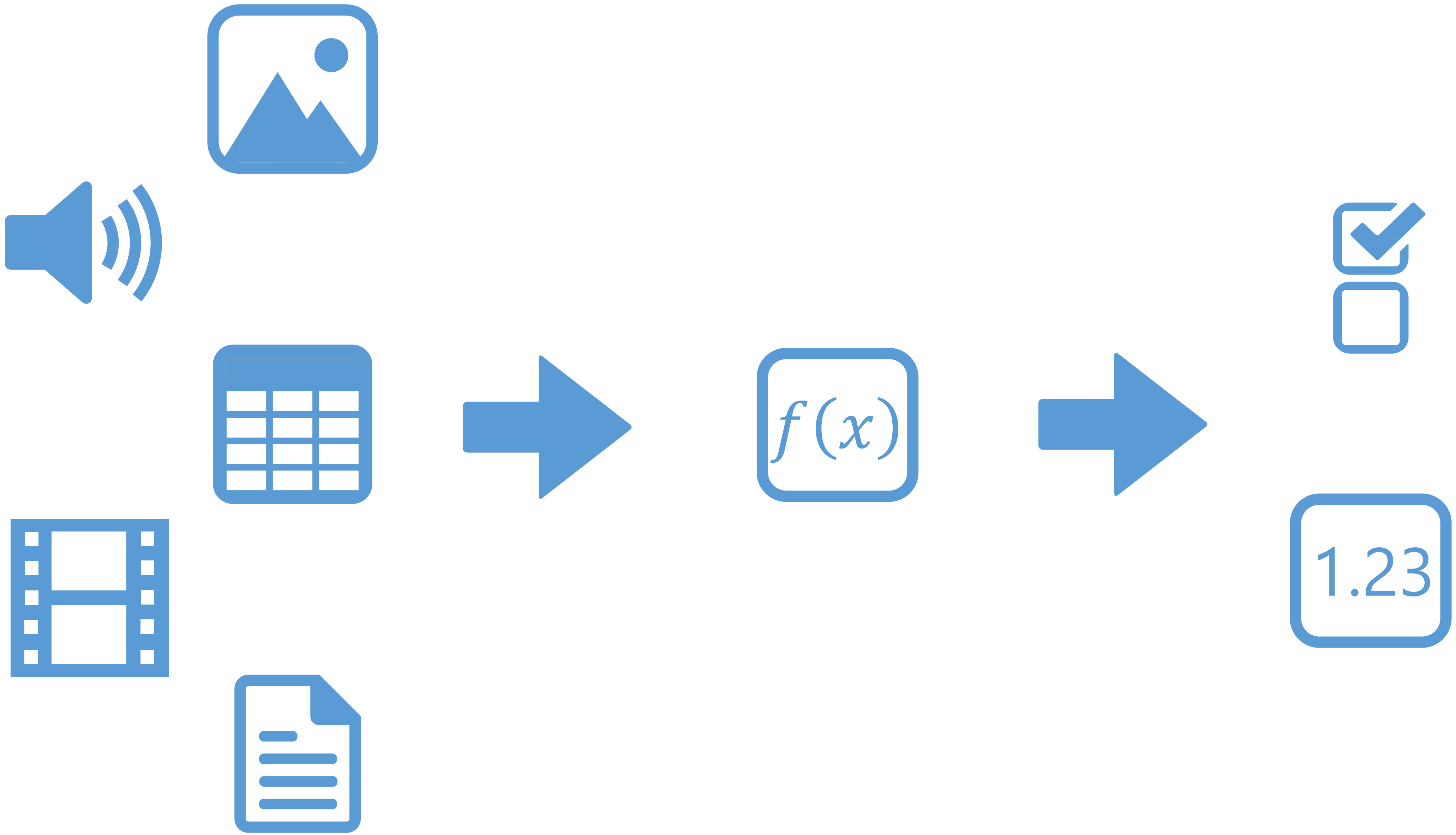


Super simplified version
of machine learning!

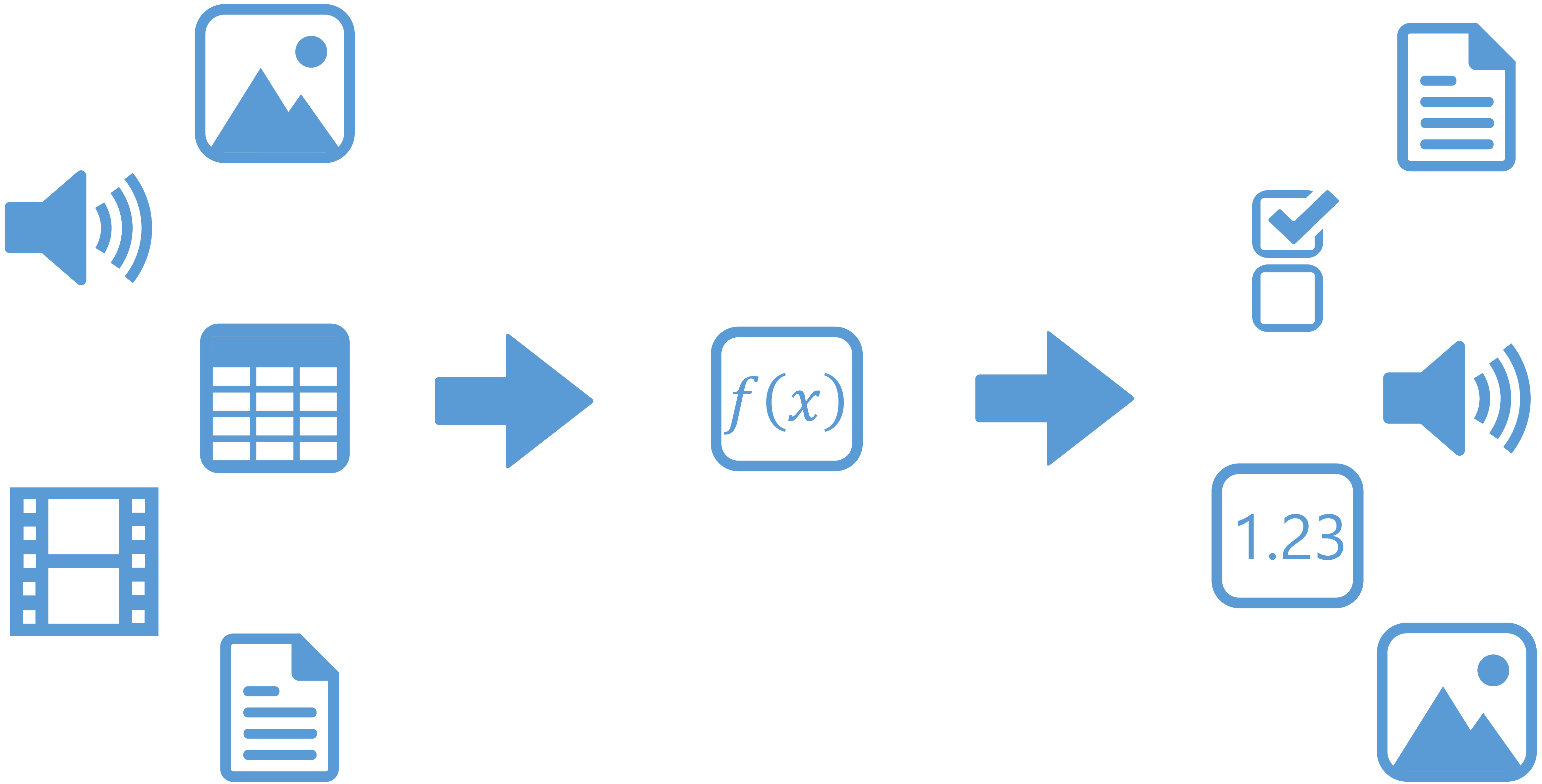
What can machine learning do?



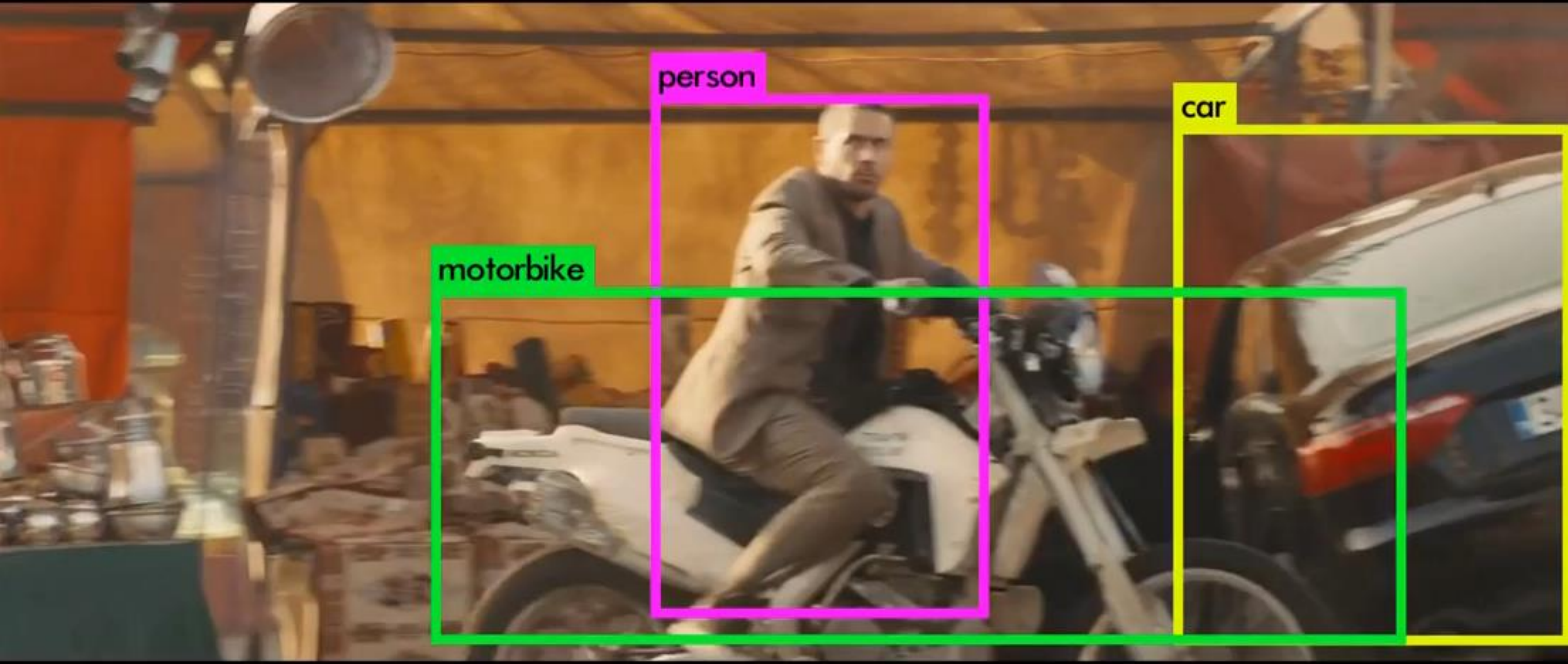
$f(x)$
1.23



f(x)
1.23



$f(x)$
1.23

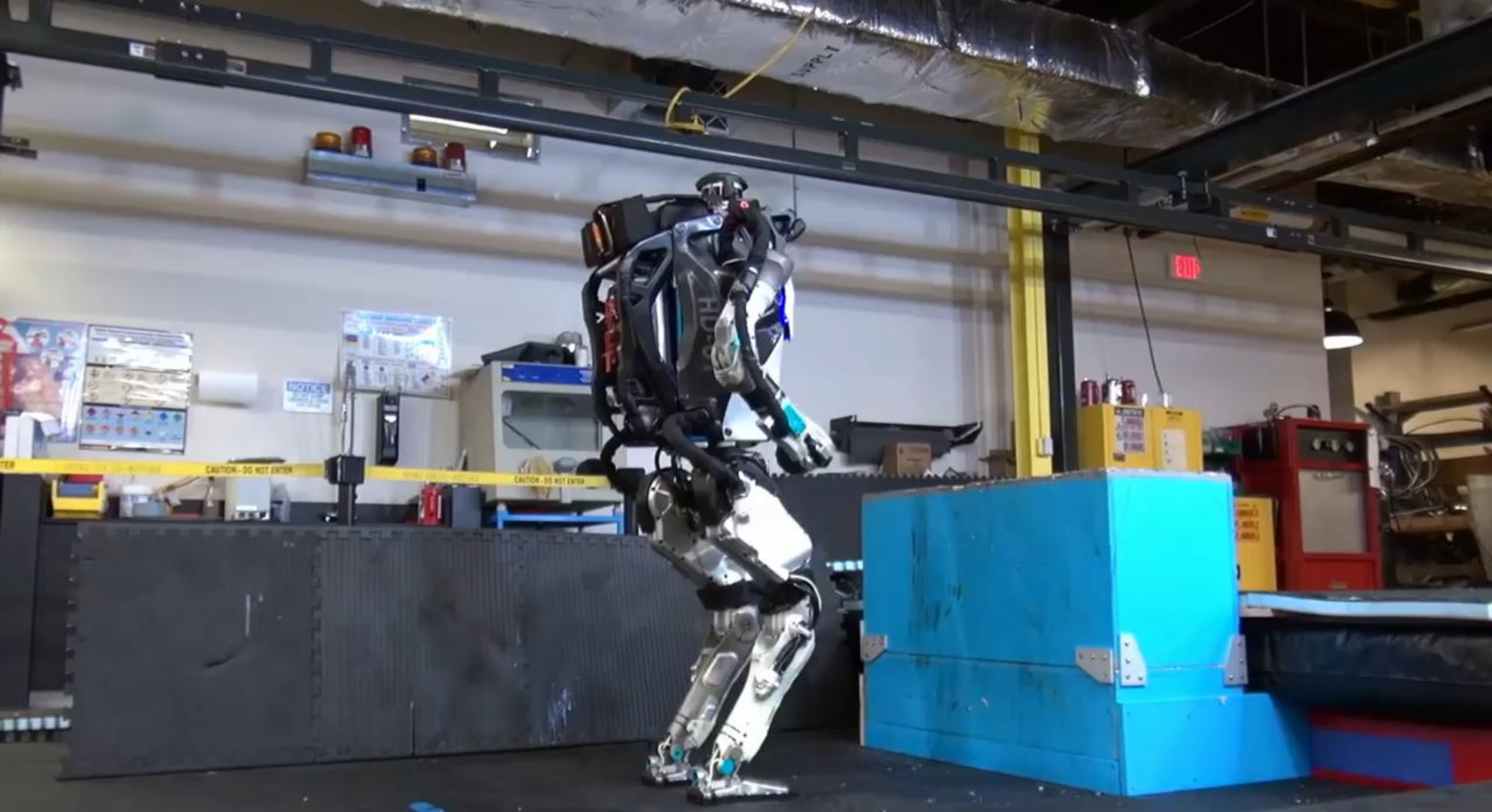


Source: YOLO: Real-Time Object Detection

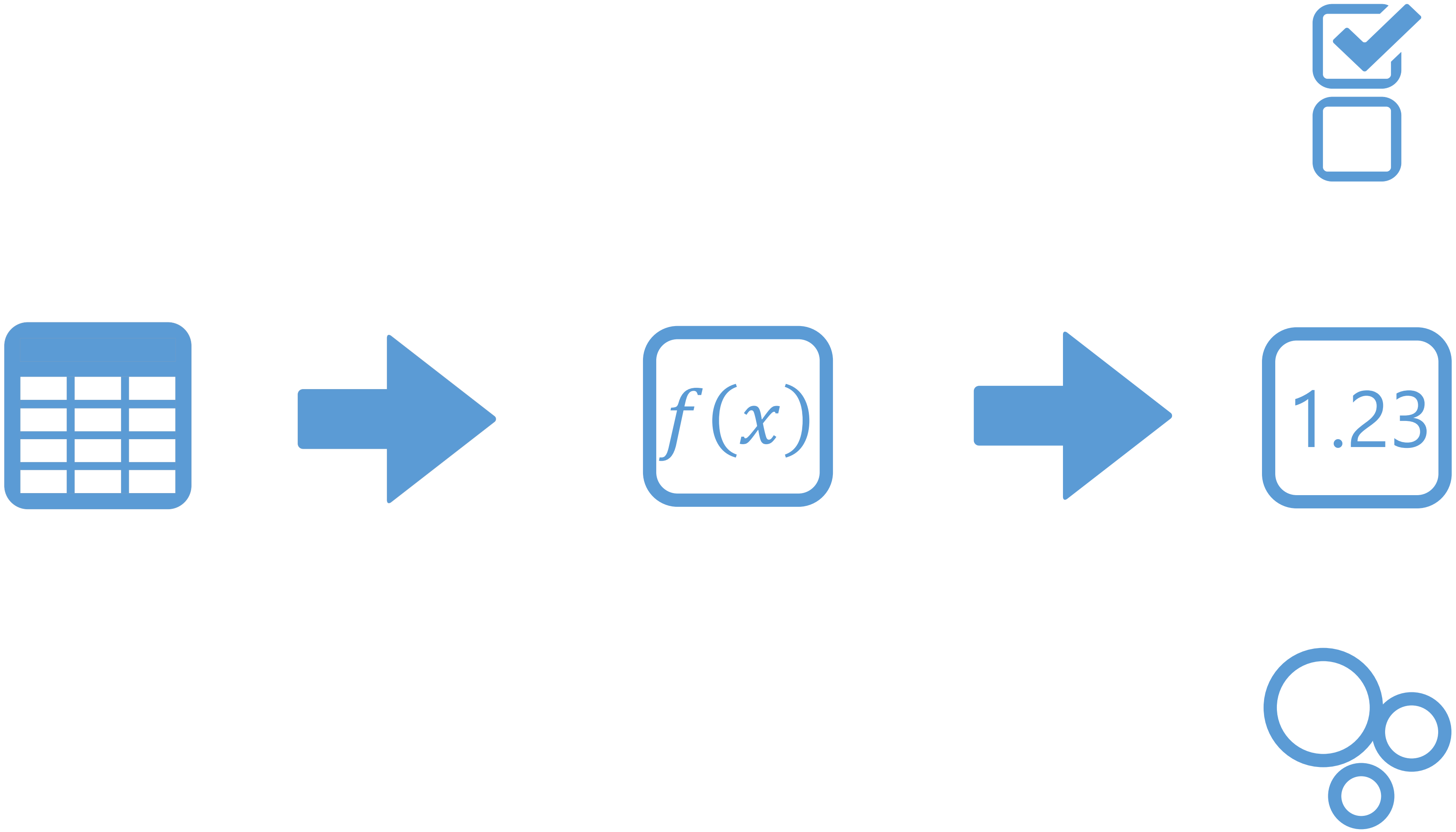
Source: Nvidia

Source: http://grail.cs.washington.edu/projects/AudioToObama/

Source: Google Deep Mind



Source: Boston Dynamics


f(x)
1.23

Disclaimer

Introduction to R

What is R?

Open source

Language and environment

Numerical and graphical

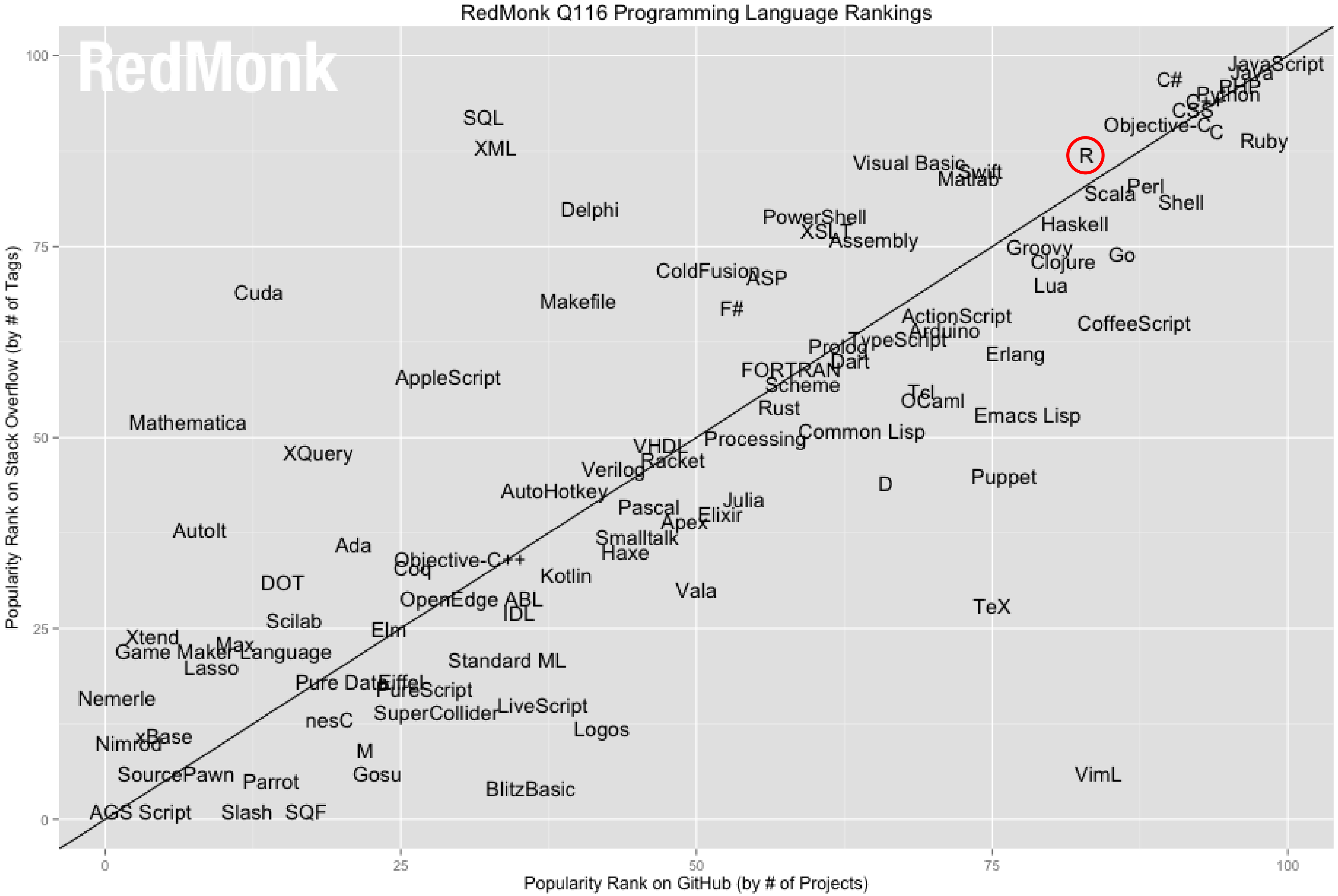
Cross platform

What is R?

Active development
Large user community
Modular and extensible
10,000+ extensions

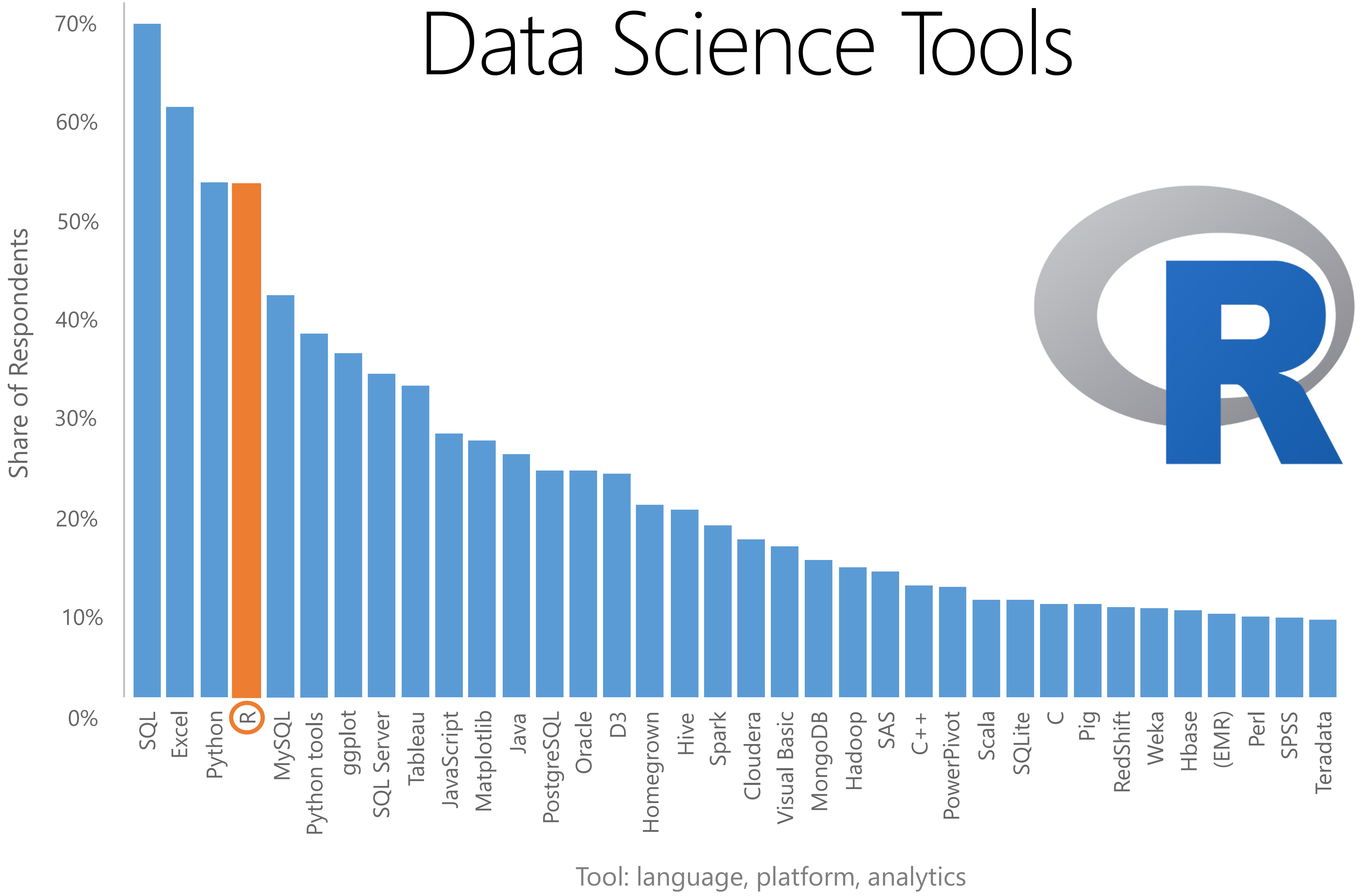

FREE


FREE

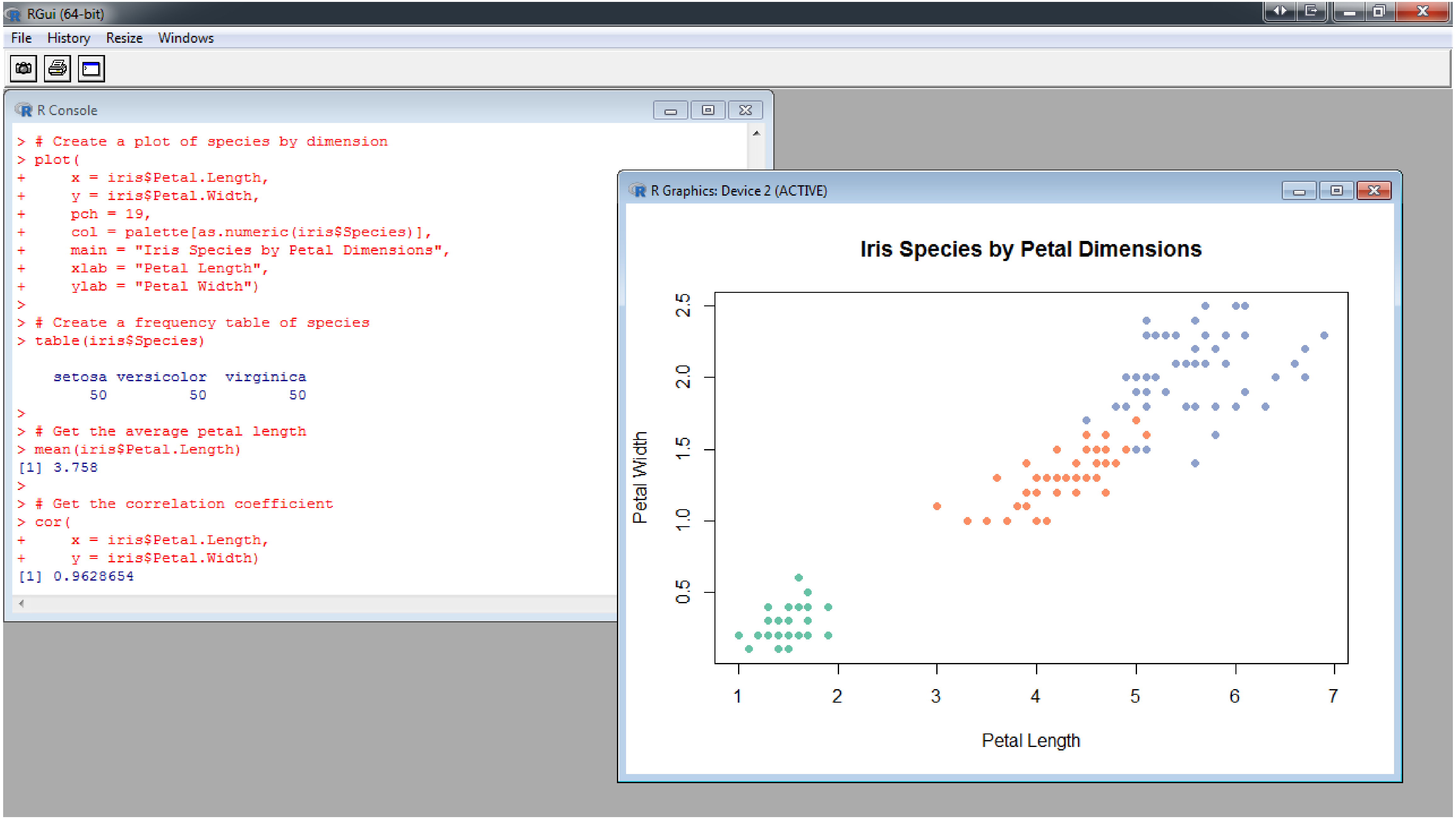


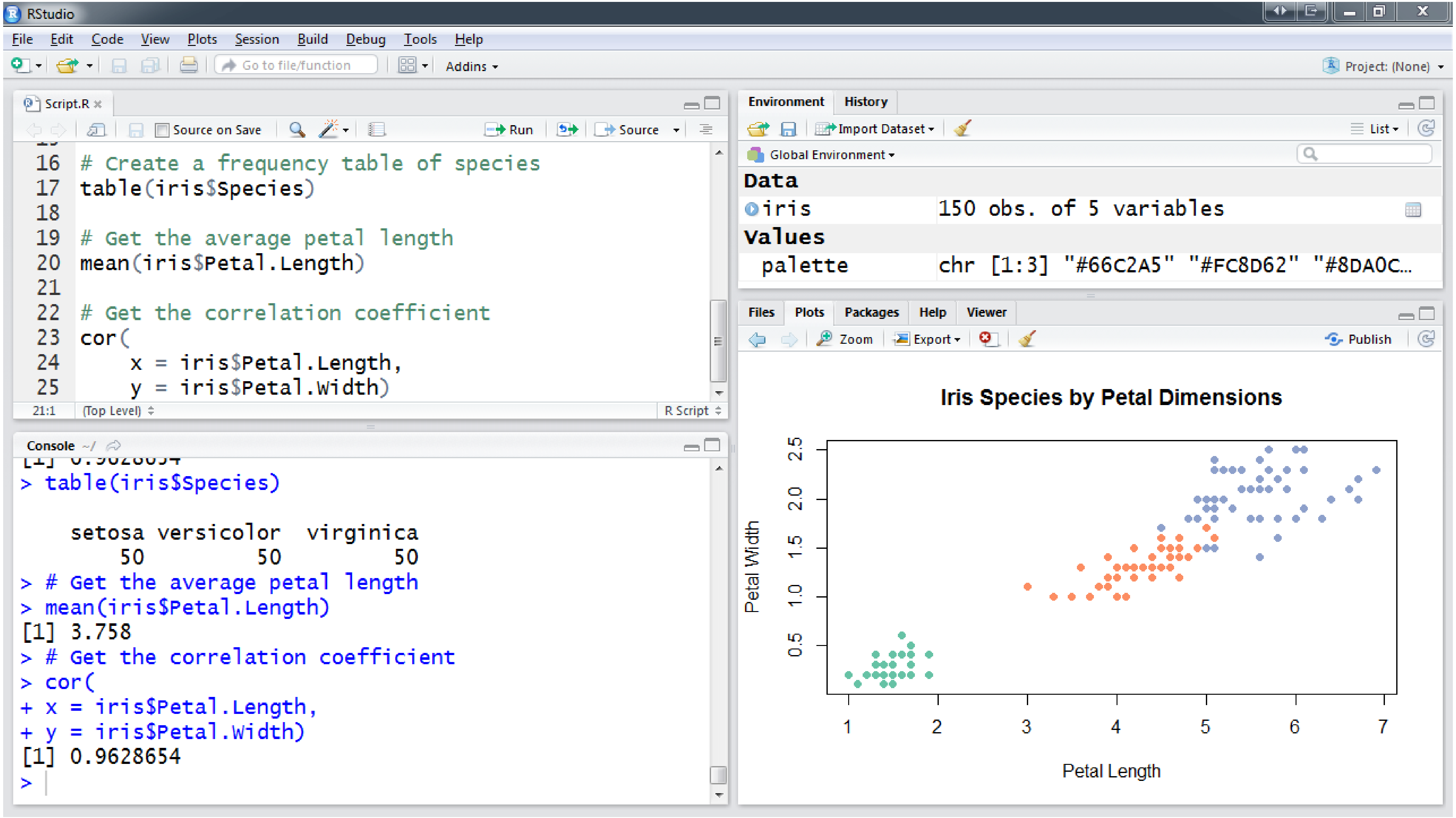
Source: http://redmonk.com/sogrady/2016/07/20/language-rankings-6-16/

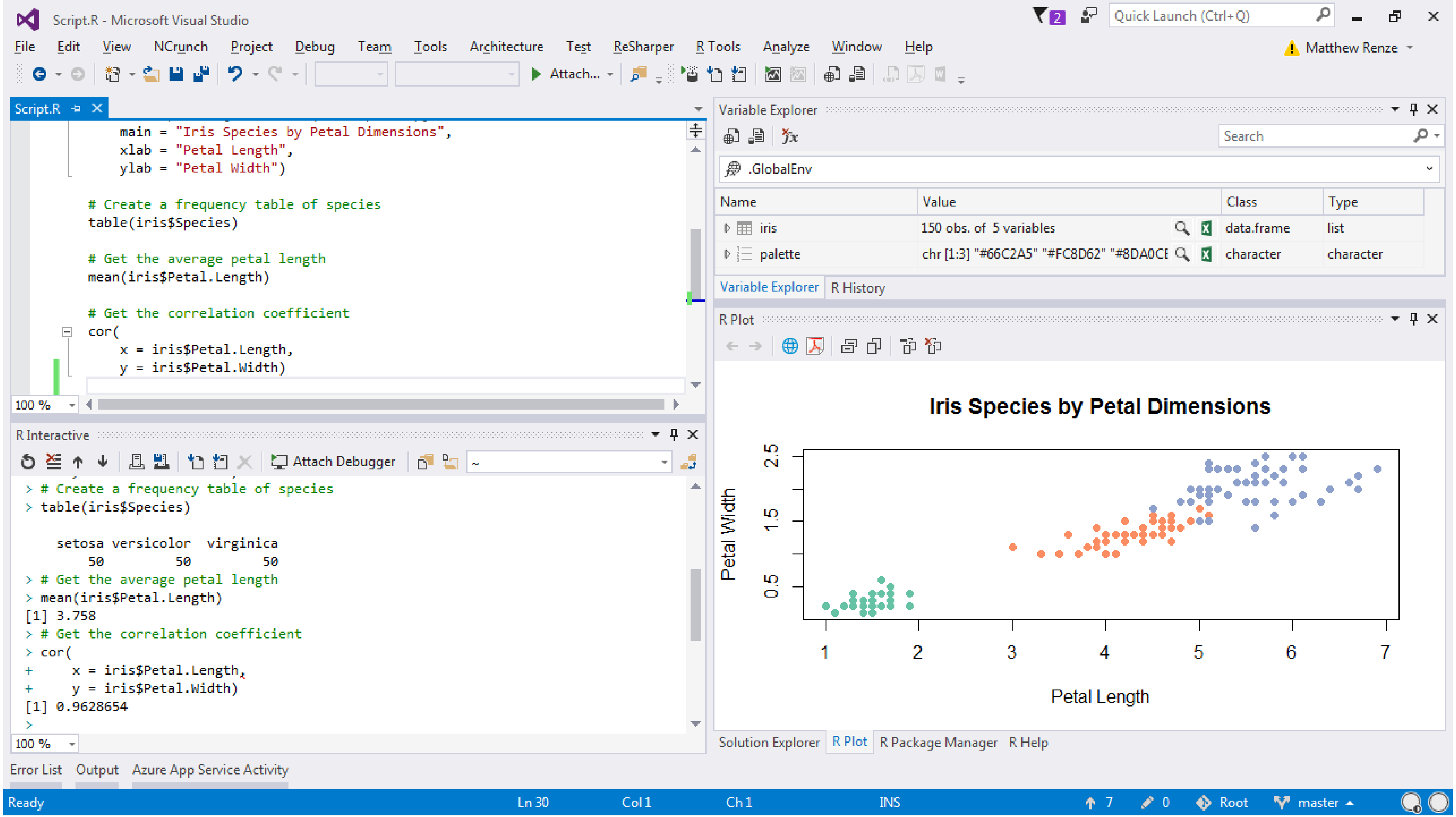
Data Science Tools



Source: O'Reilly 2015 Data Science Salary Survey


RGui (64-bit)
File History Resize Windows
R Console
> # Create a plot of species by dimension
> plot(
+ x = iris$Petal.Length,
+ y = iris$Petal.Width,
+ pch = 19,
+ col = palette[as.numeric(iris$Species)],
+ main = "Iris Species by Petal Dimensions",
+ xlab = "Petal Length",
+ ylab = "Petal Width")
>
> # Create a frequency table of species
> table(iris$Species)
setosa versicolor virginica
50 50 50
>
> # Get the average petal length
> mean(iris$Petal.Length)
[1] 3.758
>
> # Get the correlation coefficient
> cor(
+ x = iris$Petal.Length,
+ y = iris$Petal.Width)
[1] 0.9628654
R Graphics: Device 2 (ACTIVE)
Iris Species by Petal Dimensions
Petal Width
0.5
1.0
1.5
2.0
2.5
1
2
3
4
5
6
7
Petal Length

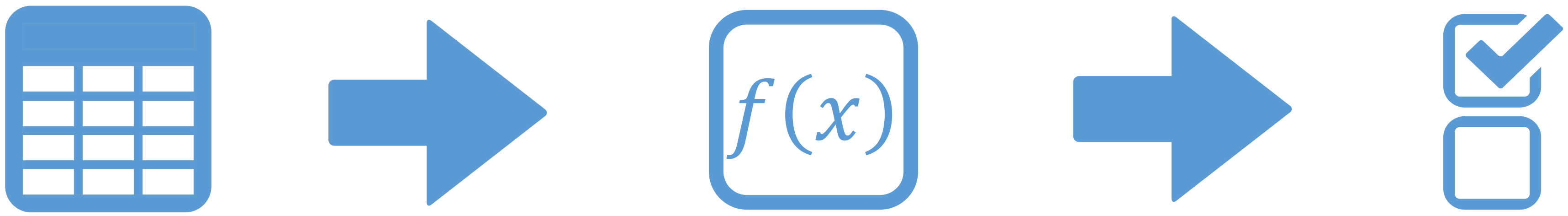

RStudio
File Edit Code View Plots Session Build Debug Tools Help
Go to file/function
Addins
Project: (None)
Script.R
Source on Save
Run
Source
# Create a frequency table of species
table(iris$Species)
# Get the average petal length
mean(iris$Petal.Length)
# Get the correlation coefficient
cor(
x = iris$Petal.Length,
y = iris$Petal.Width)
21:1
(Top Level)
R Script
Console ~/
> table(iris$Species)
setosa versicolor virginica
50 50 50
> # Get the average petal length
> mean(iris$Petal.Length)
[1] 3.758
> # Get the correlation coefficient
> cor(
+ x = iris$Petal.Length,
+ y = iris$Petal.Width)
[1] 0.9628654
>
Environment History
Import Dataset
List
Global Environment
Data
iris 150 obs. of 5 variables
Values
palette chr [1:3] "#66C2A5" "#FC8D62" "#8DA0C…
Files Plots Packages Help Viewer
Zoom
Export
Publish
Iris Species by Petal Dimensions
Petal Width
0.5
1.0
1.5
2.0
2.5
1
2
3
4
5
6
7
Petal Length

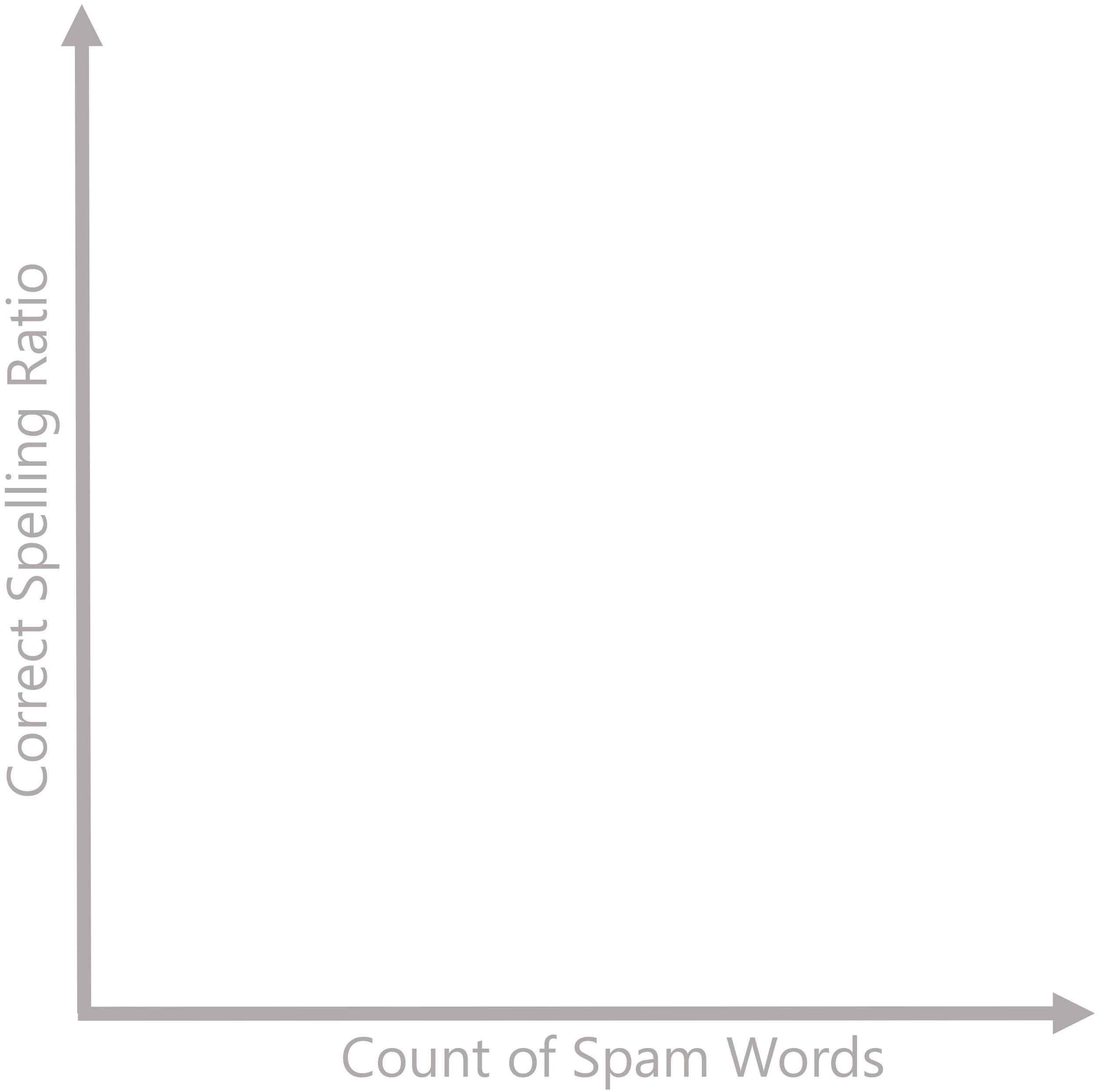

Script.R - Microsoft Visual Studio
Quick Launch (Ctrl+Q)
File Edit View NCrunch Project Debug Team Tools Architecture Test ReSharper R Tools Analyze Window Help
Matthew Renze
Attach...
Script.R
main = "Iris Species by Petal Dimensions",
xlab = "Petal Length",
ylab = "Petal Width")
# Create a frequency table of species
table(iris$Species)
# Get the average petal length
mean(iris$Petal.Length)
# Get the correlation coefficient
cor(
x = iris$Petal.Length,
y = iris$Petal.Width)
100 %
R Interactive
Attach Debugger
> # Create a frequency table of species
> table(iris$Species)
setosa versicolor virginica
50 50 50
> # Get the average petal length
> mean(iris$Petal.Length)
[1] 3.758
> # Get the correlation coefficient
> cor(
+ x = iris$Petal.Length,
+ y = iris$Petal.Width)
[1] 0.9628654
>
100 %
Error List Output Azure App Service Activity
Variable Explorer
Search
.GlobalEnv
Name Value Class Type
iris 150 obs. of 5 variables data.frame list
palette chr [1:3] "#66C2A5" "#FC8D62" "#8DA0CE character character
Variable Explorer R History
R Plot
Iris Species by Petal Dimensions
Petal Width
0.5 1.5 2.5
1 2 3 4 5 6 7
Petal Length
Solution Explorer R Plot R Package Manager R Help
Ready Ln 30 Col 1 Ch 1 INS 7 0 Root master

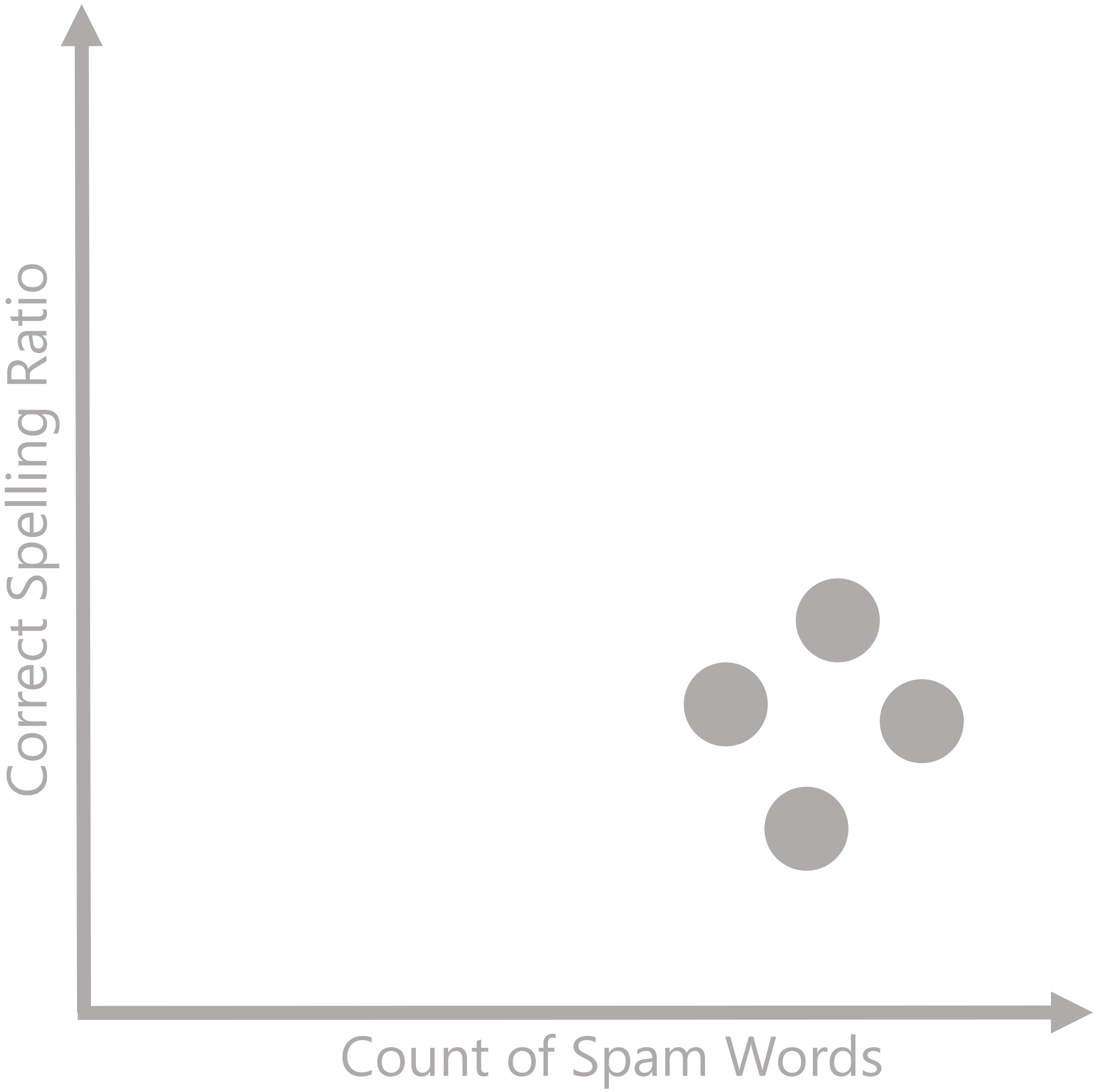
Demo 1
R Language Basics

Lab 1
R Language Basics

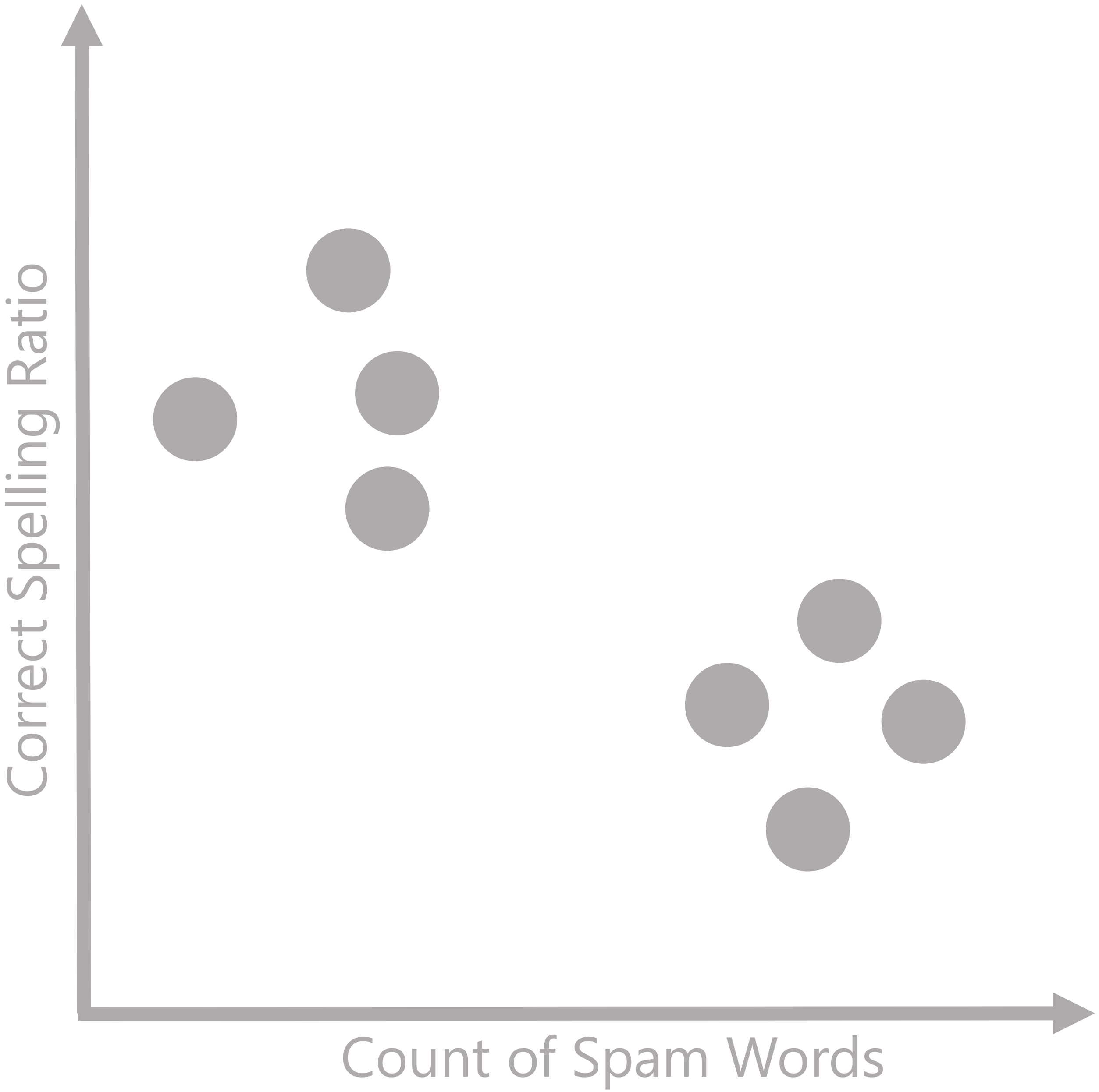
Classification

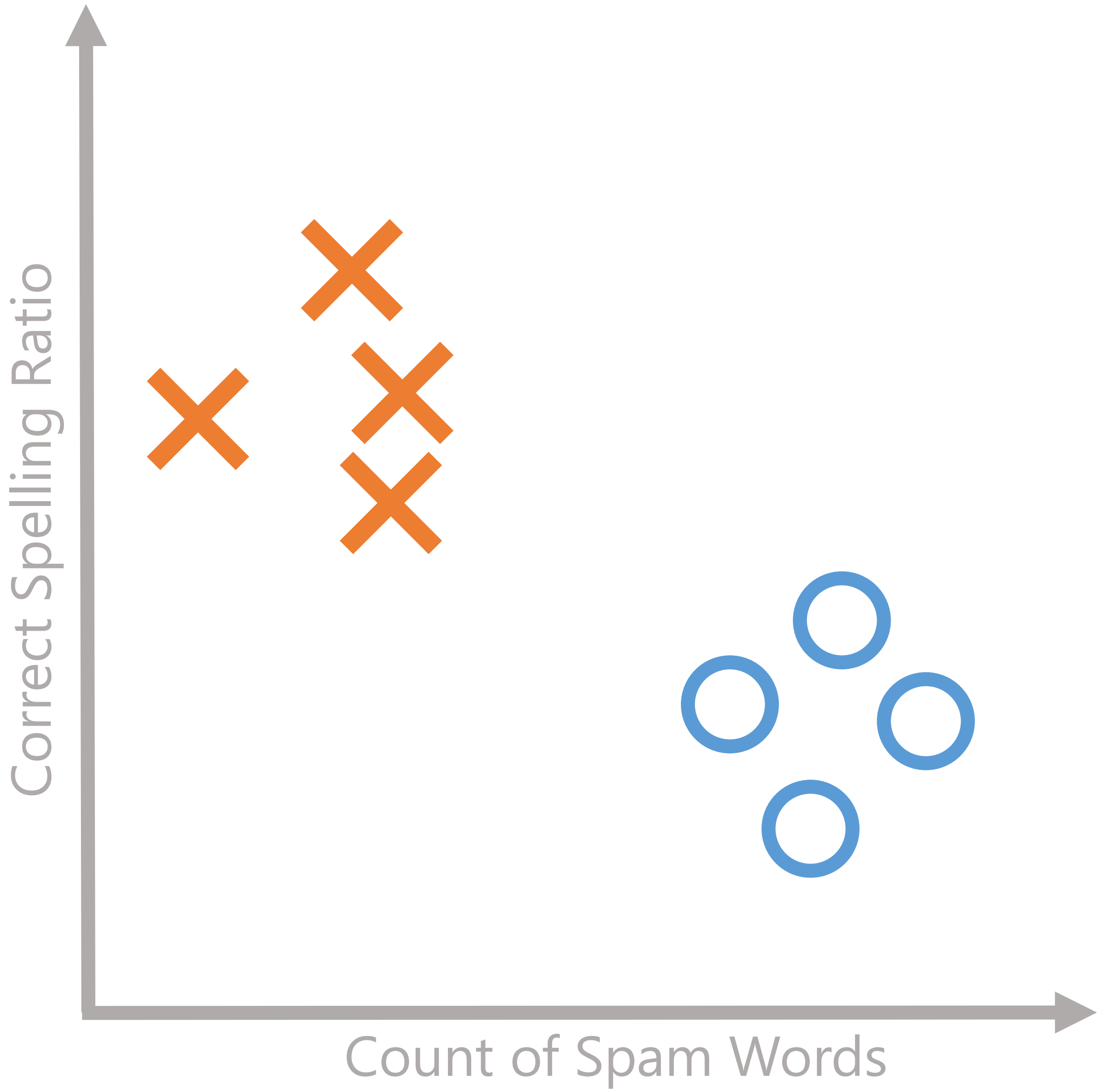

f(x)

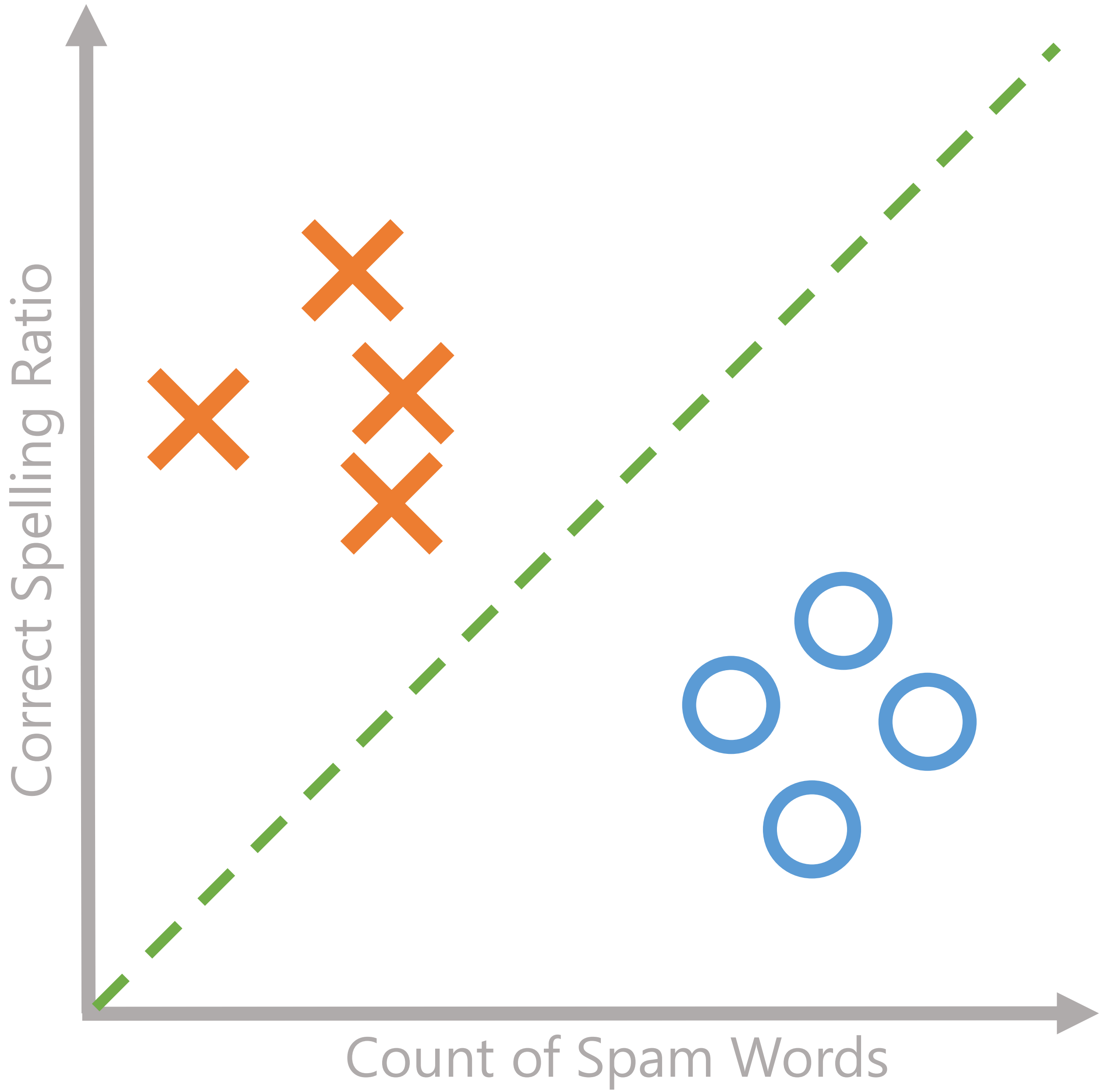

Correct Spelling Ratio
Count of Spam Words

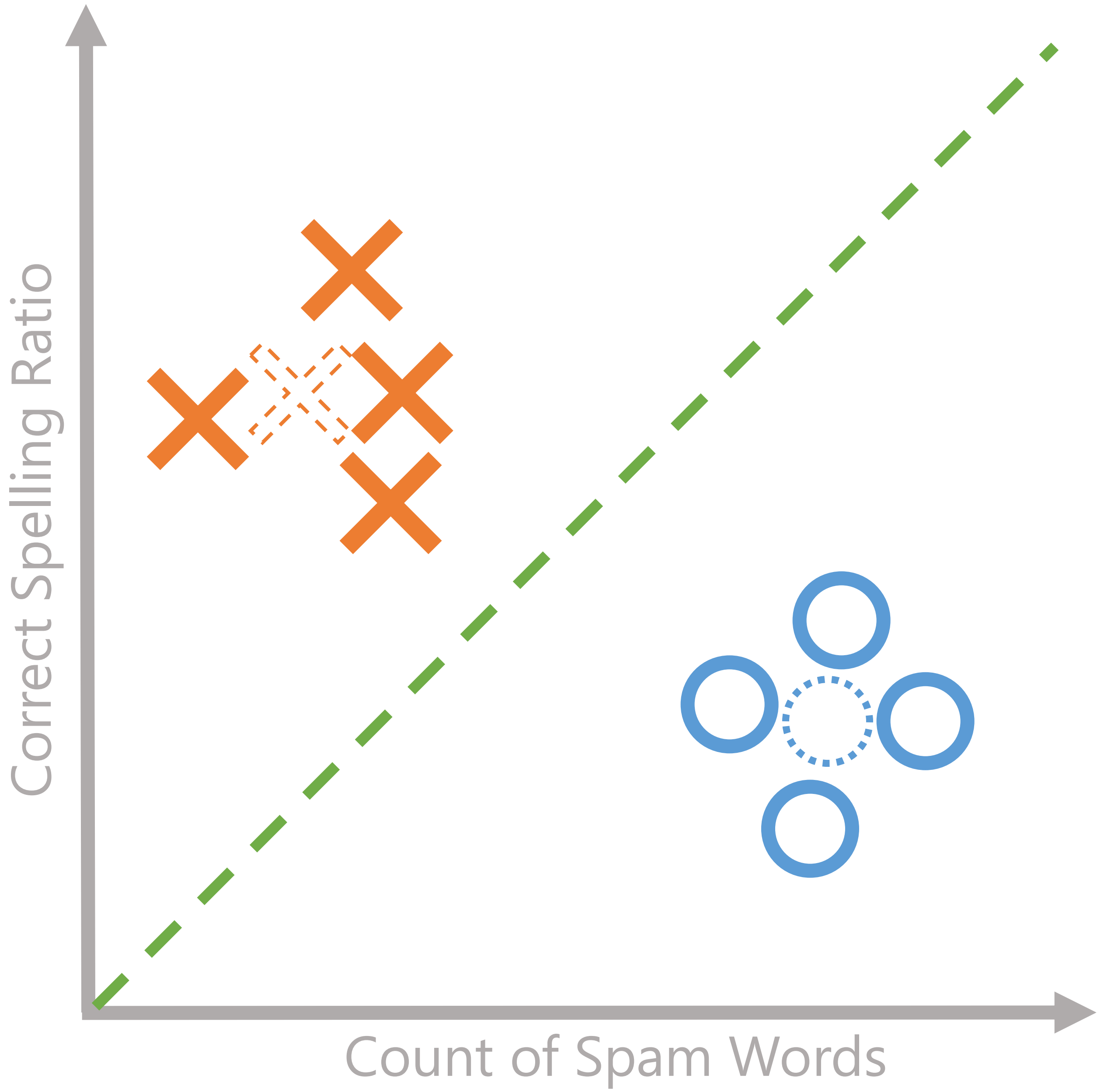


Correct Spelling Ratio
Count of Spam Words


Correct Spelling Ratio
Count of Spam Words


Correct Spelling Ratio
Count of Spam Words

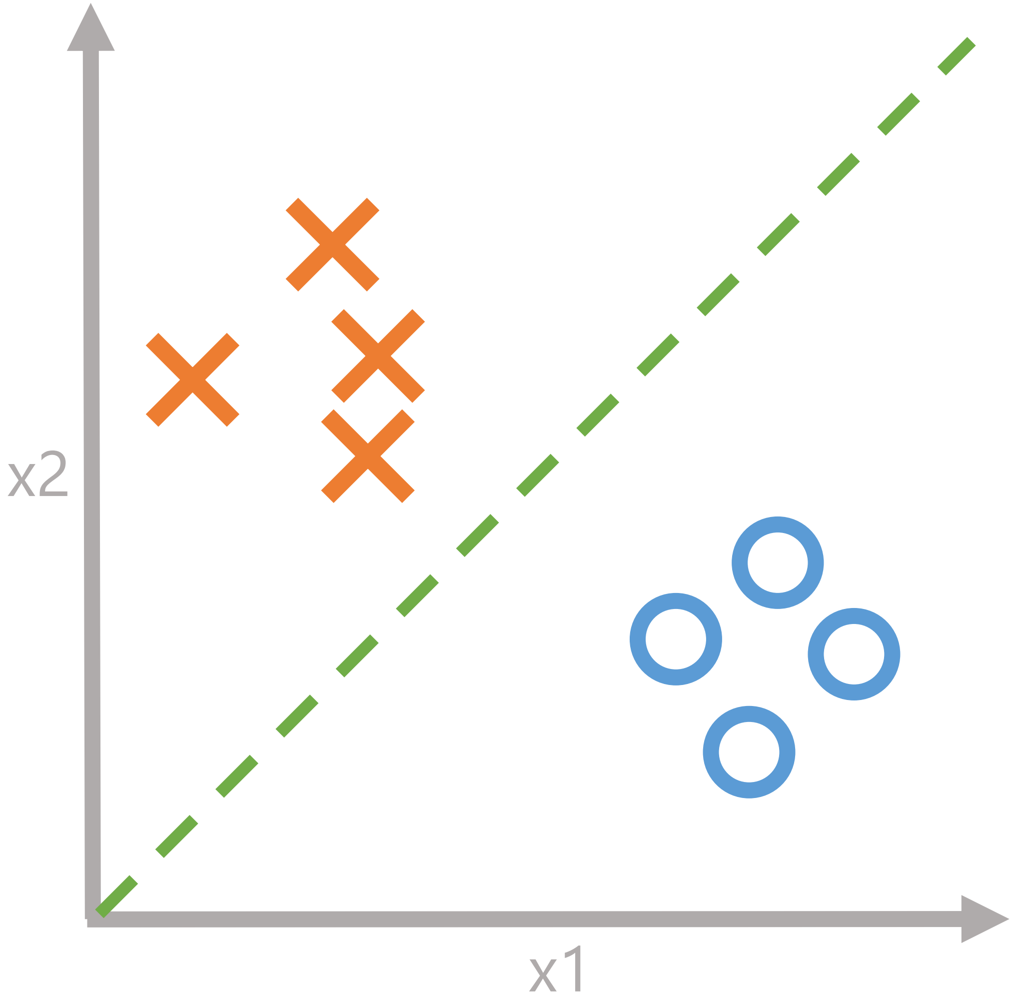

Correct Spelling Ratio
Count of Spam Words



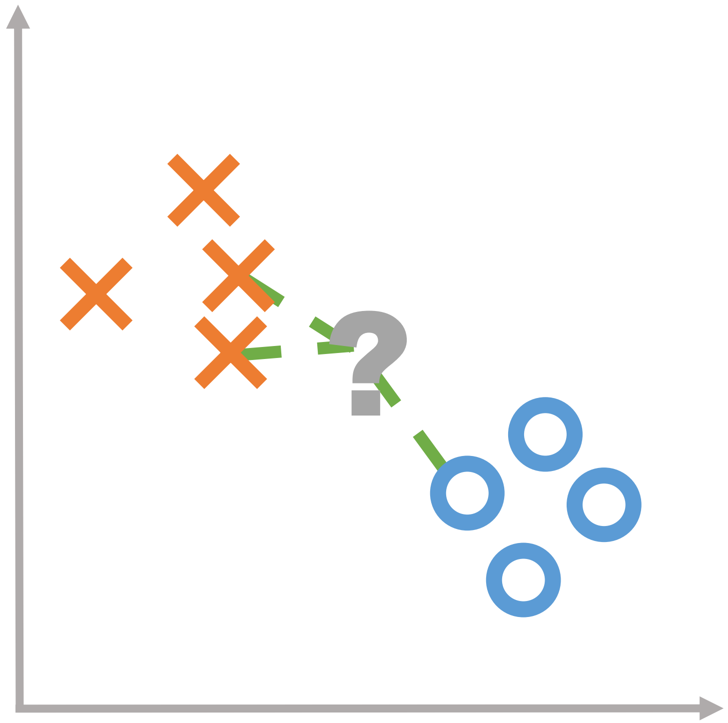
Correct Spelling Ratio
Count of Spam Words

Classification Algorithms

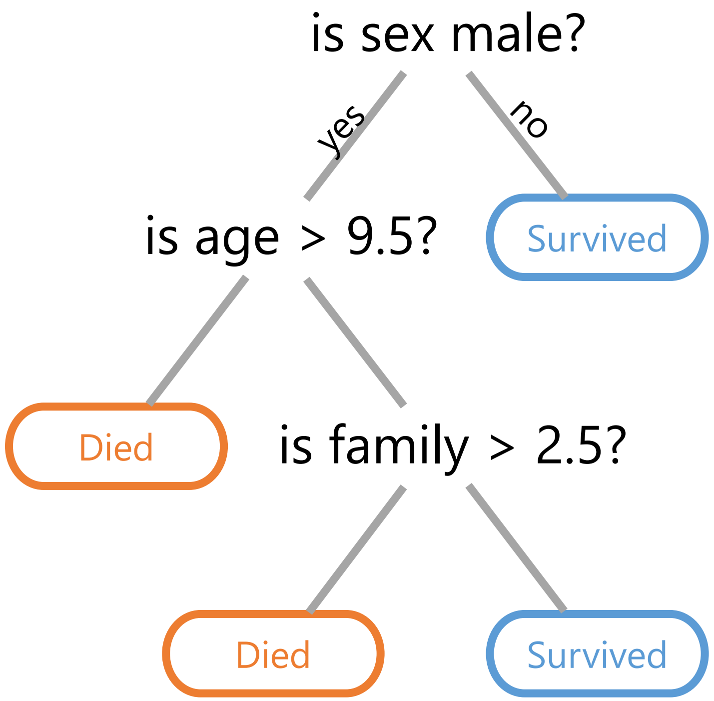
k-Nearest Neighbors
Decision Tree Classifier
Naïve Bayes Classifier
Support Vector Machine
Neural Network Classifier



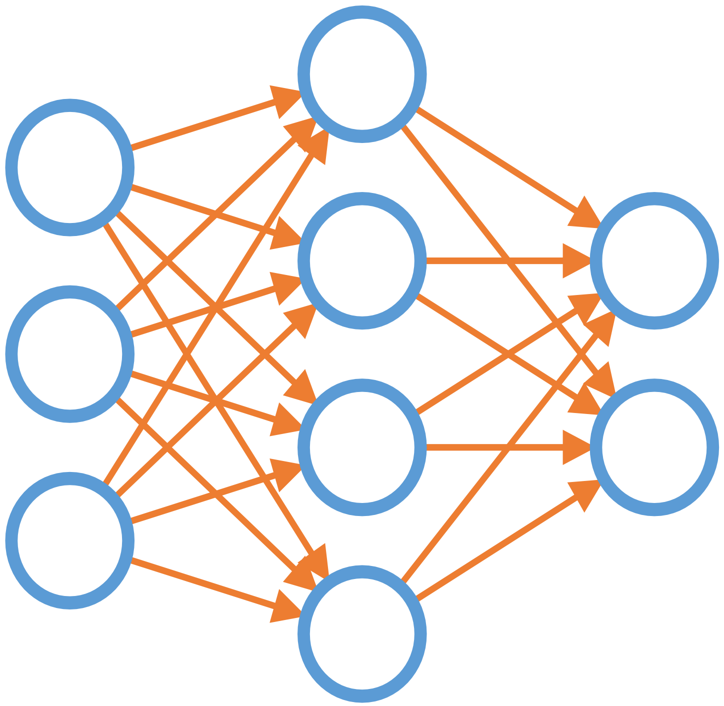
Classification Algorithms



k-Nearest Neighbors



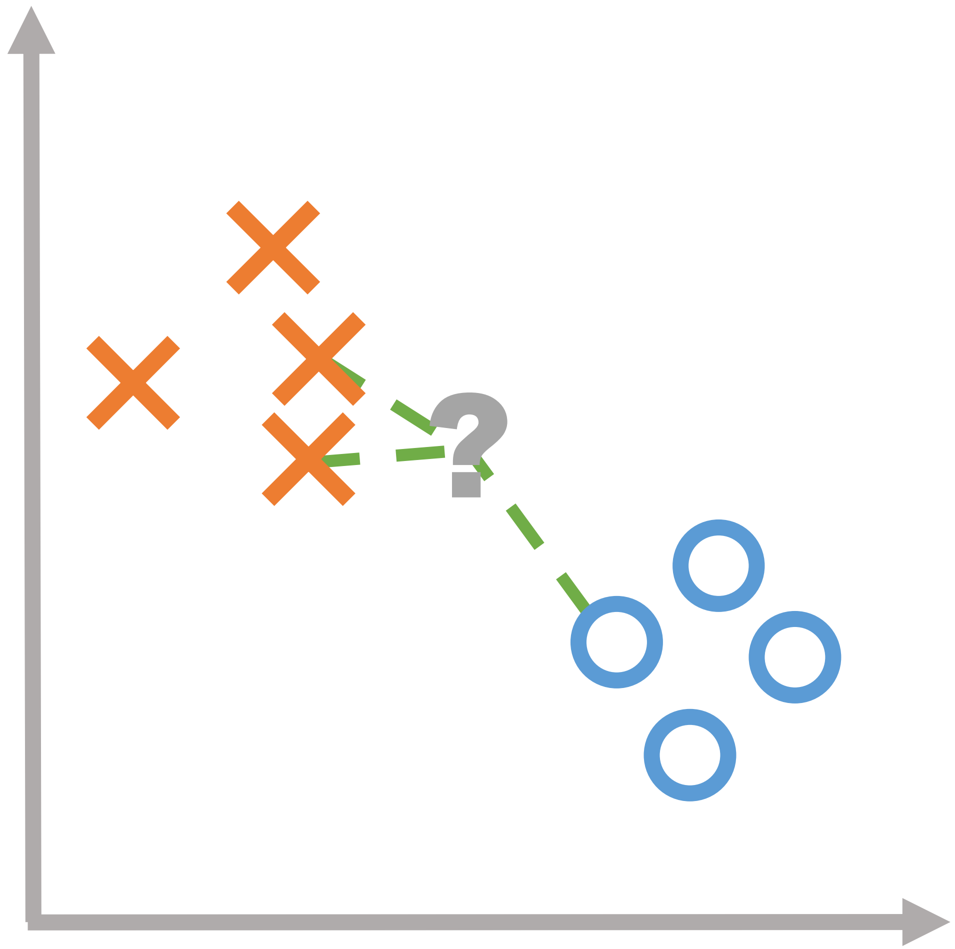
Decision Tree



Neural Network

K-Nearest Neighbors Classifier

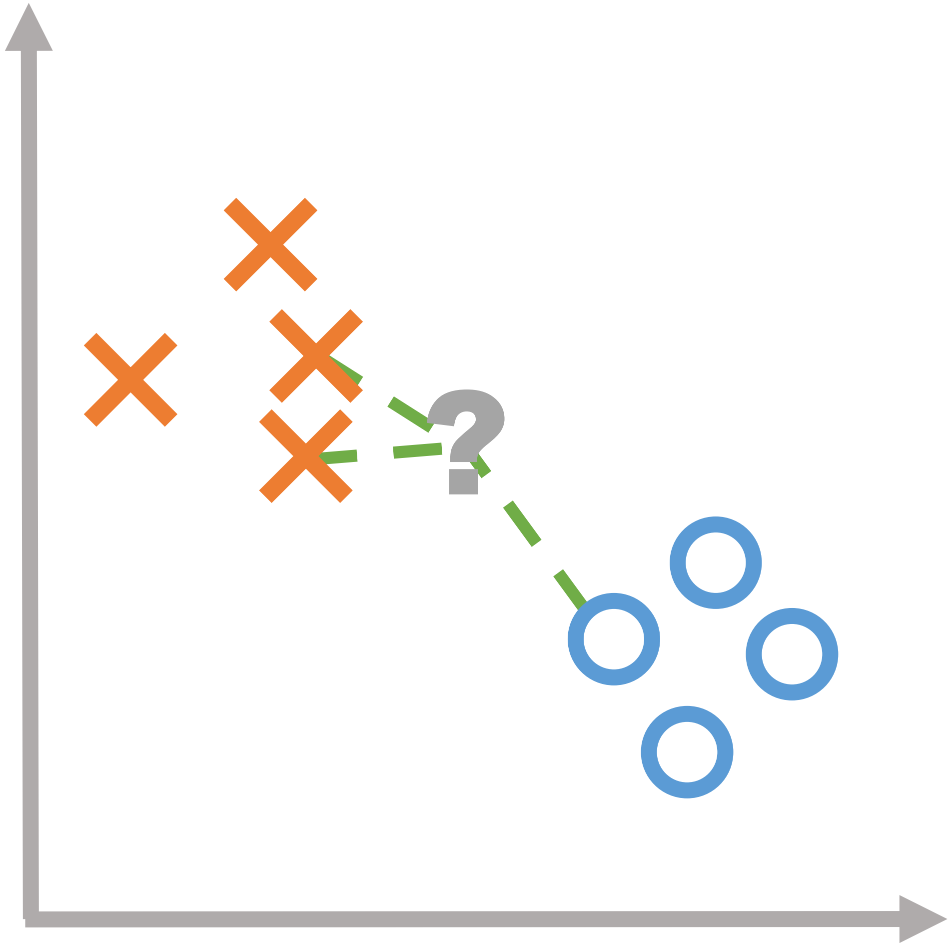
Supervised learning



K-Nearest Neighbors Classifier

Supervised learning

Uses class of neighbors

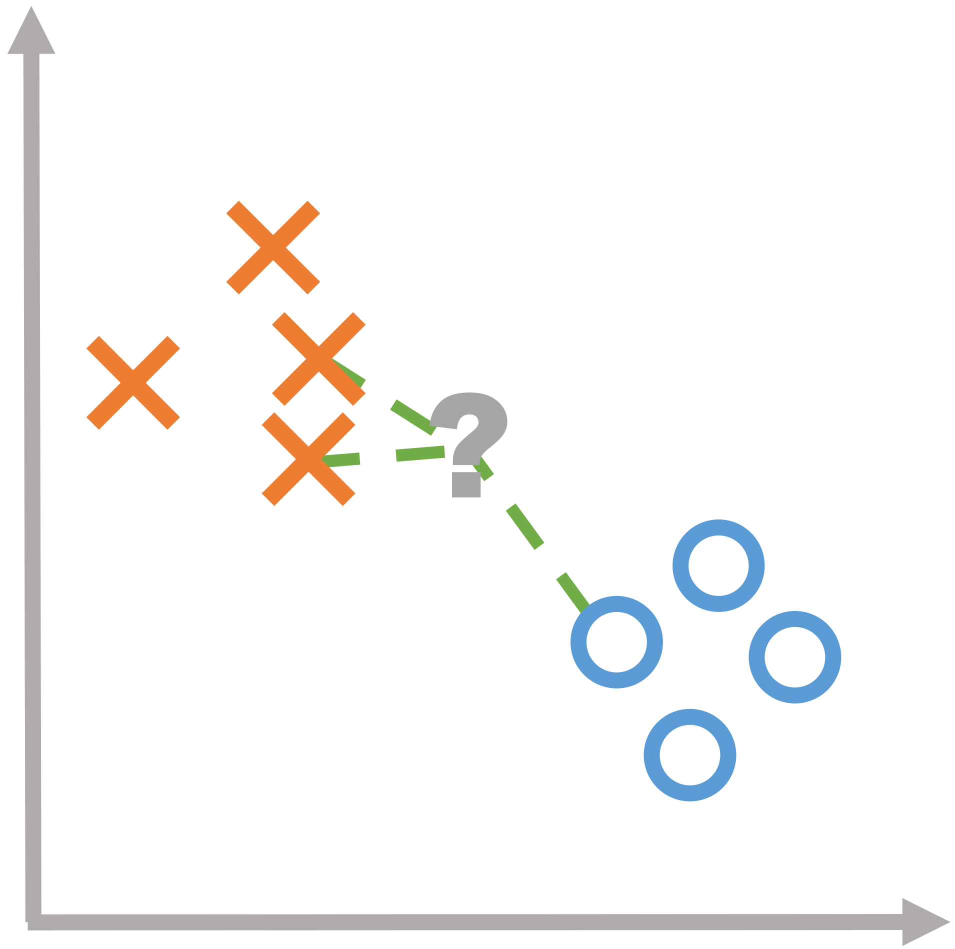


K-Nearest Neighbors Classifier

Supervised learning

Uses class of neighbors

k specifies how many



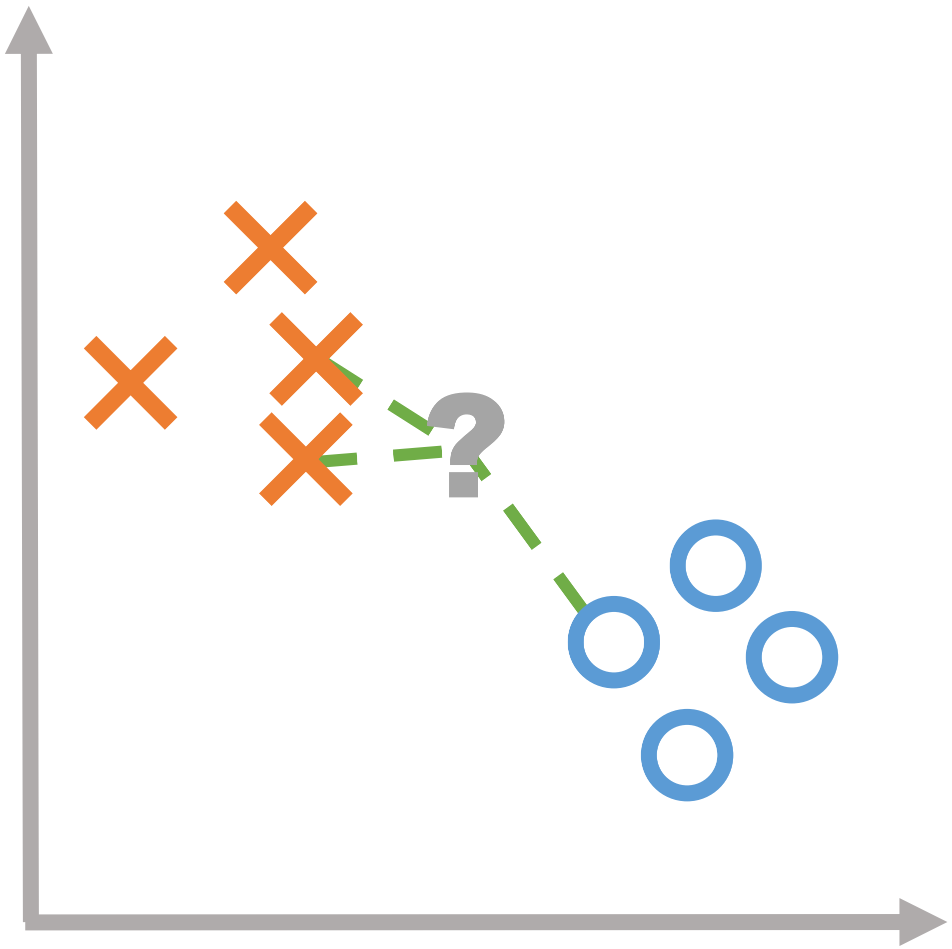
K-Nearest Neighbors Classifier

Supervised learning

Uses class of neighbors

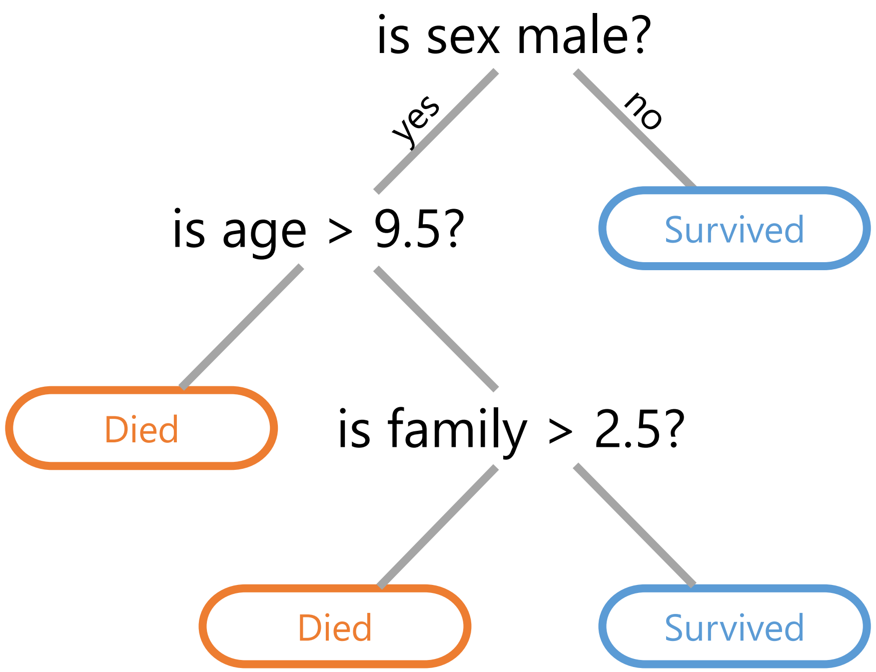
k specifies how many

Simple and easy



Decision Tree Classifier

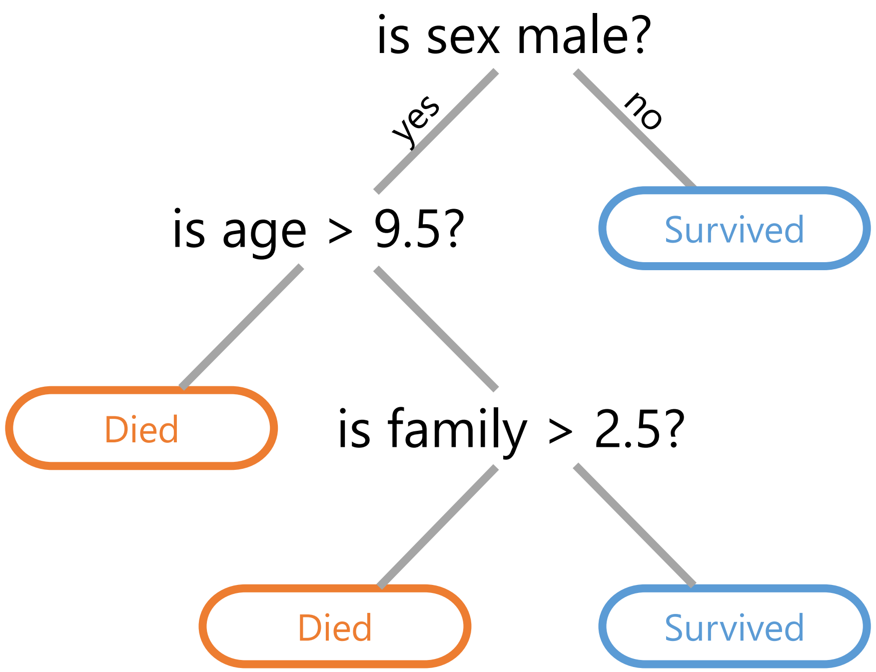
Supervised learning



Decision Tree Classifier

Supervised learning

Tree of decisions

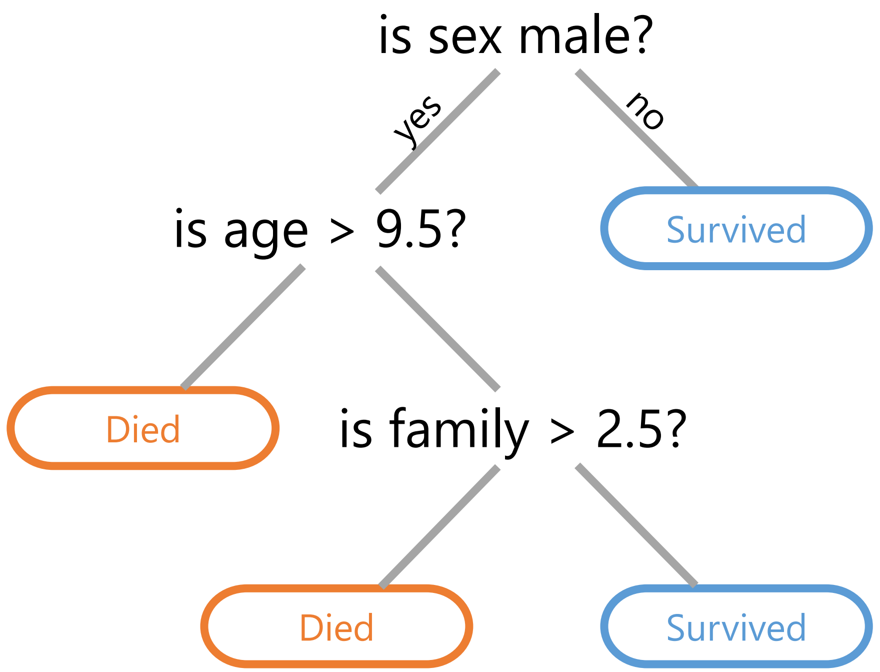


Decision Tree Classifier

Supervised learning

Tree of decisions

Information gain



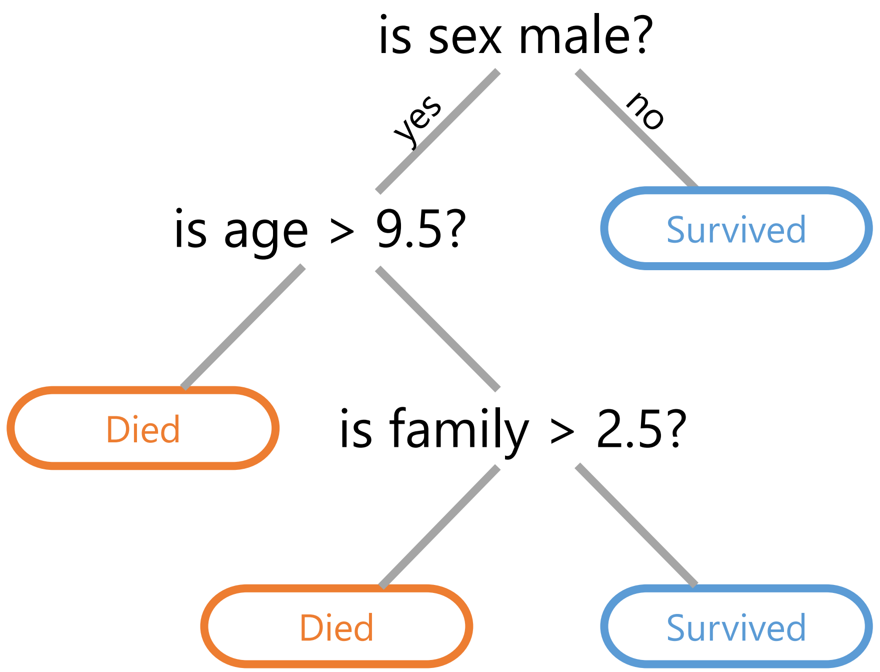
Decision Tree Classifier

Supervised learning

Tree of decisions

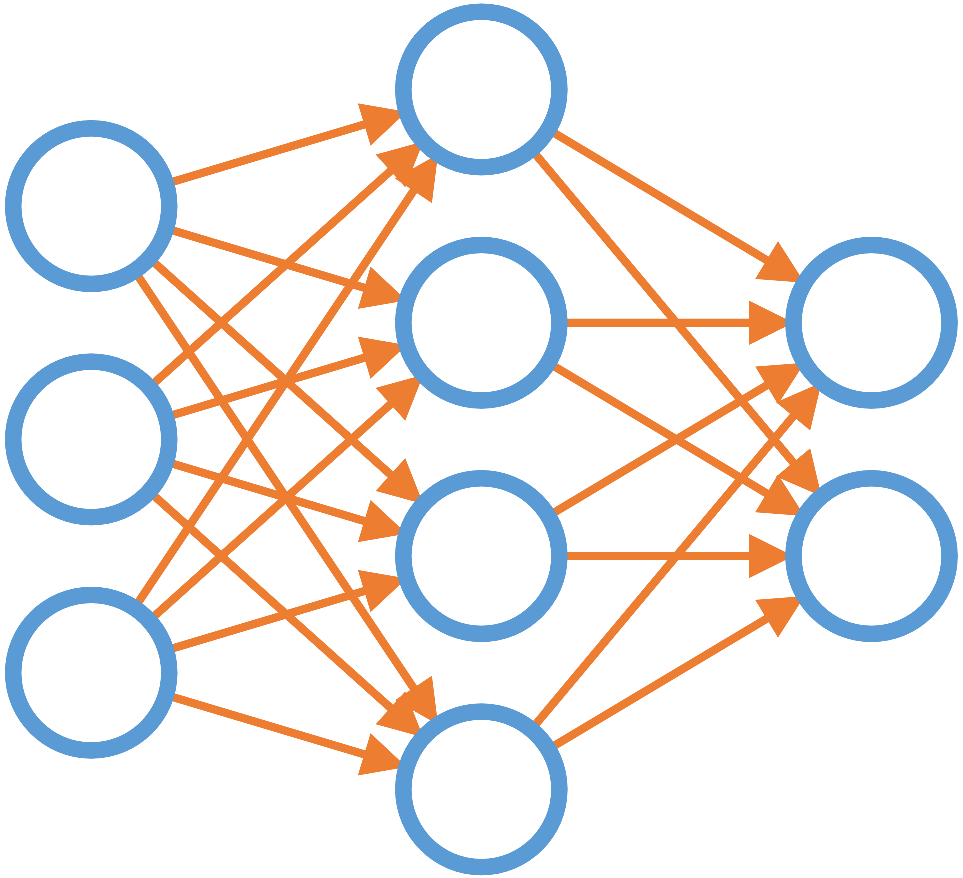
Information gain

Simple and easy



Neural Network Classifier

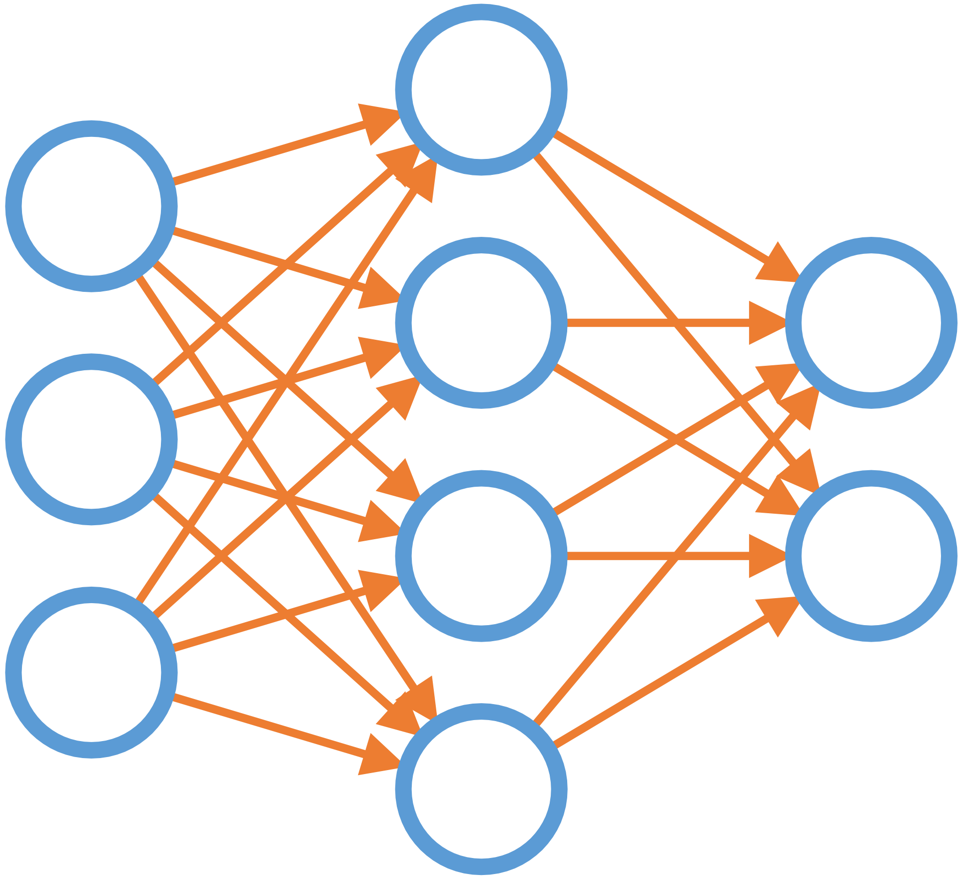
Supervised learning



Neural Network Classifier

Supervised learning

Neurons in a brain

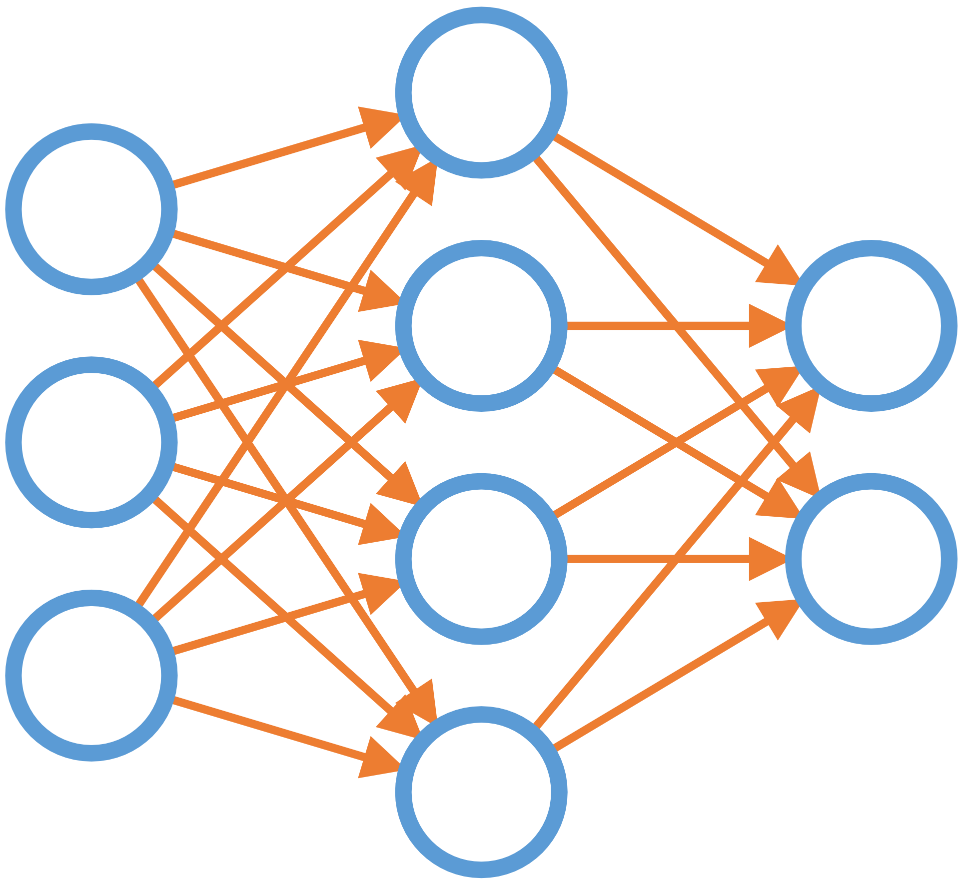


Neural Network Classifier

Supervised learning

Neurons in a brain

Weighted connections



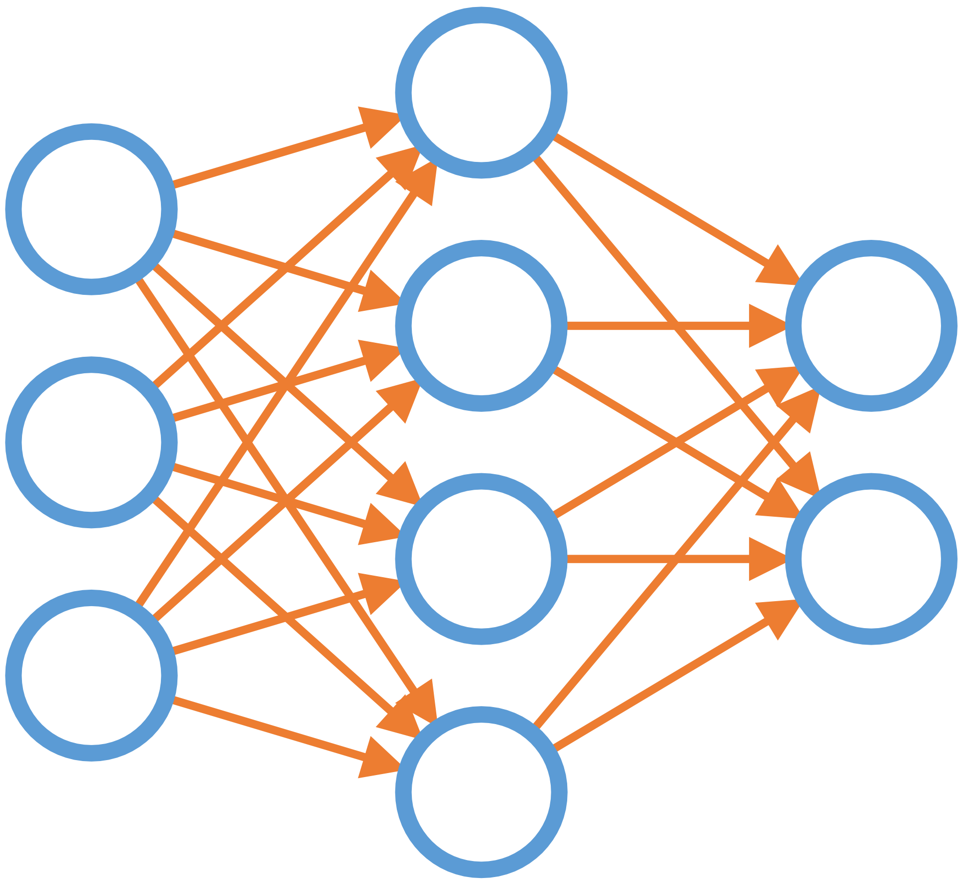
Neural Network Classifier

Supervised learning

Neurons in a brain

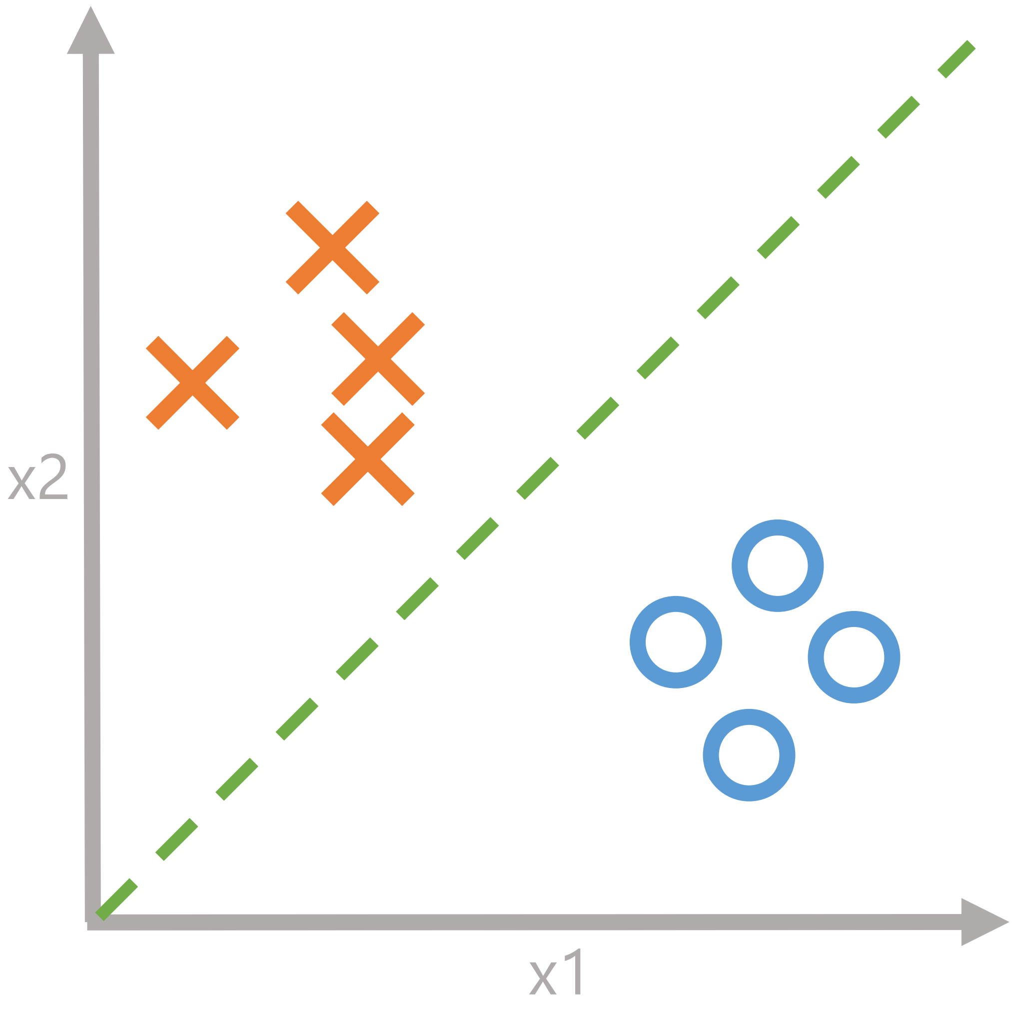
Weighted connections

Complex



Real-World Examples

Should we approve this loan?
Will this customer buy from us?
Should we replace this part?
Does this person have cancer?



Iris Data Set

Iris Setosa

Iris Versicolor

Iris Virginica

Photos by Radomił Binek, Danielle Langlois, and Frank Mayfield

Iris Data Set

Fisher's Iris Data				
Species	Petal Length	Petal Width	Sepal Length	Sepal Width
setosa	1.1	0.1	4.3	3
setosa	1.4	0.2	4.4	2.9
setosa	1.3	0.2	4.4	3
setosa	1.3	0.2	4.4	3.2
setosa	1.3	0.3	4.5	2.3
…		…	…	…

Demo 2 - Classification

Goal: Predict species based on petal and sepal measurements

Insurance Policy Data Set

Insurance Policy Risk							
Gender	State	State Rate	Height	Weight	BMI	Age	Risk
Male	MA	0.01	184	67.8	20.0	77	High
Male	VA	0.14	163	89.4	33.6	82	High
Female	NY	0.09	170	81.2	28.1	31	Low
Male	TN	0.12	175	99.7	32.6	39	Low
Female	FL	0.11	184	72.1	21.3	68	High
…		…	…	…	…	…	…

Insurance Policy Data Set

Insurance Policy Rates							
Gender	State	State Rate	Height	Weight	BMI	Age	Rate
Male	MA	0.01	184	67.8	20.0	77	0.33
Male	VA	0.14	163	89.4	33.6	82	0.87
Female	NY	0.09	170	81.2	28.1	31	0.01
Male	TN	0.12	175	99.7	32.6	39	0.02
Female	FL	0.11	184	72.1	21.3	68	0.15
…		…	…	…	…	…	…

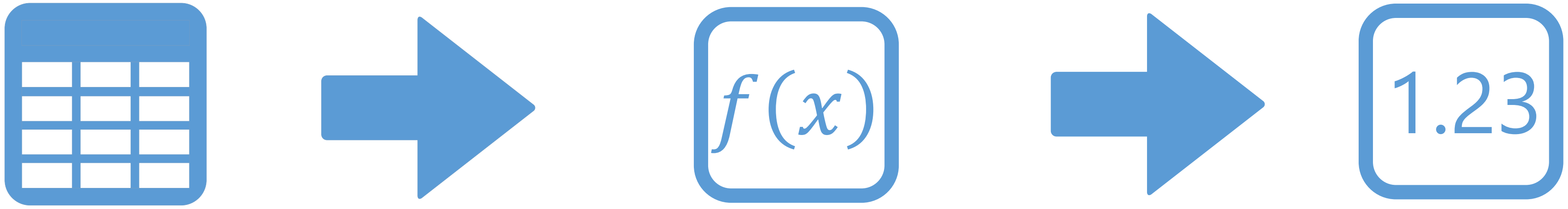
Lab 2A – Classification (Easy)

Goal: Predict species based on petal and sepal measurements

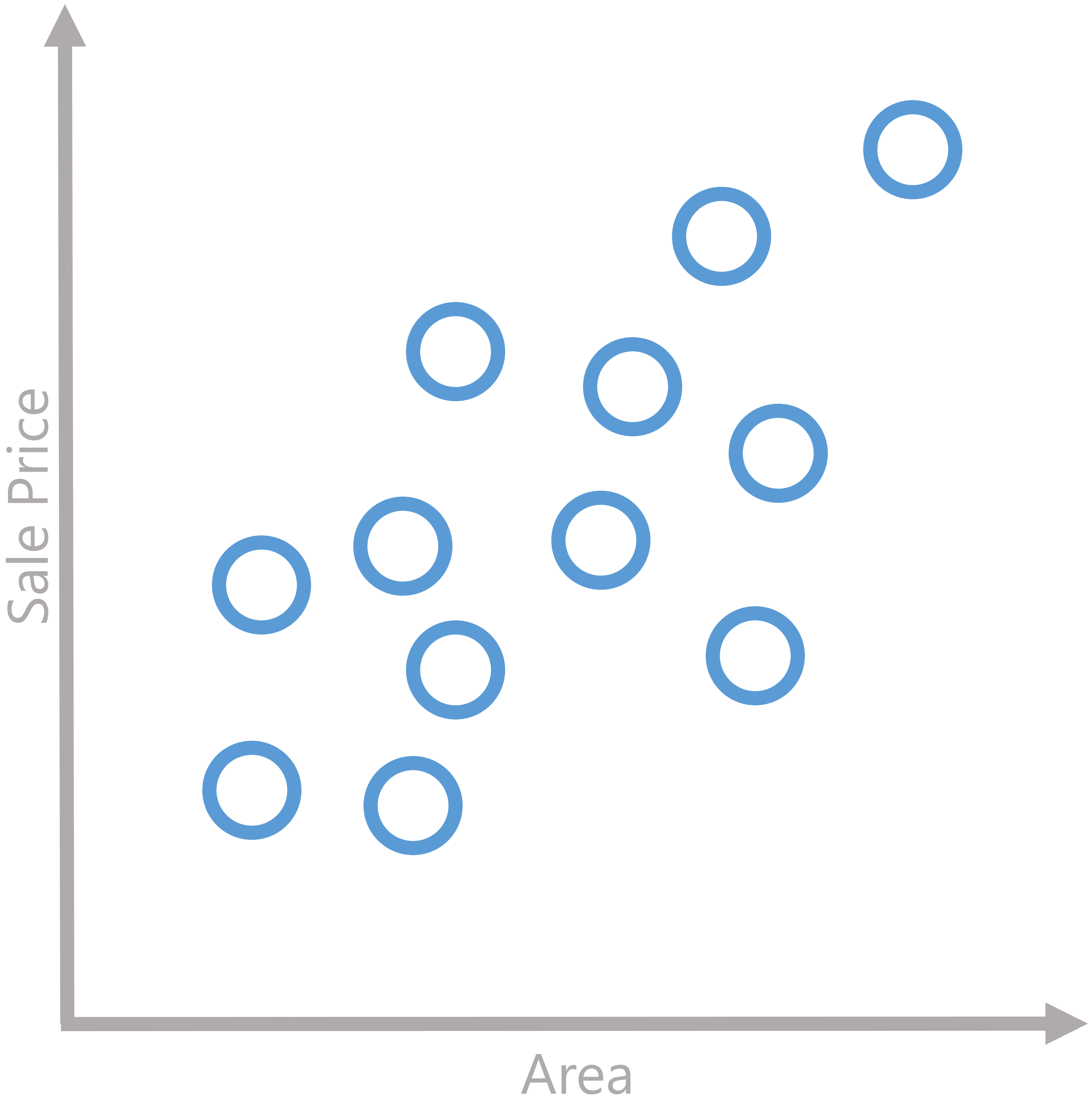
Lab 2B – Classification (Hard)

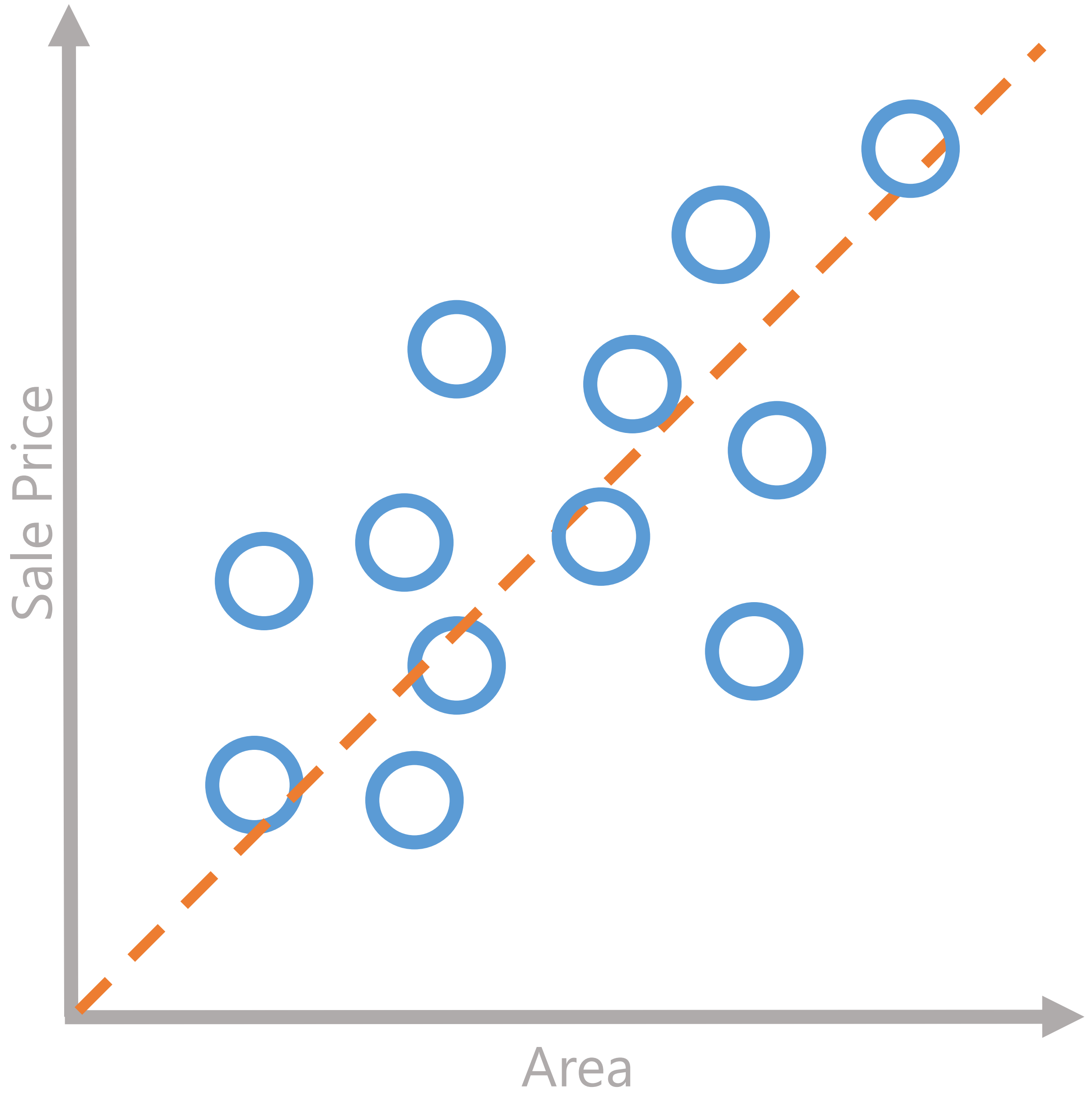
Goal: Predict the risk of
an insurance policy

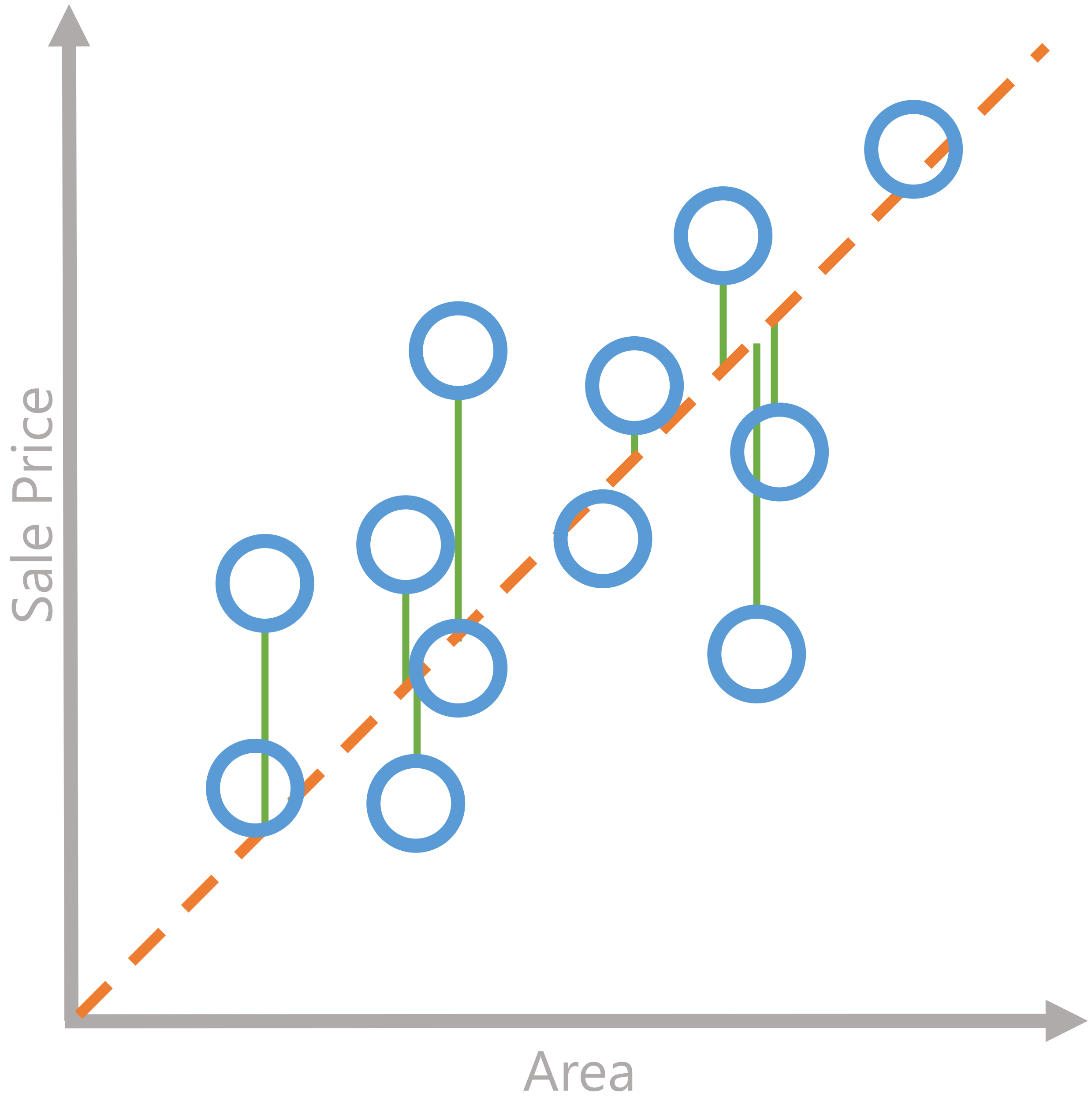
Regression

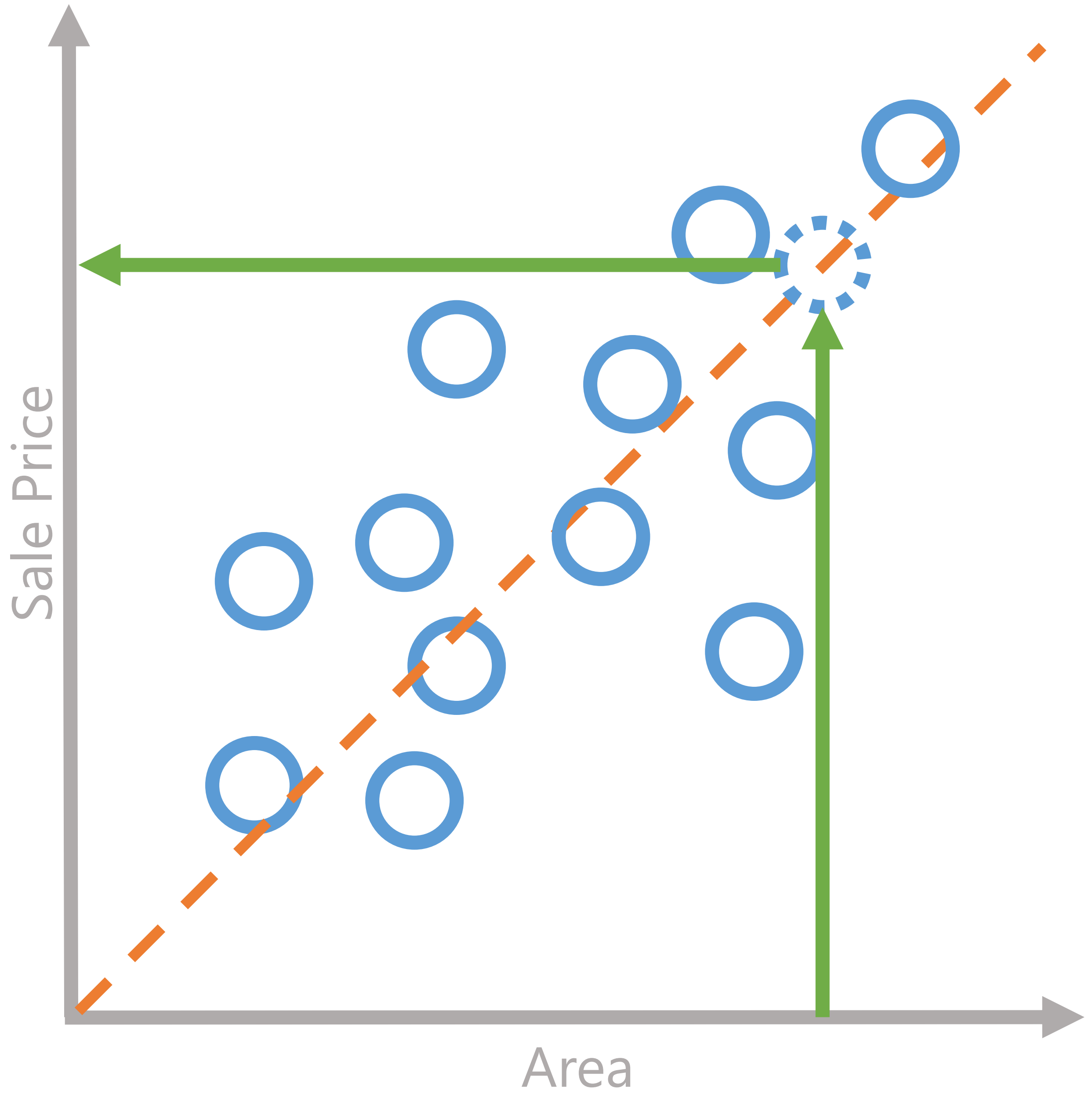


f(x)
1.23


Sale Price
Area


Sale Price
Area

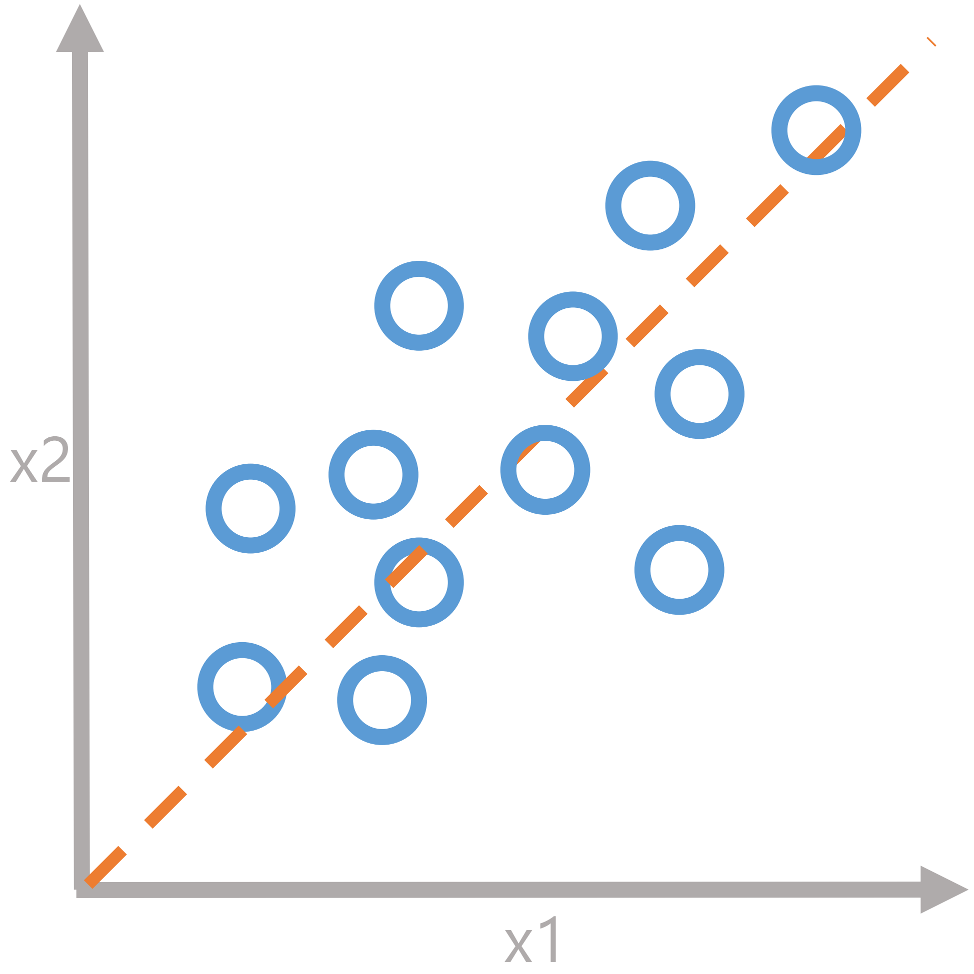

Sale Price
Area



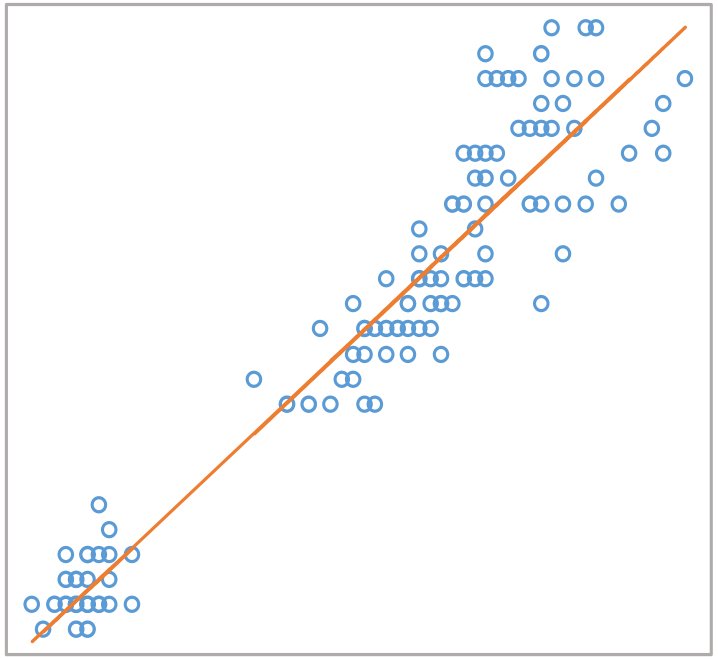
Sale Price
Area

Regression Algorithms

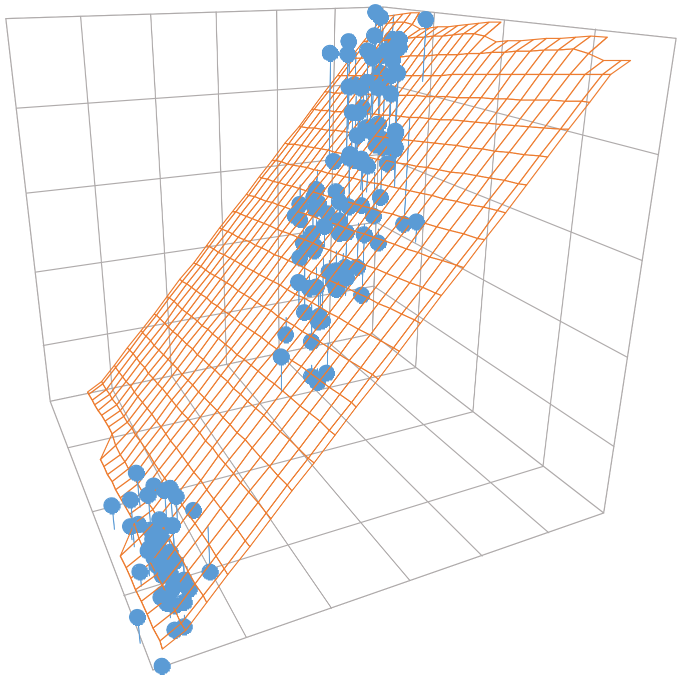
Linear Regression
Polynomial Regression
Lasso Regression
ElasticNet Regression
Neural Network Regression



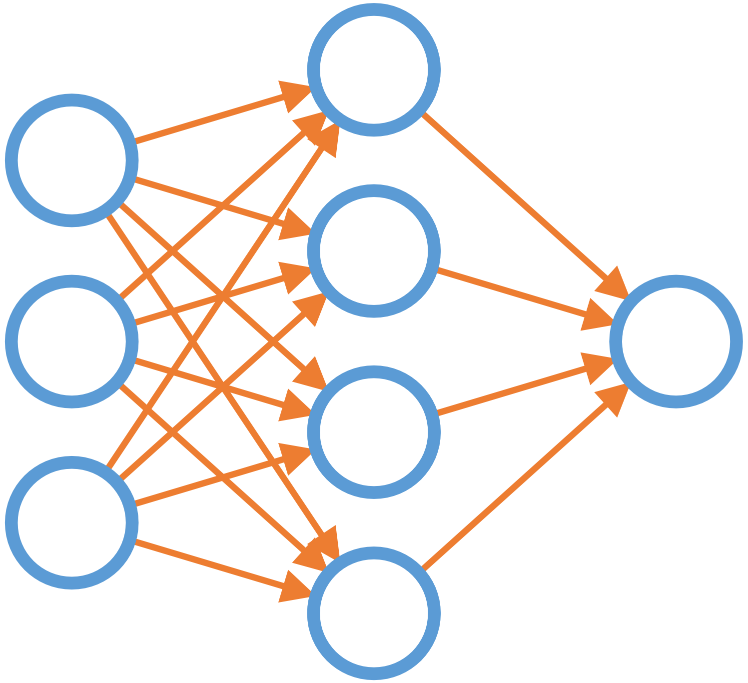
Regression Algorithms



Simple Linear



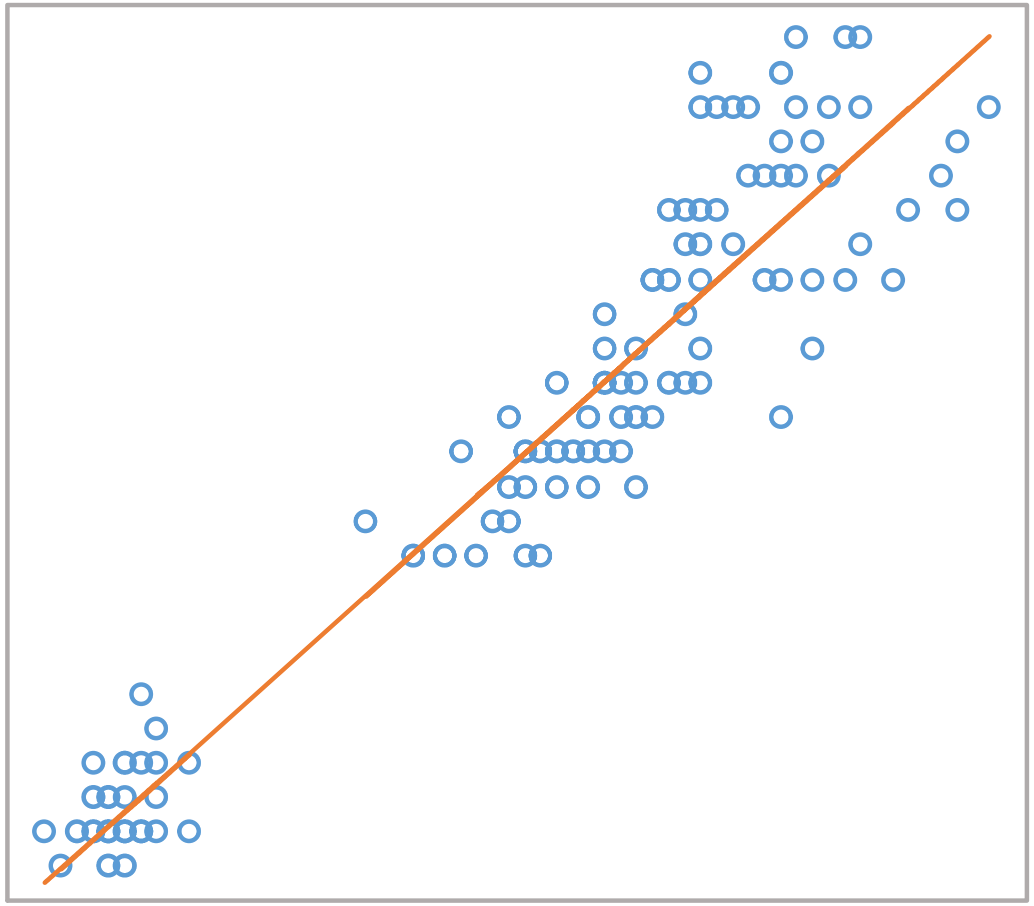
Multiple Linear



Neural Network

Simple Linear Regression

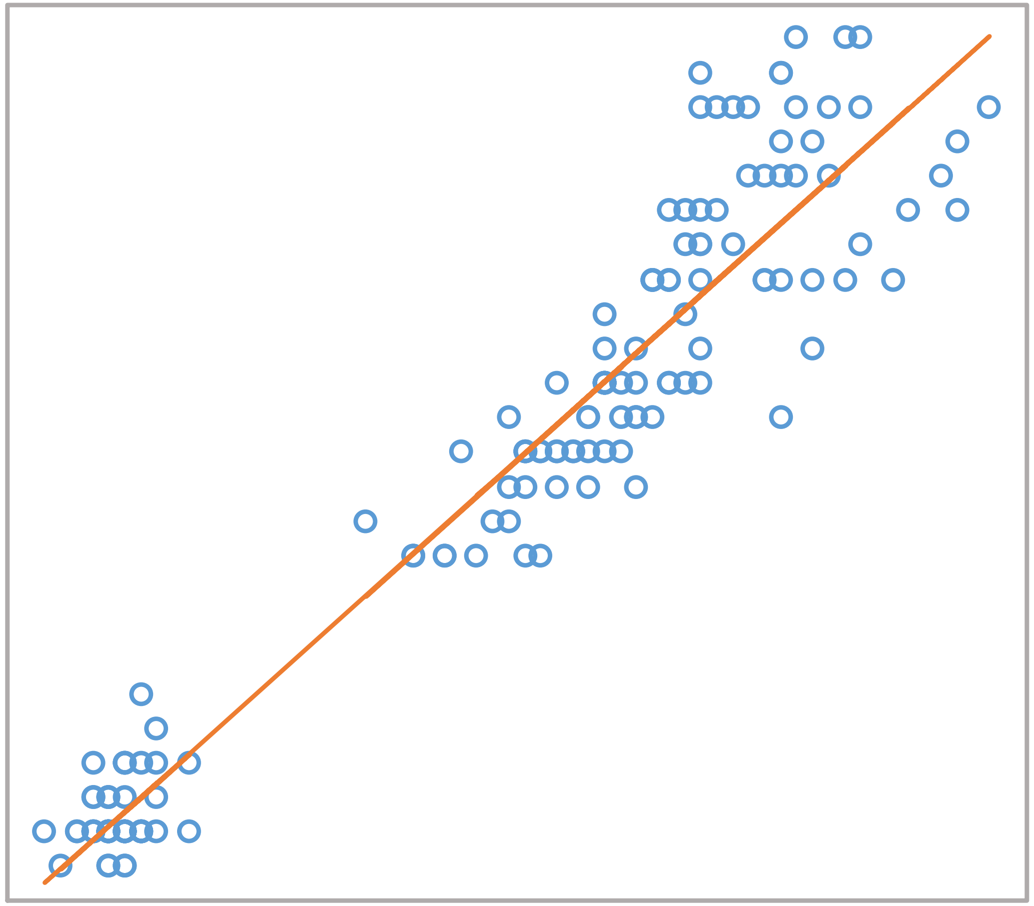
Relationship



Simple Linear Regression

Relationship

Linear model

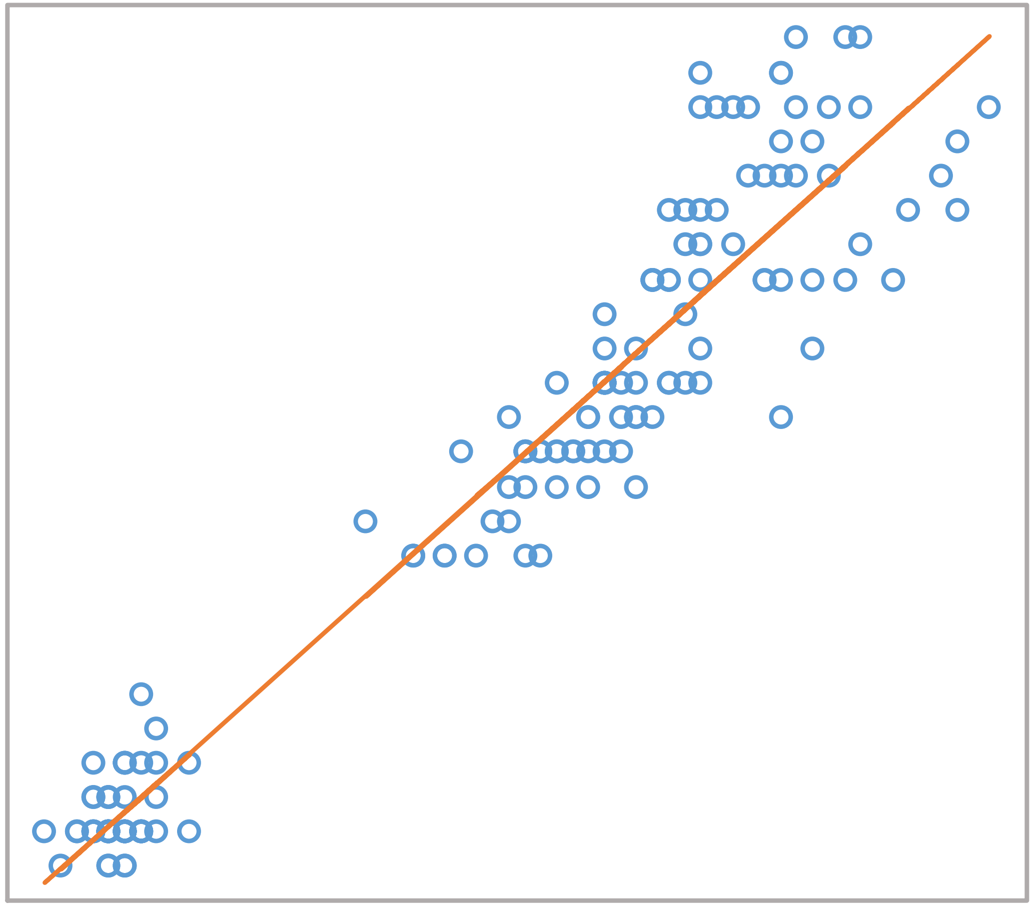


Simple Linear Regression

Relationship

Linear model

$y = m \cdot x + b$



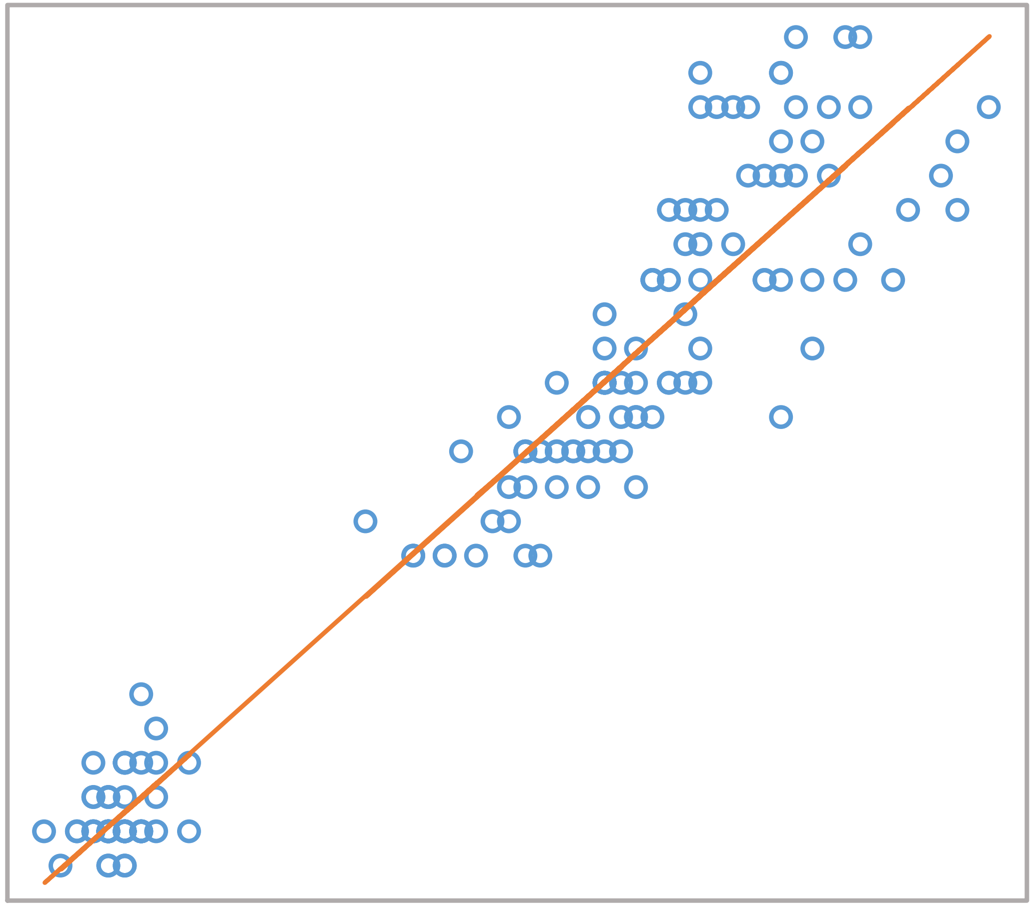
Simple Linear Regression

Relationship

Linear model

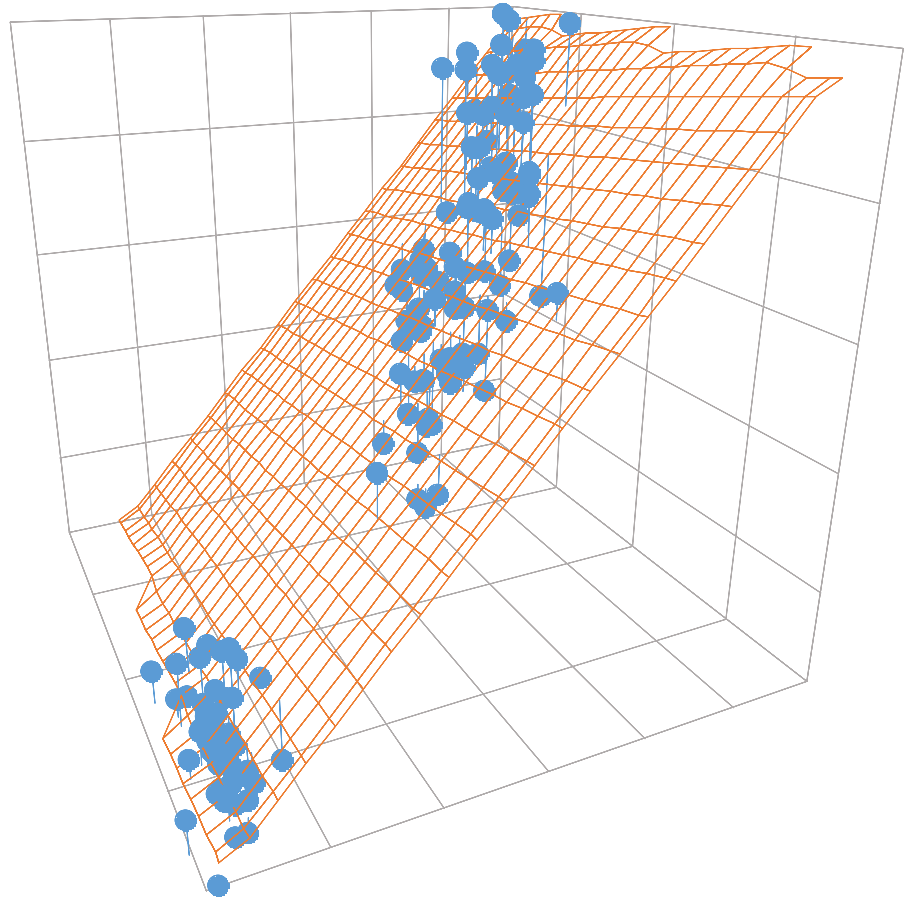
$y = m \cdot x + b$

Parameters estimated



Multiple Linear Regression

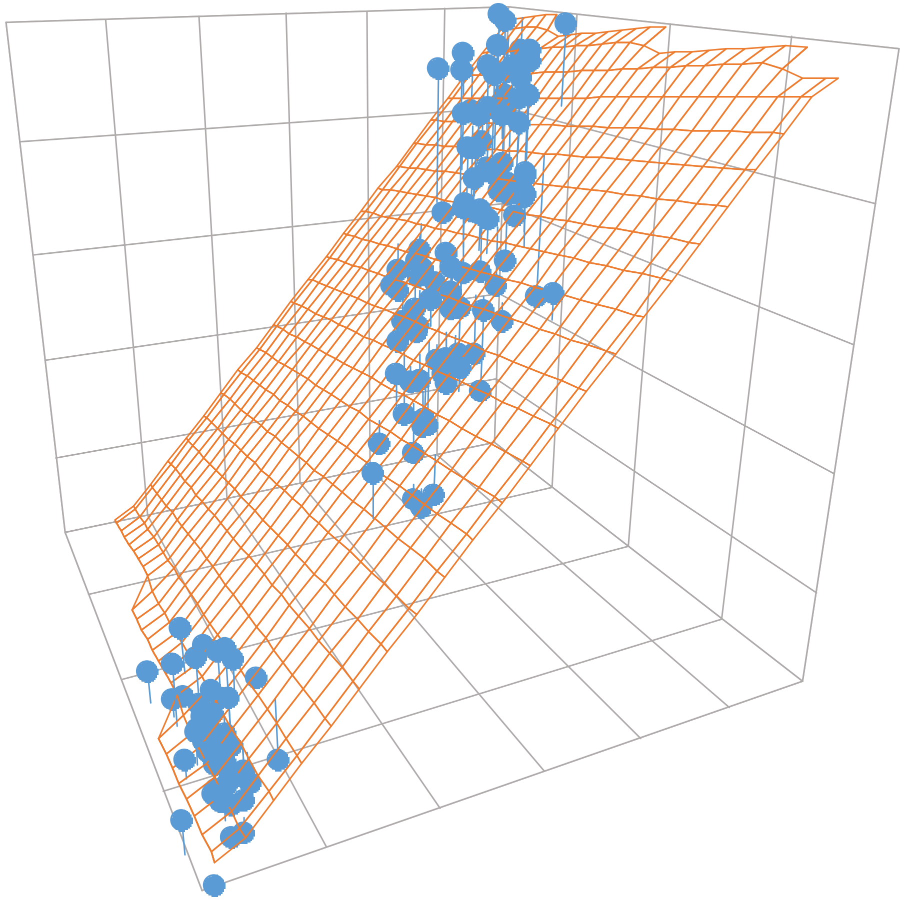
Similar to SLR



Multiple Linear Regression

Similar to SLR

Multiple variables

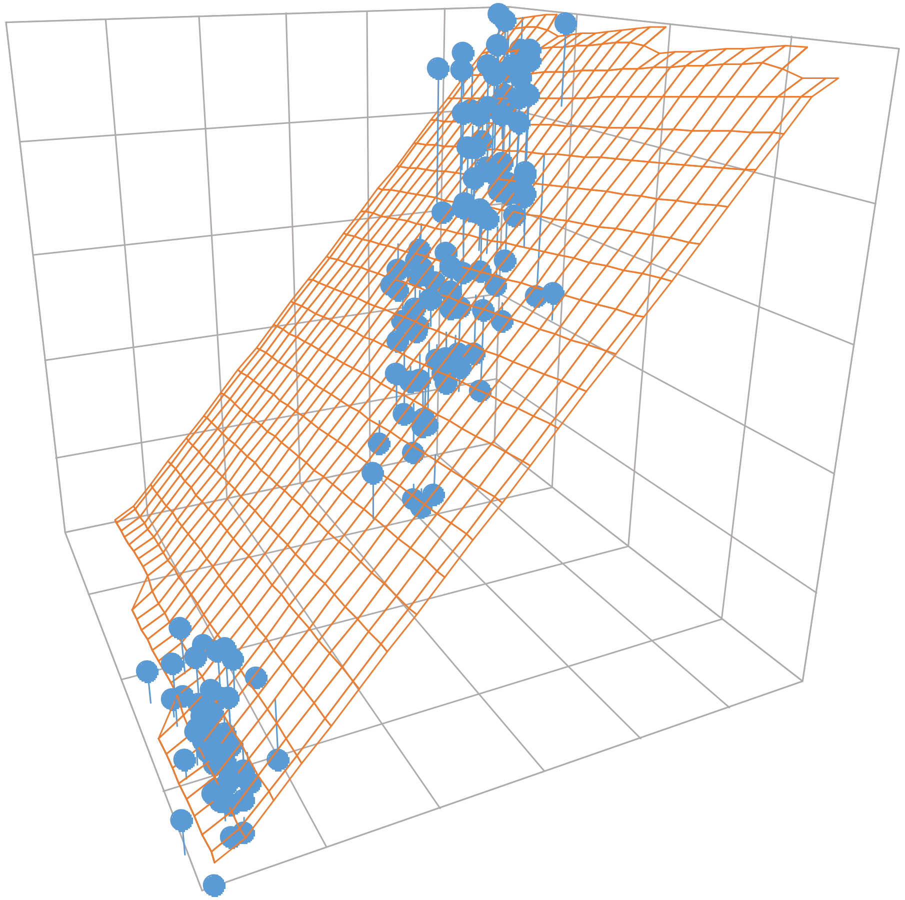


Multiple Linear Regression

Similar to SLR

Multiple variables

Multiple slopes



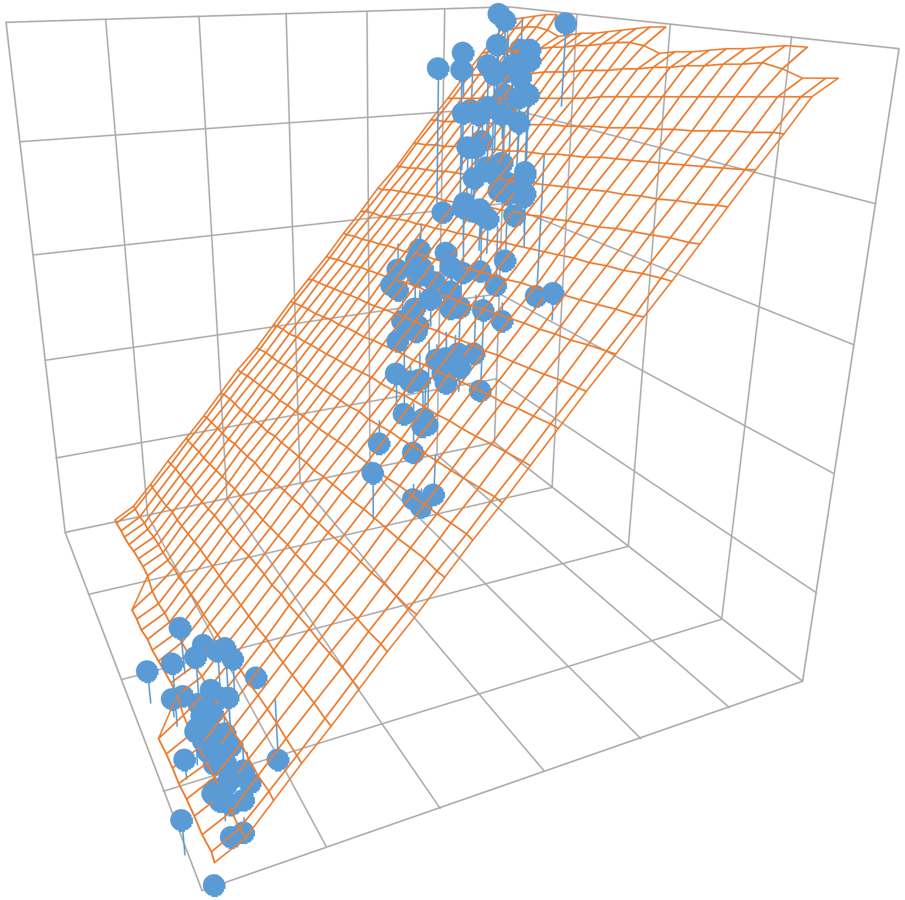
Multiple Linear Regression

Similar to SLR

Multiple variables

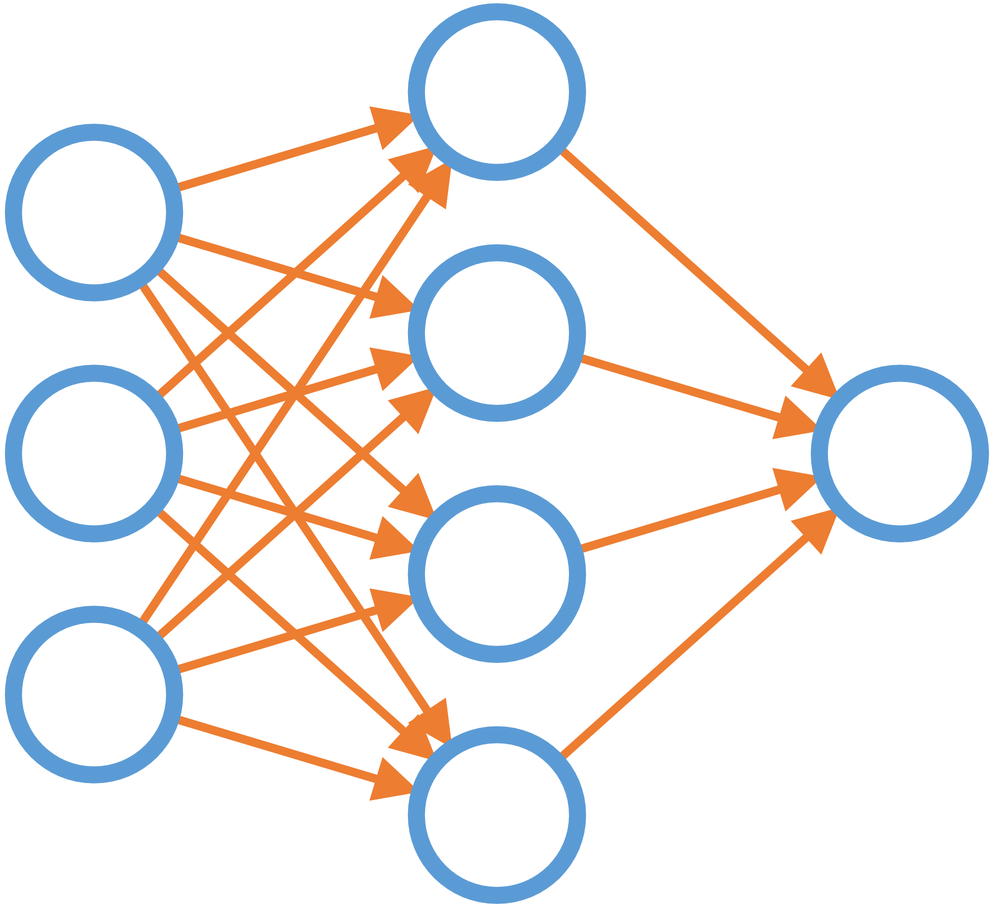
Multiple slopes

Categorical variables



Neural Network Regression

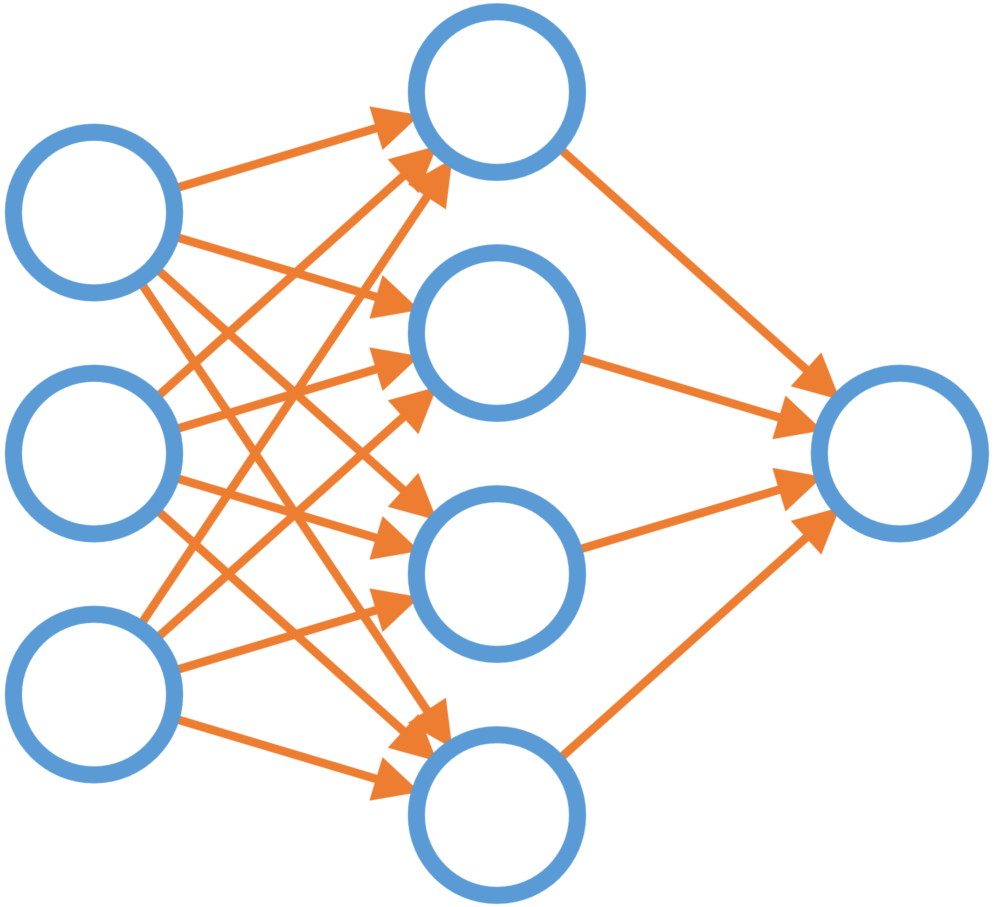
Similar to NN classifier



Neural Network Regression

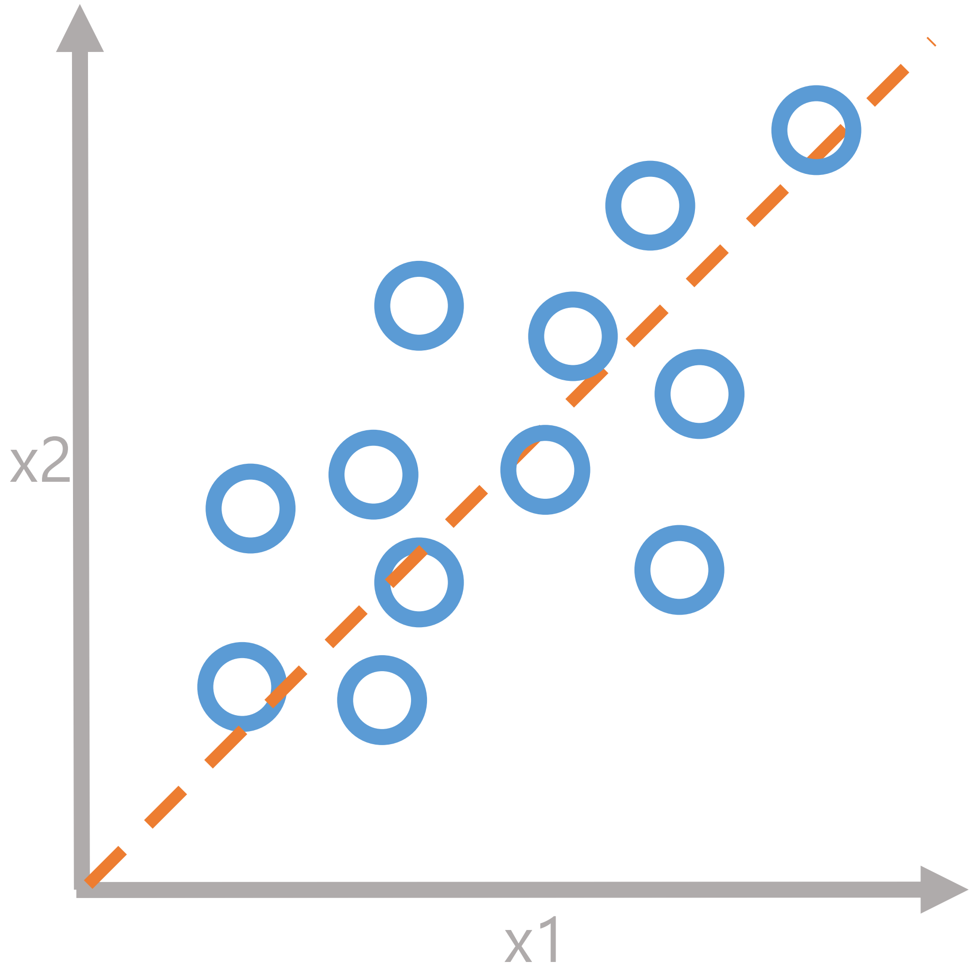
Similar to NN classifier

Numeric output



Real-World Examples

How much profit will we make?
What will the price be tomorrow?
How many units will they buy?
How long until this part fails?



Demo 3 - Regression

Goal: Predict petal width
based on petal length

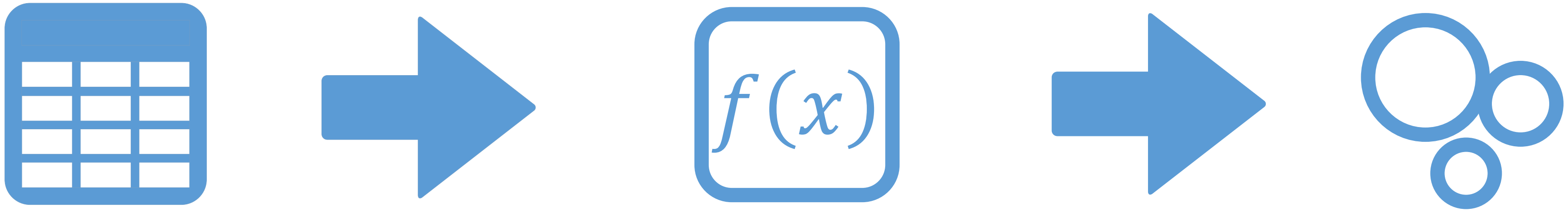
Lab 3A – Regression (Easy)

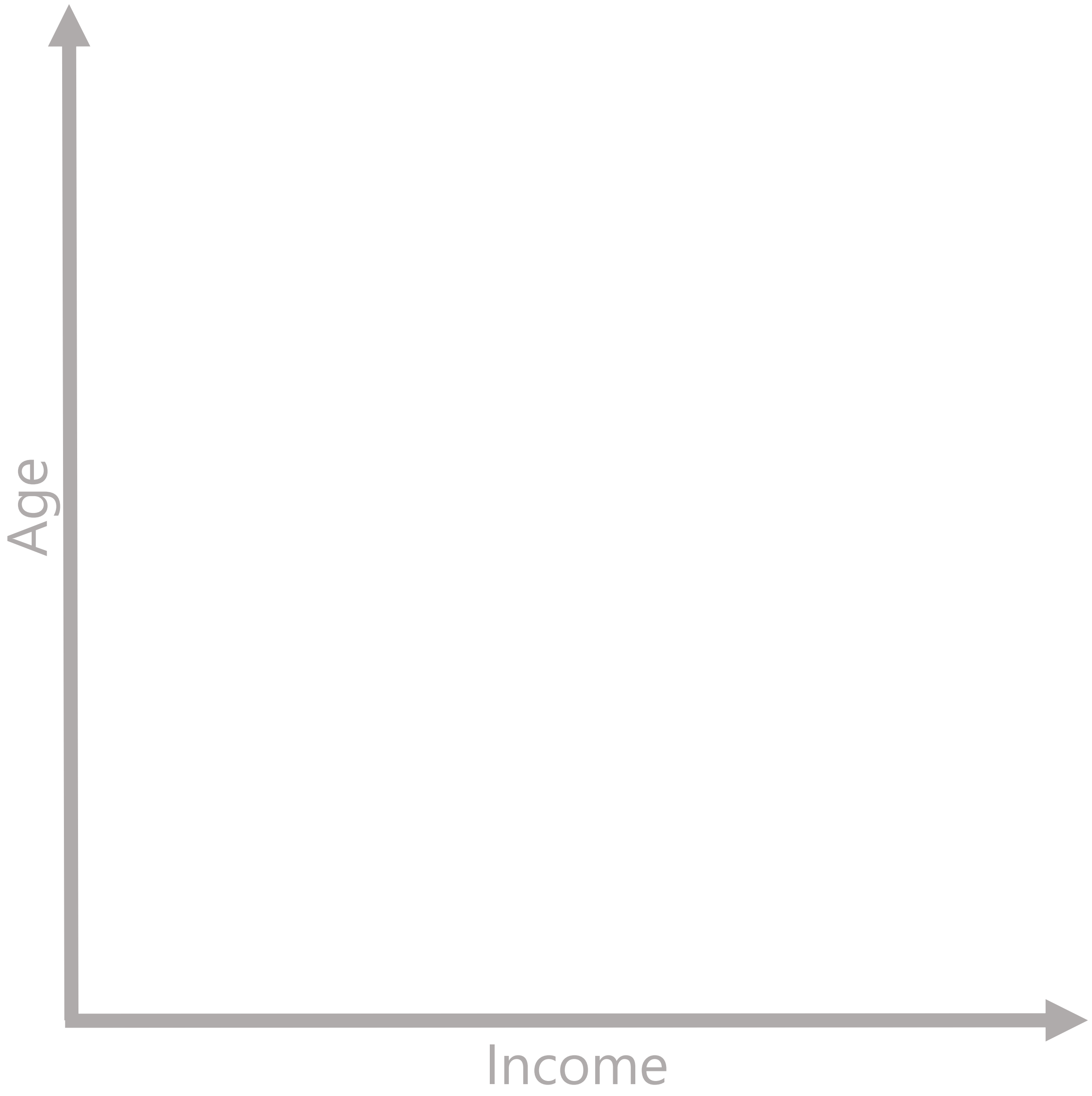
Goal: Predict petal width

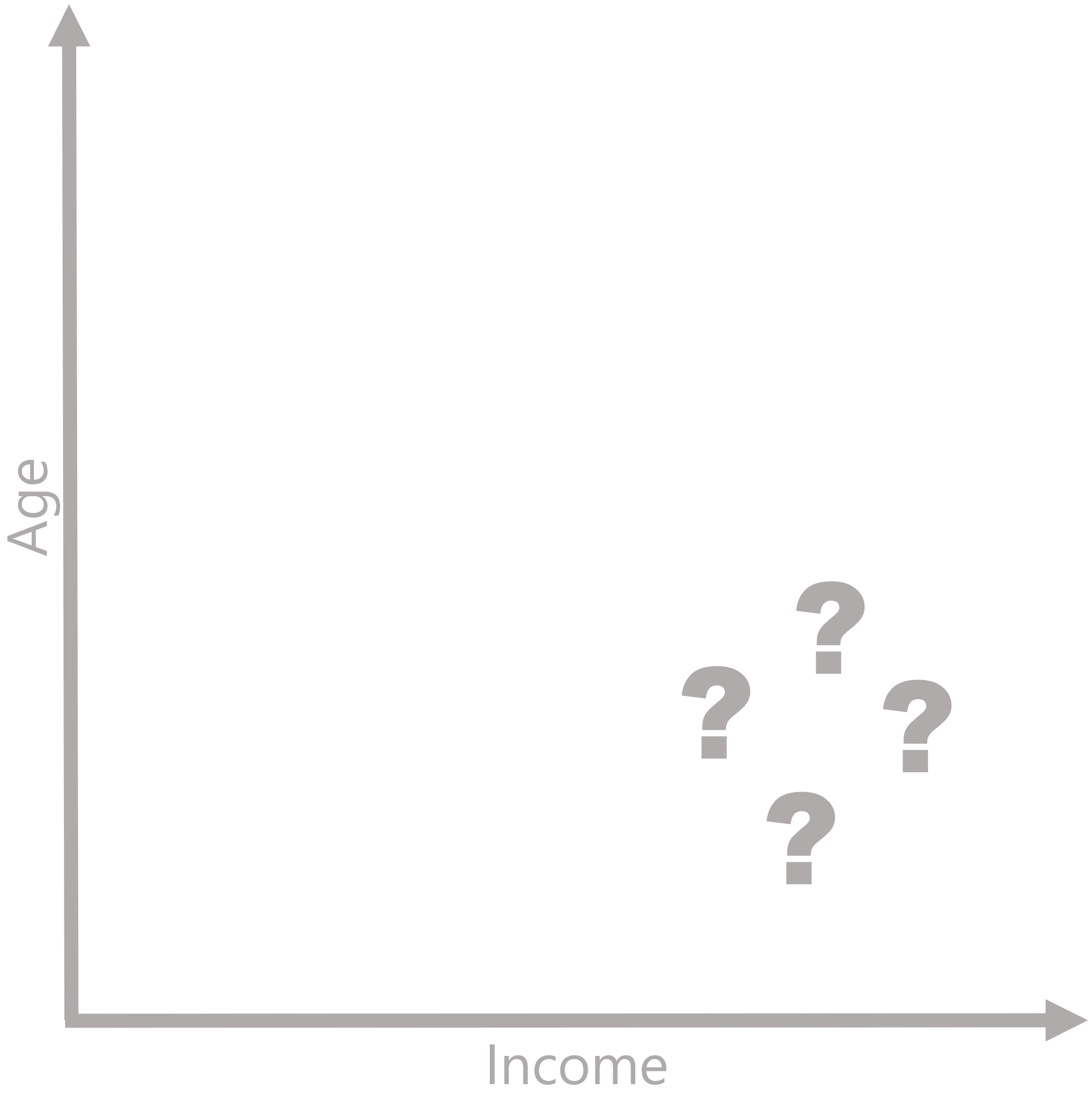
Lab 3B – Regression (Hard)

Goal: Predict mortality rate

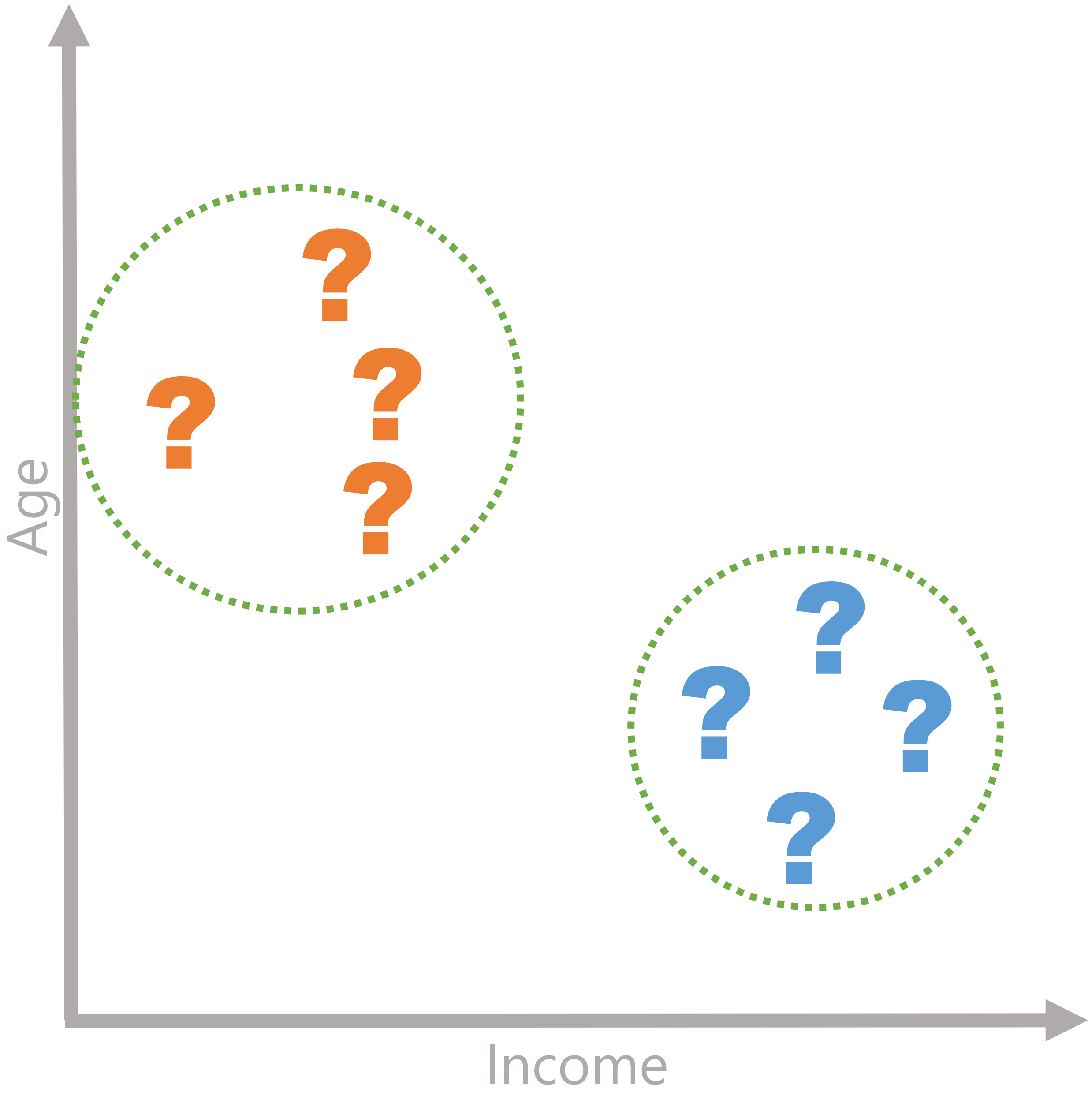
Clustering

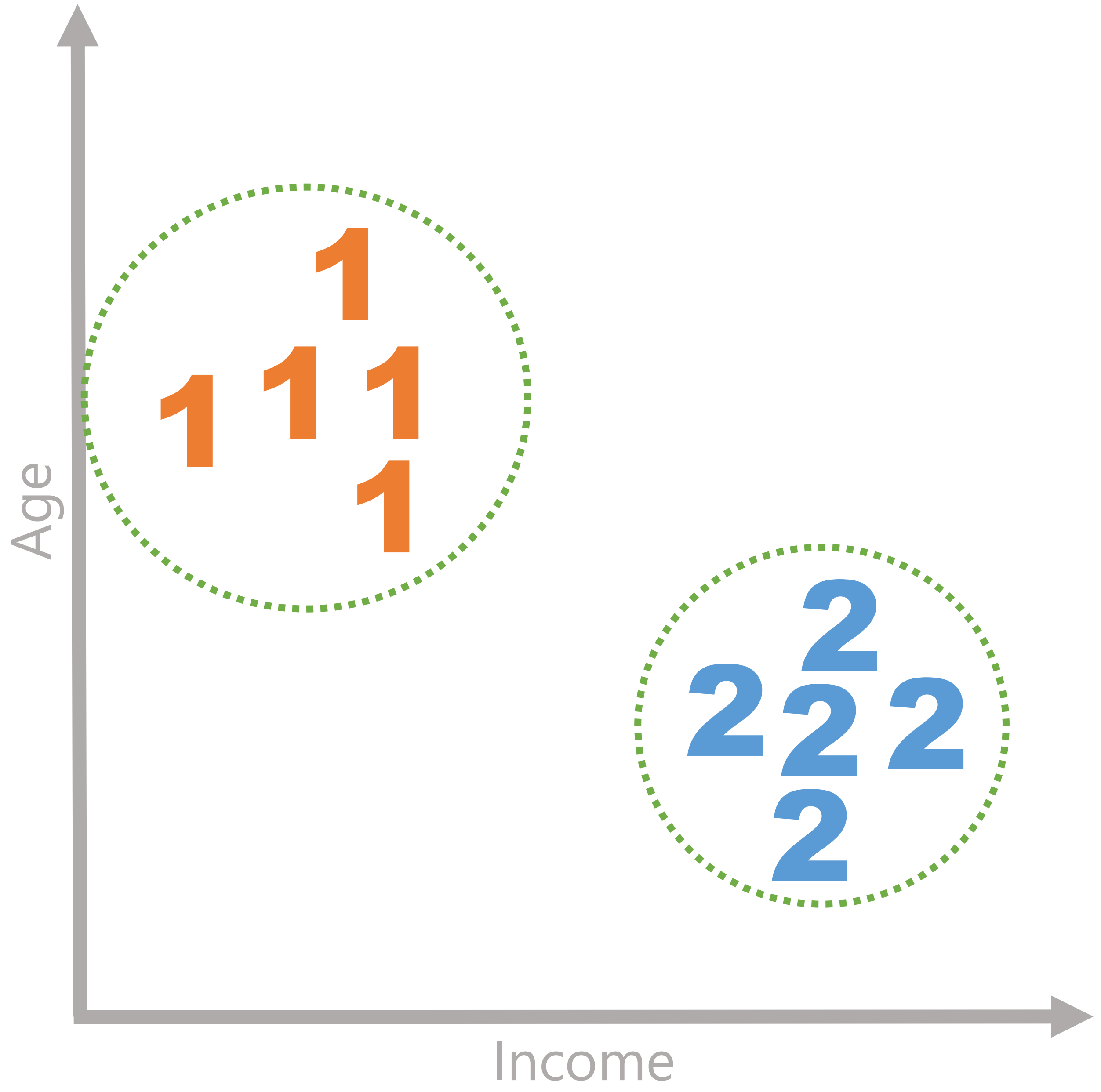

f(x)


Age
Income


Age
Income
?
?
?
?


Age
Income


Age
Income

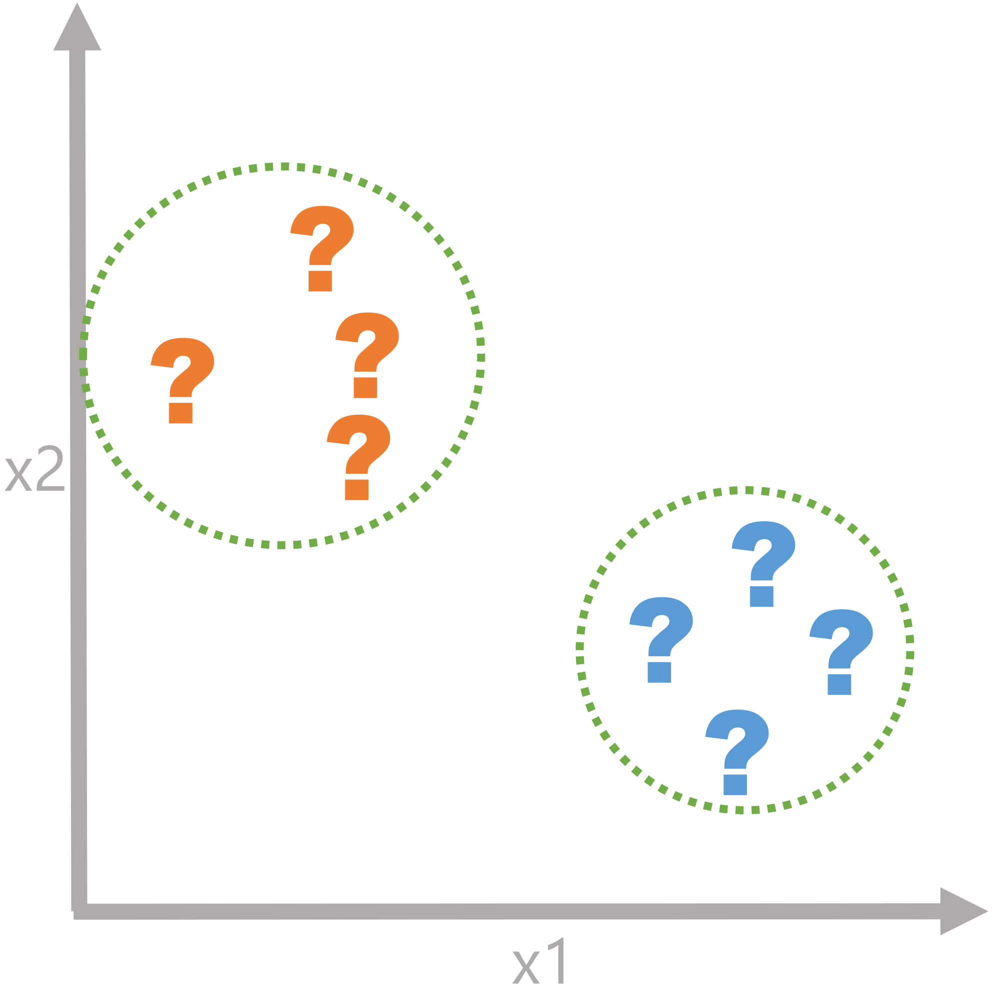

Age
1
1
1
1
1
2
2
2
2
2
Income

Clustering Algorithms

K-means

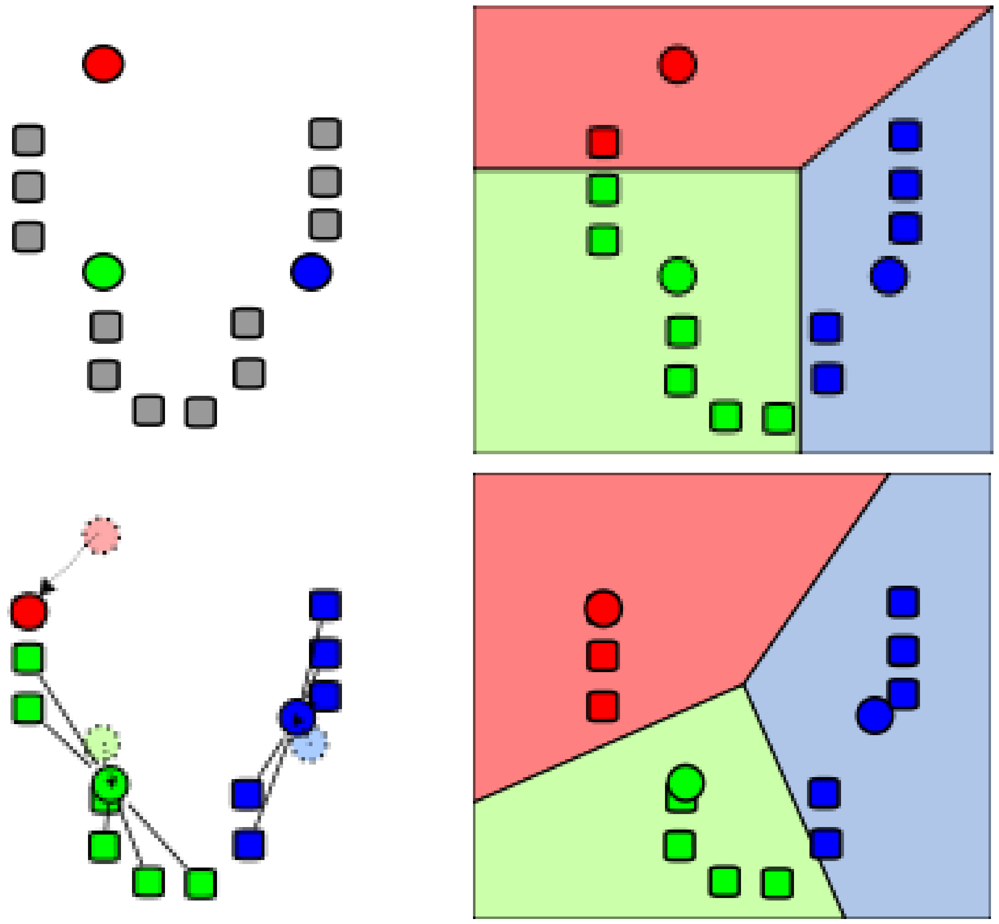
Hierarchical clustering

Expectation maximization



k-Means Clustering

Unsupervised learning

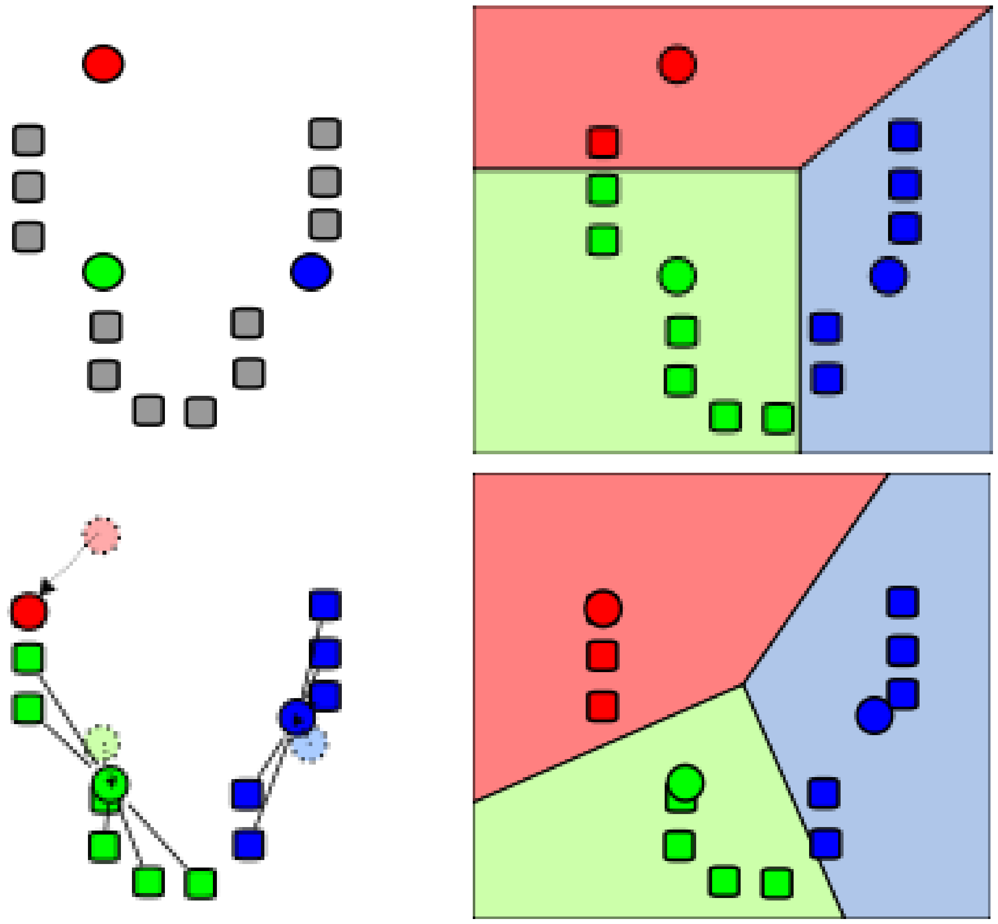


Source: Wikipedia

k-Means Clustering

Unsupervised learning

Specify k (# of clusters)



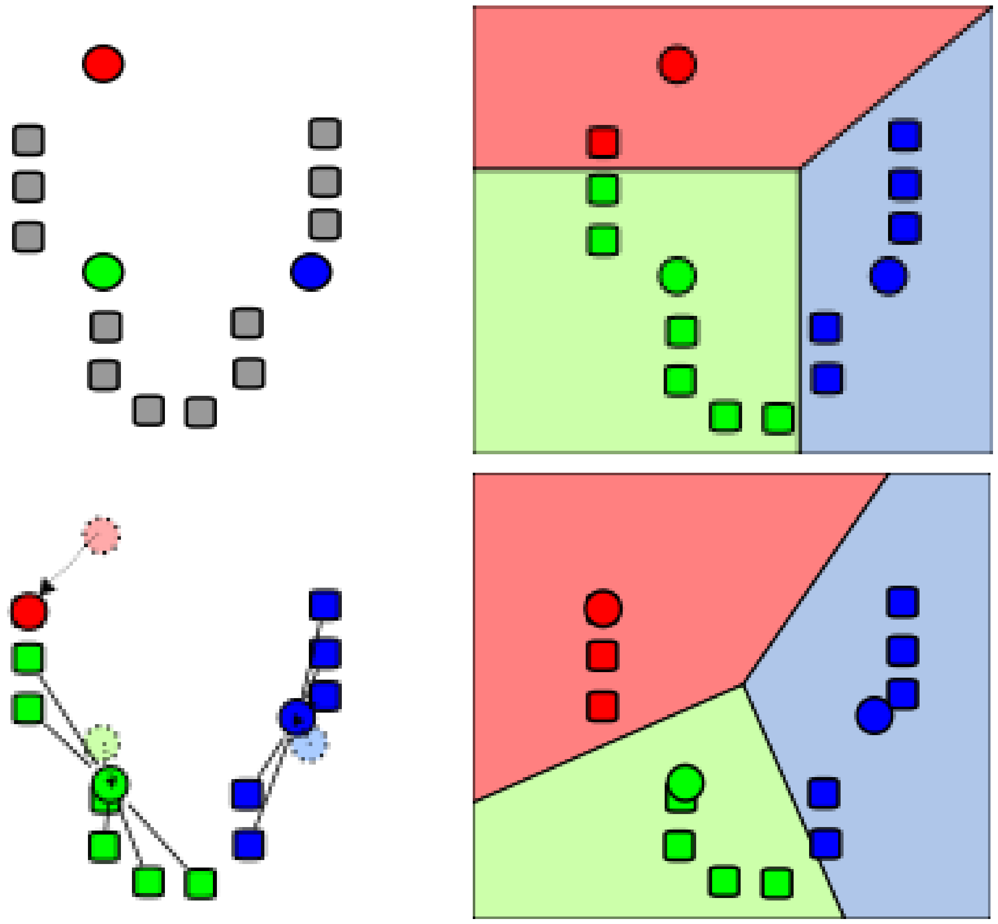
Source: Wikipedia

k-Means Clustering

Unsupervised learning

Specify k (# of clusters)

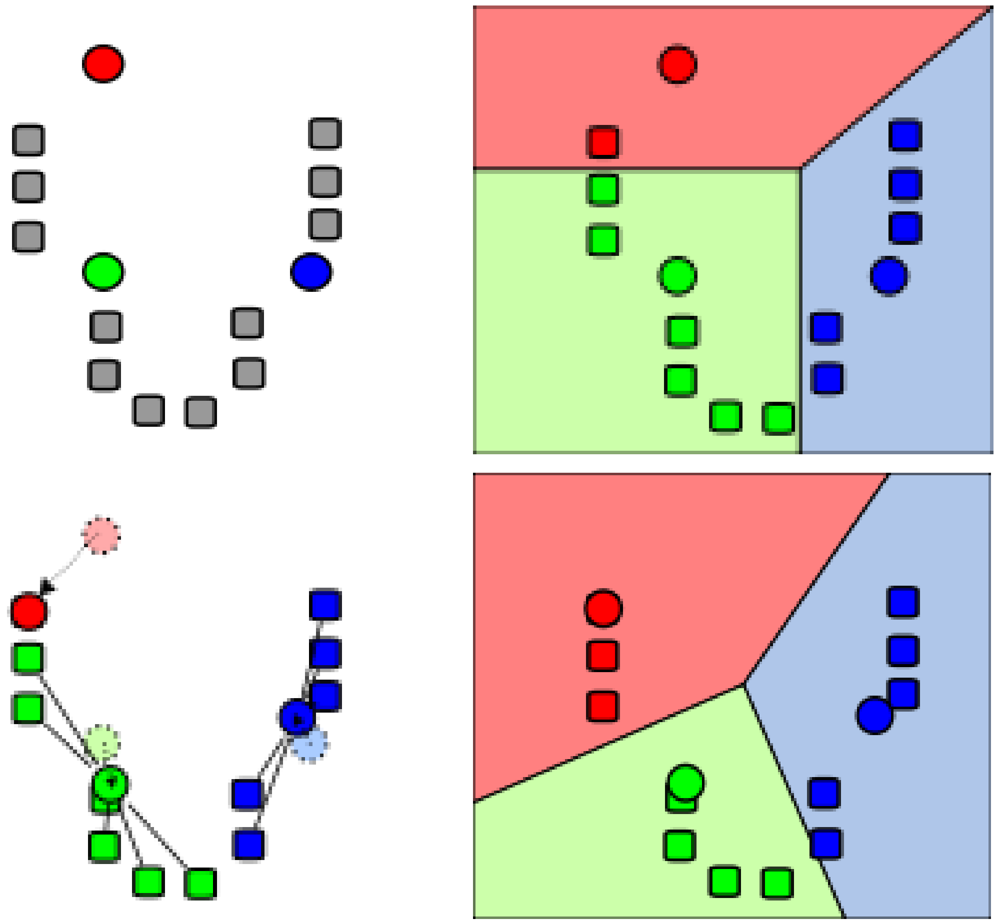
Algorithm finds centers



Source: Wikipedia

k-Means Clustering

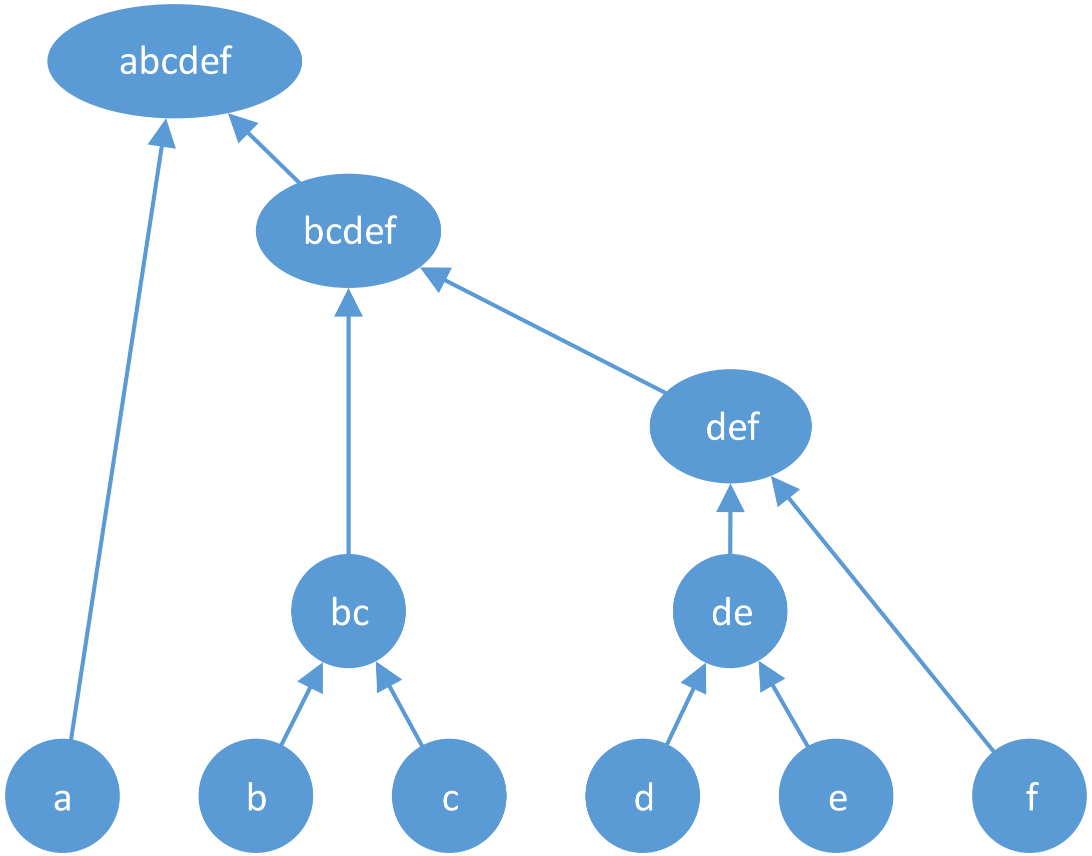
- Unsupervised learning
- Specify k (# of clusters)
- Algorithm finds centers
- Random restarts



Source: Wikipedia

Hierarchical Clustering

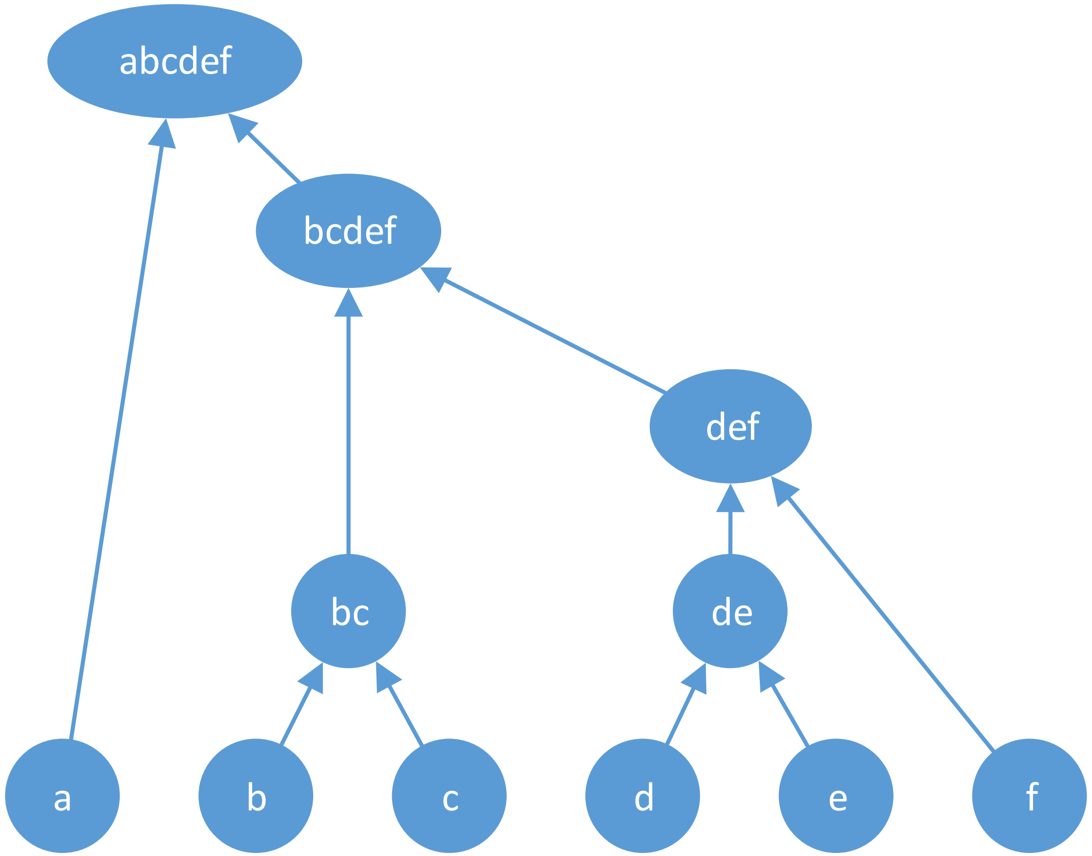
Unsupervised learning



Hierarchical Clustering

Unsupervised learning

Tree of connectedness

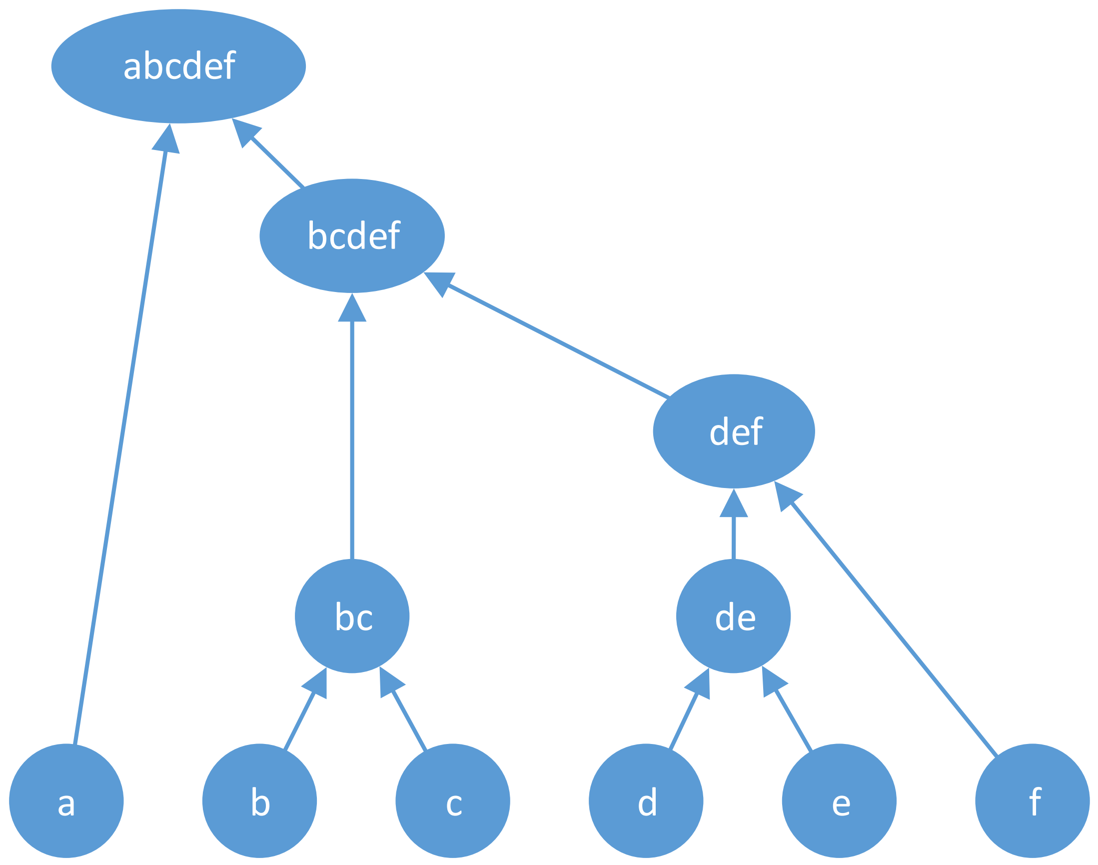


Hierarchical Clustering

Unsupervised learning

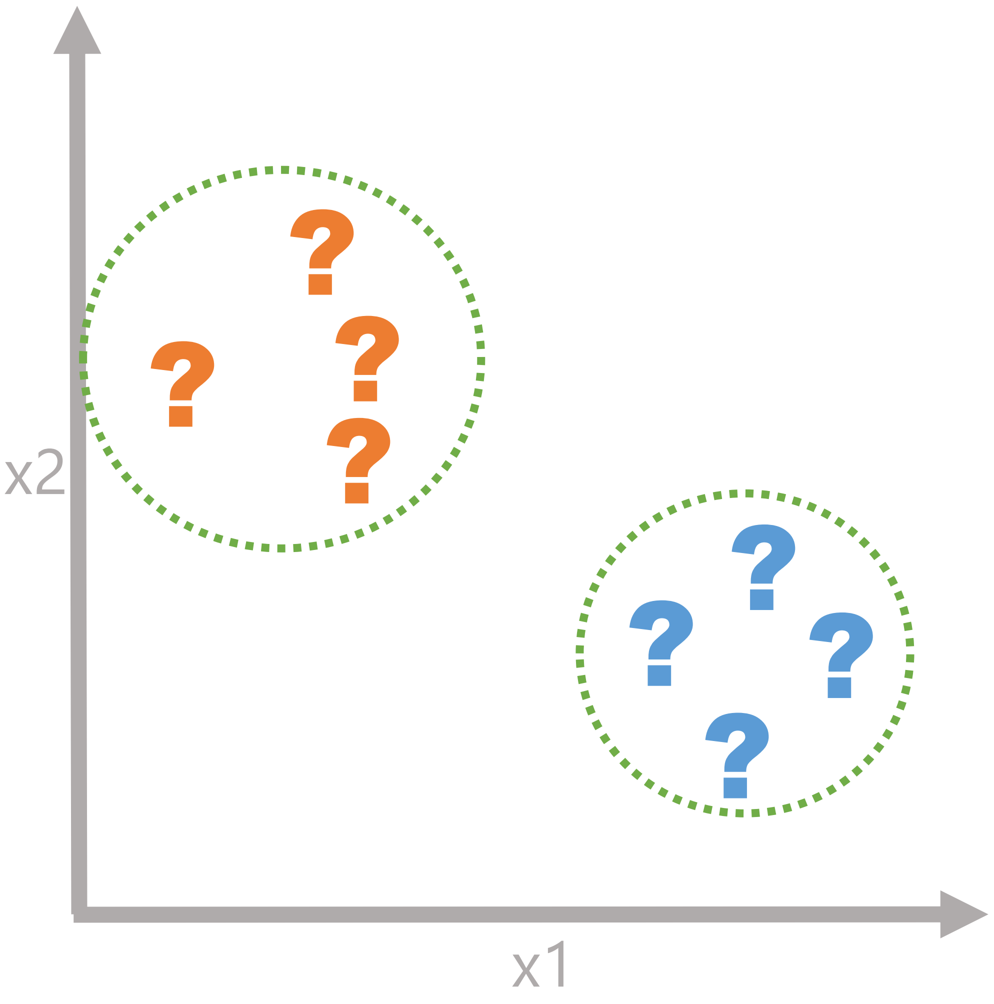
Tree of connectedness

Cuts create clusters



Real-world Examples

What are our market segments?
How to group our documents?
Which products to recommend?



Demo 4 - Clustering

Goal: Group flowers by similarity

Lab 4A – Clustering (Easy)

Goal: Group flowers by similarity

Lab 4B – Clustering (Hard)

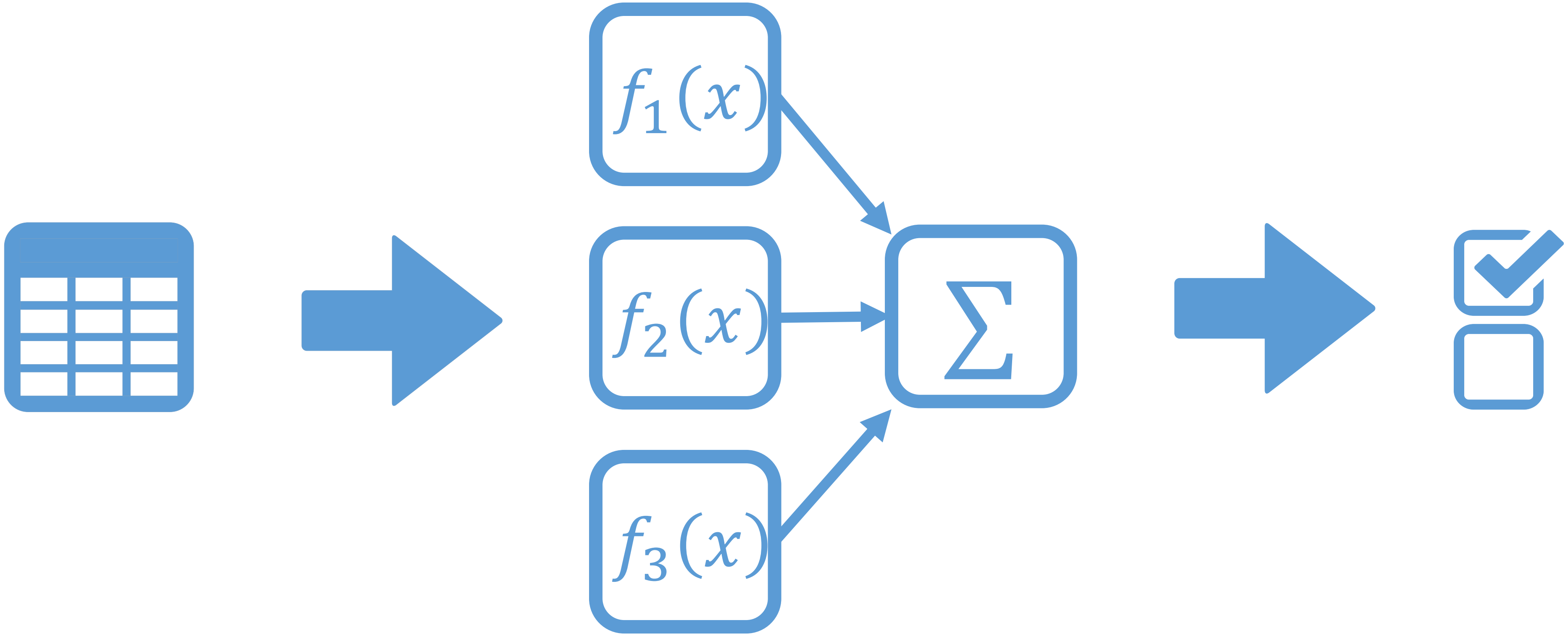
Goal: Group insurance policies

Ensemble Learning

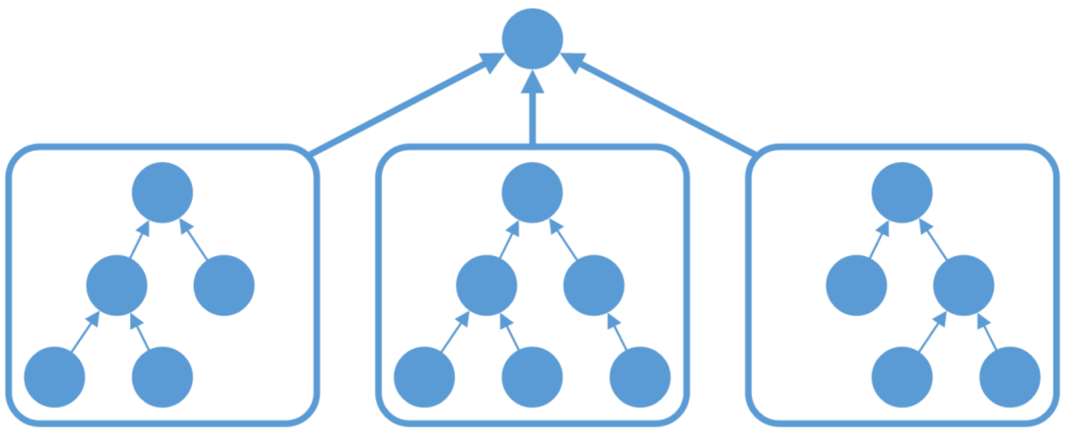
Wisdom of the Crowds


BRAVE NEW FILMS
HOORAY! OUR SPORTS TEAM WON!
LET'S CELEBRATE
BY DESTROYING OUR OWN CITY!

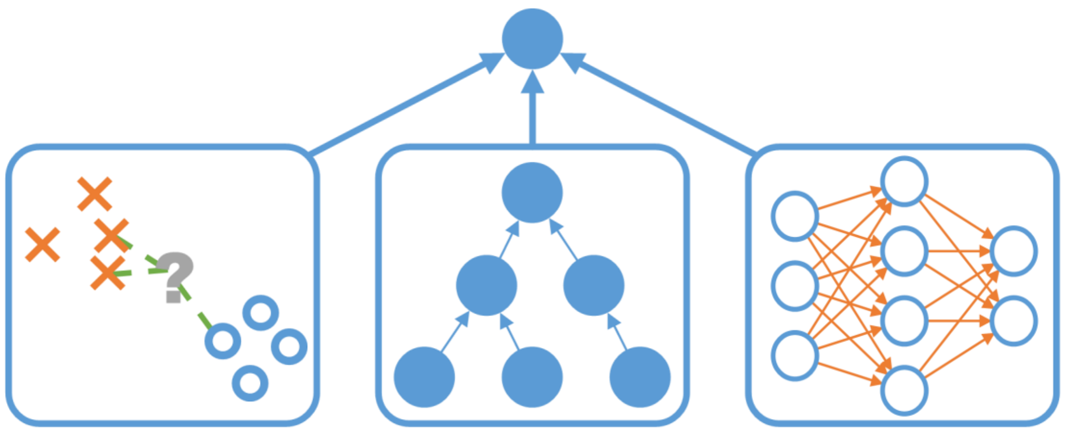
Ensemble Learning



Types of Ensembles



Same Type of Model



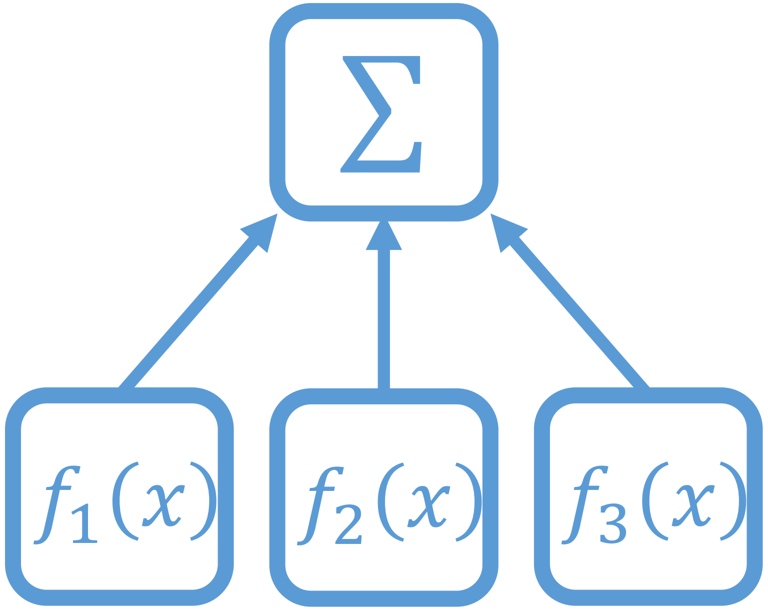
Different Types of Models

Ensemble Creation Techniques

Bagging

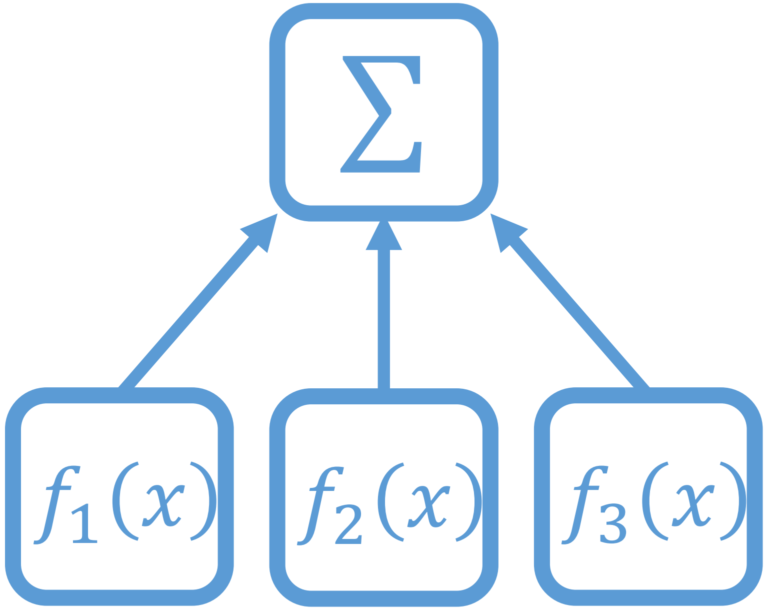
Boosting

Stacking

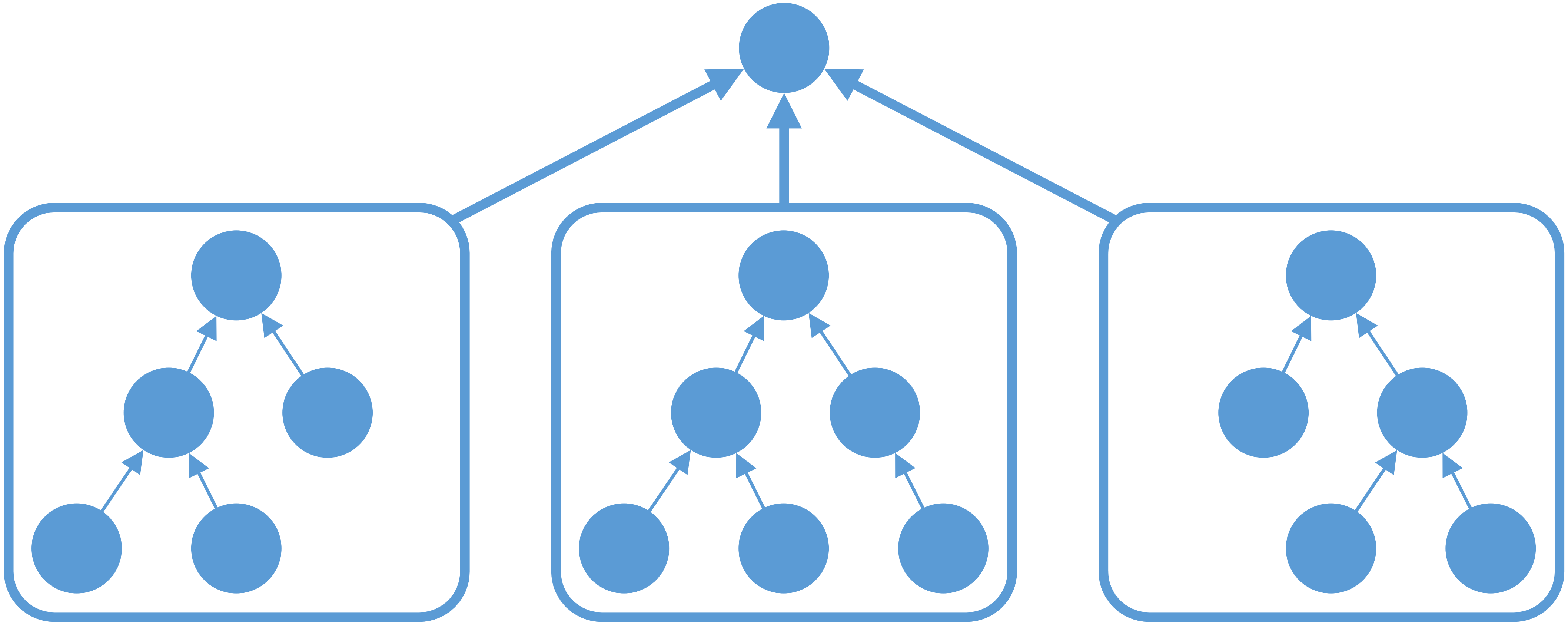


Ensemble Aggregation Techniques

Averaging
Majority Vote
Weighted Average
Weighted Majority Vote

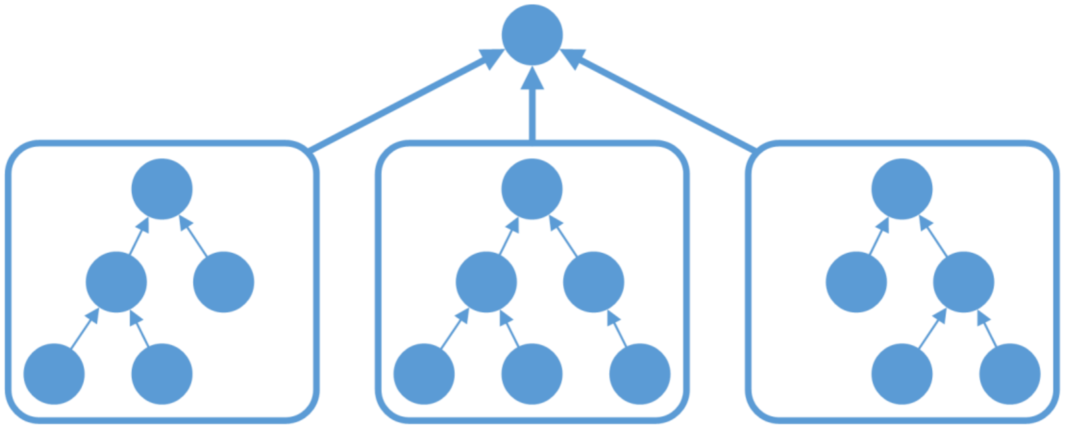


Random Forest Classifier



Random Forest Classifier

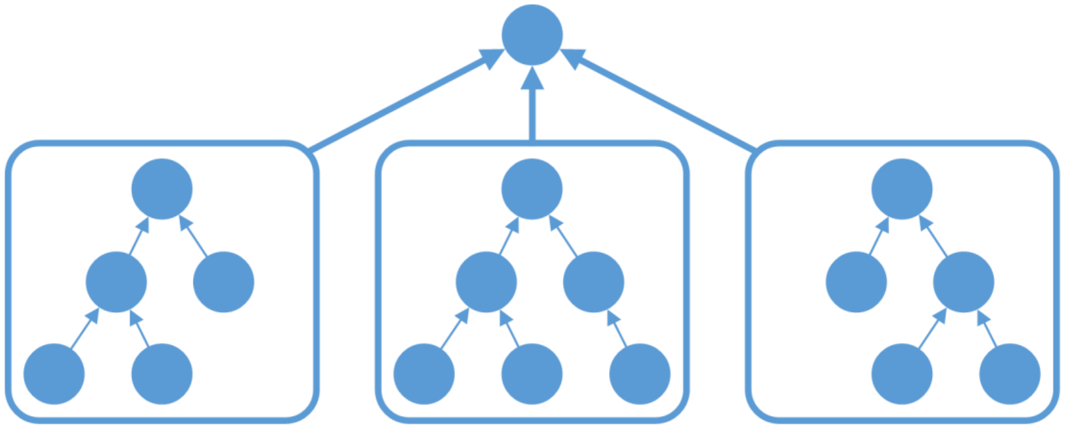
Multiple trees



Random Forest Classifier

Multiple trees

Created by bagging

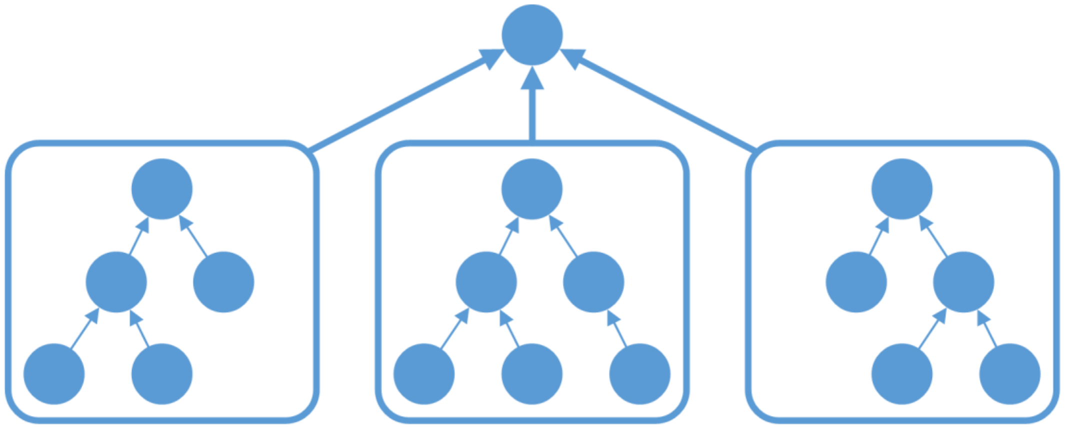


Random Forest Classifier

Multiple trees

Created by bagging

Majority vote



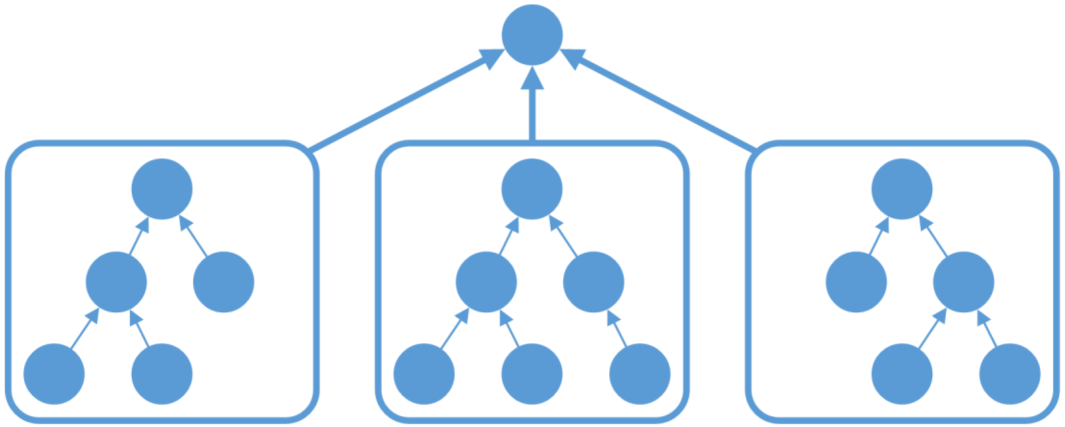
Random Forest Classifier

Multiple trees

Created by bagging

Majority vote

More robust



Why Use Ensemble Learning?

Pros

More accurate

More robust

More stable

Why Use Ensemble Learning?

Pros

More accurate

More robust

More stable

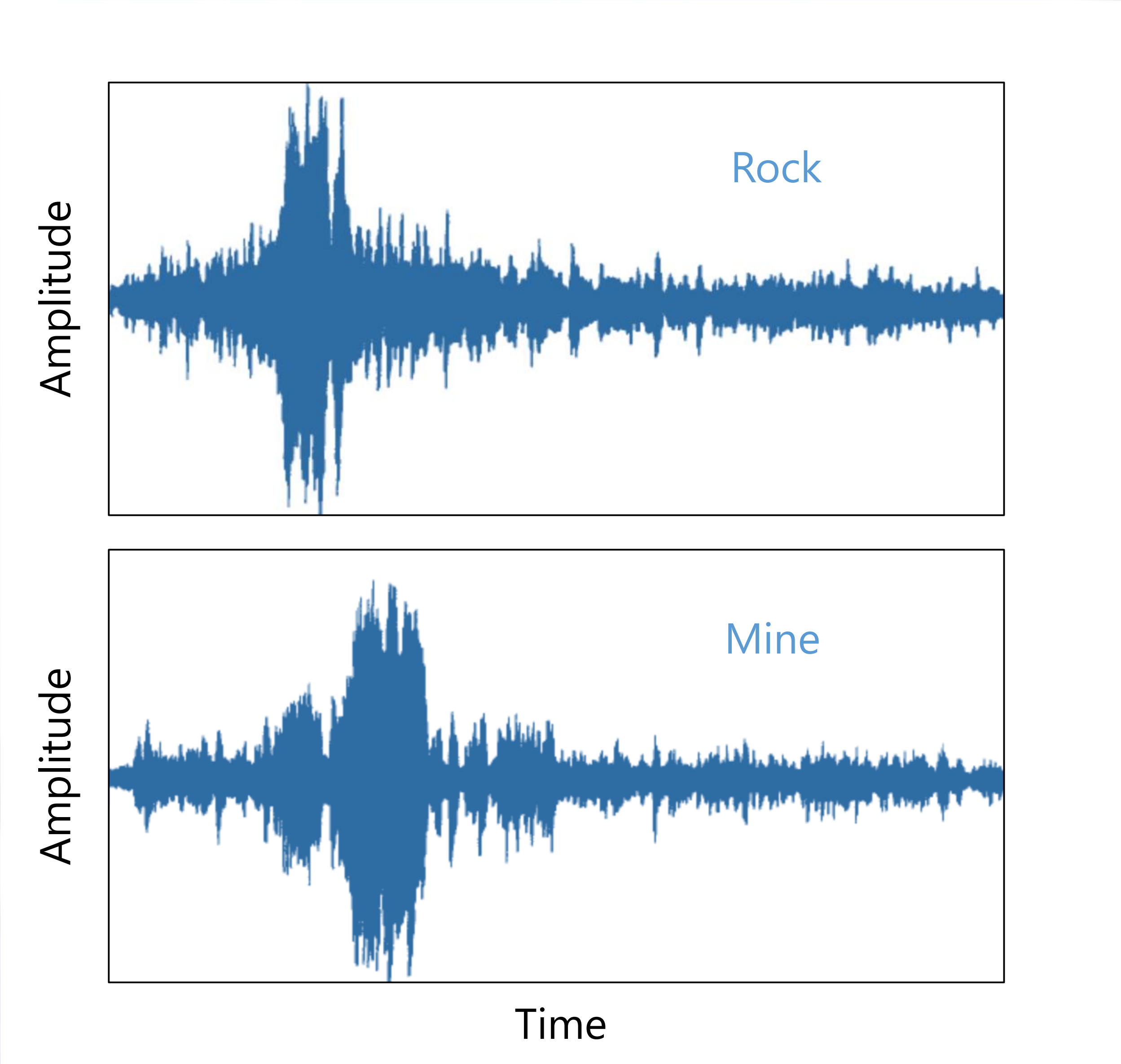
Cons

More complex

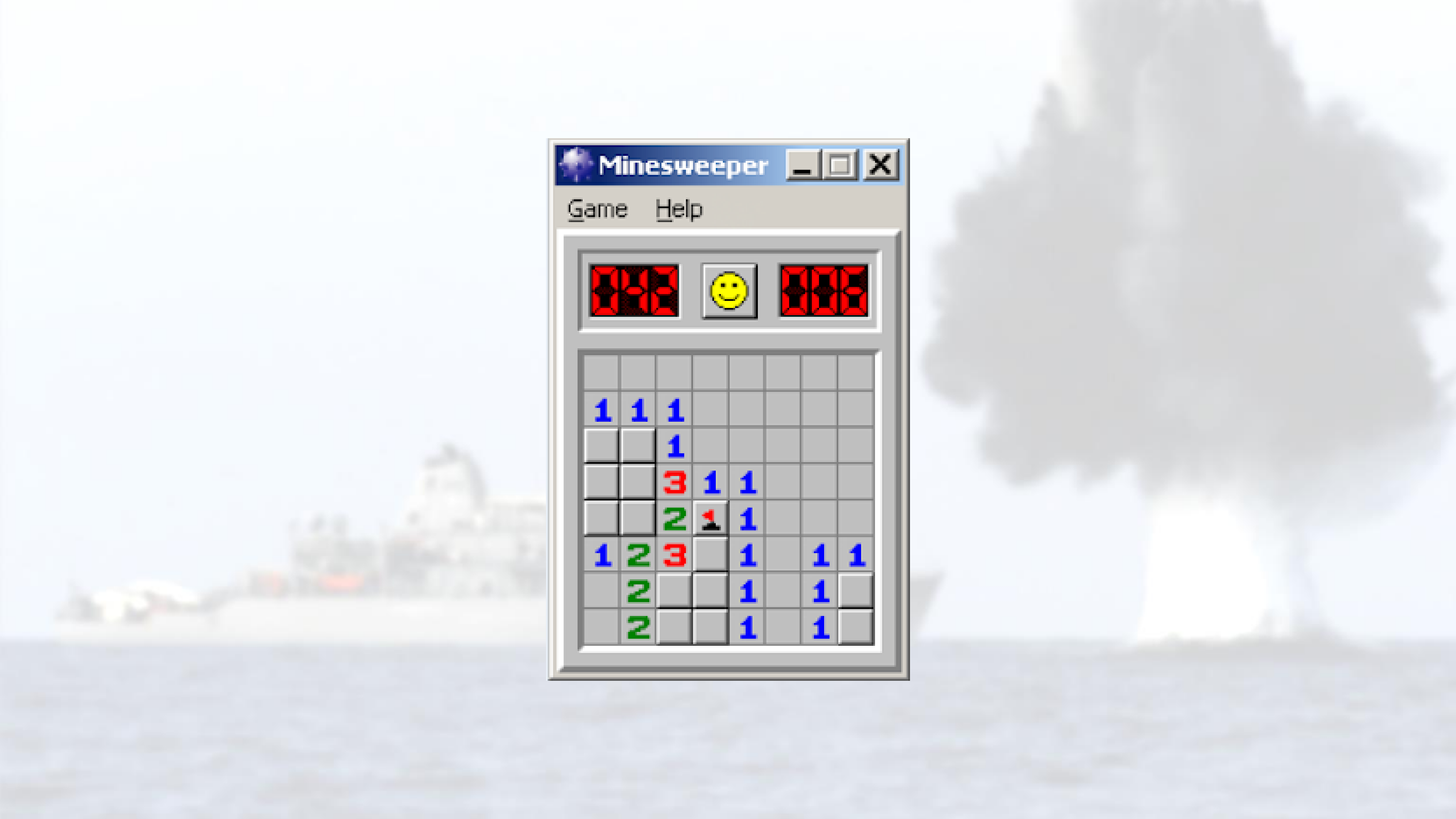
More CPU time

More art than science

Ensemble Learning Demo


Rock
Mine
Amplitude
Amplitude
Time

Sonar							
V1	V2	V3	...	V58	V59	V60	Class
0.02	0.03	0.04	...	0.00	0.01	0.00	rock
0.04	0.05	0.08	...	0.00	0.01	0.00	mine
0.02	0.05	0.10	...	0.01	0.01	0.01	rock
0.01	0.01	0.06	...	0.00	0.00	0.01	rock
0.07	0.06	0.04	...	0.00	0.01	0.01	mine
0.02	0.04	0.02	...	0.00	0.01	0.00	rock
...	...	...	...	...	...	...	...


Minesweeper
Game
Help
042
006

Demo 5 – ML in Practice

Goal: Predict rock or mine

Lab 5A – ML in Practice (Easy)

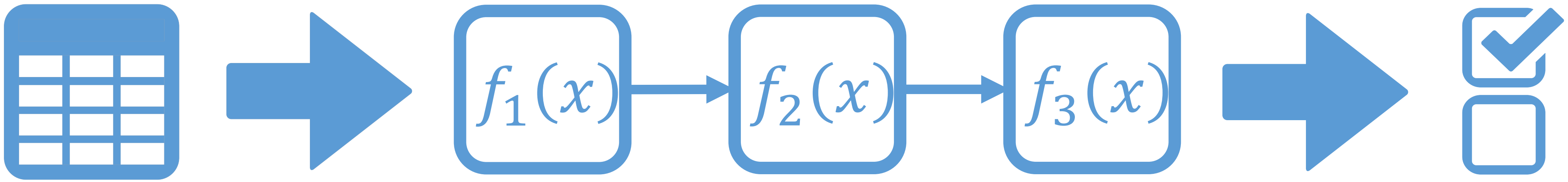
Goal: Predict rock or mine

Lab 5B – ML in Practice (Hard)

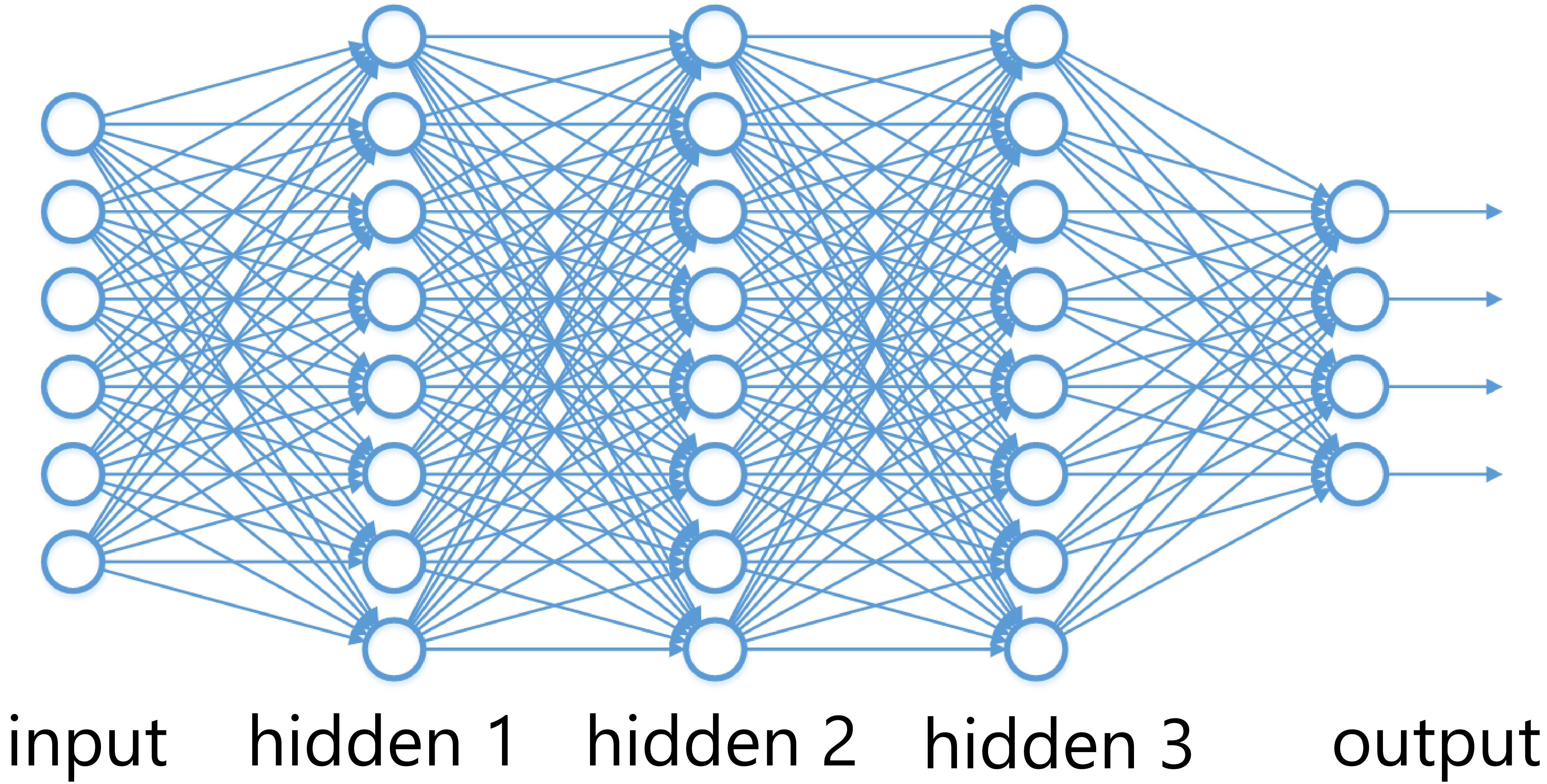
Goal: Predict risk class

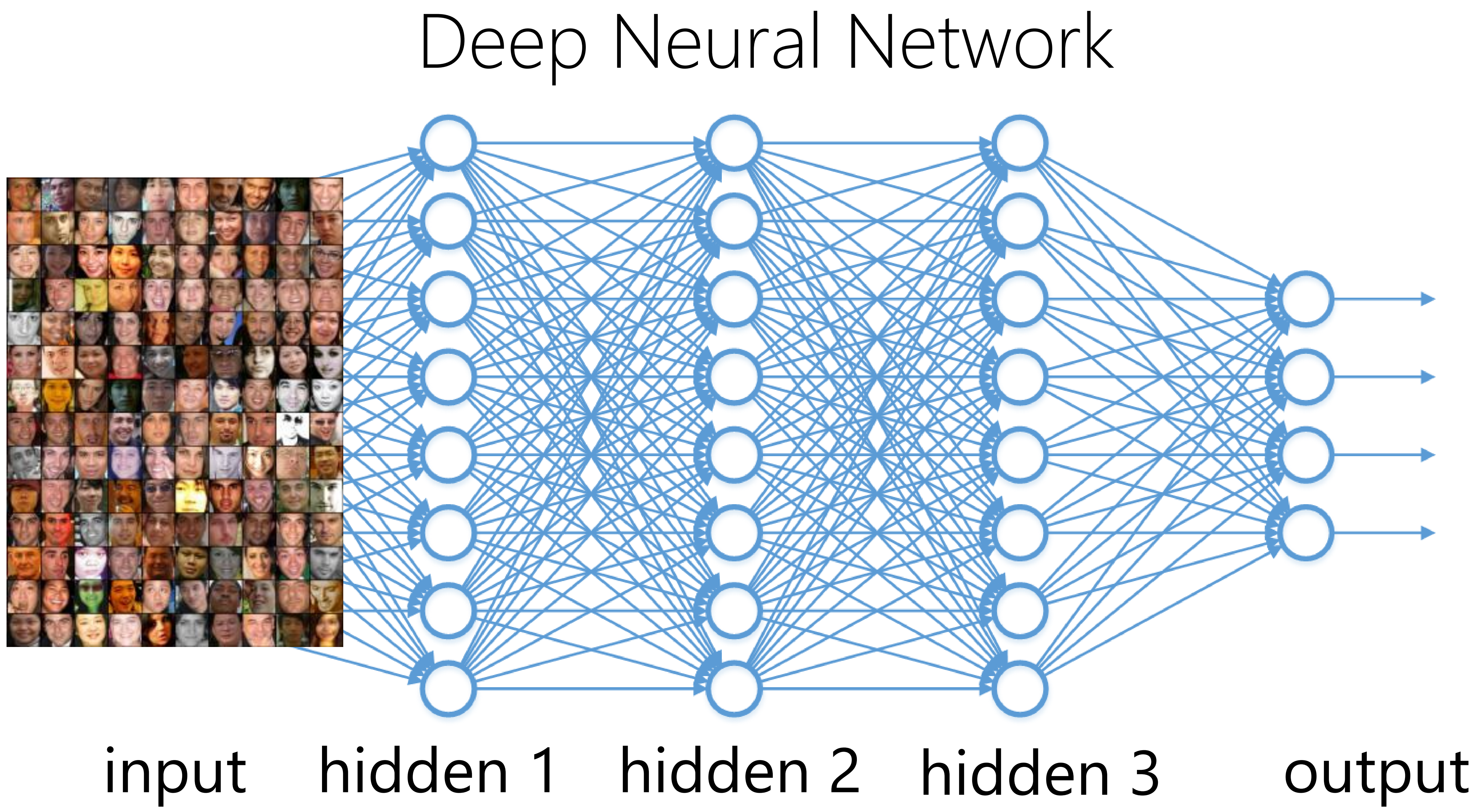
Deep Learning

Deep Learning

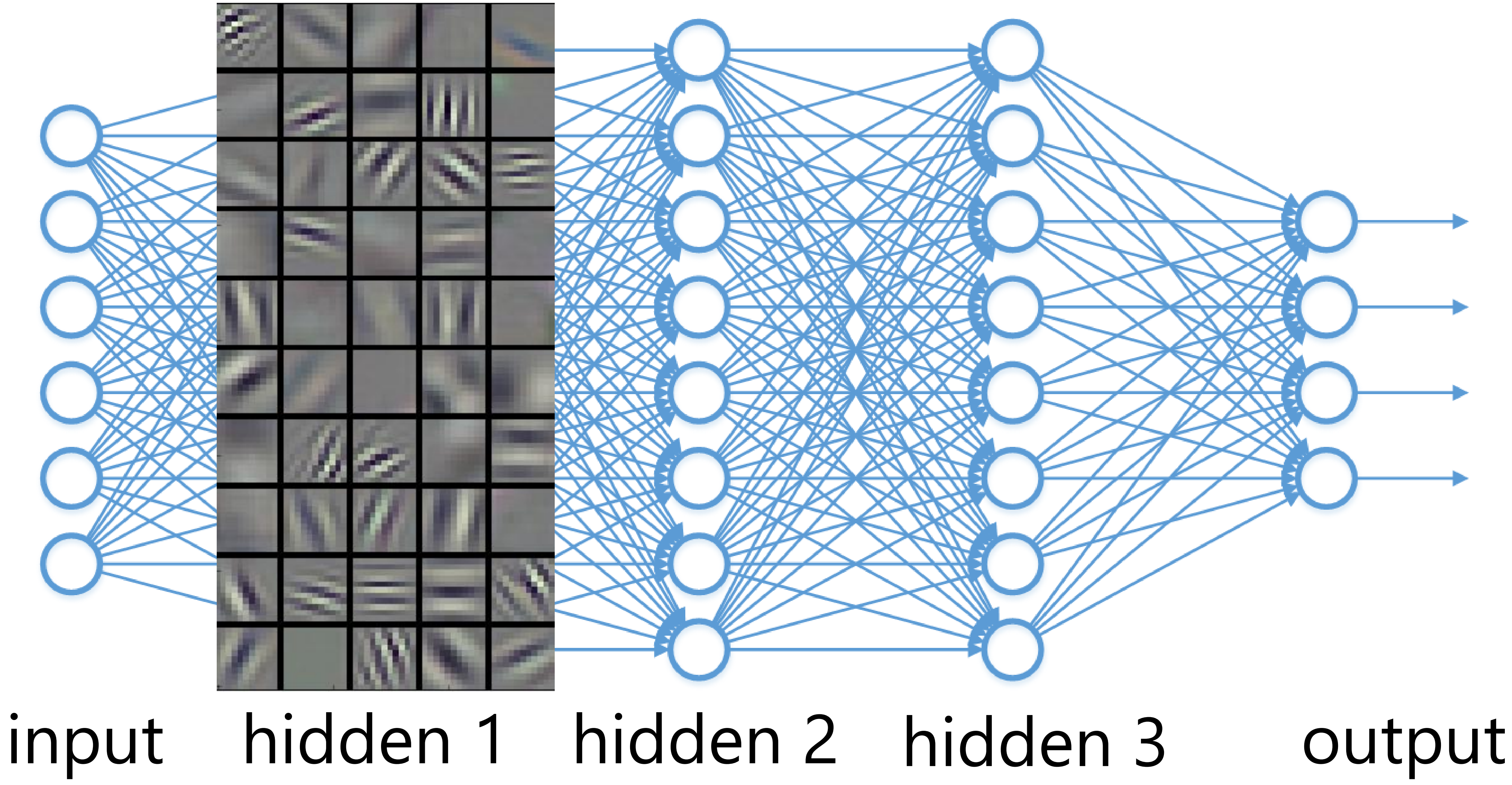


Deep Neural Network

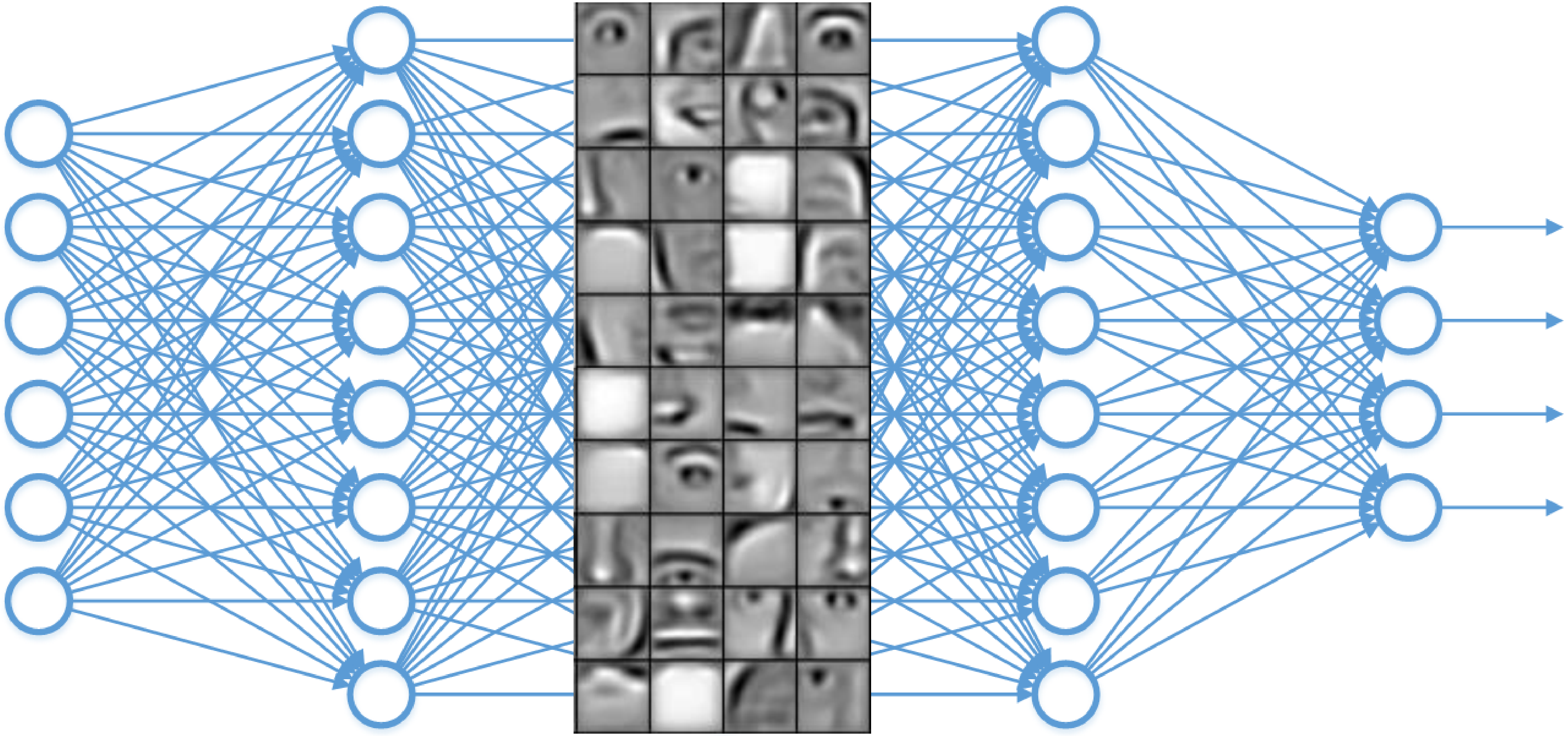



Deep Neural Network
input
hidden 1
hidden 2
hidden 3
output

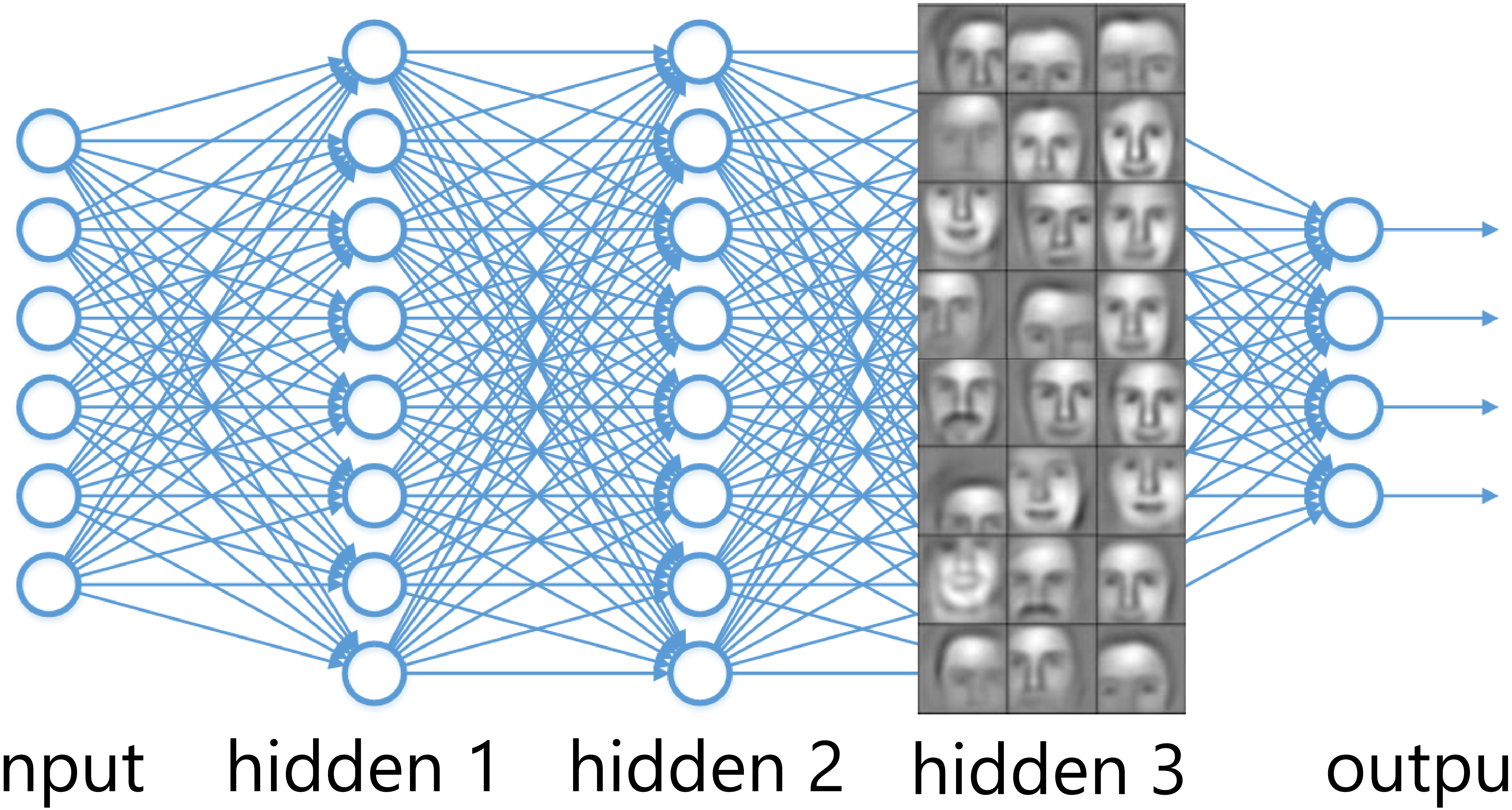
Deep Neural Network

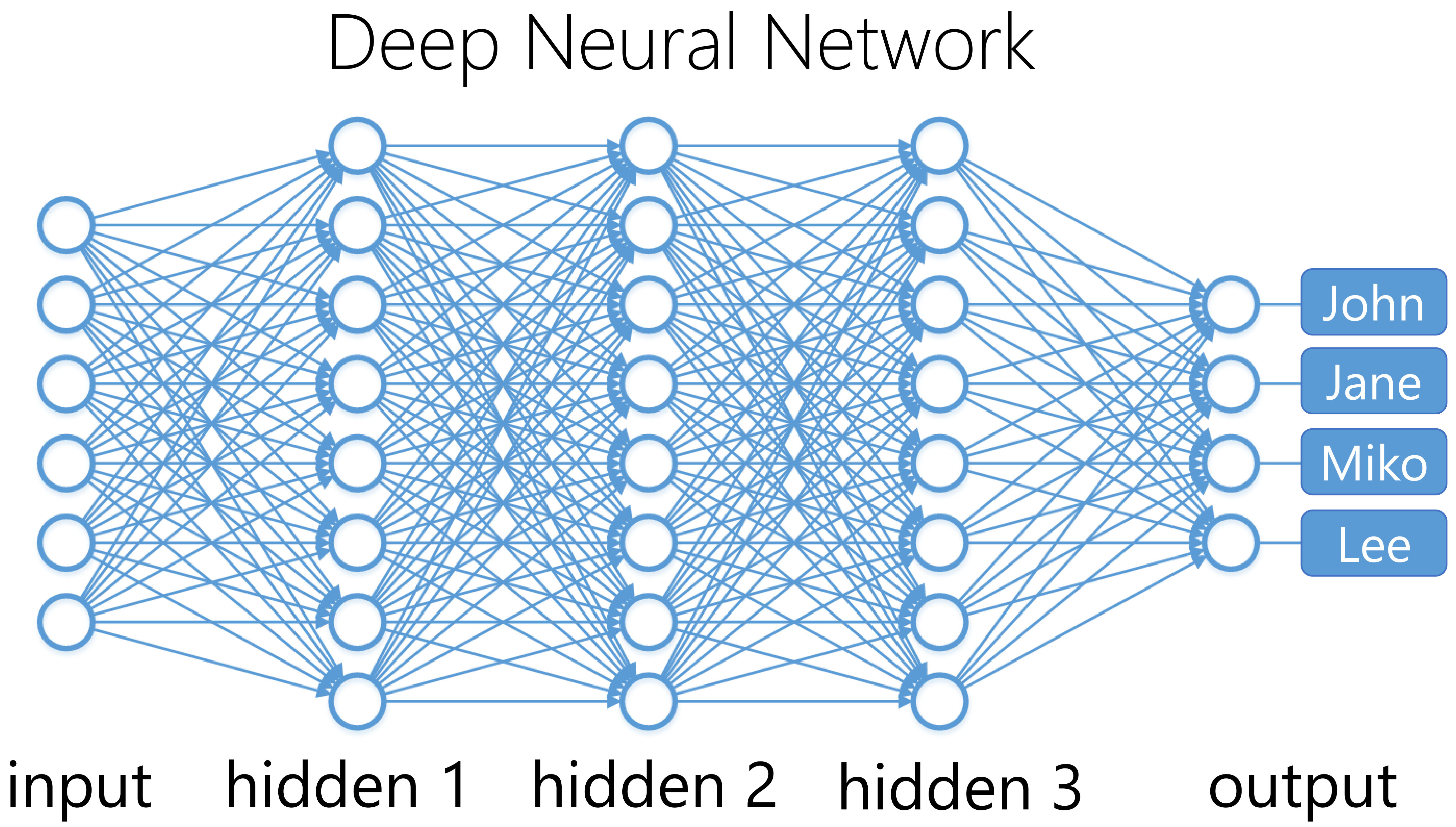


Deep Neural Network

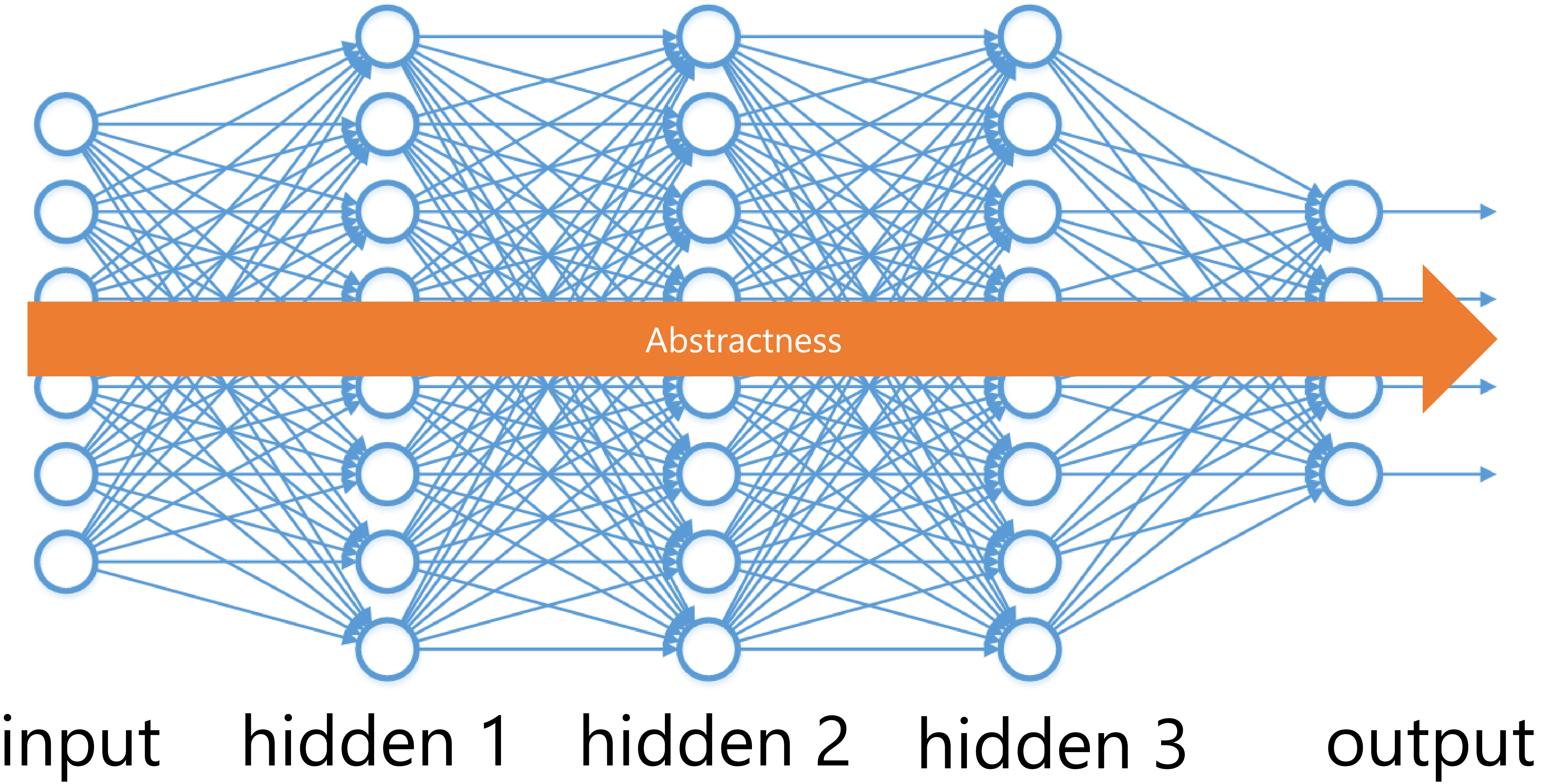


Deep Neural Network



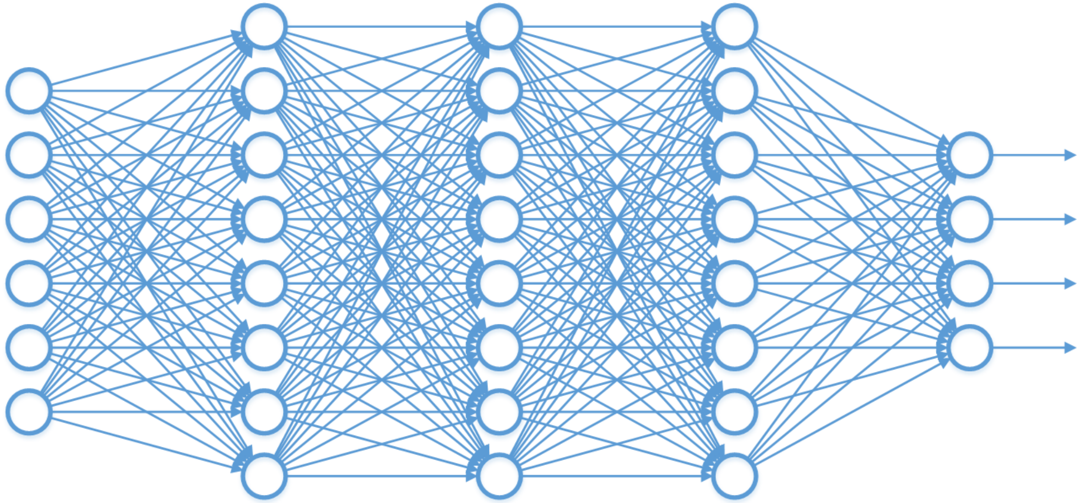

Deep Neural Network
John
Jane
Miko
Lee
input
hidden 1
hidden 2
hidden 3
output

Deep Neural Network

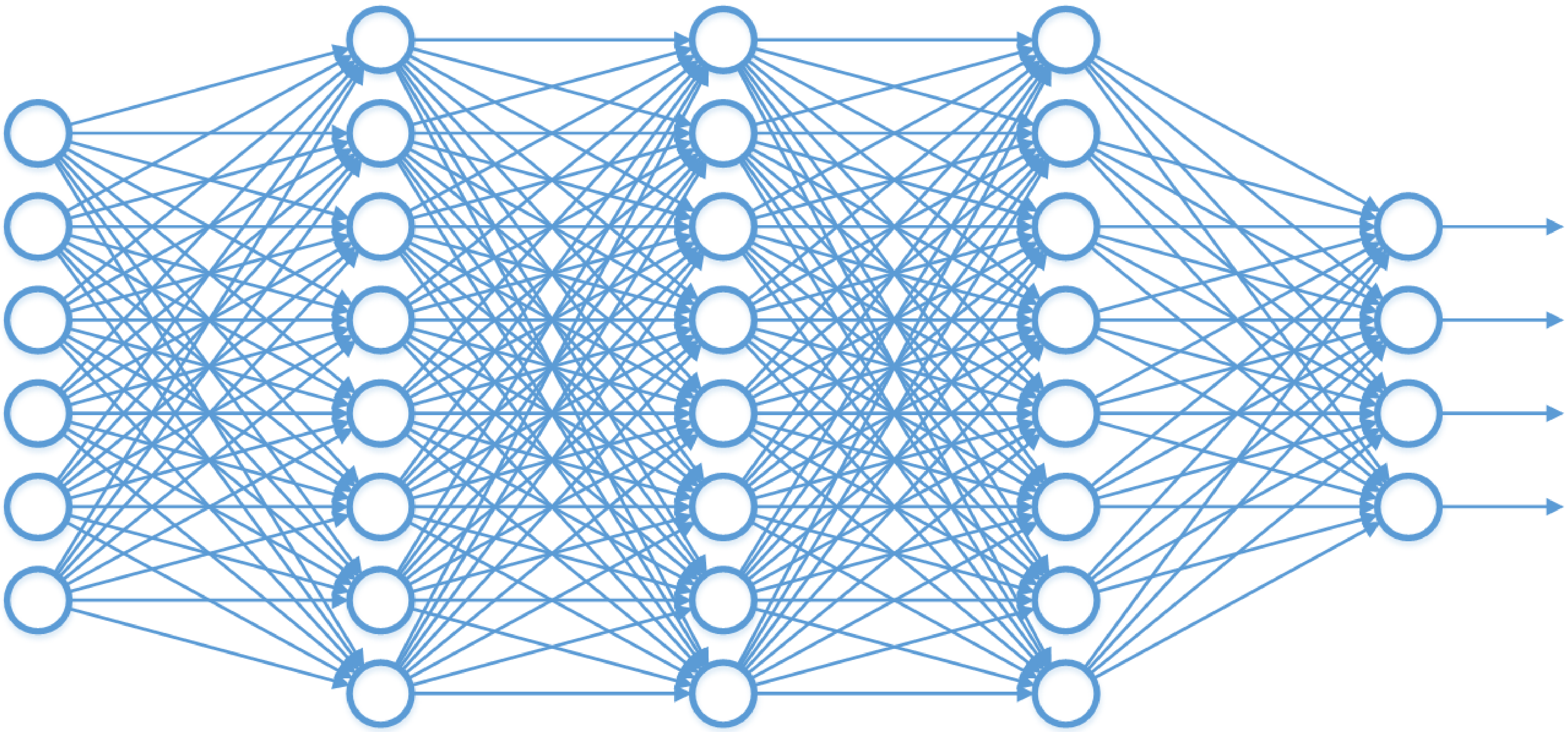


Deep Learning Techniques

Fully connected (DNN)
Convolutional (CNN)
Recurrent (RNN)
Generative Adversarial (GAN)
Deep Q Learning (DQN)

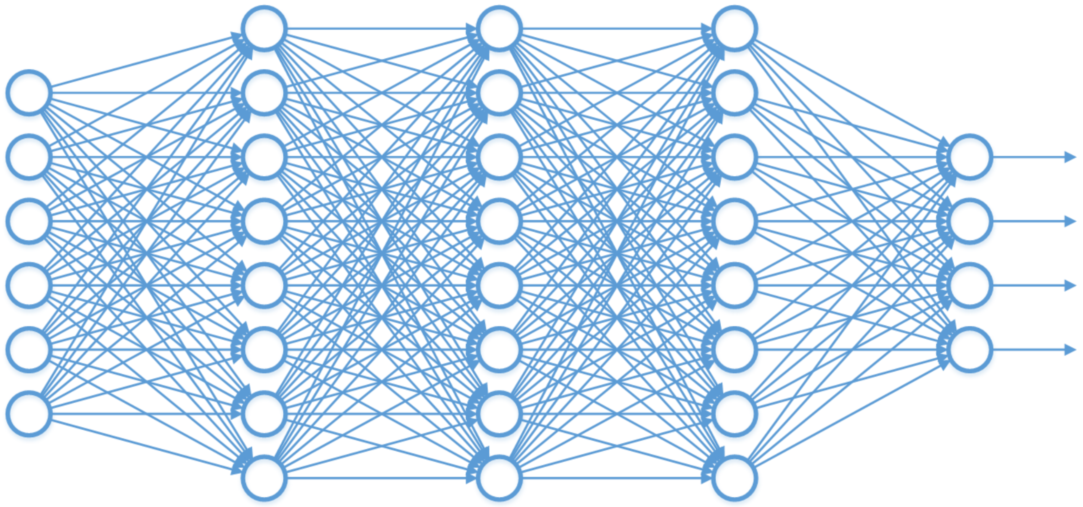


Deep Neural Network



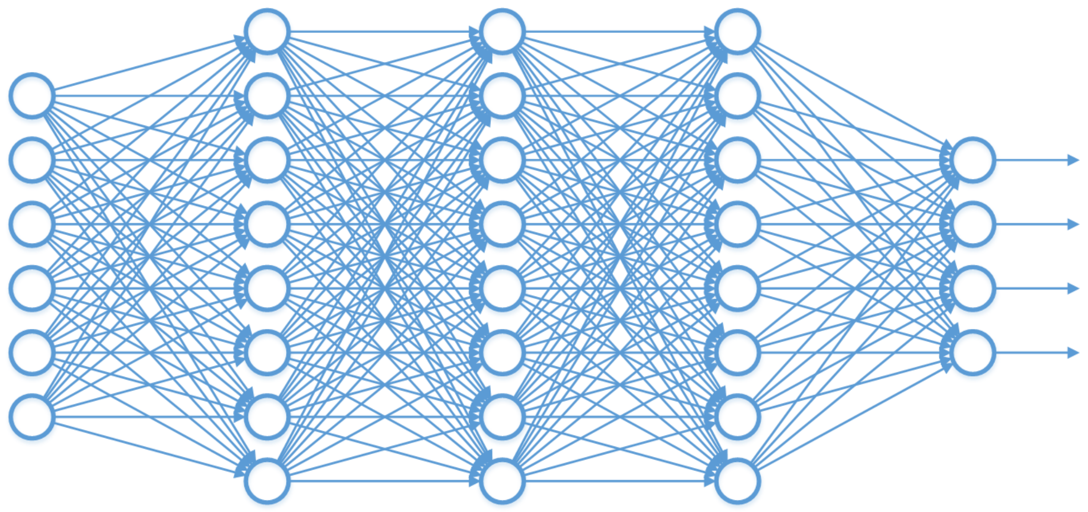
Deep Neural Network

Neural network



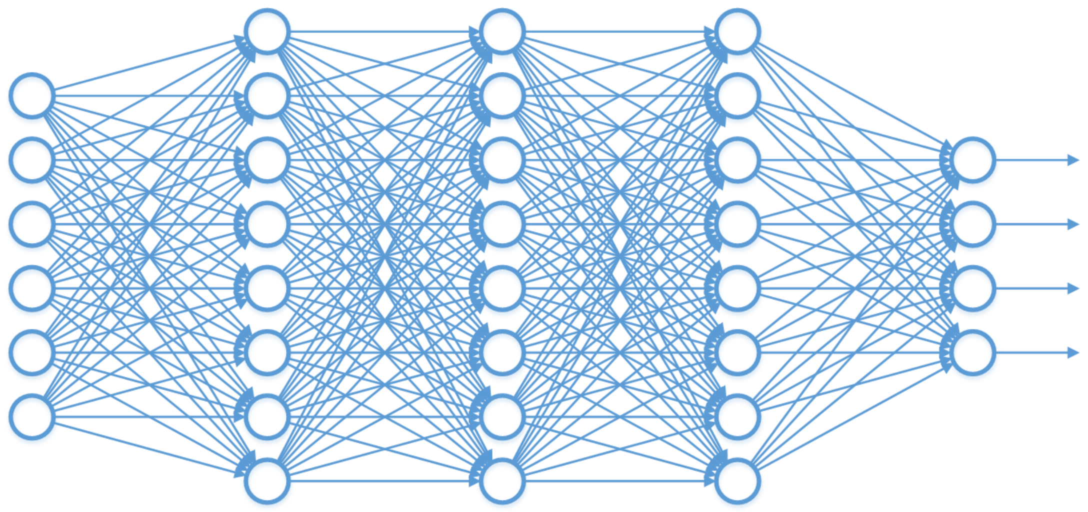
Deep Neural Network

Neural network
Multiple hidden layers



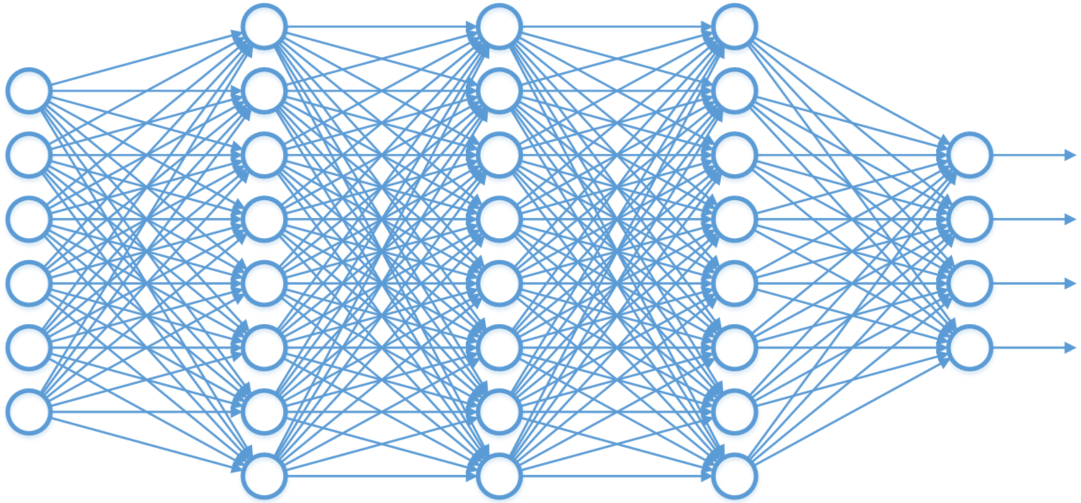
Deep Neural Network

Neural network
Multiple hidden layers
Non-linear activation

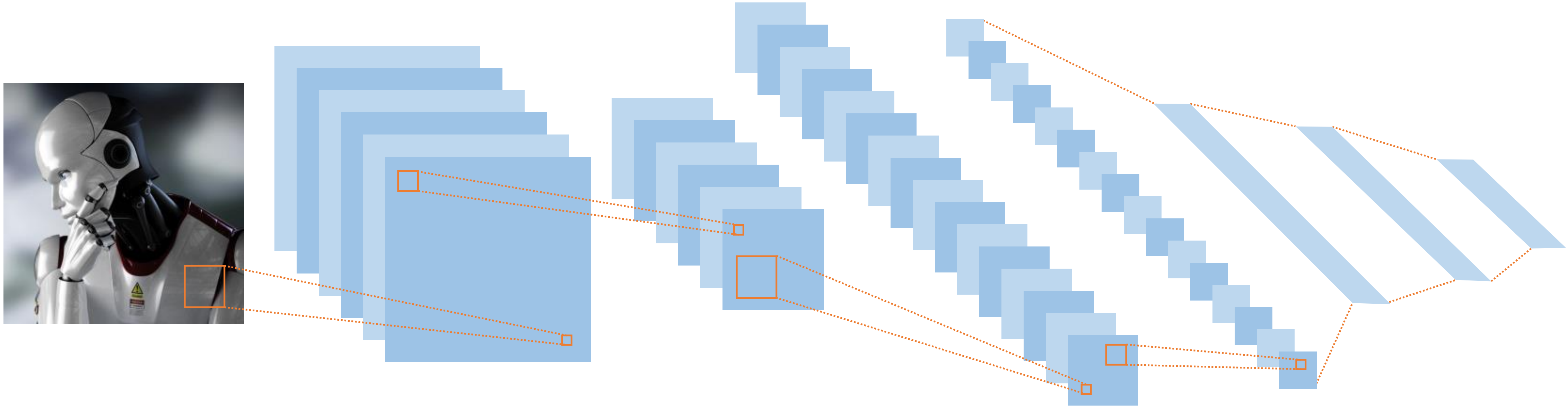


Deep Neural Network

Neural network
Multiple hidden layers
Non-linear activation
Fully connected

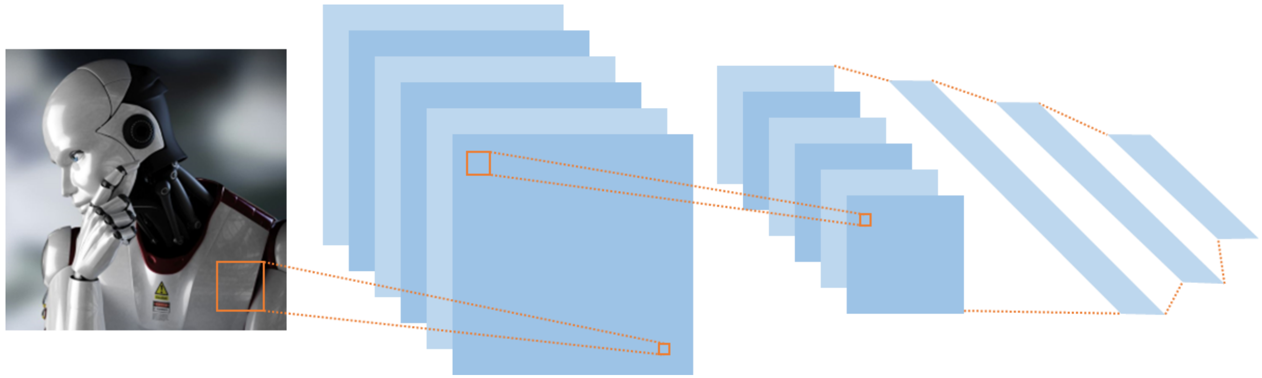


Convolutional Neural Networks (CNN)



Convolutional Neural Network (CNN)

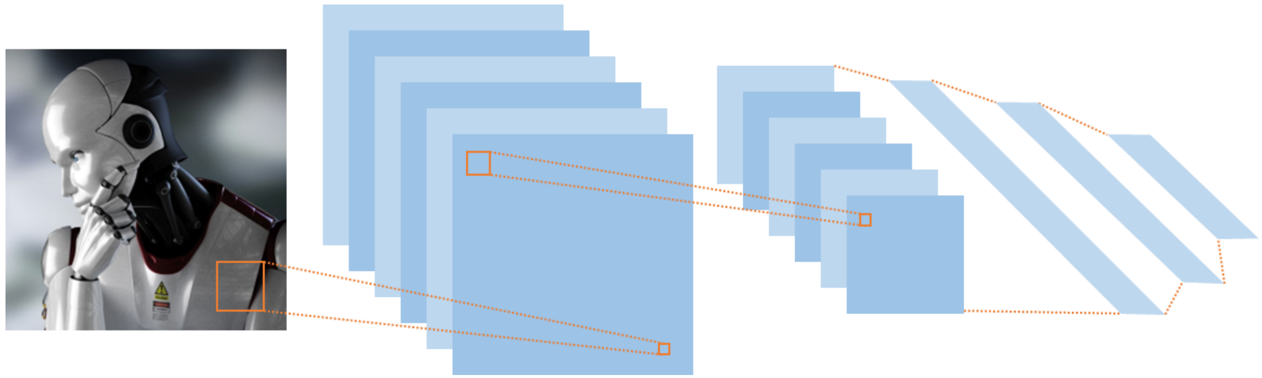
Sparse



Convolutional Neural Network (CNN)

Sparse

Convolutions

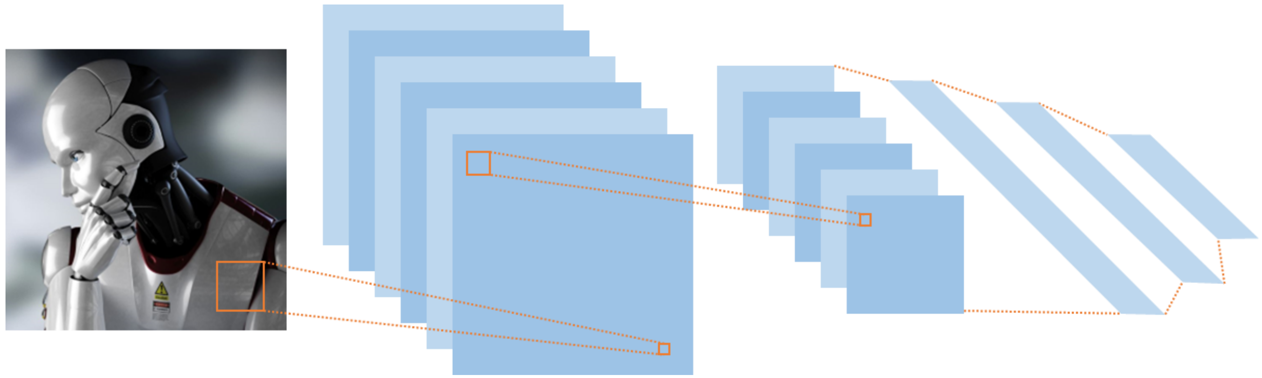


Convolutional Neural Network (CNN)

Sparse

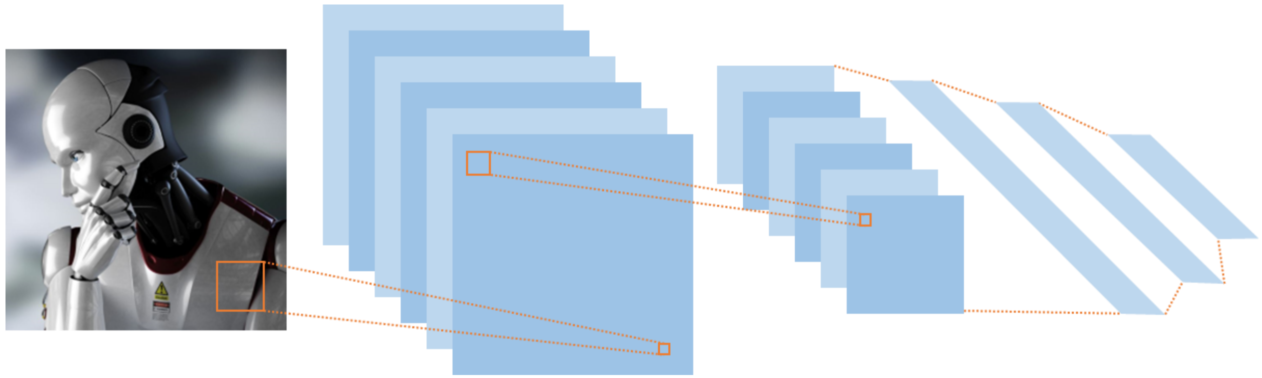
Convolutions

Filters

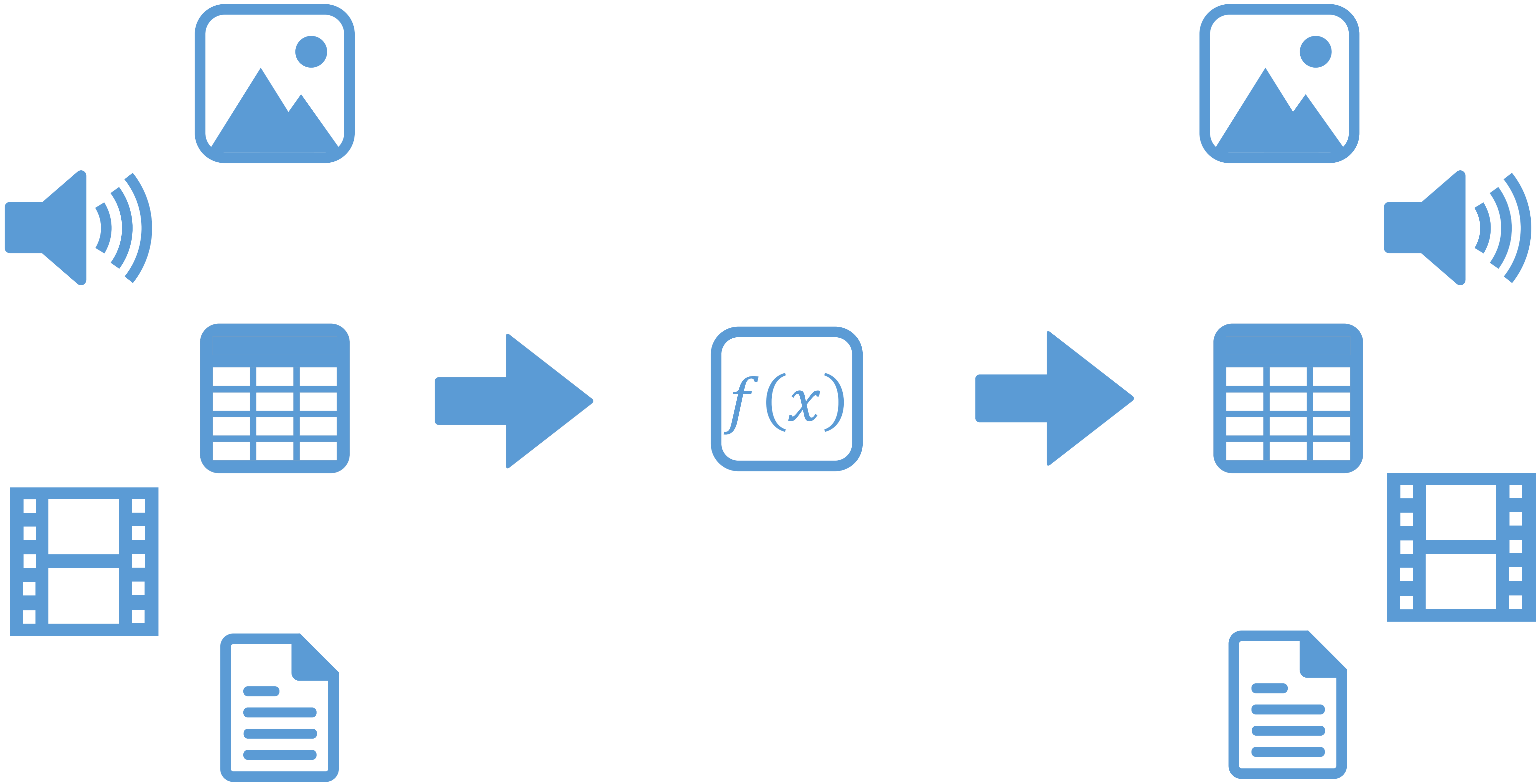


Convolutional Neural Network (CNN)

Sparse
Convolutions
Filters
Pooling




f(x)


f(x)

Why Use Deep Learning?

Pros

More powerful

More accurate

Data synthesis

Why Use Deep Learning?

Pros

More powerful

More accurate

Data synthesis

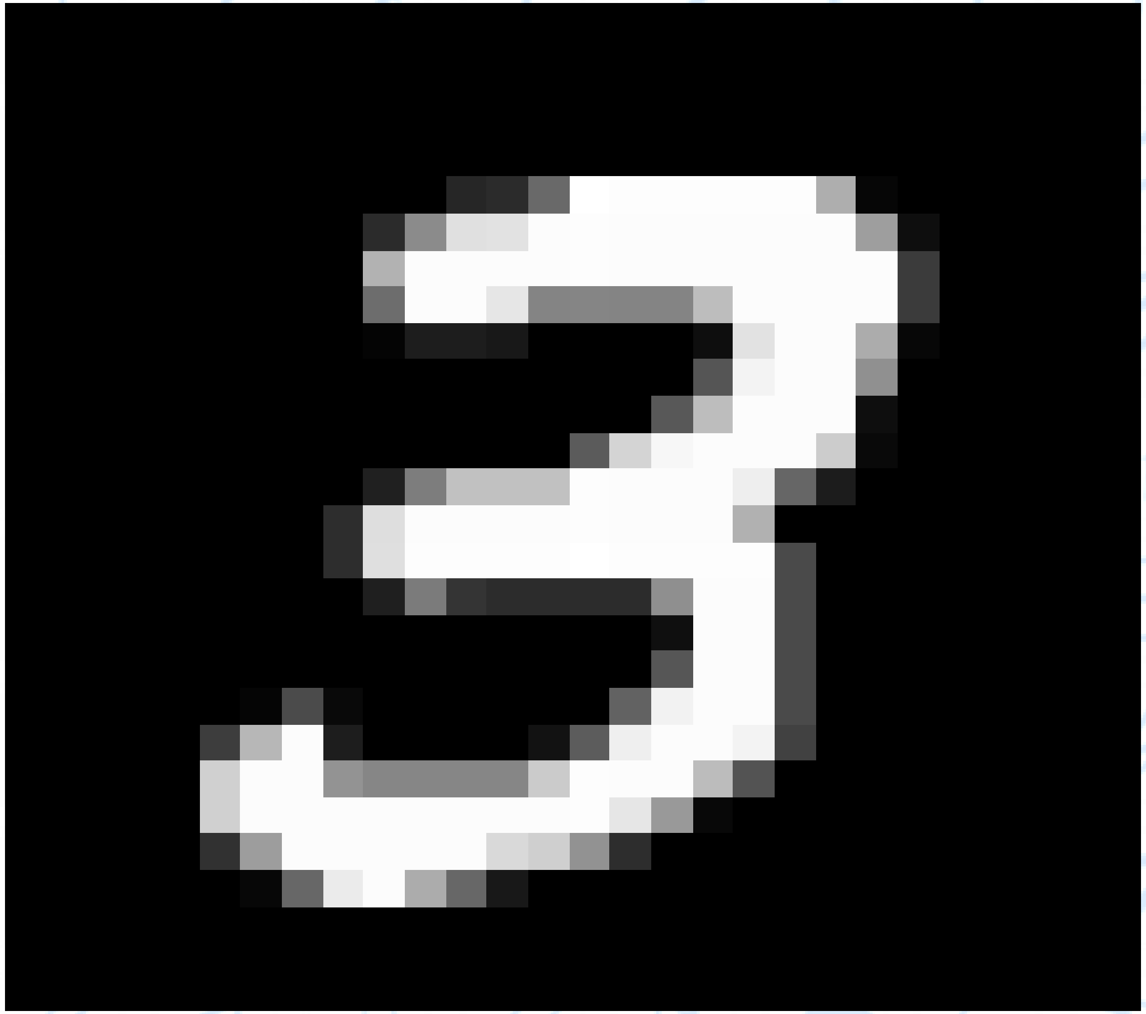
Cons

More complex

More training

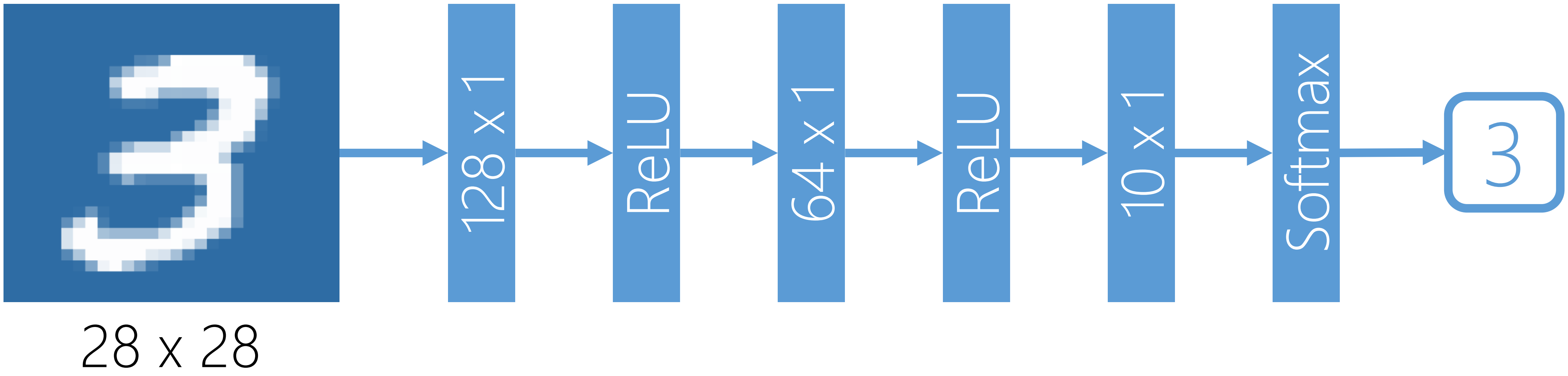
Less transparent

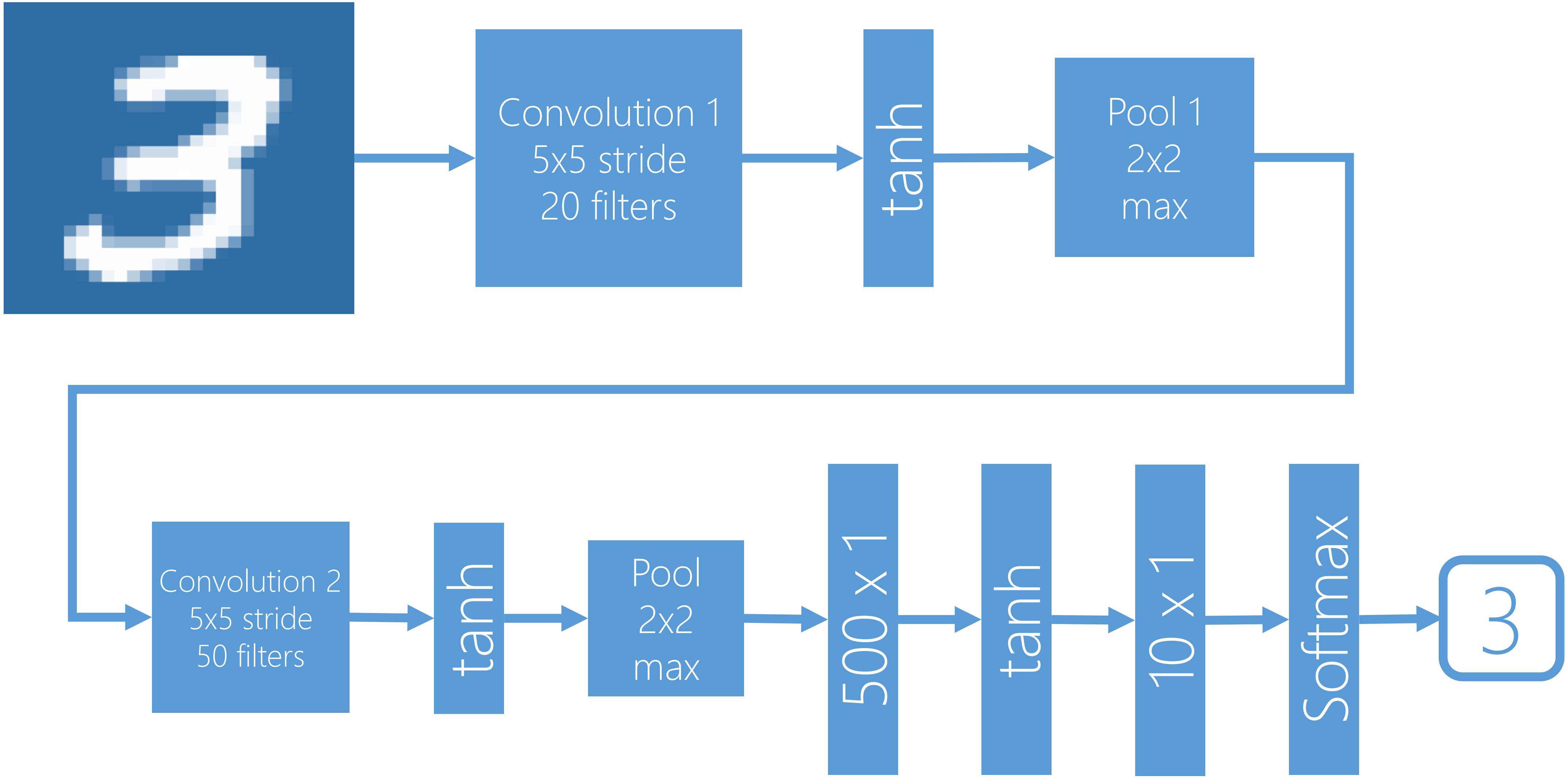
Deep Learning Demo



28 x 28

MNIST							
Label	Pixel 0	Pixel 1	Pixel 2	...	Pixel 781	Pixel 782	Pixel 783
3	0	0	0	...	0	0	0
5	0	0	0	...	0	0	0
0	0	0	0	...	0	0	0
4	0	0	0	...	0	0	0
1	0	0	0	...	0	0	0
9	0	0	0	...	0	0	0
...	...	...	...	...	...	...	...


28 x 28
128 x 1
ReLU
64 x 1
ReLU
10 x 1
Softmax
3



Convolution 1
5x5 stride
20 filters
tanh
Pool 1
2x2
max
Convolution 2
5x5 stride
50 filters
tanh
Pool
2x2
max
500 x 1
tanh
10 x 1
Softmax
3

Demo 6 – Deep Learning

Goal: Predict handwritten digits
with a deep neural network

Lab 6A – Deep Learning (Easy)

Goal: Predict handwritten digits with a deep neural network

Lab 6B – ML in Practice (Hard)

Goal: Predict handwritten digits
with CNN (LeNet)

Lab 6B – ML in Practice (Hard)

Goal: Predict handwritten digits
with CNN (LeNet)

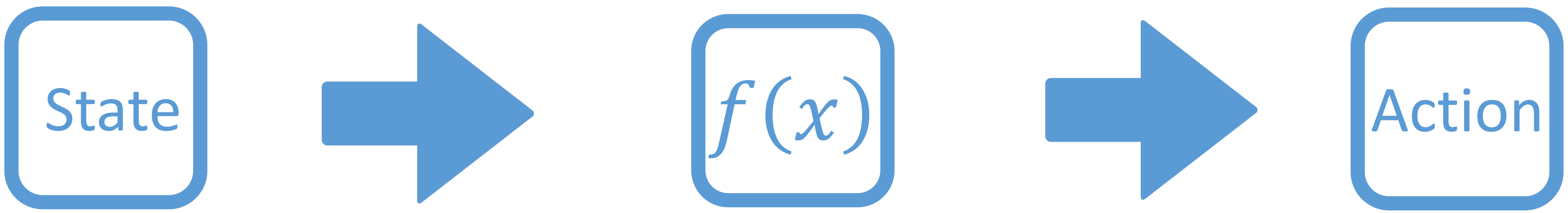
Lab 6B – ML in Practice (Hard)

Goal: Predict handwritten digits with CNN (LeNet)

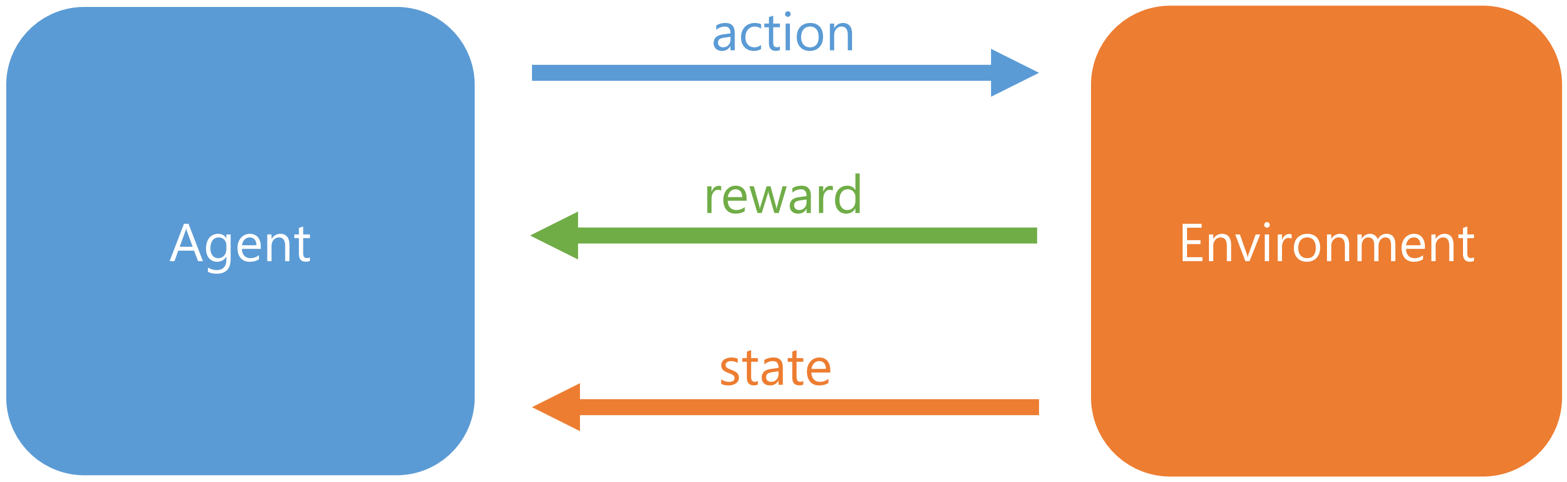
Reinforcement Learning

NOTE: Add video of RL playing video game

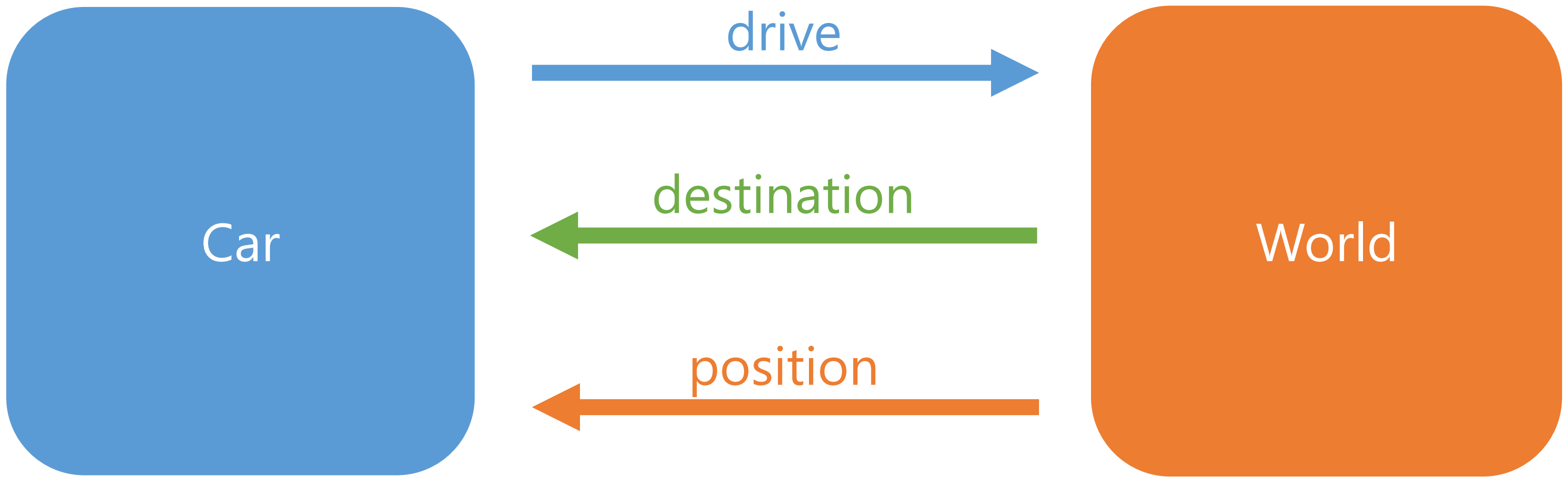
Reinforcement Learning



Reinforcement Learning

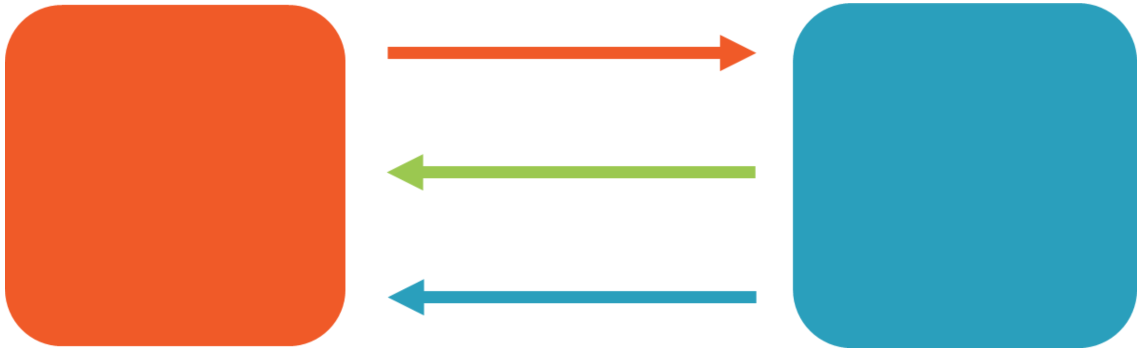


Reinforcement Learning



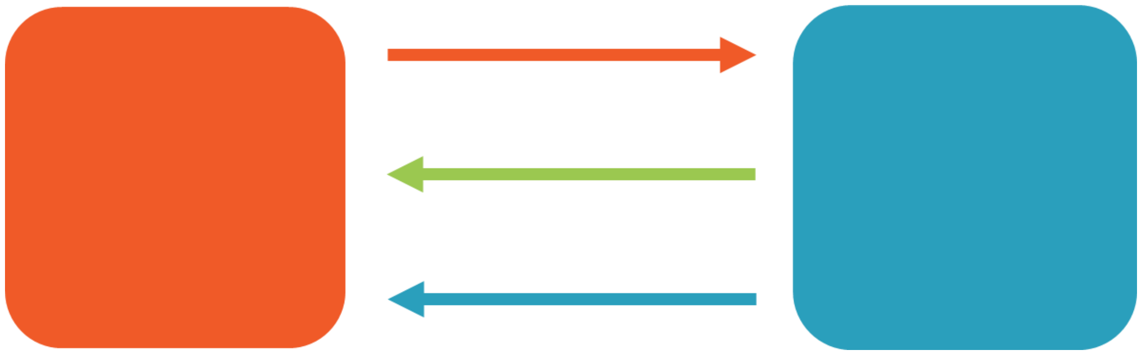
Reinforcement Learning

Action replay



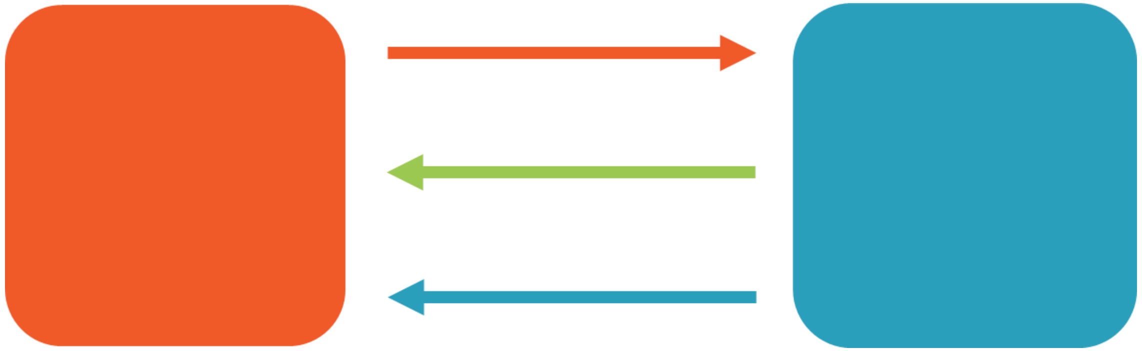
Reinforcement Learning

Action replay
Optimal policy



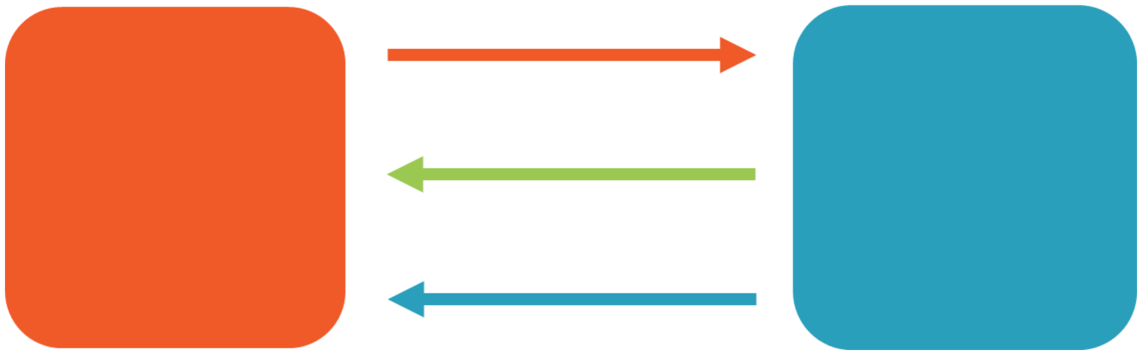
Reinforcement Learning

Action replay
Optimal policy
Discounted reward



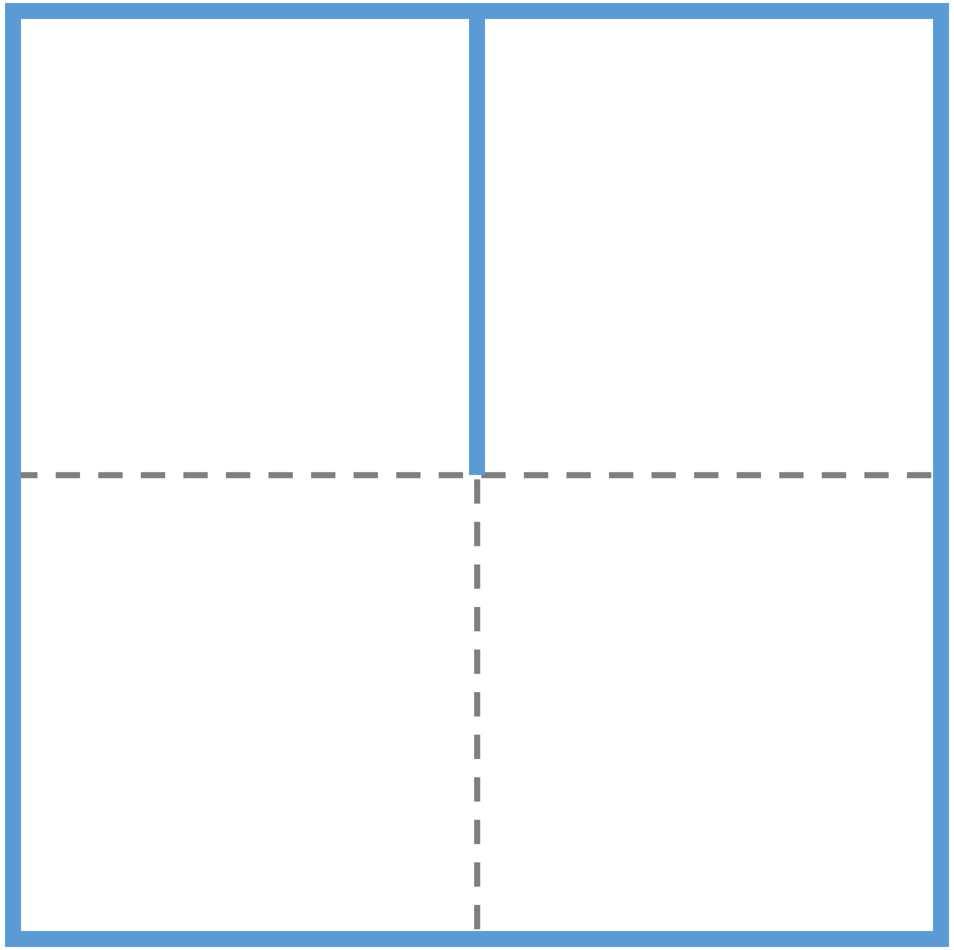
Reinforcement Learning

Action replay
Optimal policy
Discounted reward
Markov decision process

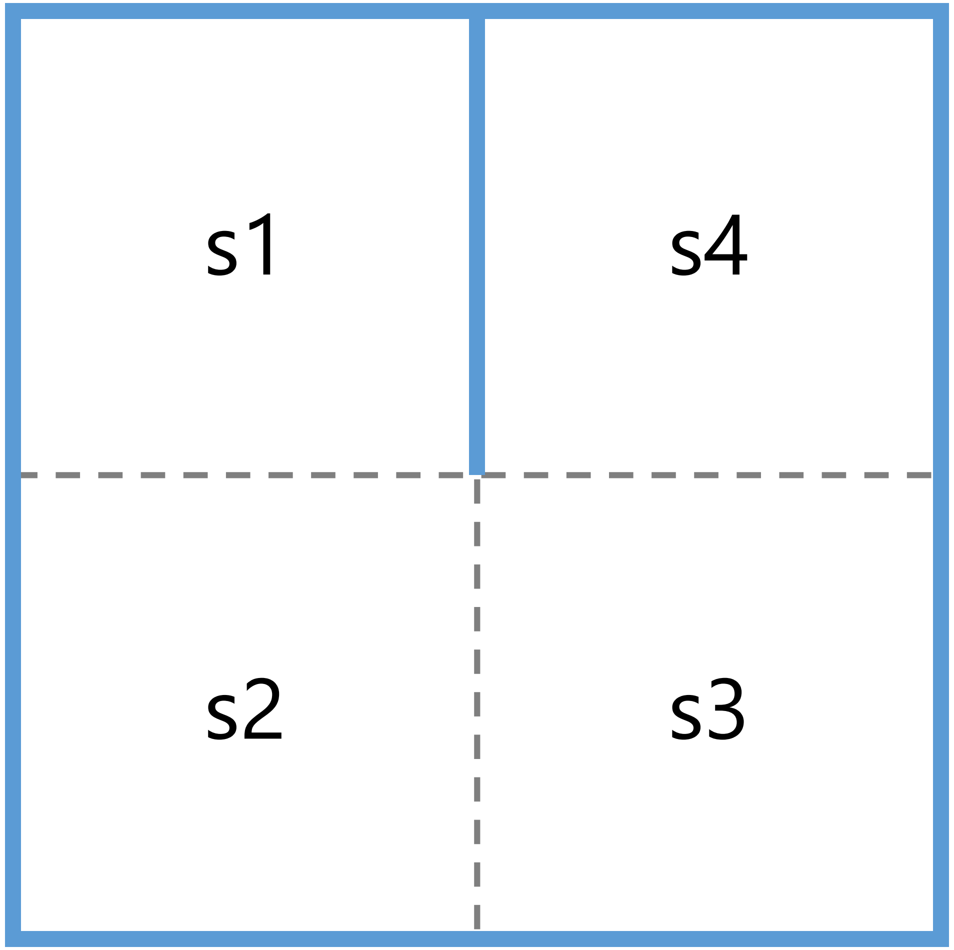


Reinforcement Learning Demo

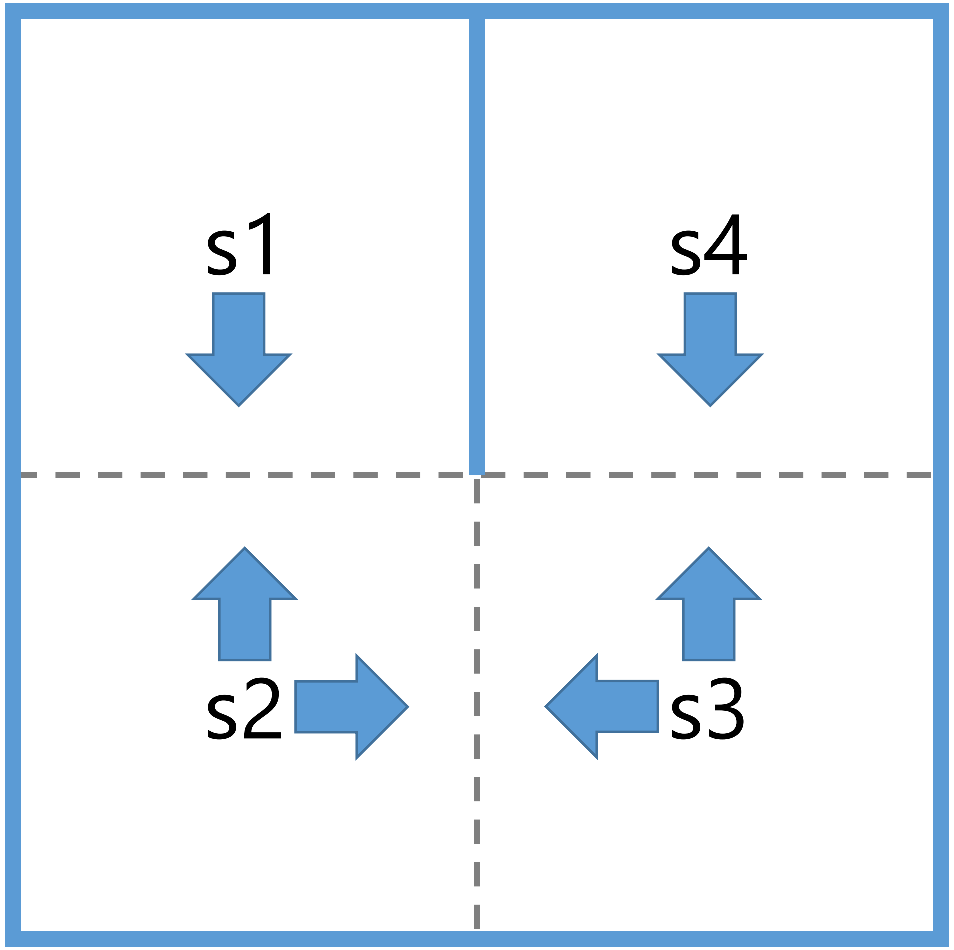
Grid World



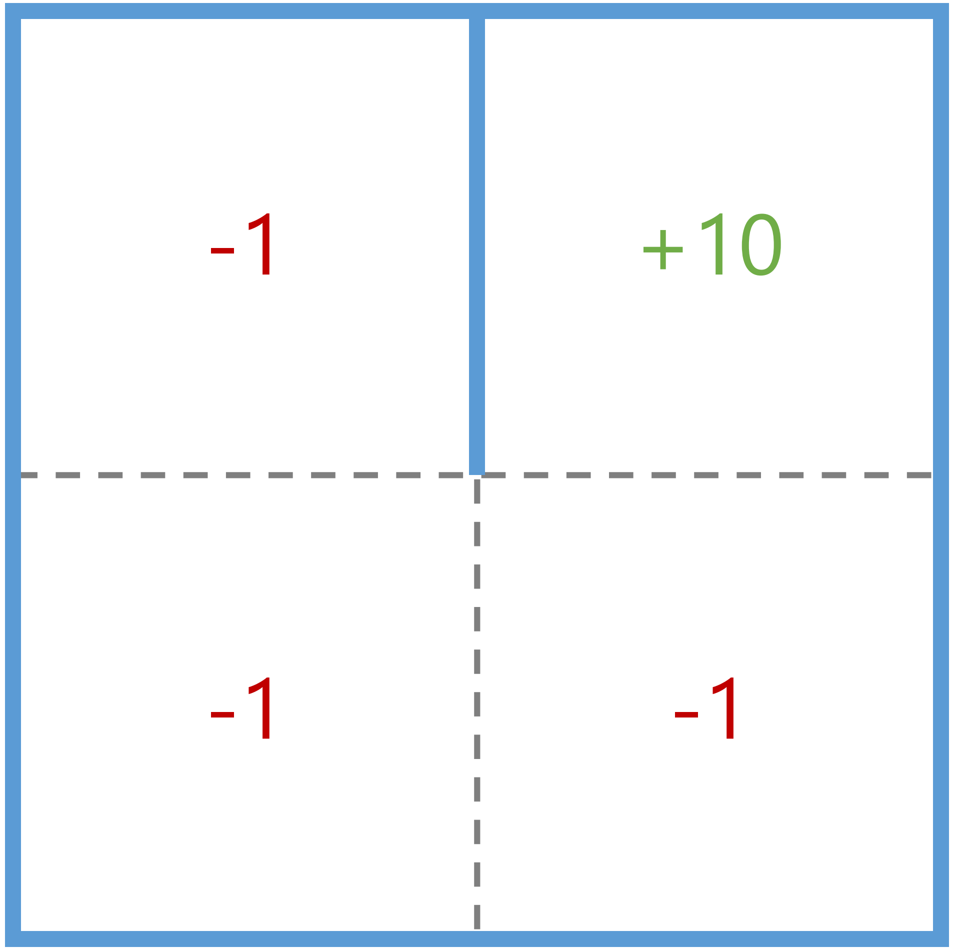
States



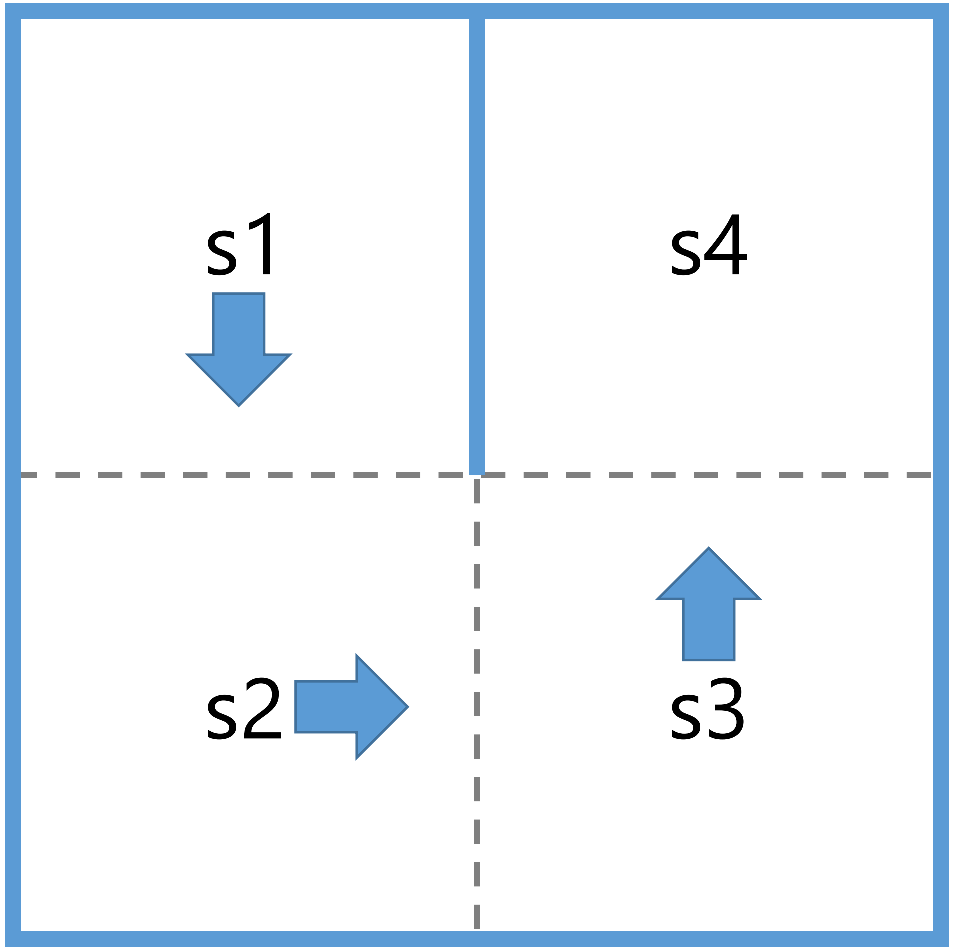
Actions



Rewards



Optimal Policy



Recap

States: s1, s2, s3, s4

Actions: up, down, left, right

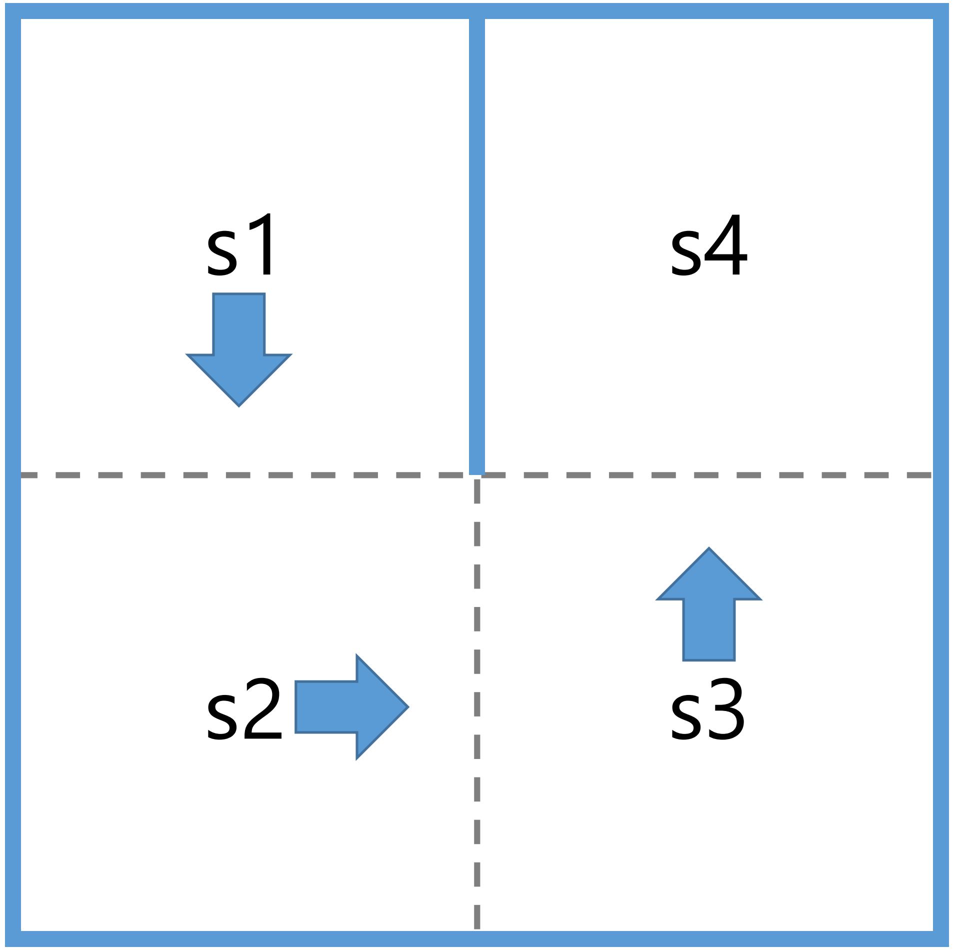
Rewards: s1, s3, s3 = -1;

s4 = 10

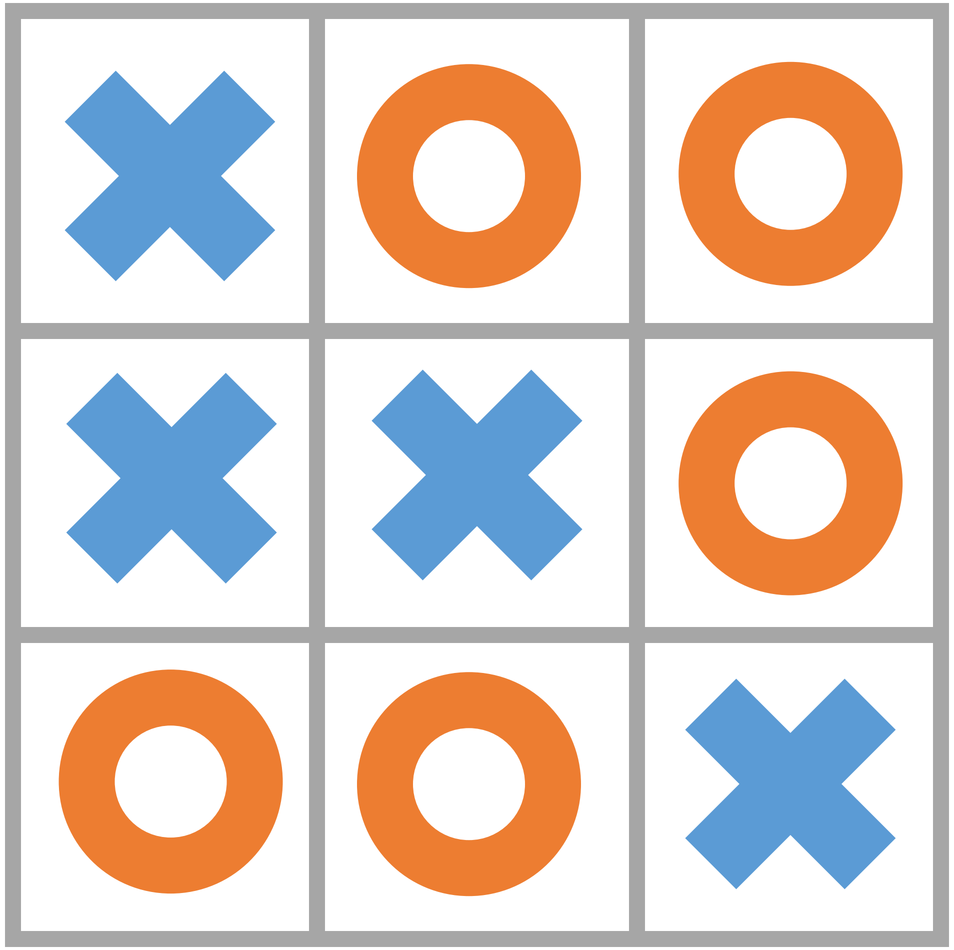
Policy: s1 = down

s2 = right

s3 = up



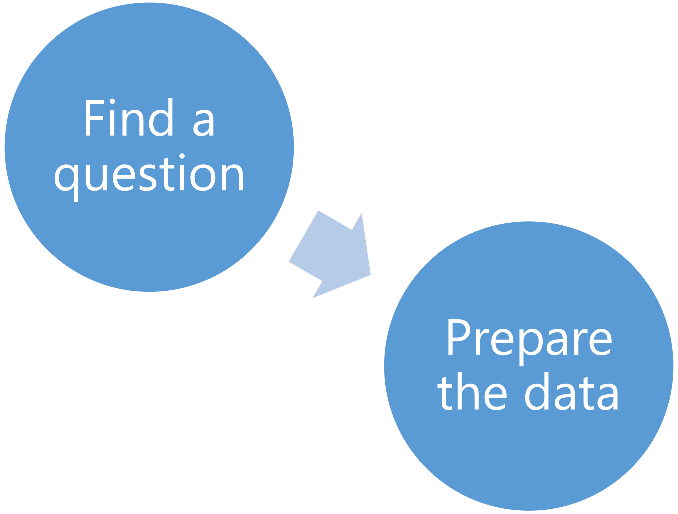
Tic-Tac-Toe

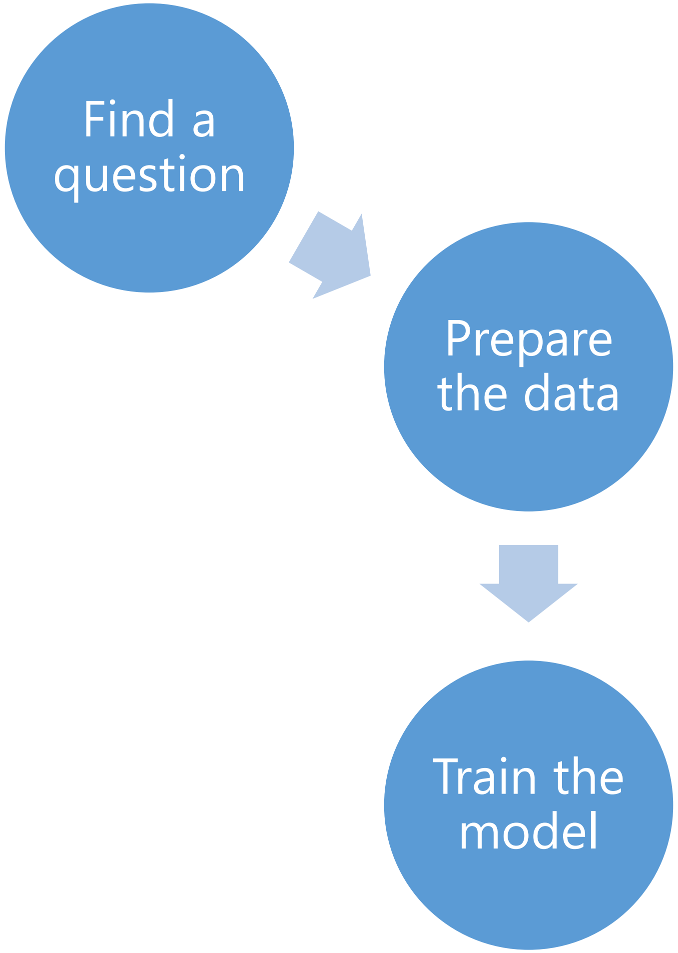


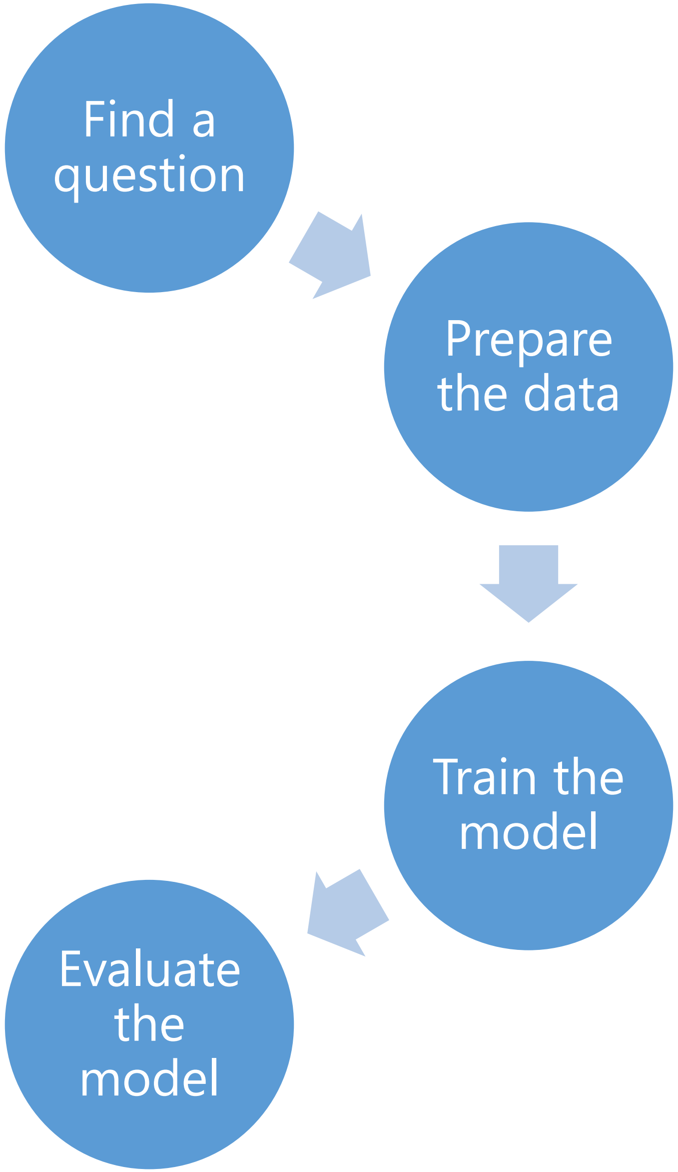
ML in Practice

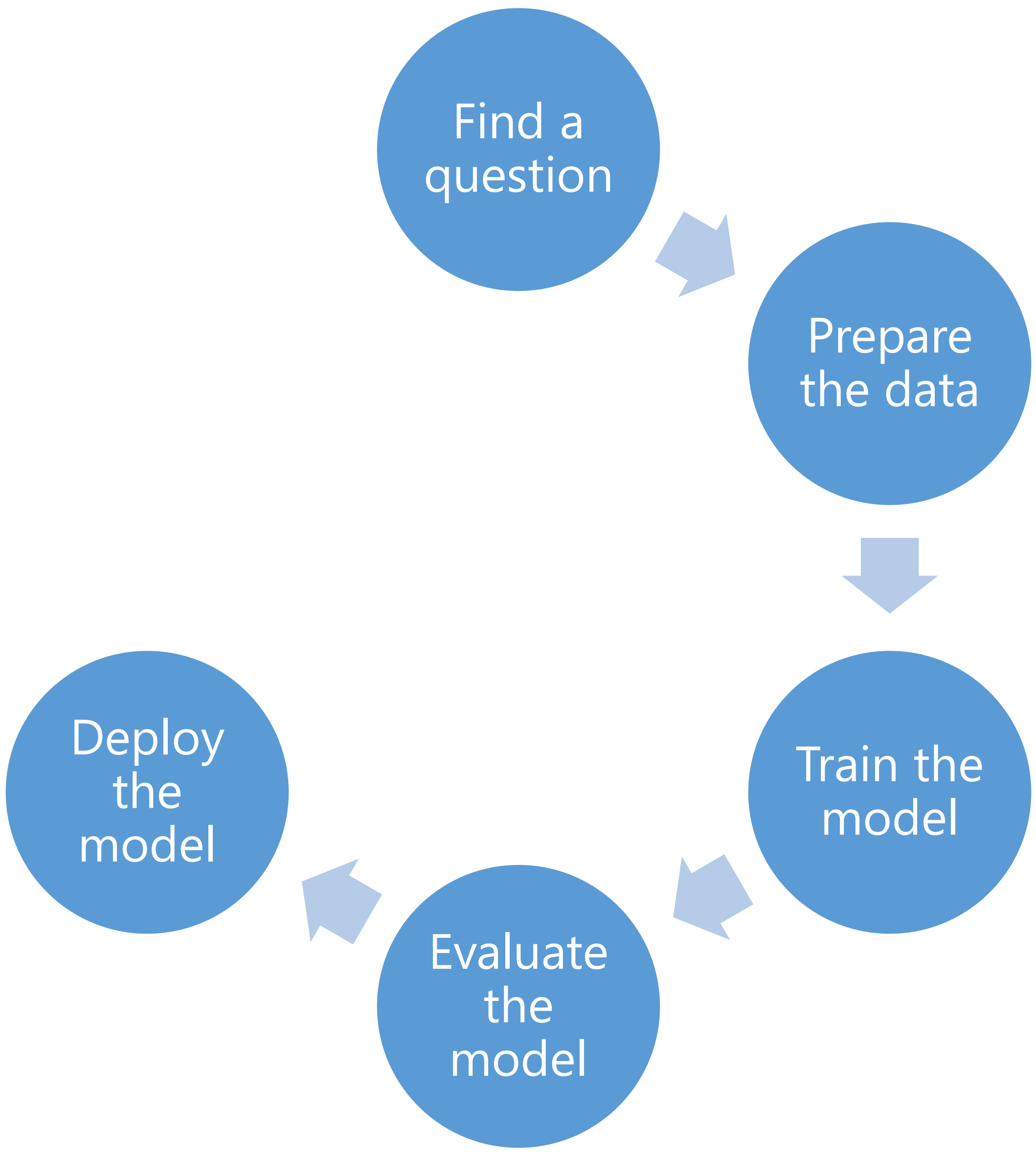
What is the machine learning process?

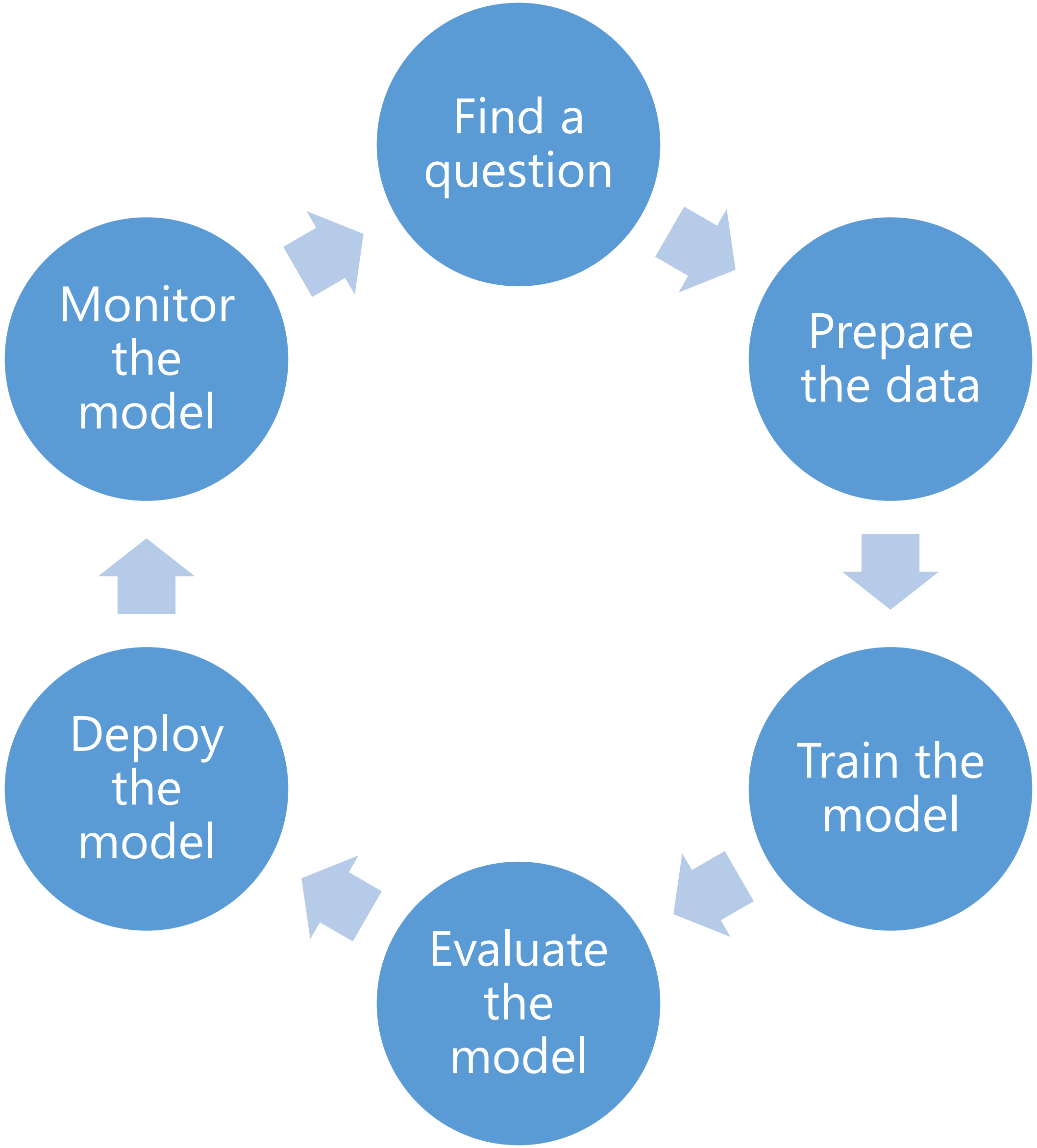
Find a question

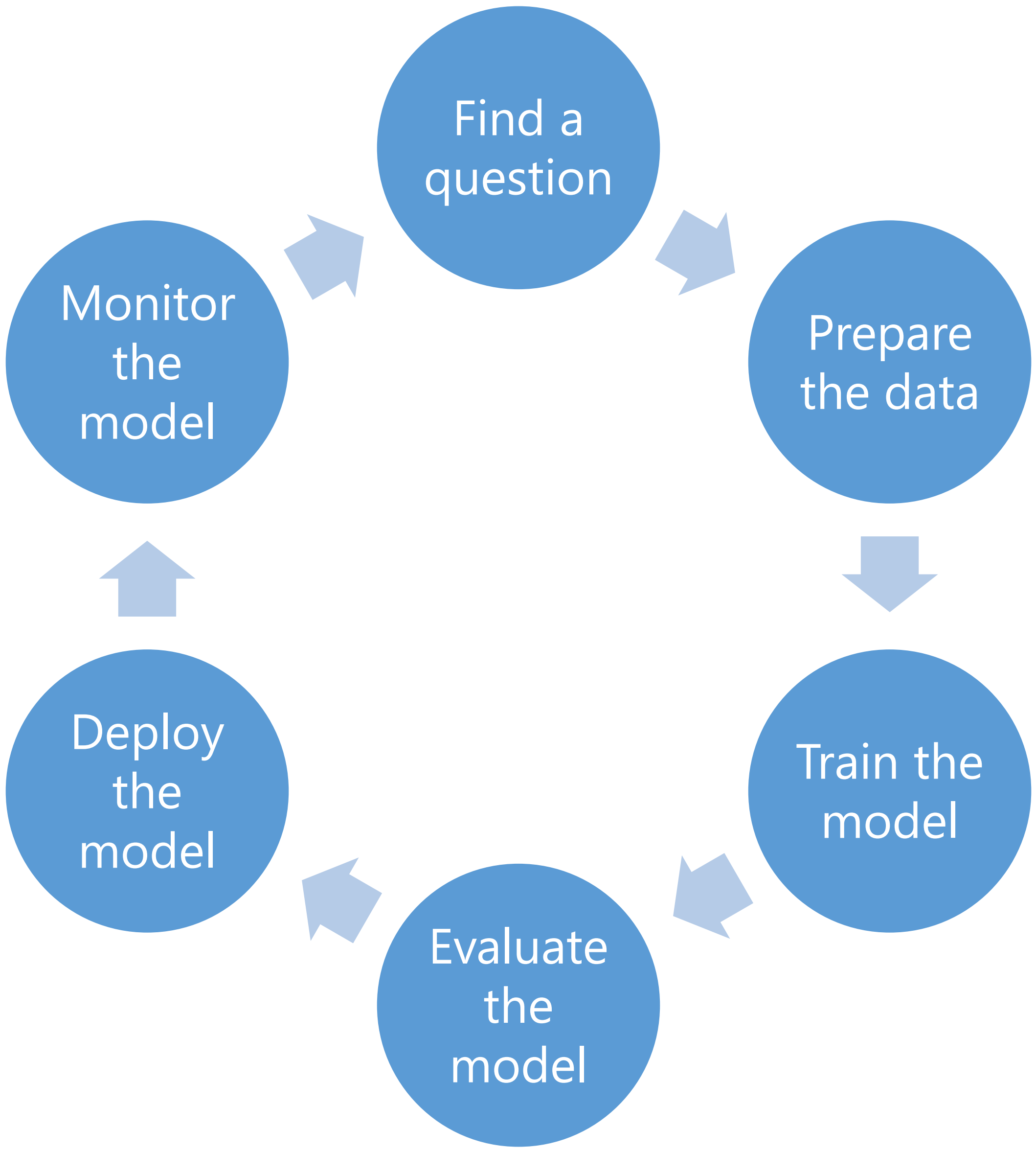

Find a question
Prepare the data


Find a question
Prepare the data
Train the model

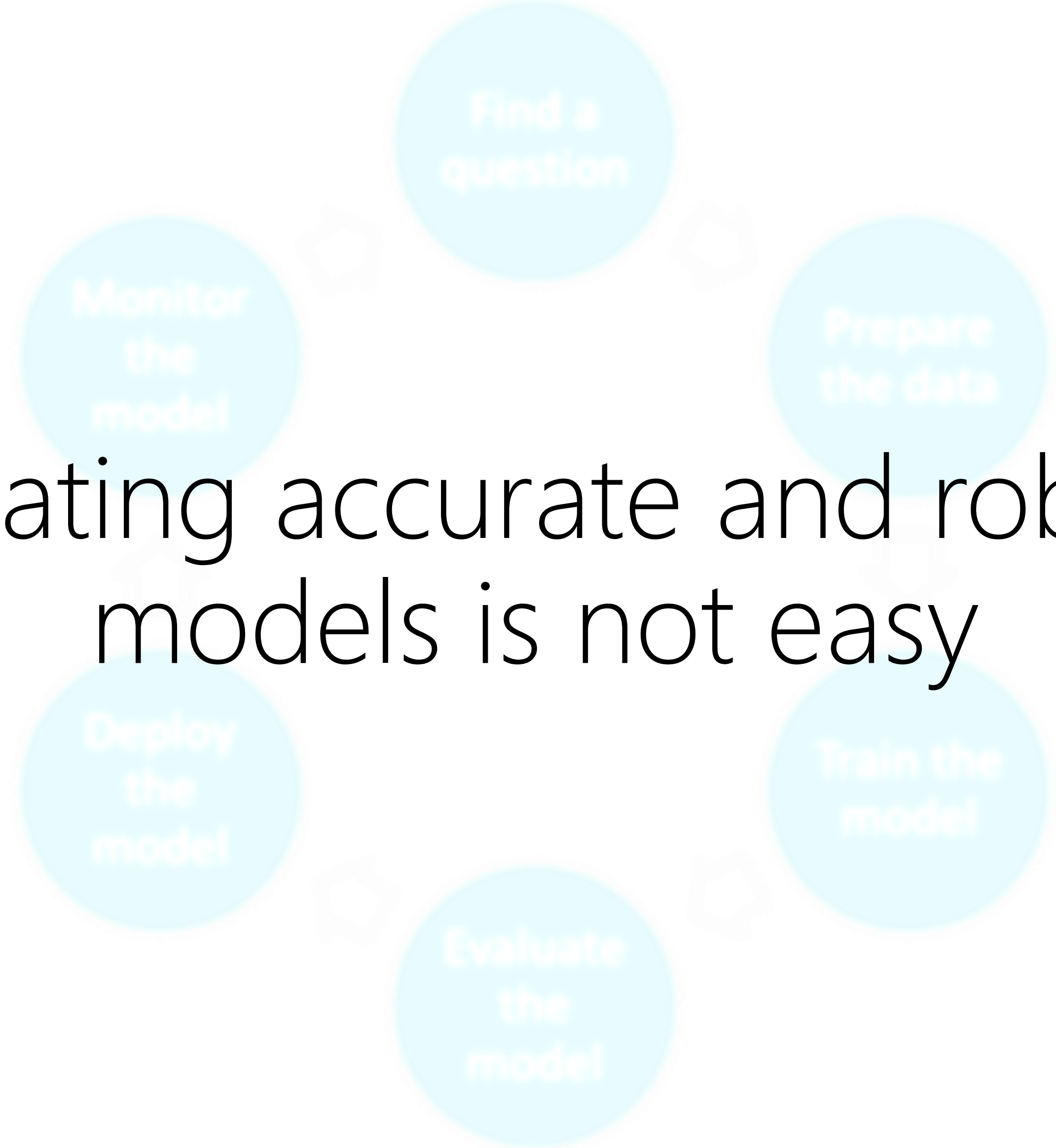

Find a question
Prepare the data
Train the model
Evaluate the model


Find a question
Prepare the data
Train the model
Evaluate the model
Deploy the model


Find a question
Prepare the data
Train the model
Evaluate the model
Deploy the model
Monitor the model

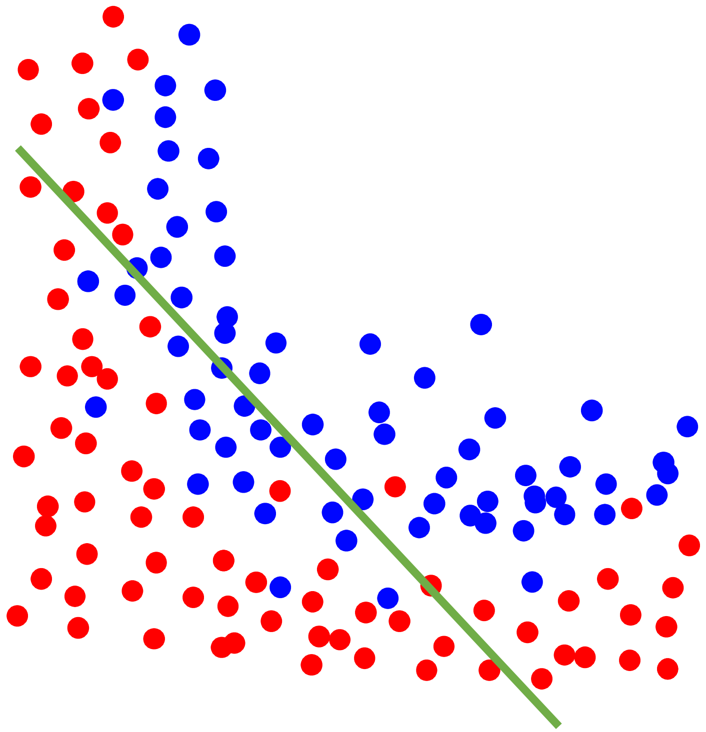

Find a question
Prepare the data
Train the model
Evaluate the model
Deploy the model
Monitor the model

Creating accurate and robust models is not easy



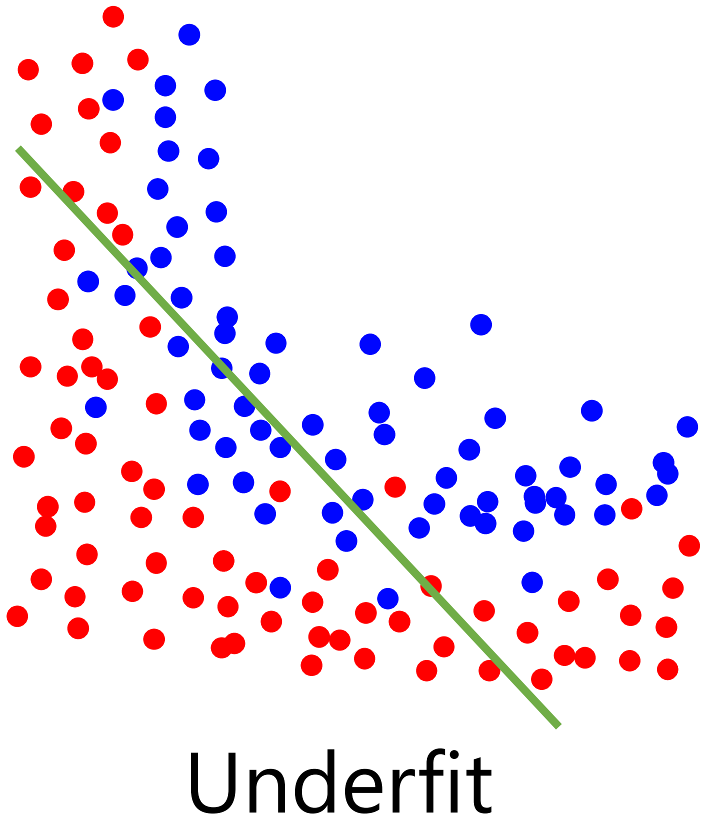
Goodness of Fit

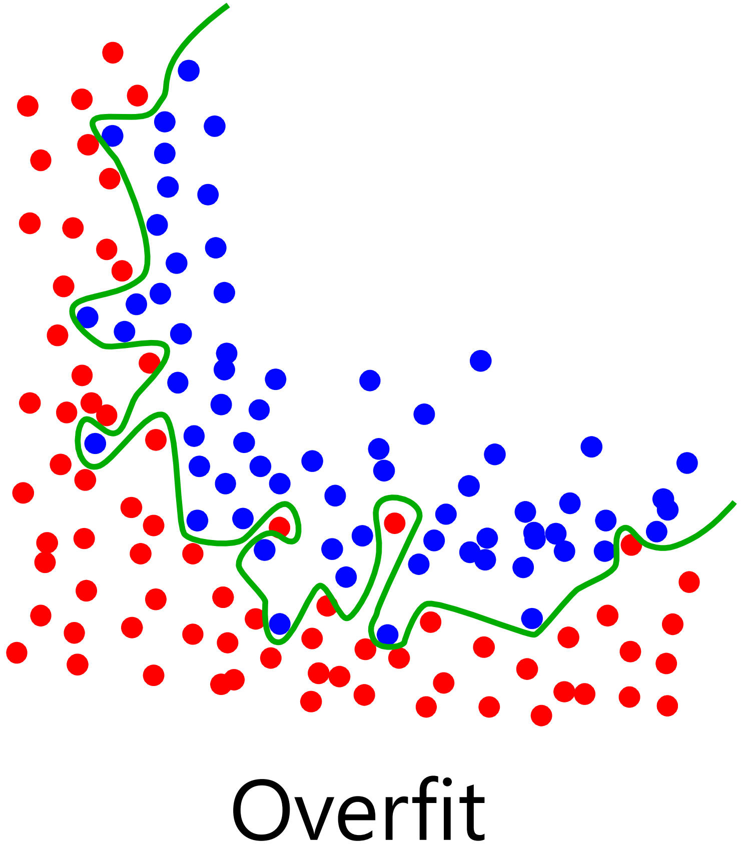
Goodness of Fit



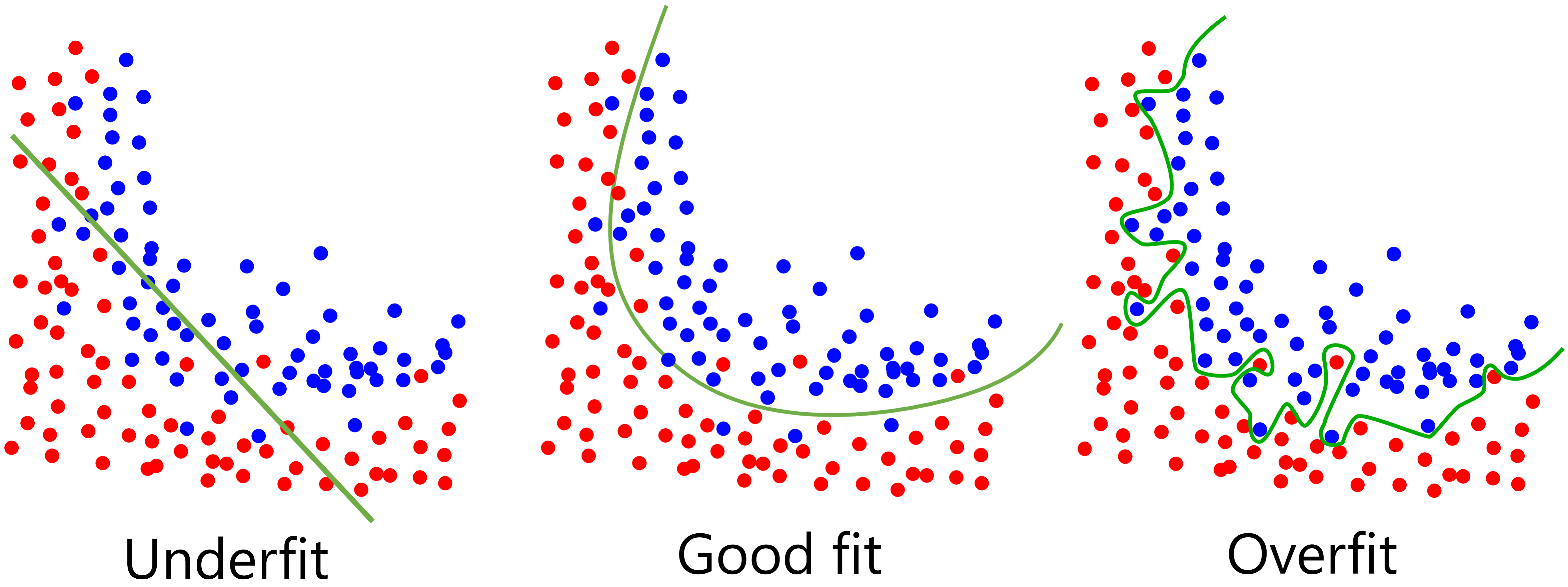
Underfit

Goodness of Fit



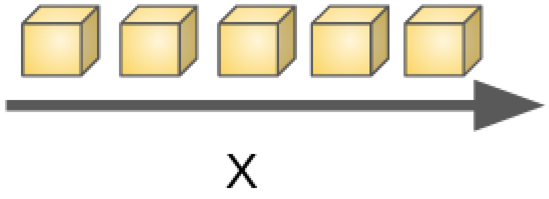


Goodness of Fit

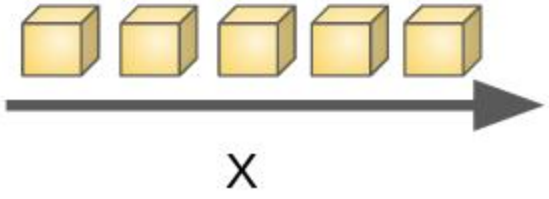


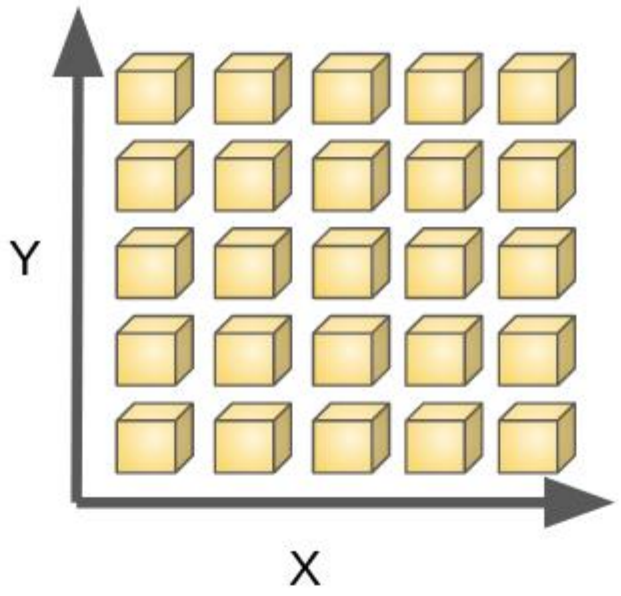
Curse of Dimensionality

Curse of Dimensionality

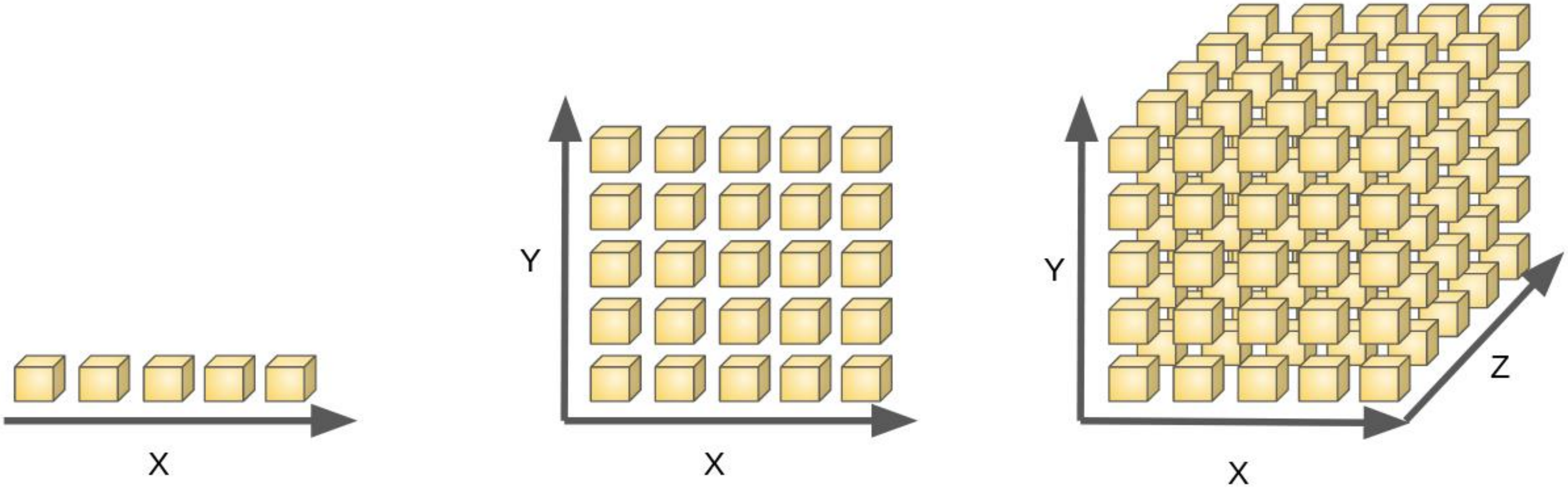


Curse of Dimensionality





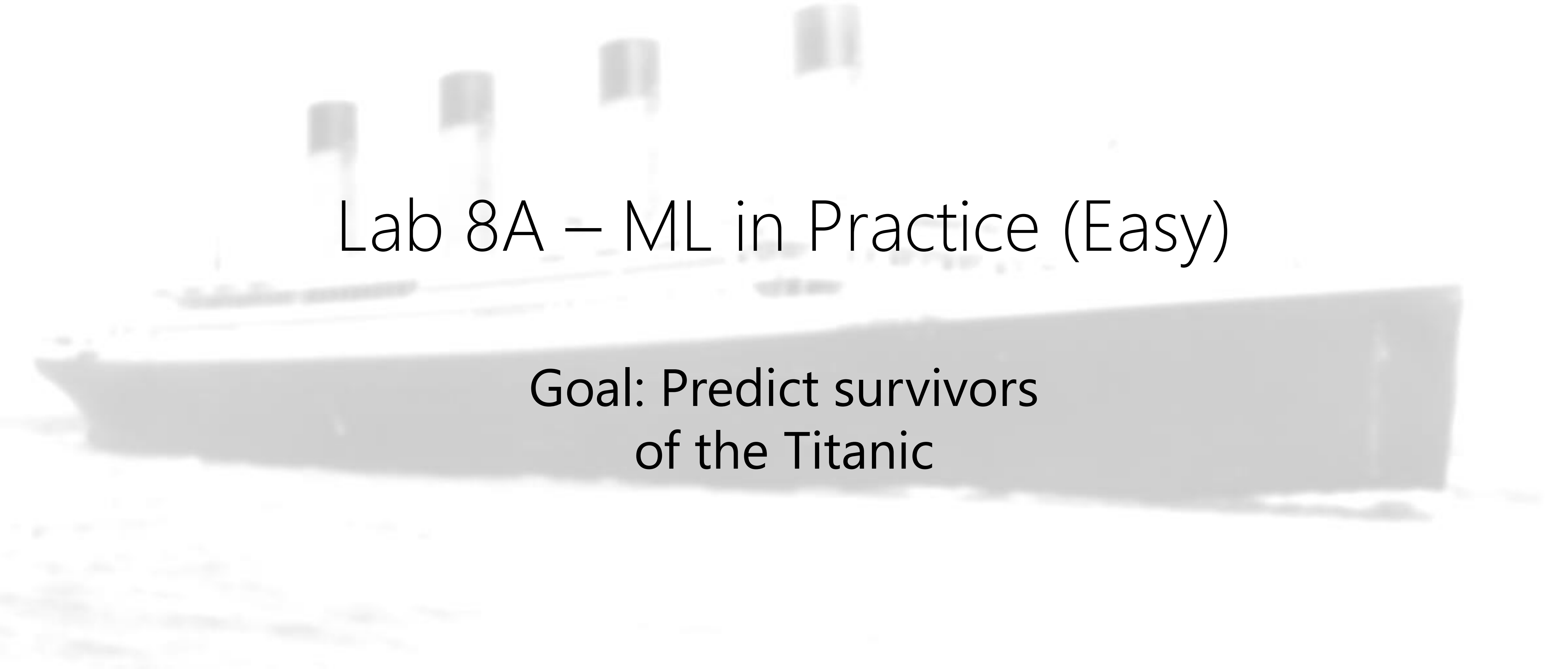
Curse of Dimensionality



Movie Break

Demo 8 – ML in Practice

Goal: Predict survivors
of the Titanic



Lab 8A – ML in Practice (Easy)

Goal: Predict survivors
of the Titanic

Lab 8B – ML in Practice (Hard)

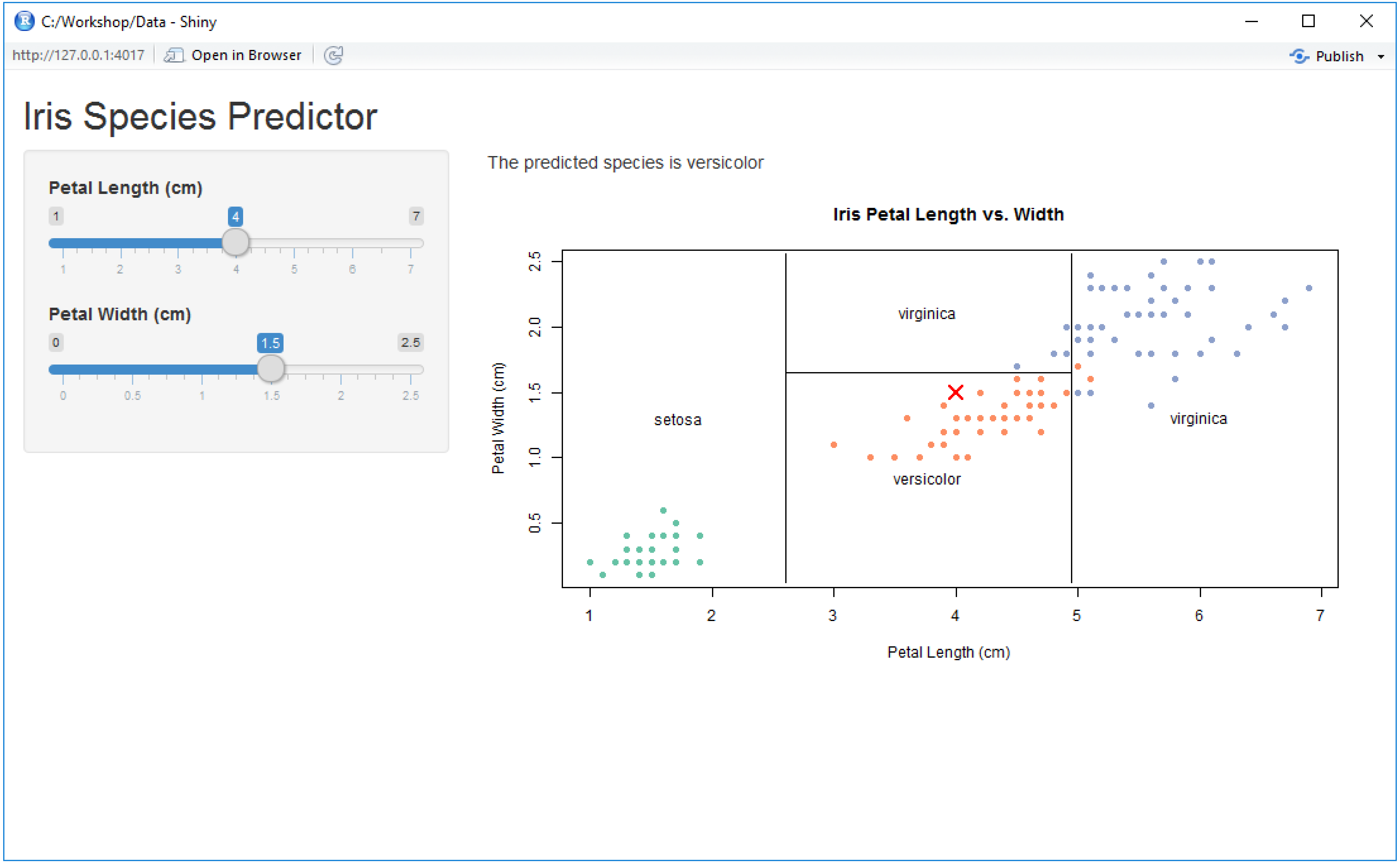
Goal: Predict risk in practice

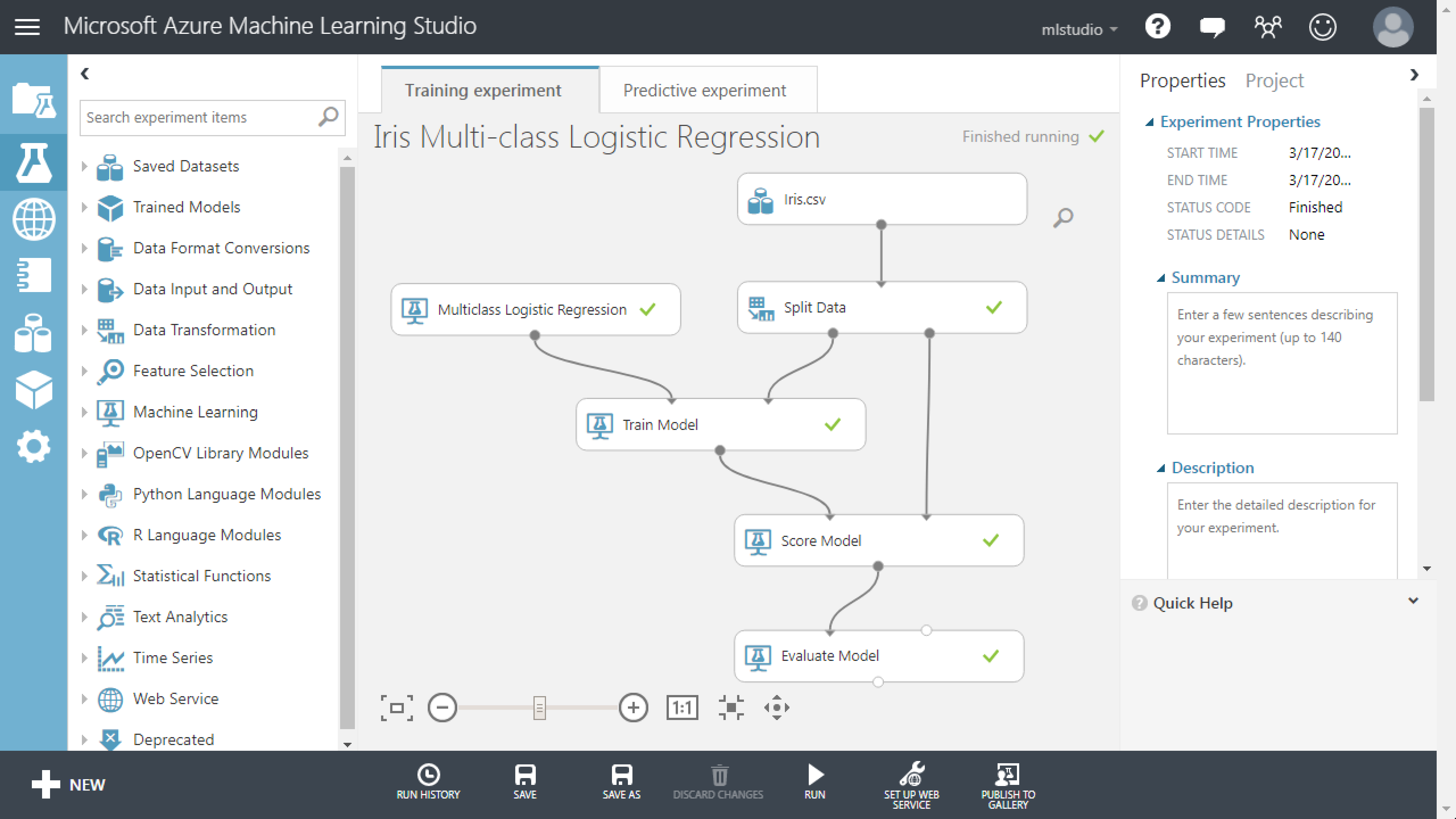
ML in Production

How to Deploy to Production

Deploy to web app (Shiny)
Deploy to cloud (Azure ML)
Deploy to server (ML Server)
Deploy to any app (ONNX)





C:/Workshop/Data - Shiny
http://127.0.0.1:4017
Open in Browser
Publish
Iris Species Predictor
Petal Length (cm)
1
4
7
1 2 3 4 5 6 7
Petal Width (cm)
0
1.5
2.5
0 0.5 1 1.5 2 2.5
The predicted species is versicolor
Iris Petal Length vs. Width
virginica
setosa
versicolor
virginica
Petal Width (cm)
0.5 1.0 1.5 2.0 2.5
1 2 3 4 5 6 7
Petal Length (cm)


Microsoft Azure Machine Learning Studio
mlstudio
Search experiment items
Saved Datasets
Trained Models
Data Format Conversions
Data Input and Output
Data Transformation
Feature Selection
Machine Learning
OpenCV Library Modules
Python Language Modules
R Language Modules
Statistical Functions
Text Analytics
Time Series
Web Service
Deprecated
Training experiment
Predictive experiment
Iris Multi-class Logistic Regression
Finished running
Iris.csv
Multiclass Logistic Regression
Split Data
Train Model
Score Model
Evaluate Model
1:1
Properties
Project
Experiment Properties
START TIME
3/17/20...
END TIME
3/17/20...
STATUS CODE
Finished
STATUS DETAILS
None
Summary
Enter a few sentences describing your experiment (up to 140 characters).
Description
Enter the detailed description for your experiment.
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

Conclusion

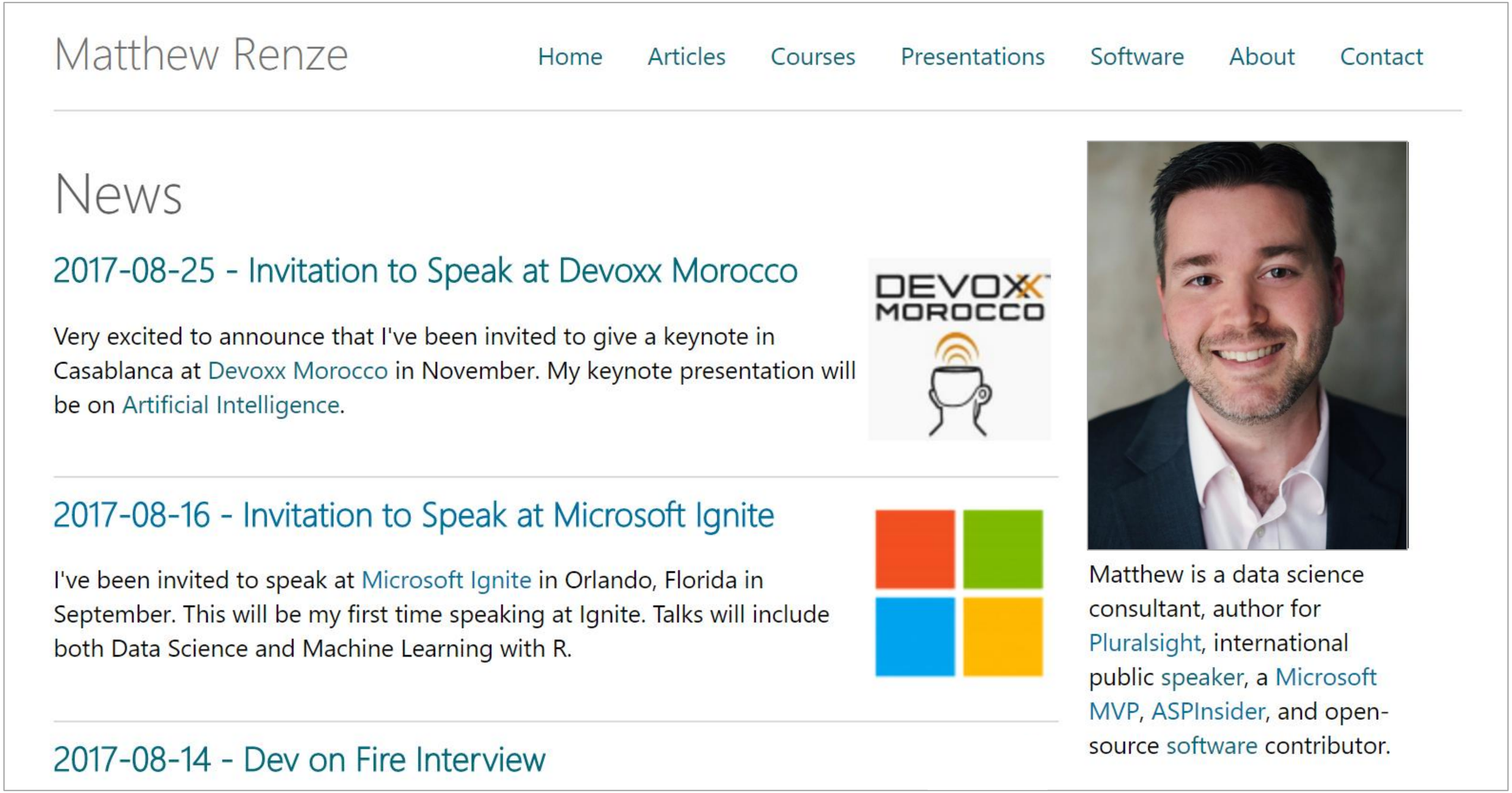

This is just the tip of the iceberg!

Where do we go from here?

Pluralsight Courses

Data Science with R
Data Science: The Big Picture
Deep Learning: The Big Picture
Exploratory Data Analysis with R
Data Visualization with R (3-part)

https://www.pluralsight.com/authors/matthew-renze



www.matthewrenze.com

Feedback

Very important to me!

What did you like?

What could I improve?

Conclusion

1. Intro to ML and R
2. Classification
3. Regression
4. Clustering
5. Ensemble Learning
6. Deep Learning
7. ML in Practice
8. ML in Production


Caution
WARNING
ELECTRICITY

Are you prepared?

Is your organization?

Is our world prepared?


Caution
WARNING
ELECTRICITY

Contact Info

Matthew Renze
Data Science Consultant
Renze Consulting

Twitter: @matthewrenze
Email: info@matthewrenze.com
Website: www.matthewrenze.com

Thank You! :)