

# Machine Learning with R

@MatthewRenze

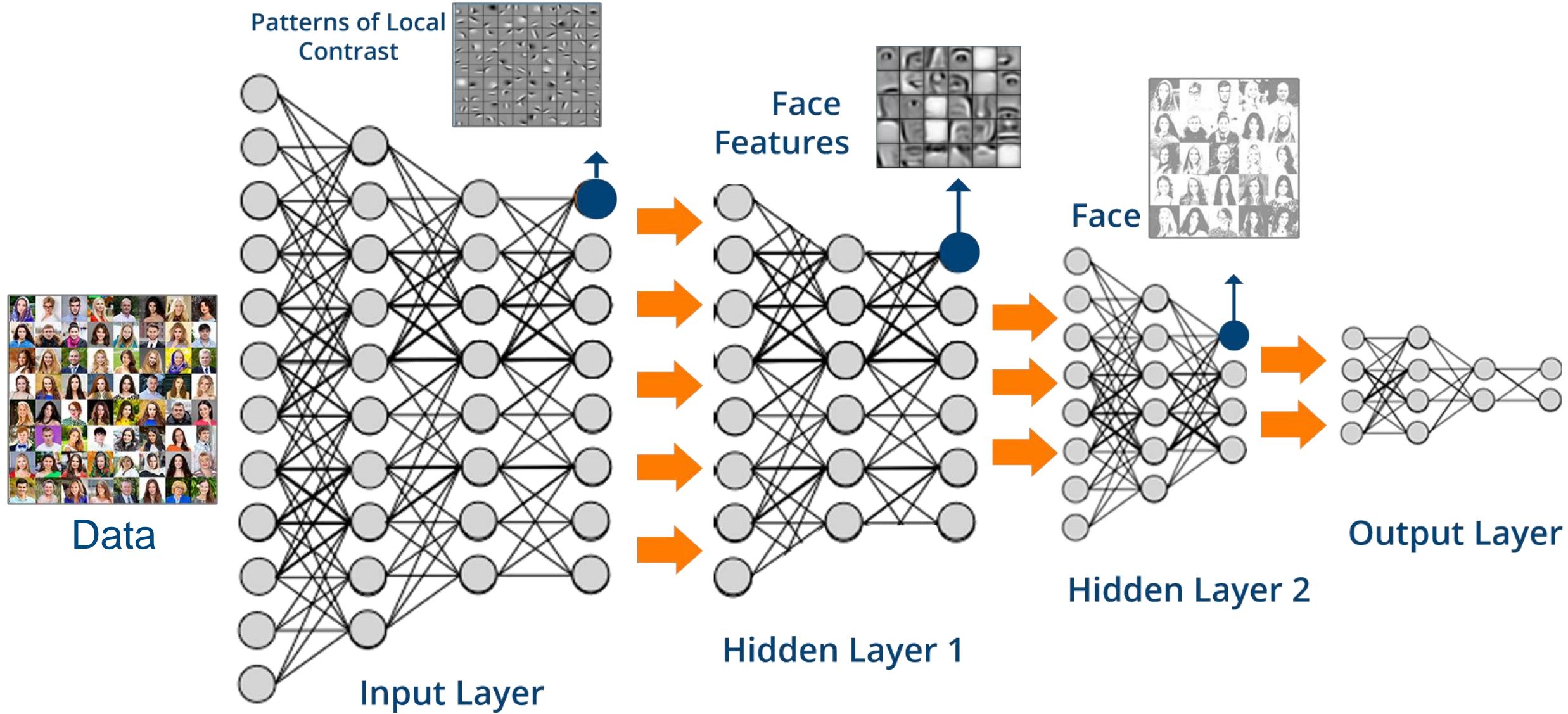
#KCDC



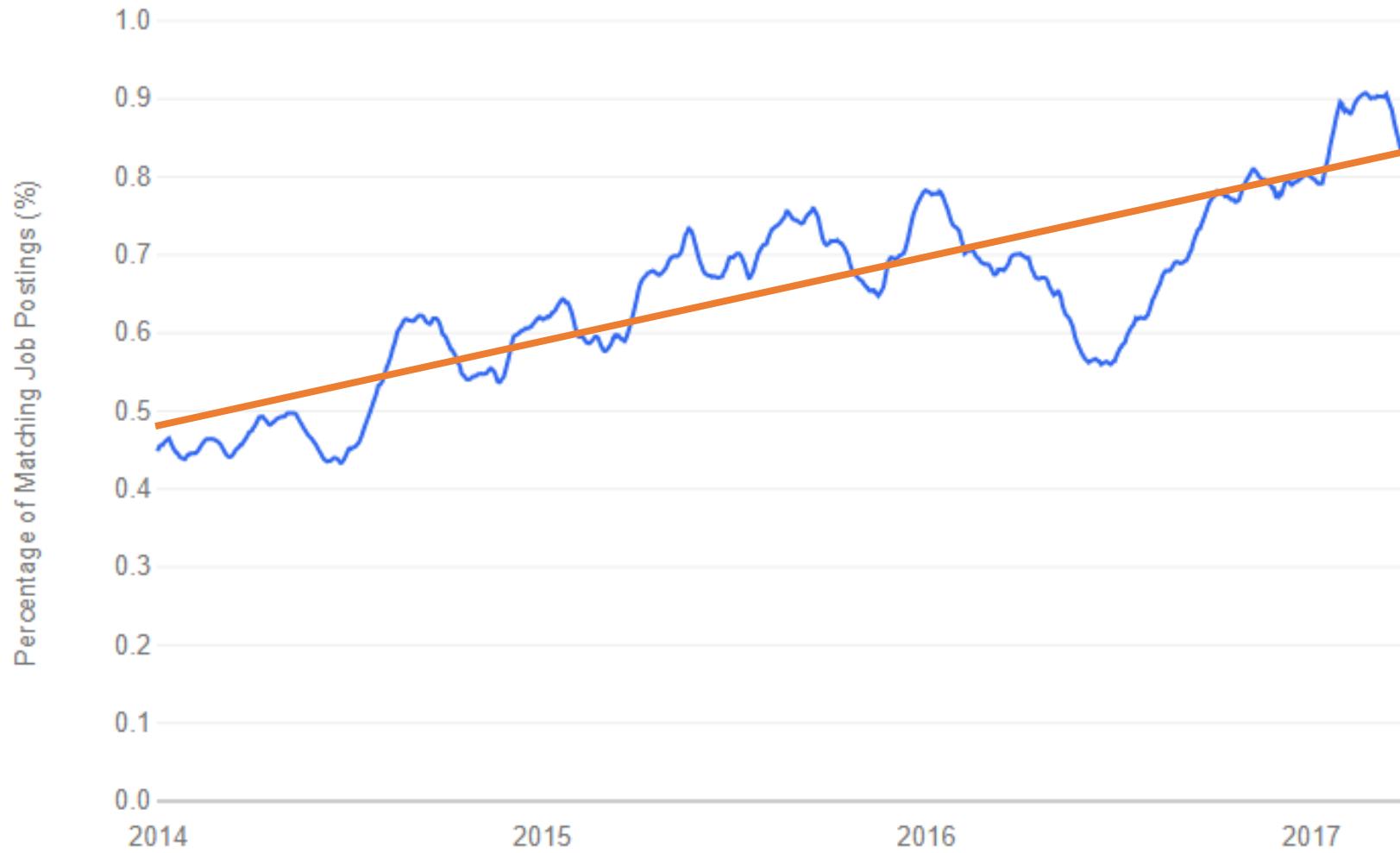


```
function updatePhotoDescription() {
    if (descriptions.length > (page * 9) + (currentImage - 1)) {
        document.getElementById('bigImageDesc').innerHTML = descriptions[currentImage - 1];
    }
}

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 = 'image/min/' + photos[i - 1];
            document.getElementById(elementIdBig).src = 'image/big/' + photos[i - 1];
        } else {
            document.getElementById(elementId).src = '';
        }
        i++;
    }
}
```

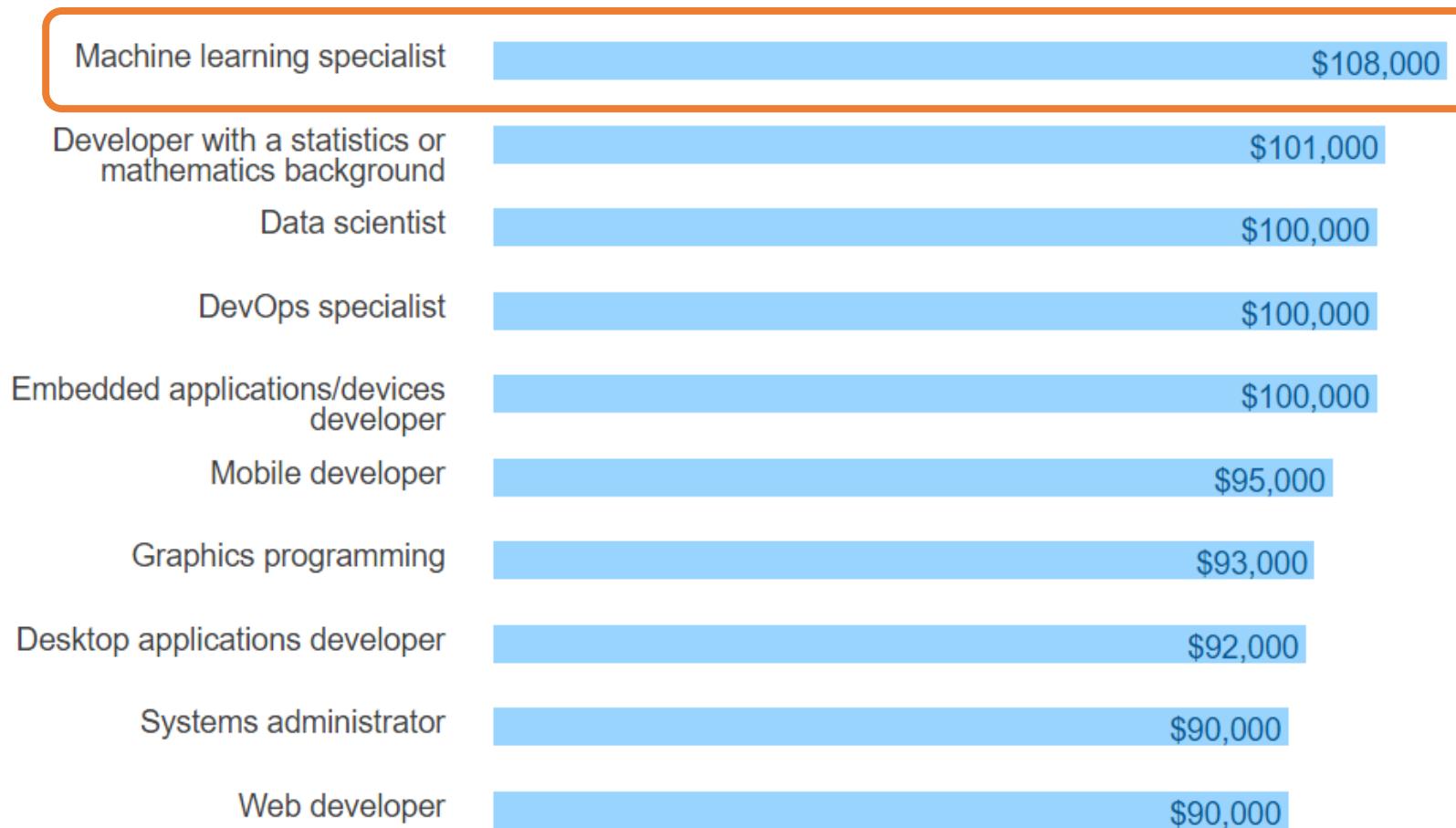


# Job Postings for Machine Learning



Source: Indeed.com

# Average Salary by Job Type (USA)





TR

# Overview

1. Introduction to ML
2. Introduction to R
3. Classification
4. Regression
5. Beyond the Basics



# About Me

Data Science Consultant  
Education

B.S. in Computer Science  
B.A. in Philosophy

## Community

Public Speaker  
Pluralsight Author  
Microsoft MVP  
ASPInsider  
Open-source Software

IOWA STATE  
UNIVERSITY

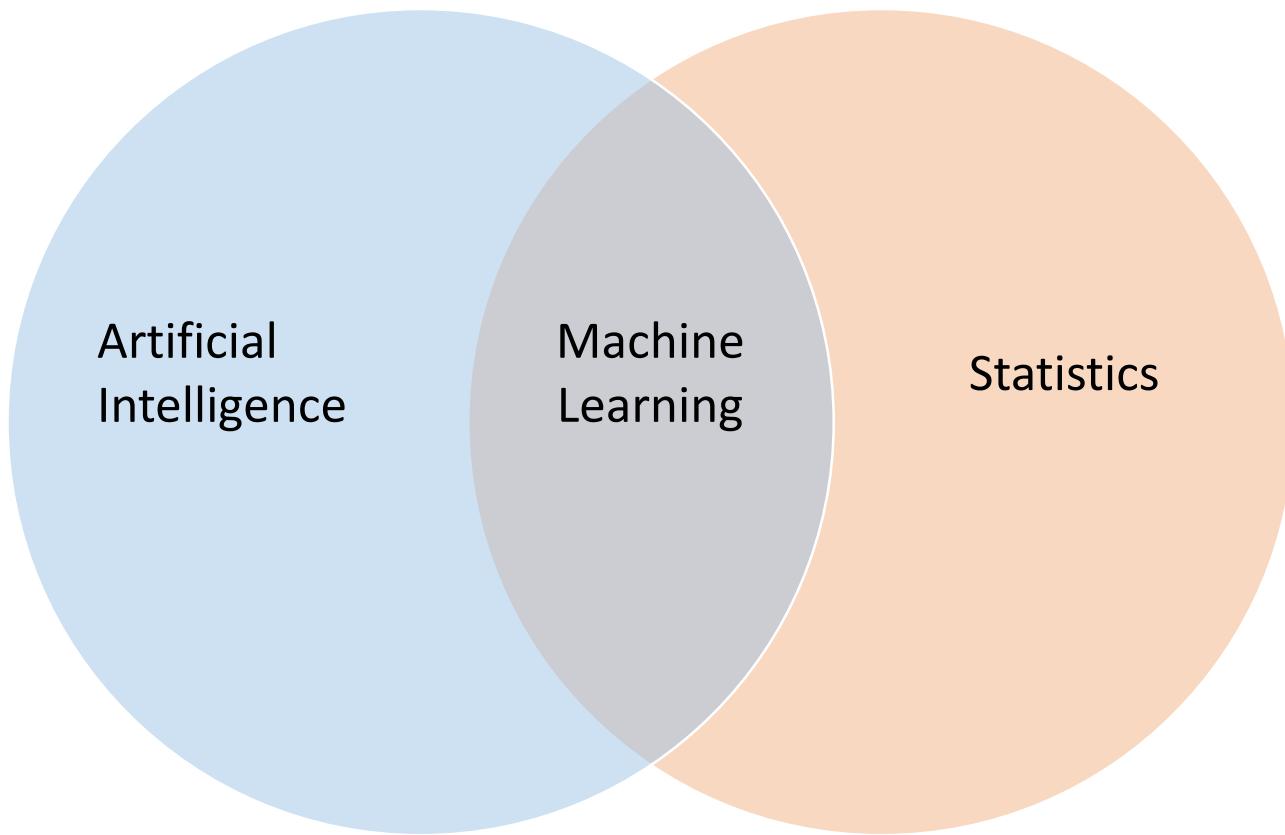


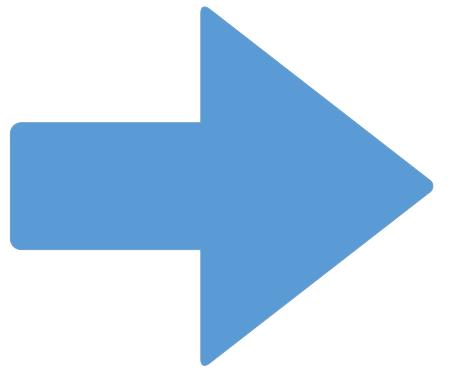
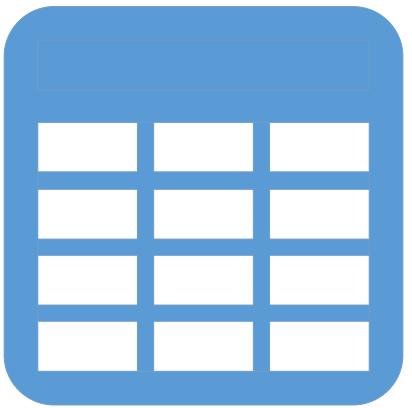
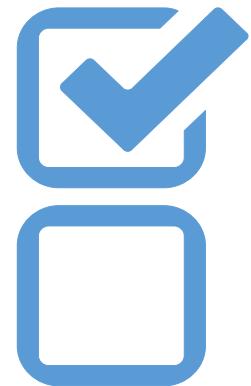
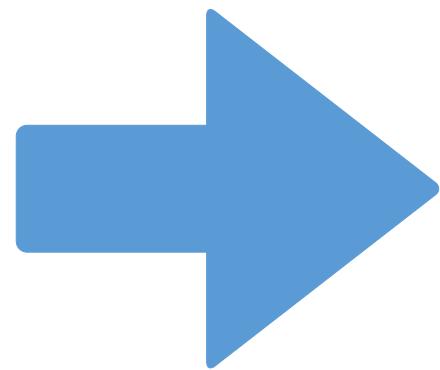
# How Does This Apply to Me?

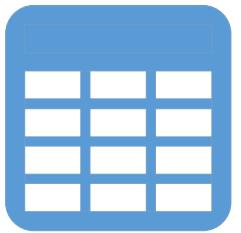
- Make decisions using data
- Make predictions using data
- Make recommendations using data
- Write code that does all these things

# Introduction to Machine Learning

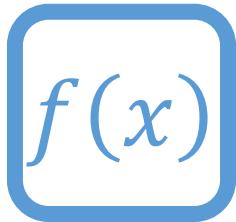
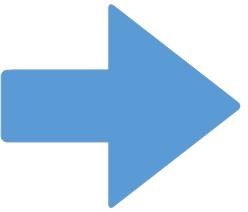
# What is Machine Learning?



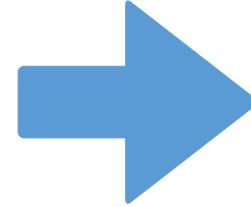
 $f(x)$ 



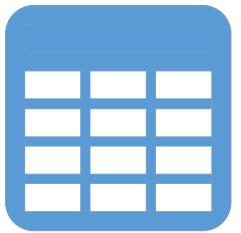
Data



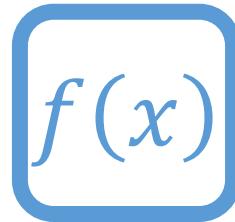
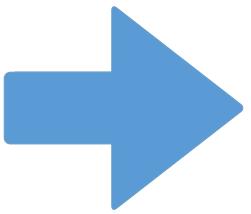
Function



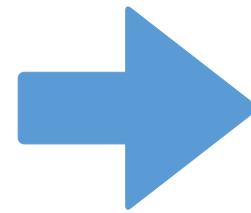
Prediction



Data



Function



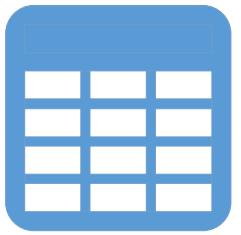
Prediction



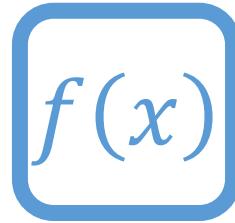
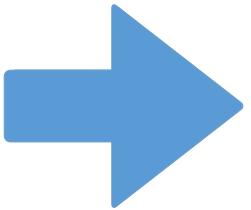
Cat



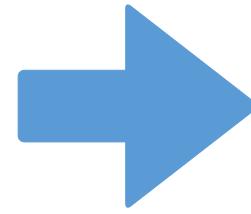
Dog



Data



Function



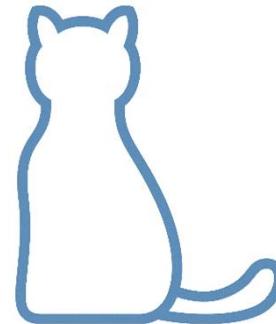
Prediction



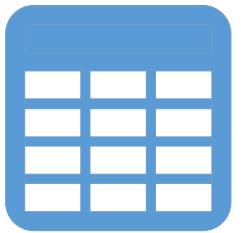
Cat



Dog



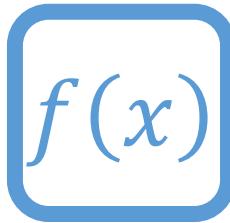
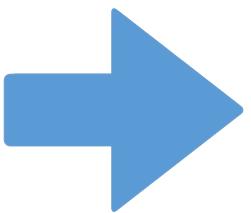
Is cat?



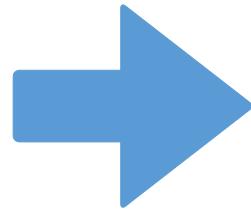
Data



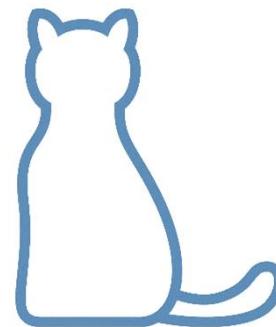
Cat



Function



Prediction



Is cat?



Dog



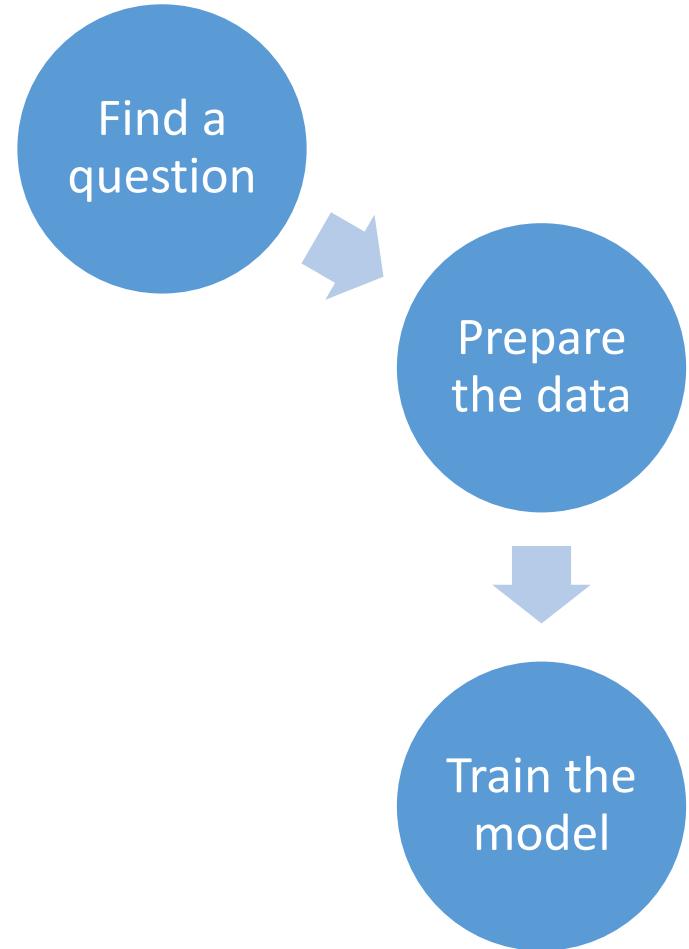
Yes

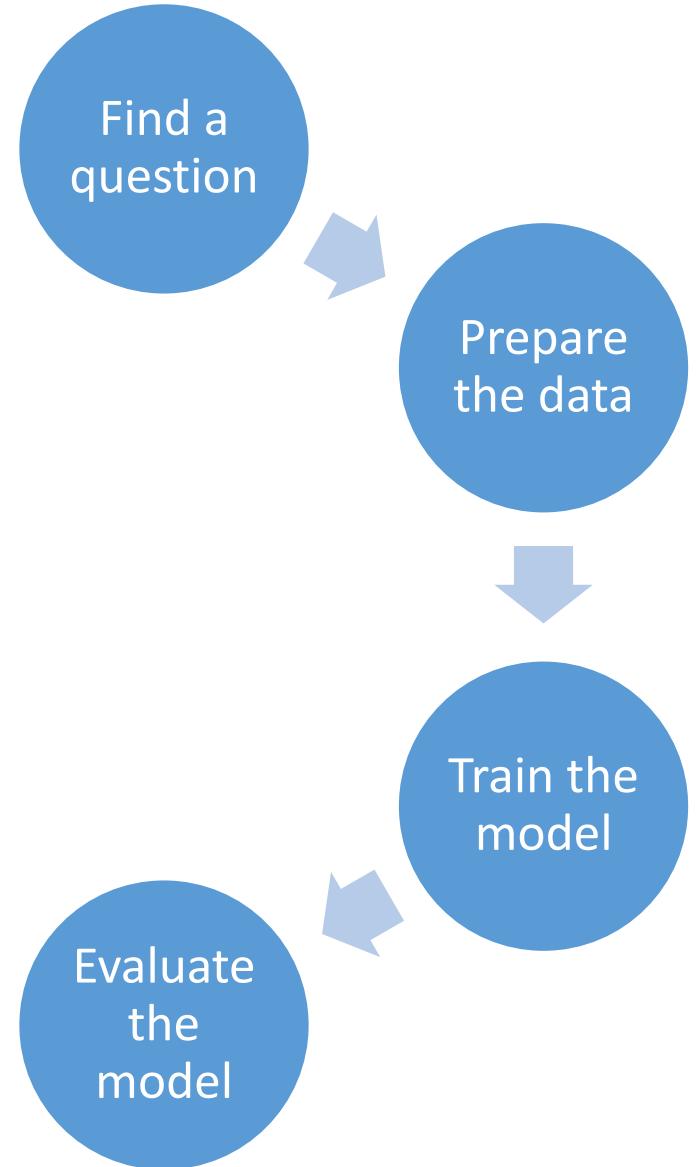
Find a  
question

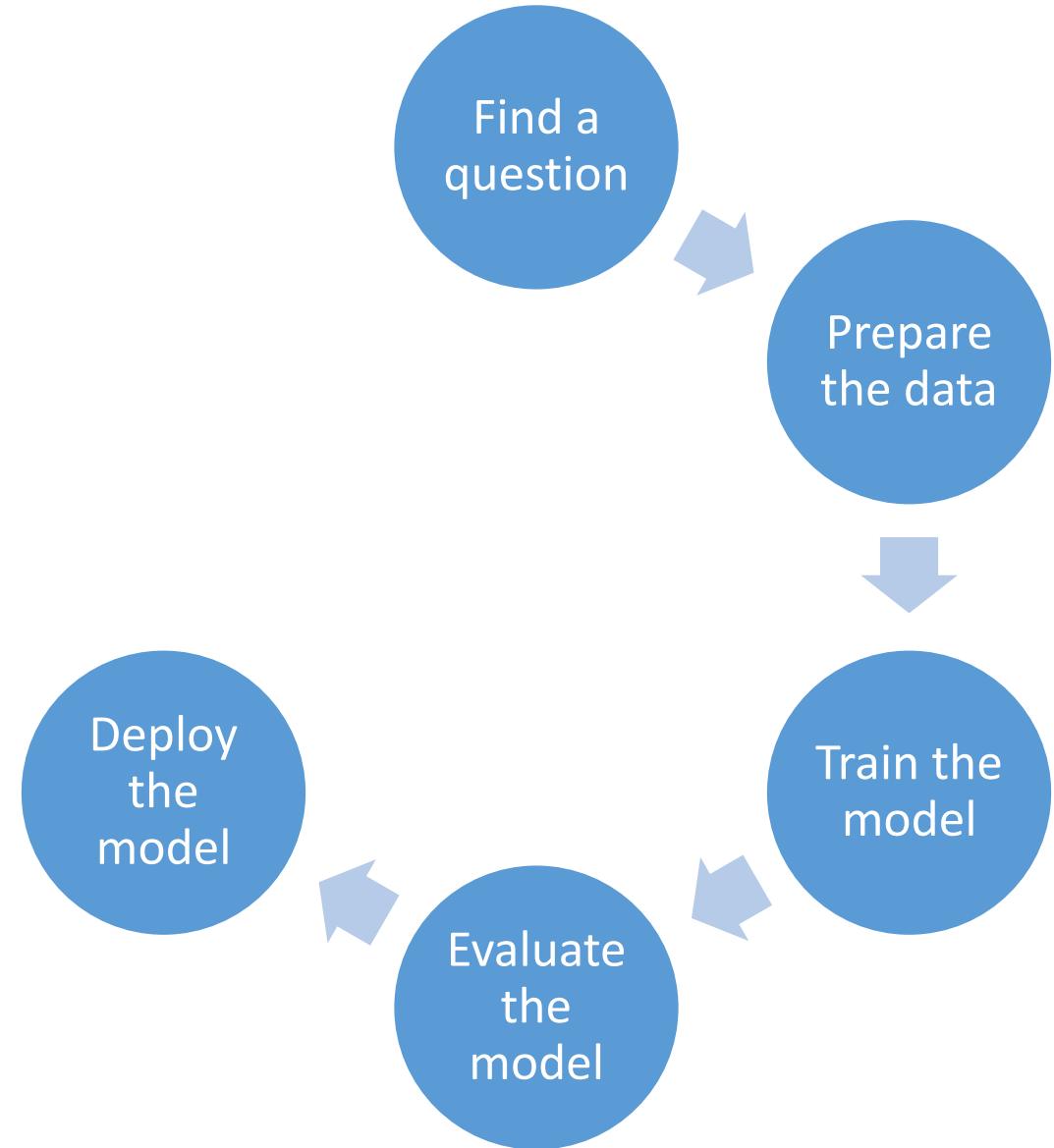
```
graph TD; A((Find a question)) --> B((Prepare the data))
```

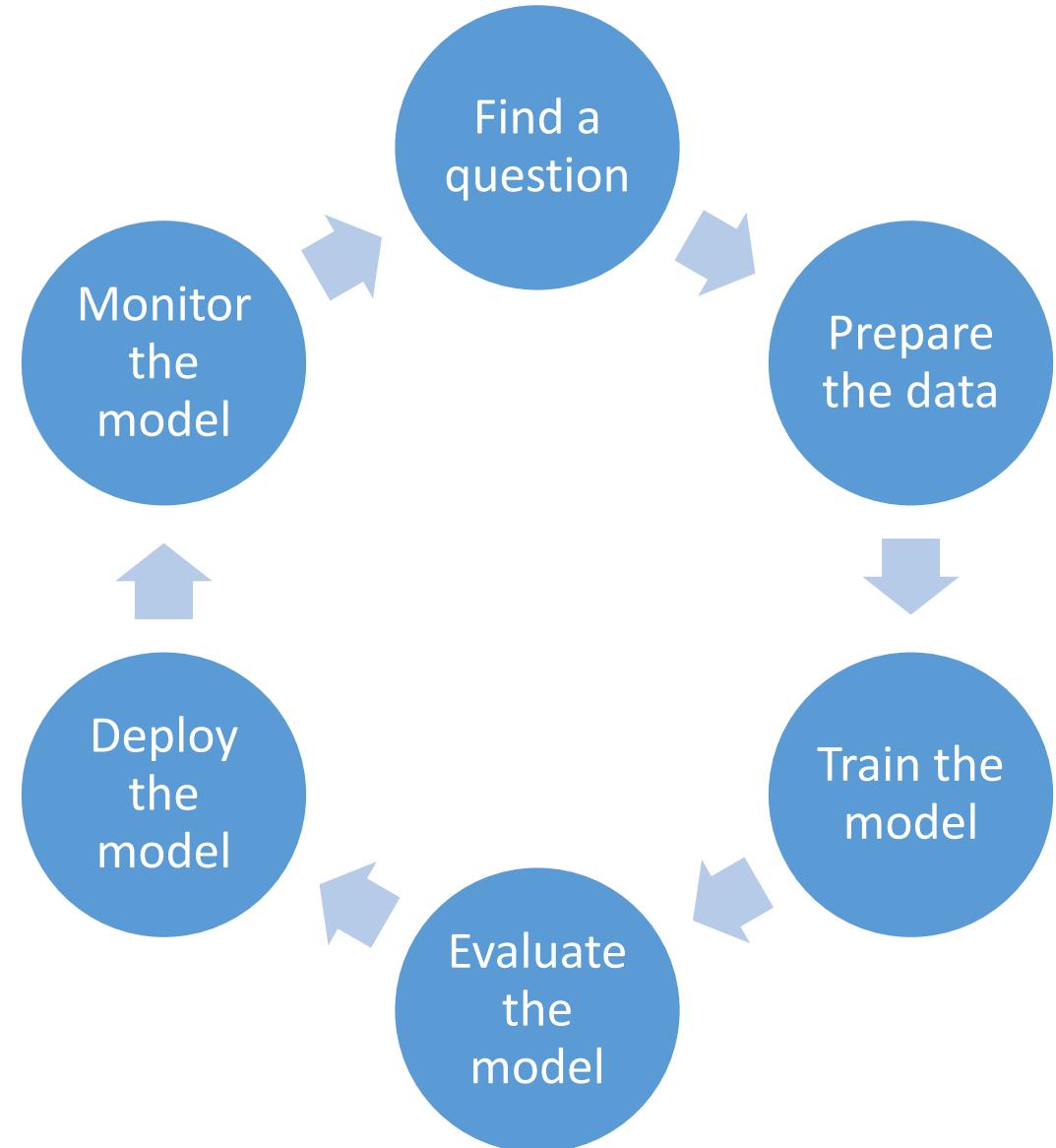
Find a  
question

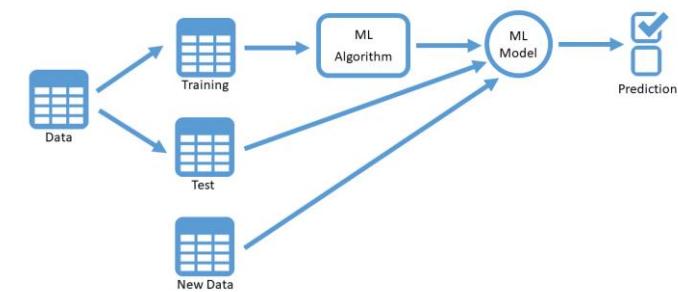
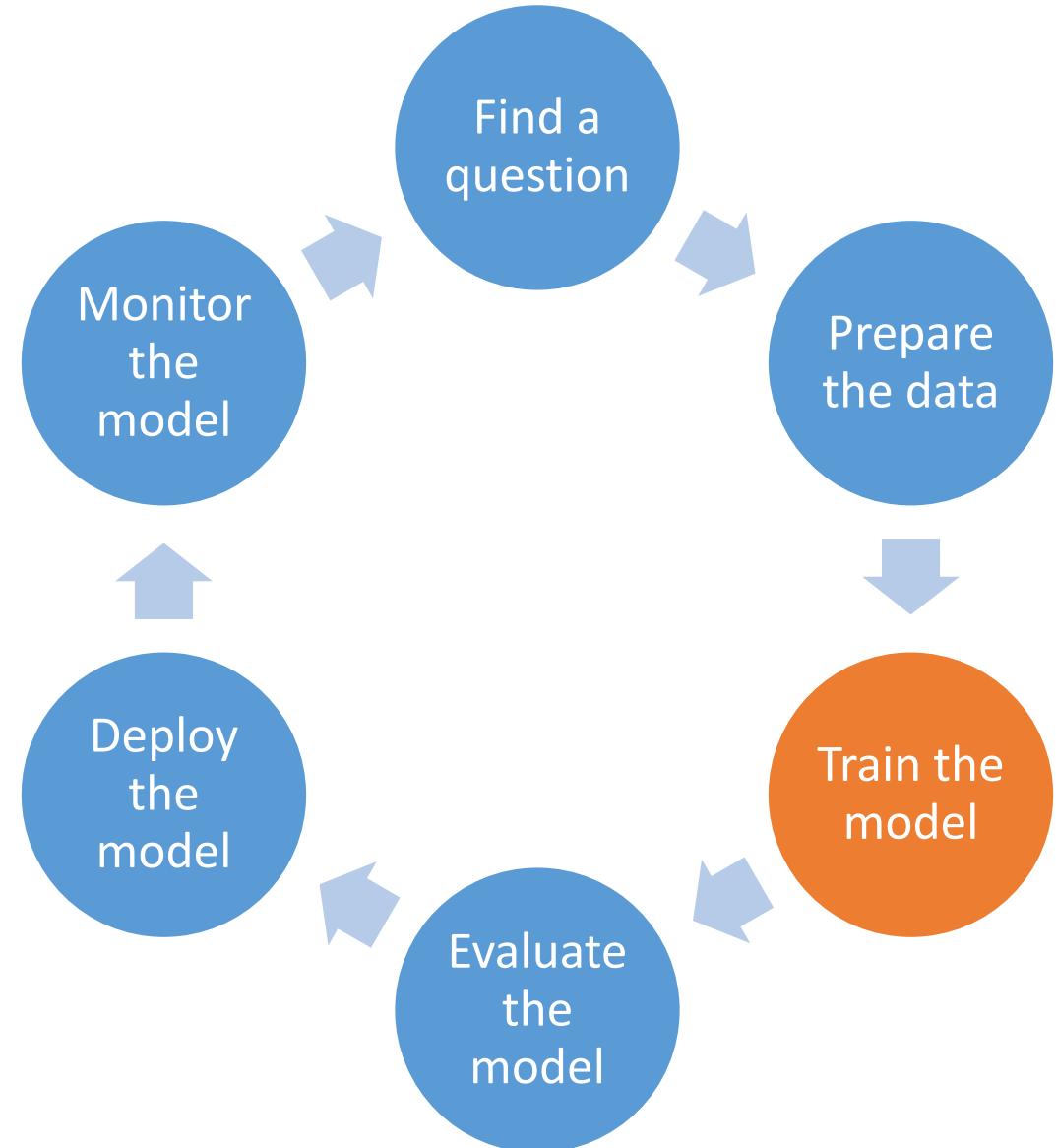
Prepare  
the data

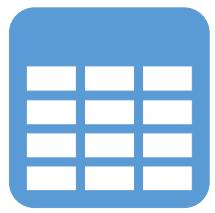




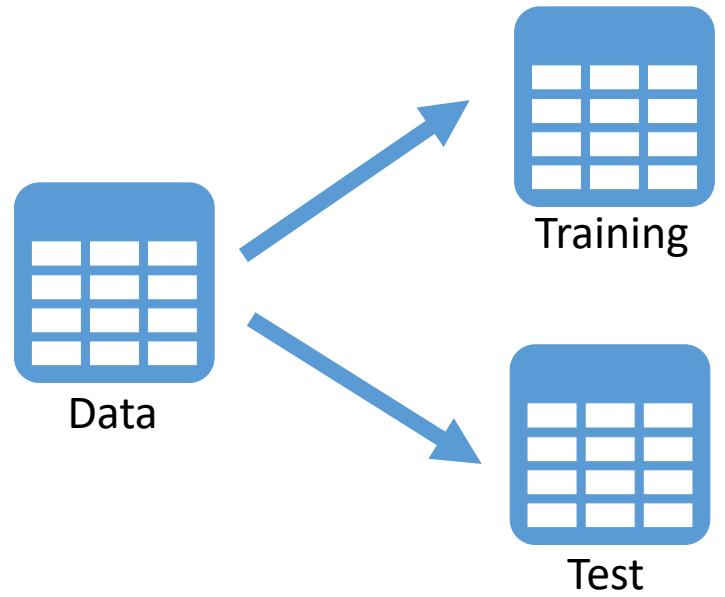


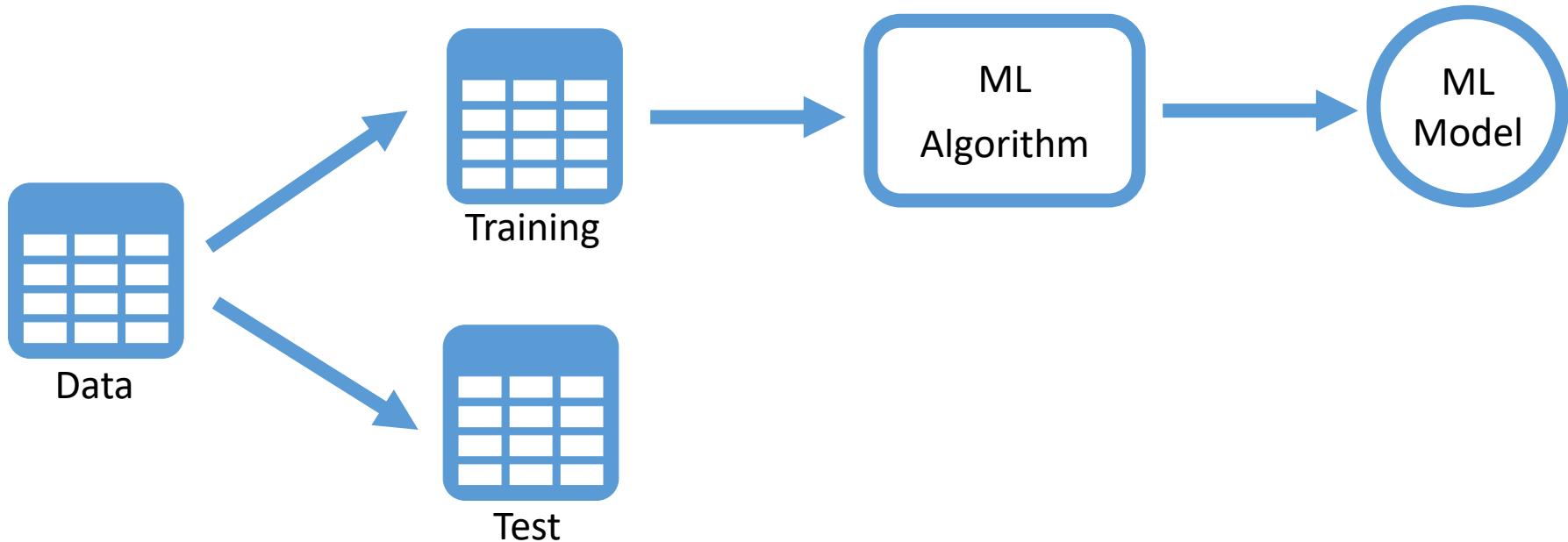


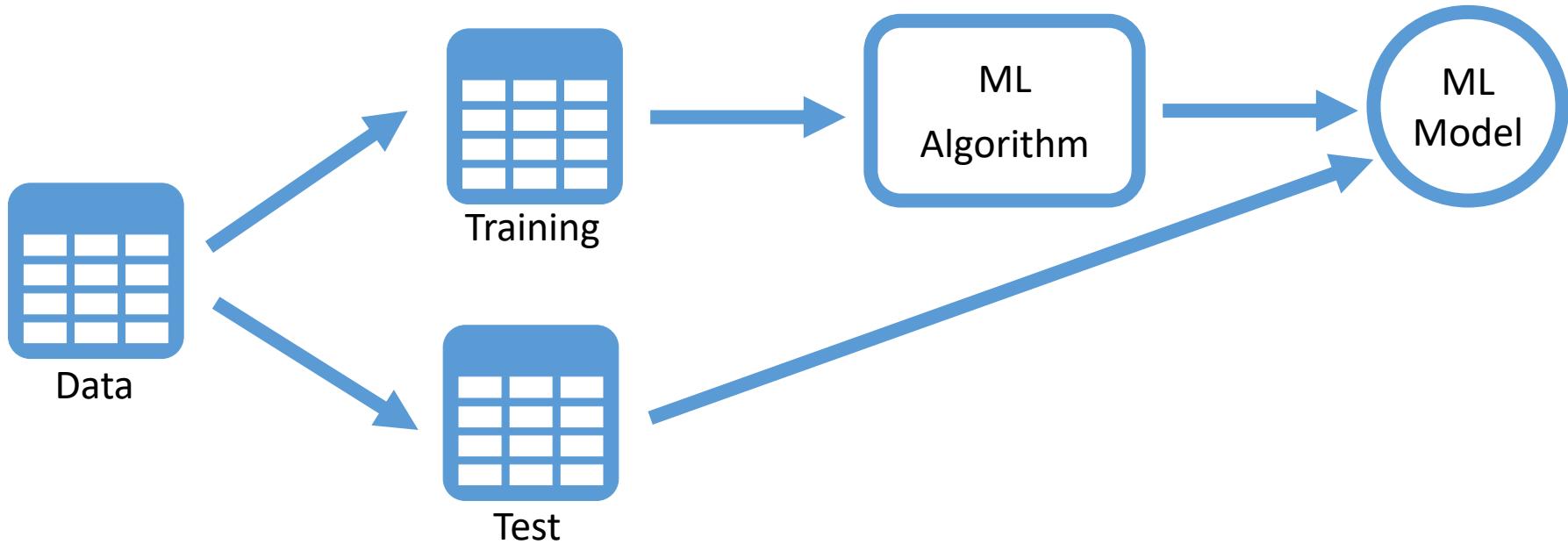


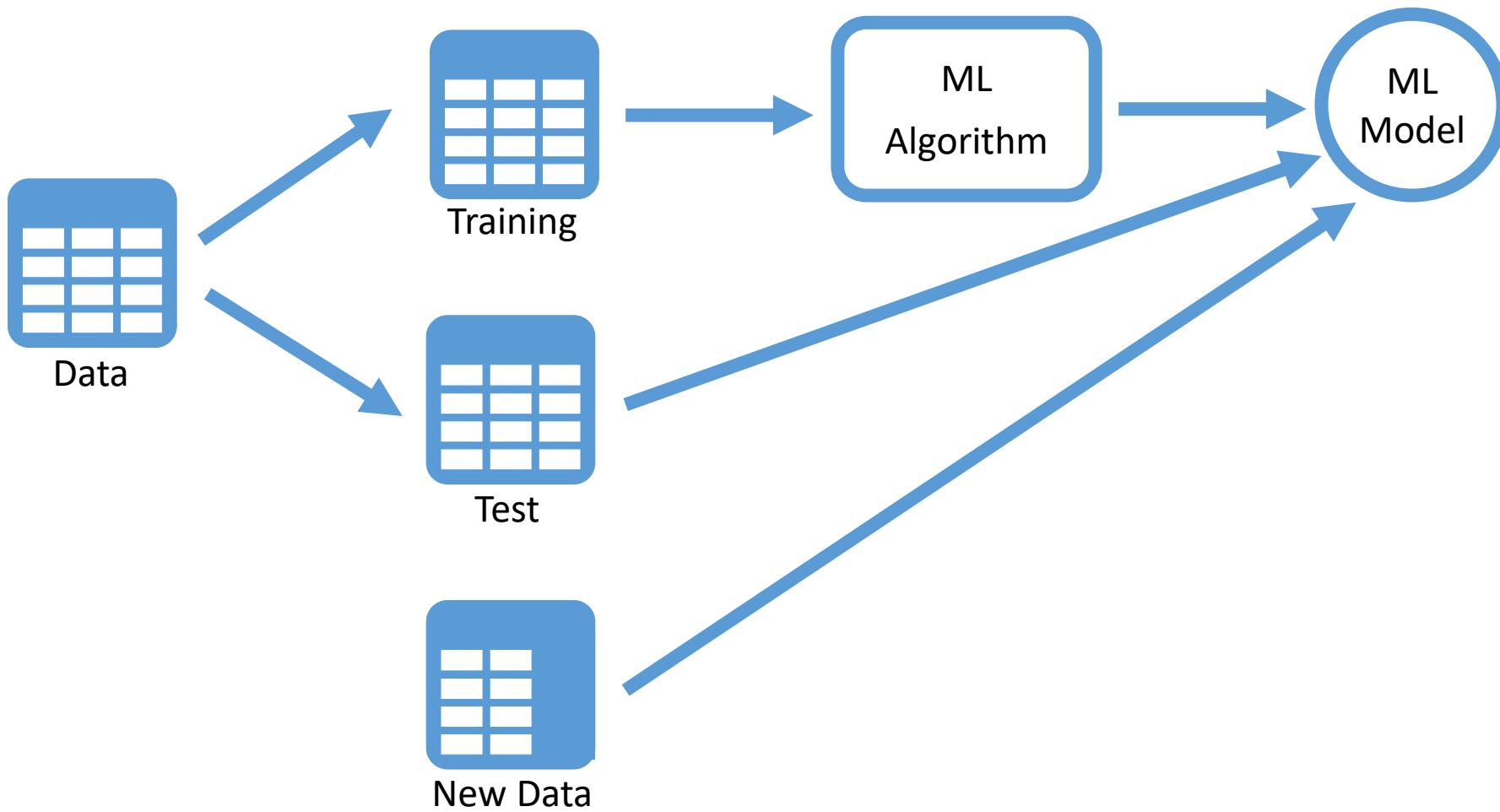


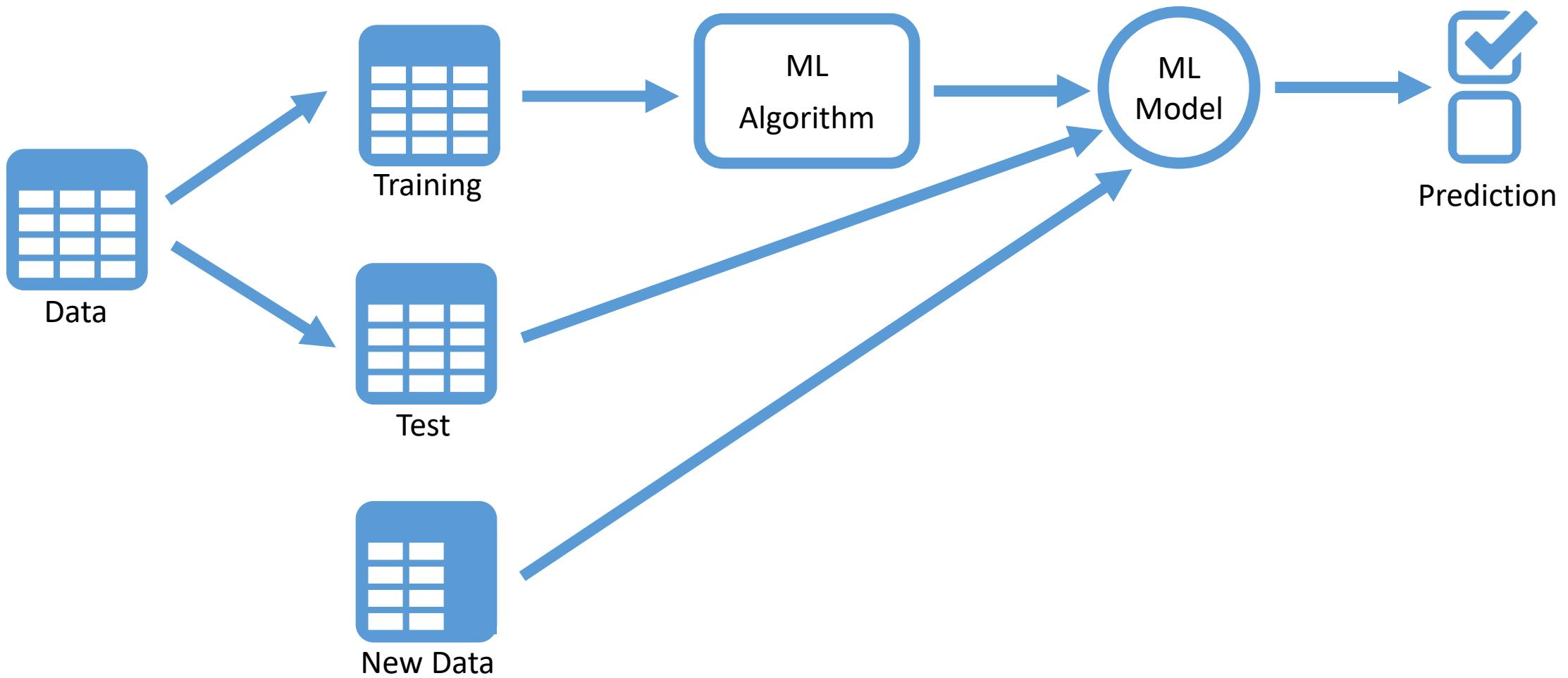
Data



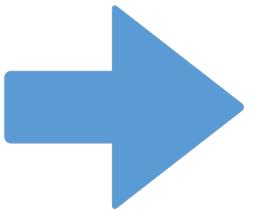
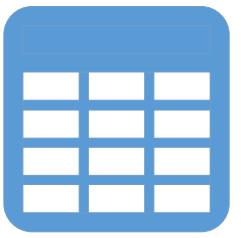
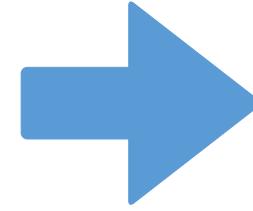






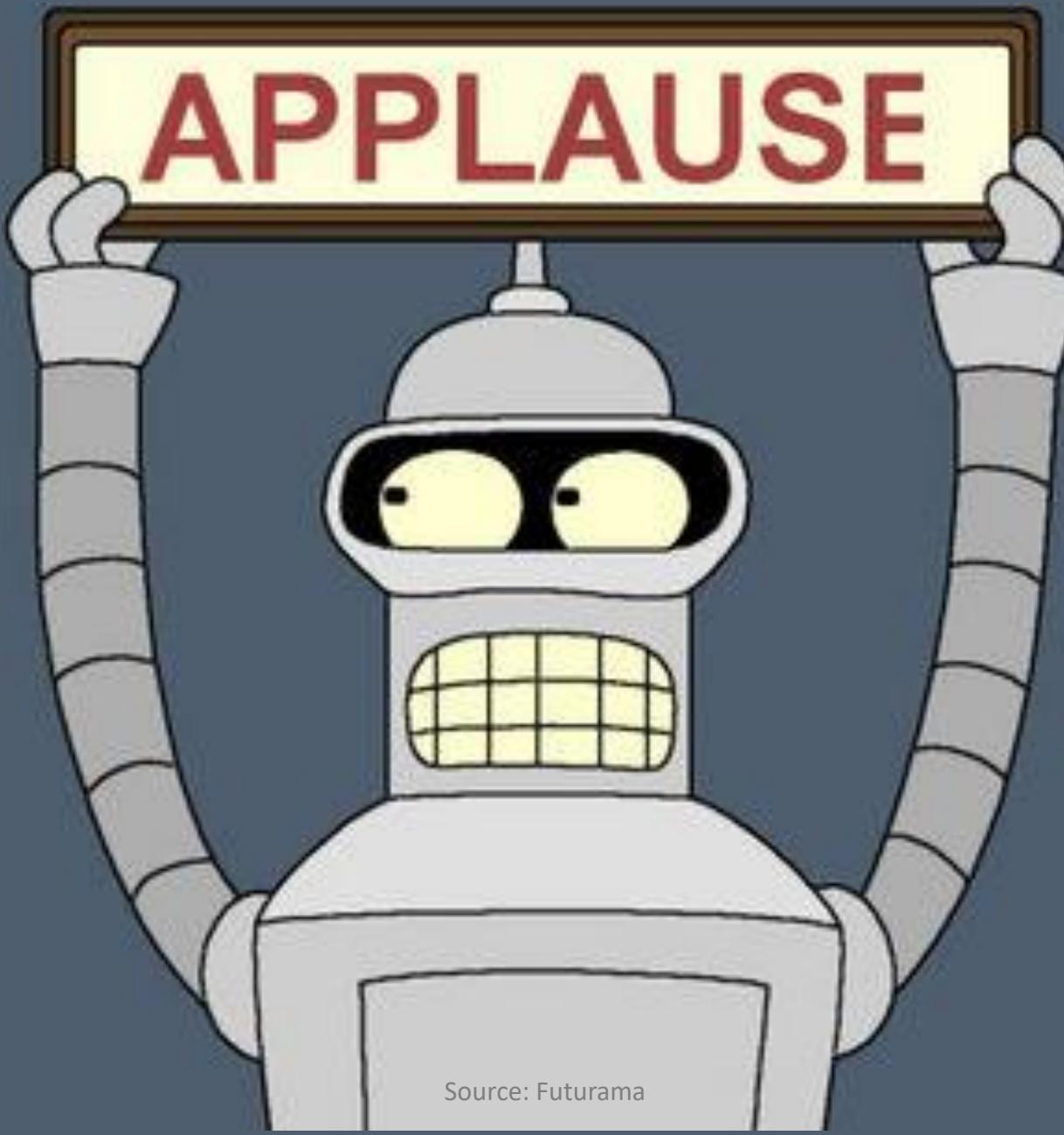


# What Can Machine Learning Do?

 $f(x)$ 

1.23





Source: Futurama

# Introduction to R

# What is R?

Open source

Language and environment

Numerical and graphical analysis

Cross platform



# What is R?

- Active development
- Large user community
- Modular and extensible
- 9000+ extensions



# FREE



A low-angle photograph of the Statue of Liberty against a clear blue sky. She is shown from the chest up, facing slightly left. Her right arm is raised high, holding a torch with a golden flame at the top. Her left arm is bent, holding a tablet or smartphone that displays the word "FREE".

FREE

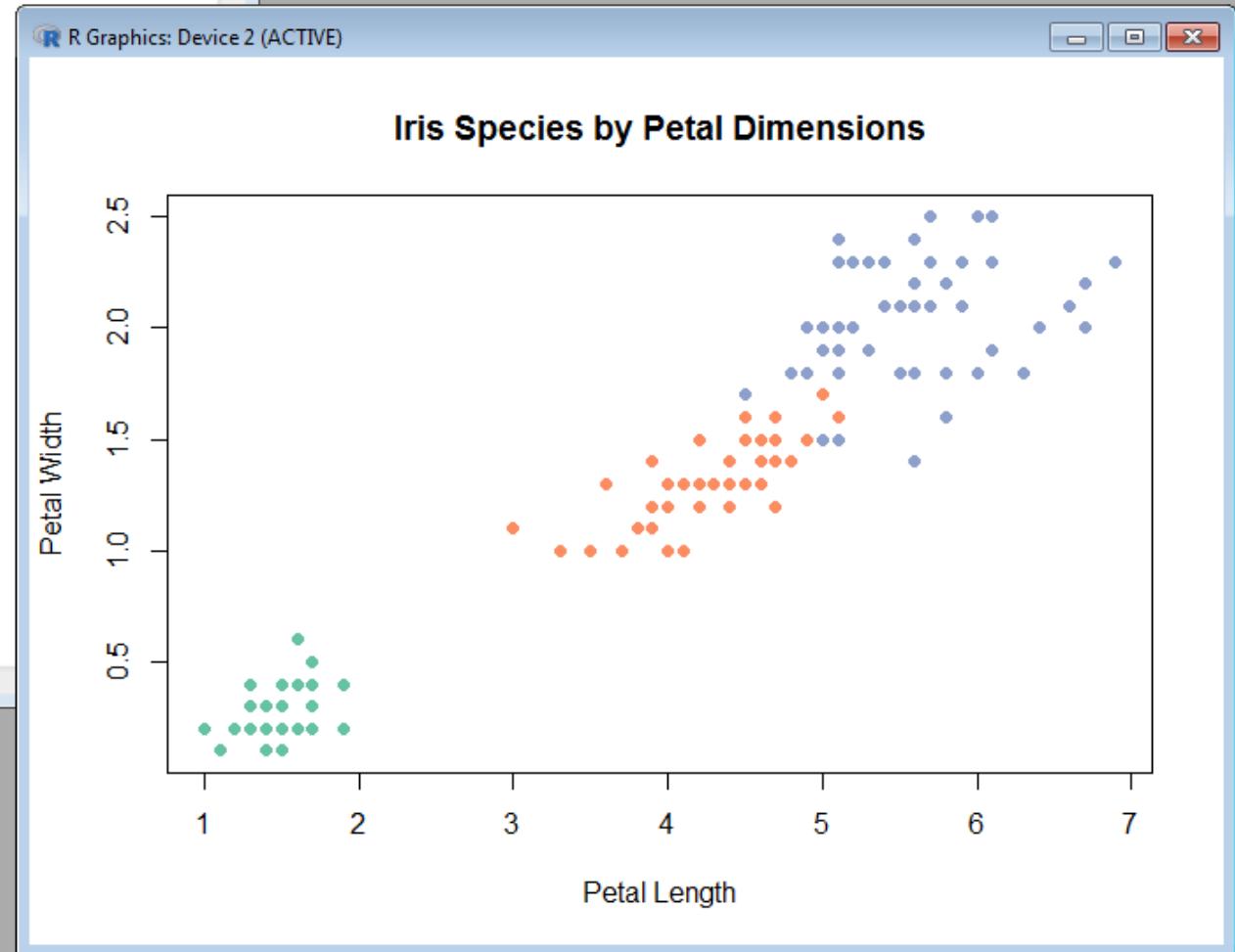


## R Console

```
> # 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
```



RStudio

File Edit Code View Plots Session Build Debug Tools Help

Script.R \* Go to file/function Addins Project: (None)

16 # Create a frequency table of species  
17 table(iris\$Species)  
18  
19 # Get the average petal length  
20 mean(iris\$Petal.Length)  
21  
22 # Get the correlation coefficient  
23 cor(  
24     x = iris\$Petal.Length,  
25     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 Global Environment Data Values palette 150 obs. of 5 variables chr [1:3] "#66C2A5" "#FC8D62" "#8DA0C... Files Plots Packages Help Viewer Publish Iris Species by Petal Dimensions Petal Width Petal Length

The figure is a scatter plot titled "Iris Species by Petal Dimensions". The vertical axis is labeled "Petal Width" and ranges from 0.5 to 2.5. The horizontal axis is labeled "Petal Length" and ranges from 1 to 7. There are three distinct clusters of data points representing different iris species: setosa (green dots), versicolor (orange dots), and virginica (blue dots). The setosa species has the lowest petal lengths and widths, ranging approximately from 1.0 to 2.0. The versicolor species has intermediate petal lengths and widths, ranging approximately from 3.0 to 5.5. The virginica species has the highest petal lengths and widths, ranging approximately from 5.0 to 7.0.

Script.R - Microsoft Visual Studio

File Edit View NCrunch Project Debug Team Tools Architecture Test ReSharper R Tools Analyze Window Help

Matthew Renze

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)
```

R Interactive

```
> # 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
>
```

Variable Explorer

Name	Value	Class	Type
iris	150 obs. of 5 variables	data.frame	list
palette	chr [1:3] "#6C2A5" "#FC8D62" "#8DA0CF	character	character

R Plot

### Iris Species by Petal Dimensions

A scatter plot titled "Iris Species by Petal Dimensions". The x-axis is labeled "Petal Length" and ranges from 1 to 7. The y-axis is labeled "Petal Width" and ranges from 0.5 to 2.5. The plot shows three distinct clusters of data points corresponding to the Iris species: Setosa (green), Versicolor (orange), and Virginica (blue). The data points are scattered across the plot area, with a general trend where Petal Length increases as Petal Width increases.

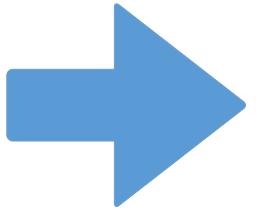
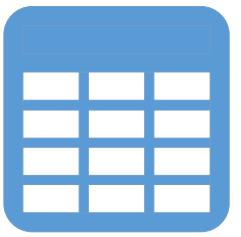
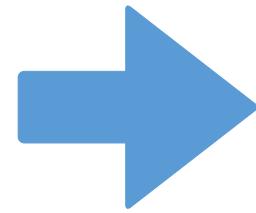
Solution Explorer R Plot R Package Manager R Help

Error List Output Azure App Service Activity

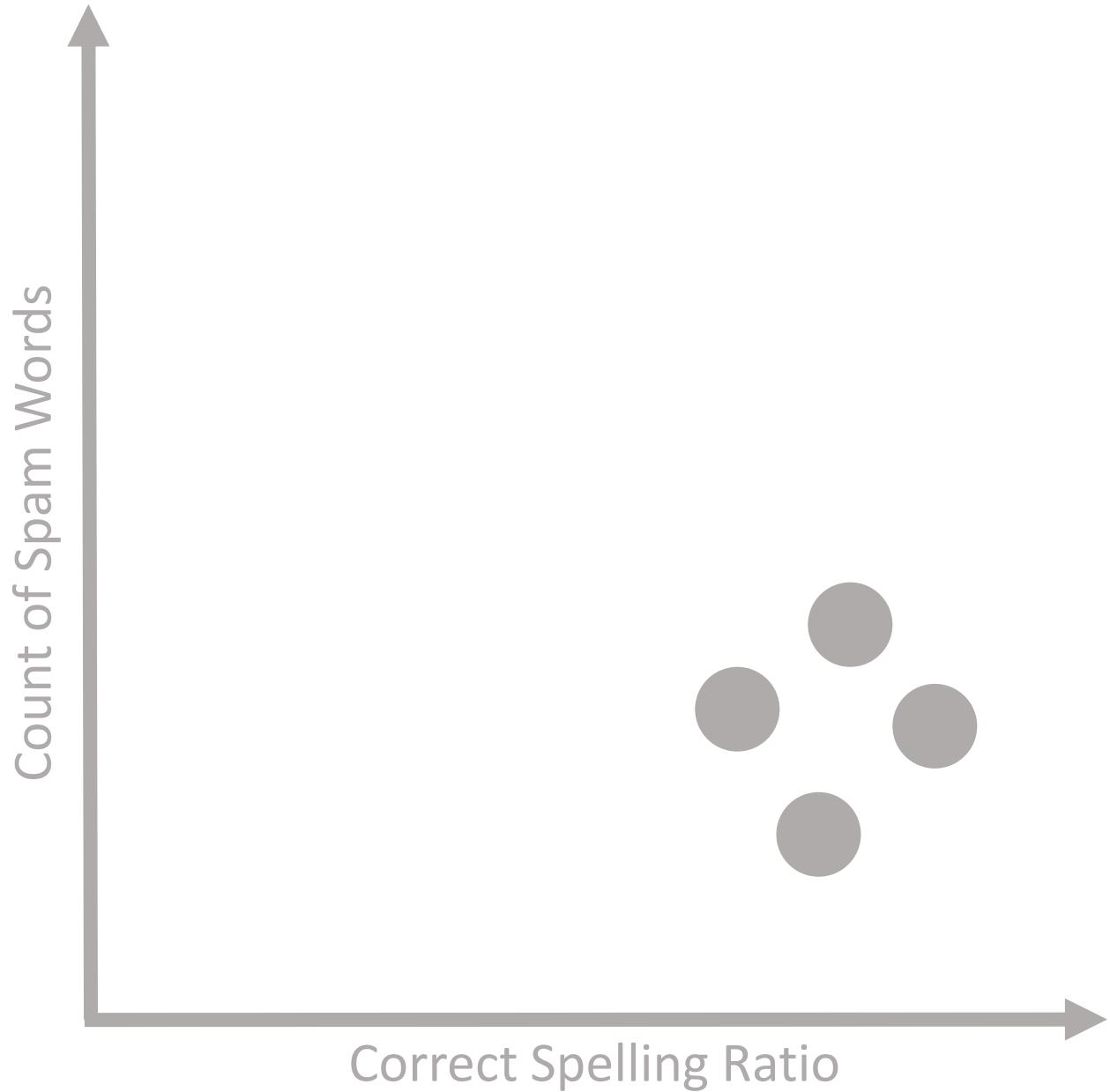
Ready Ln 30 Col1 Ch1 INS ↑ 7 ⌂ 0 ⌂ Root ⌂ master ⌂

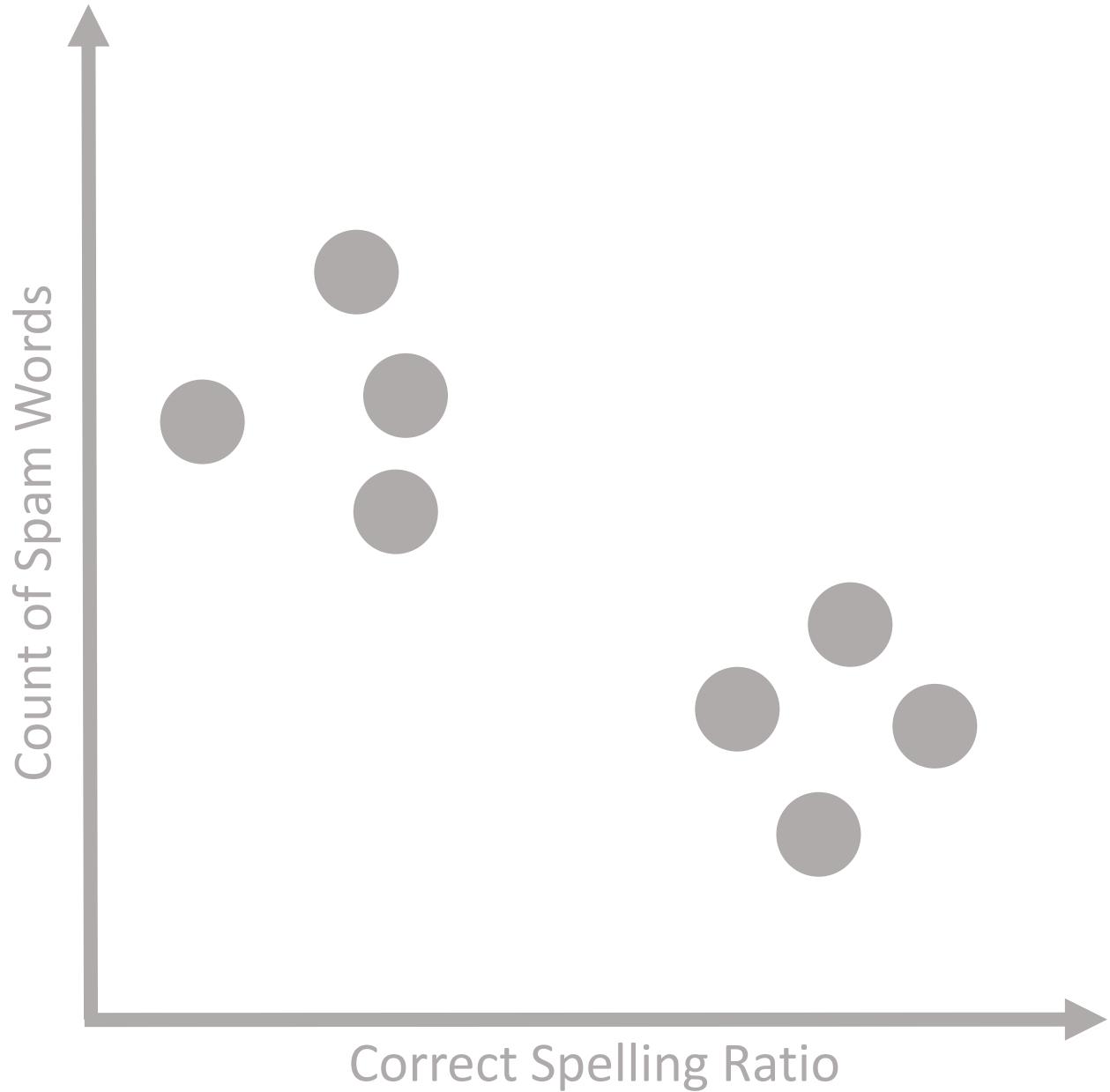
# Code Demo

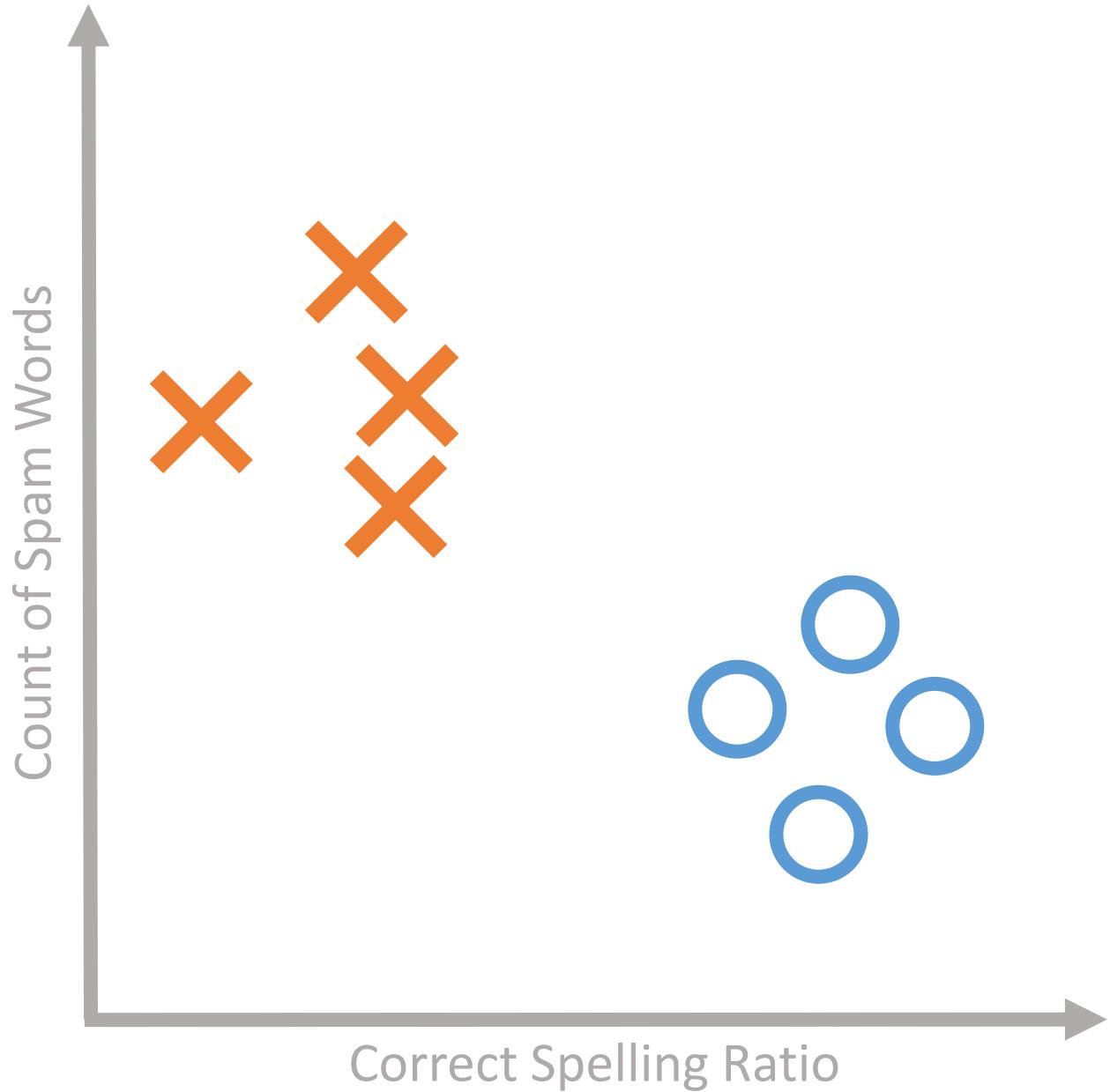
# Classification

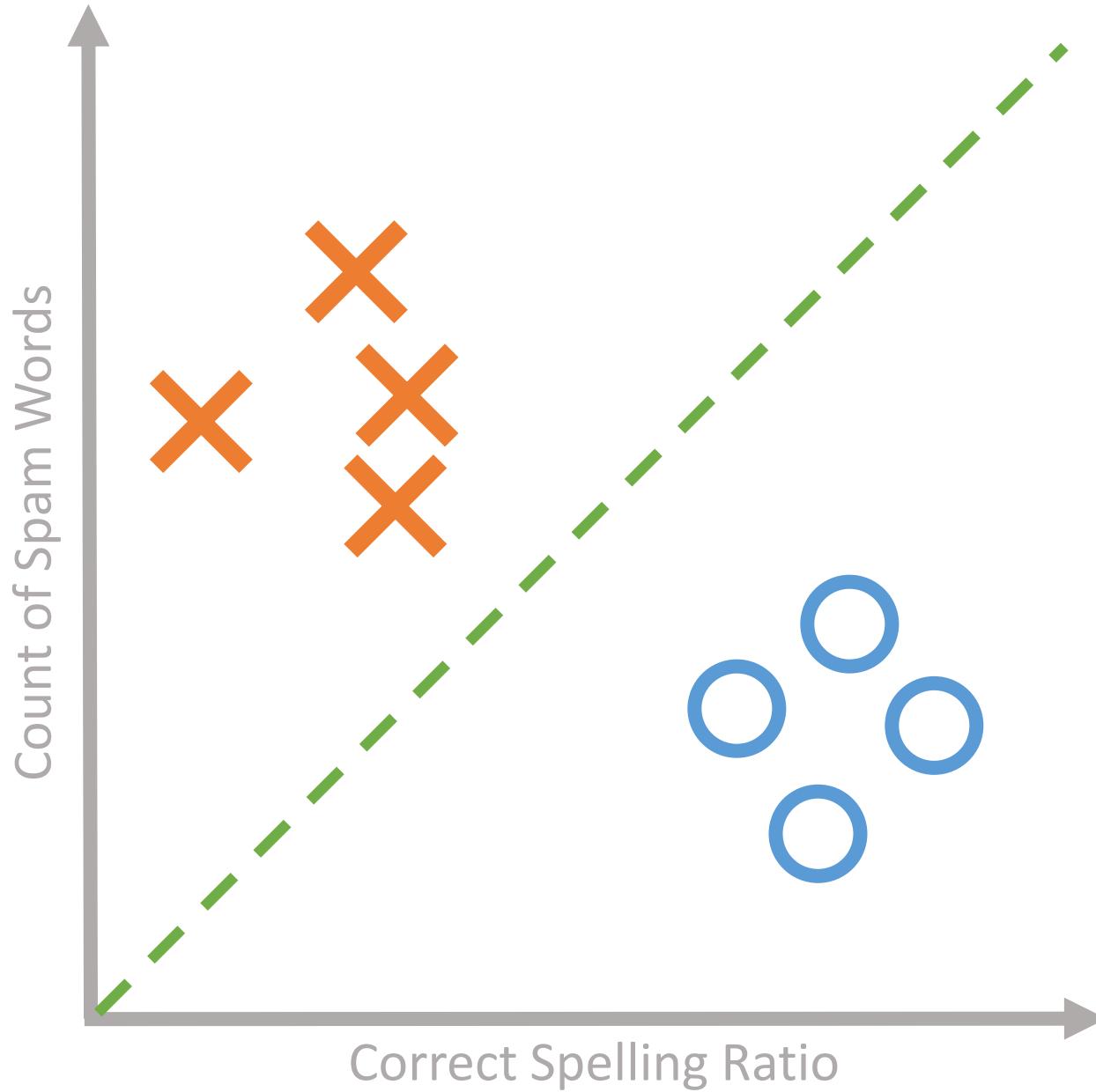
 $f(x)$ 

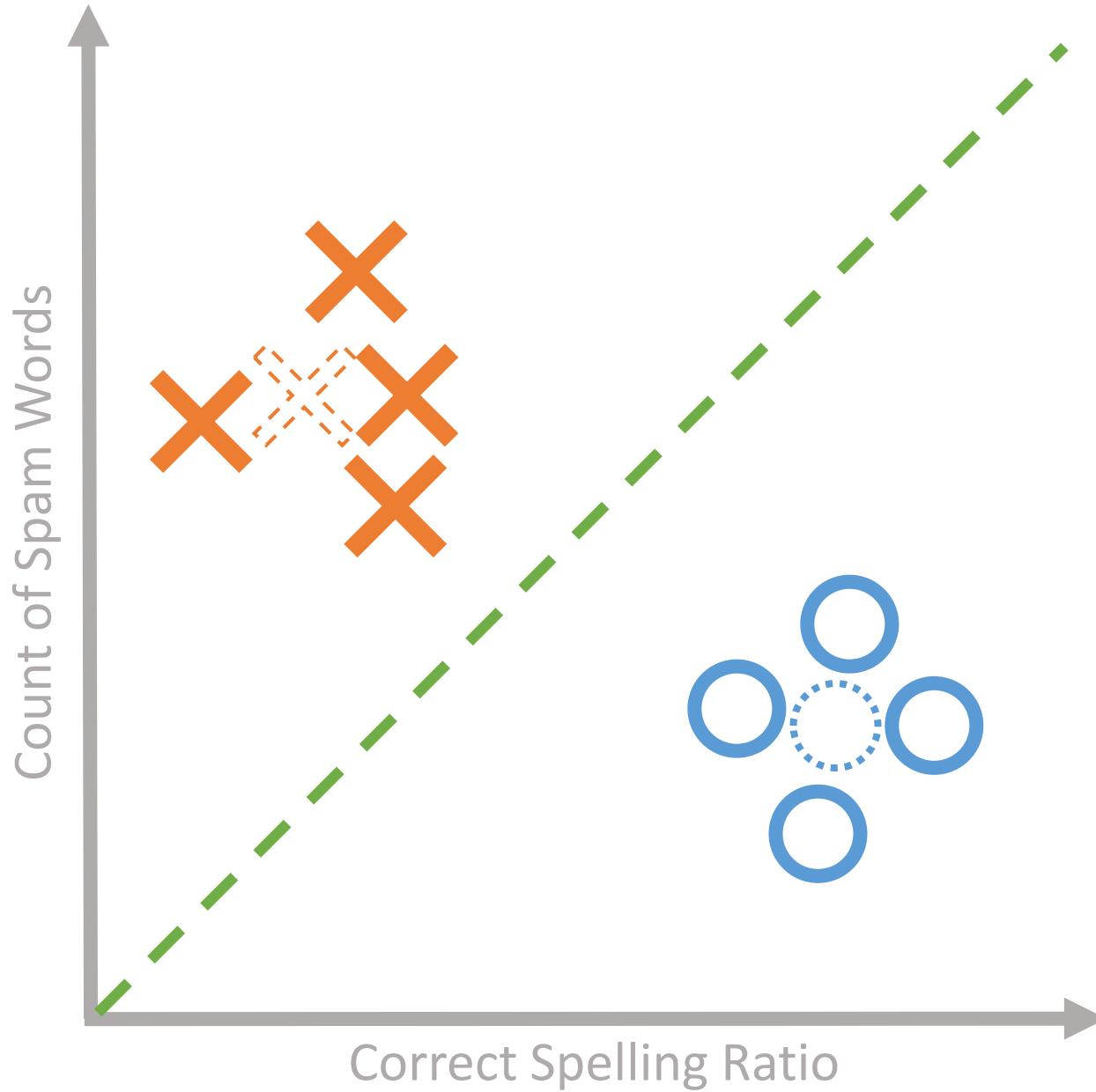






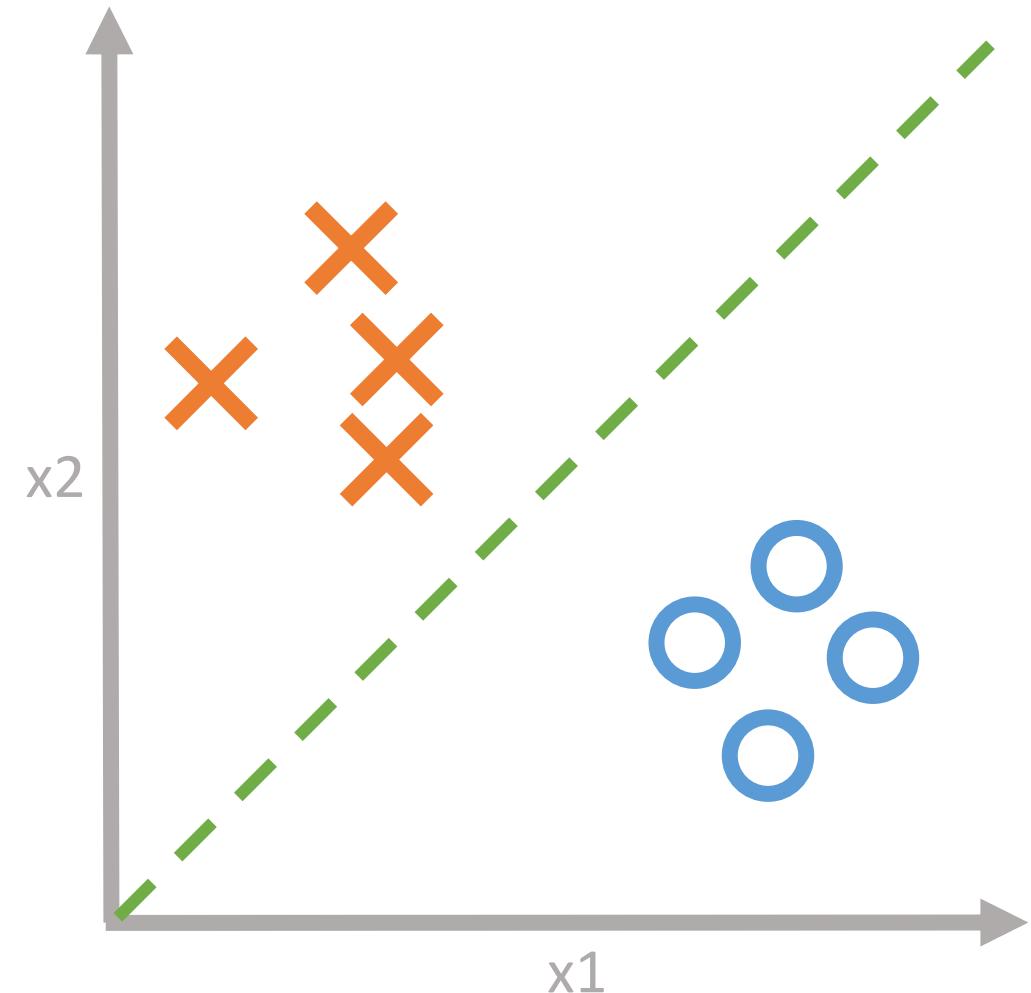






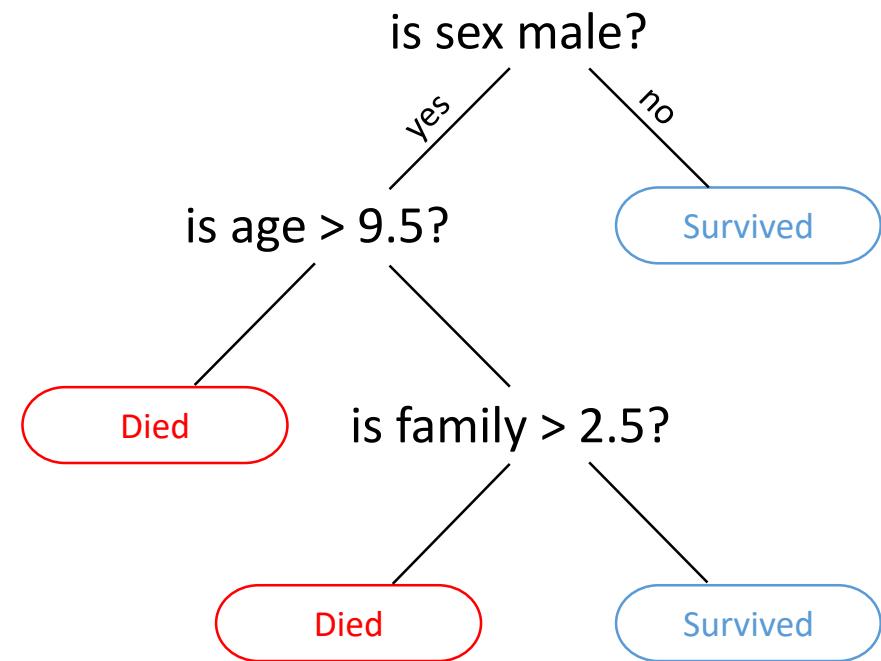
# Classification Algorithms

Decision Tree Classifier  
Naïve Bayes Classifier  
Support Vector Machine  
Neural Network



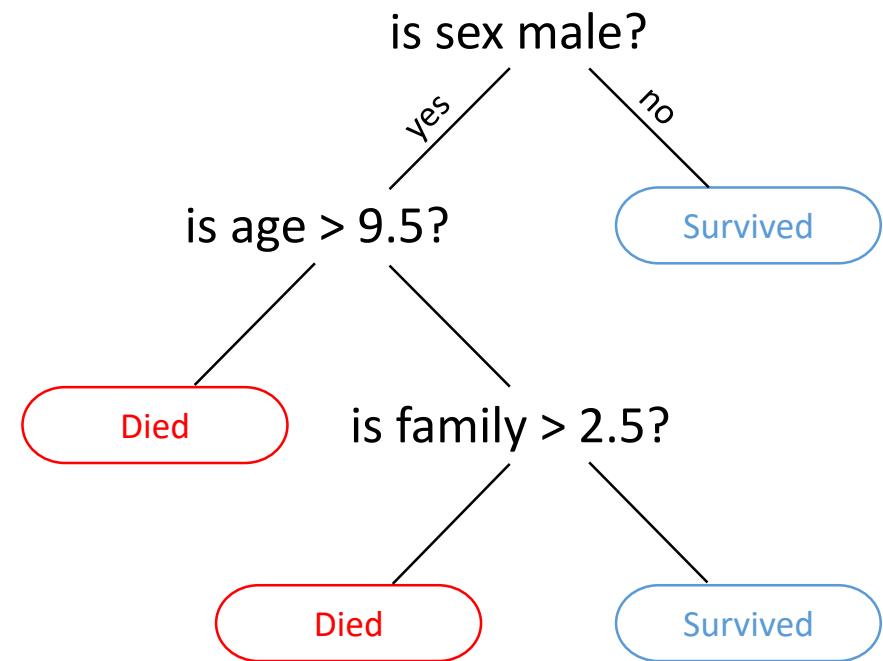
# Decision Tree Classifier

Supervised learning



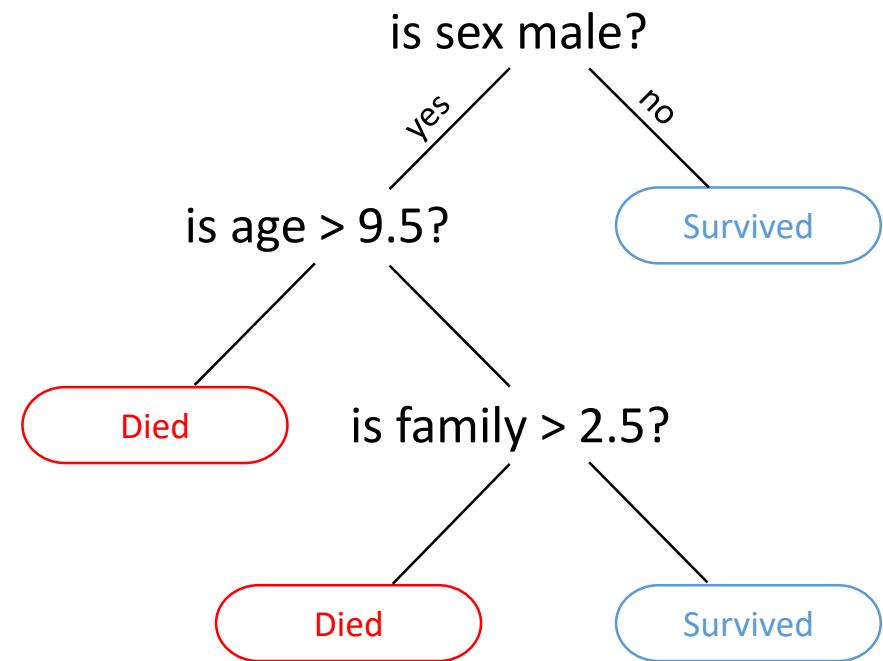
# Decision Tree Classifier

Supervised learning  
Tree of decisions



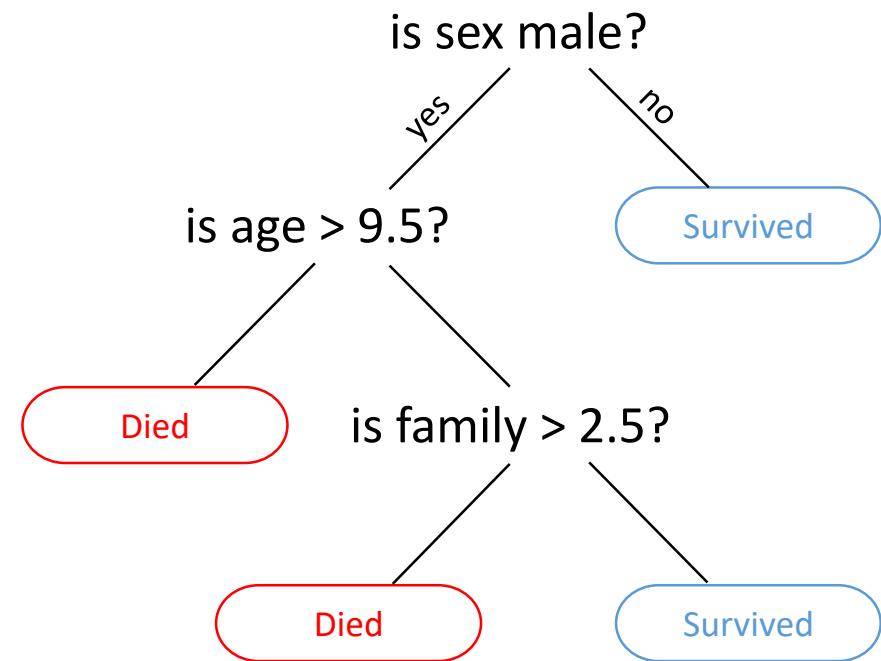
# Decision Tree Classifier

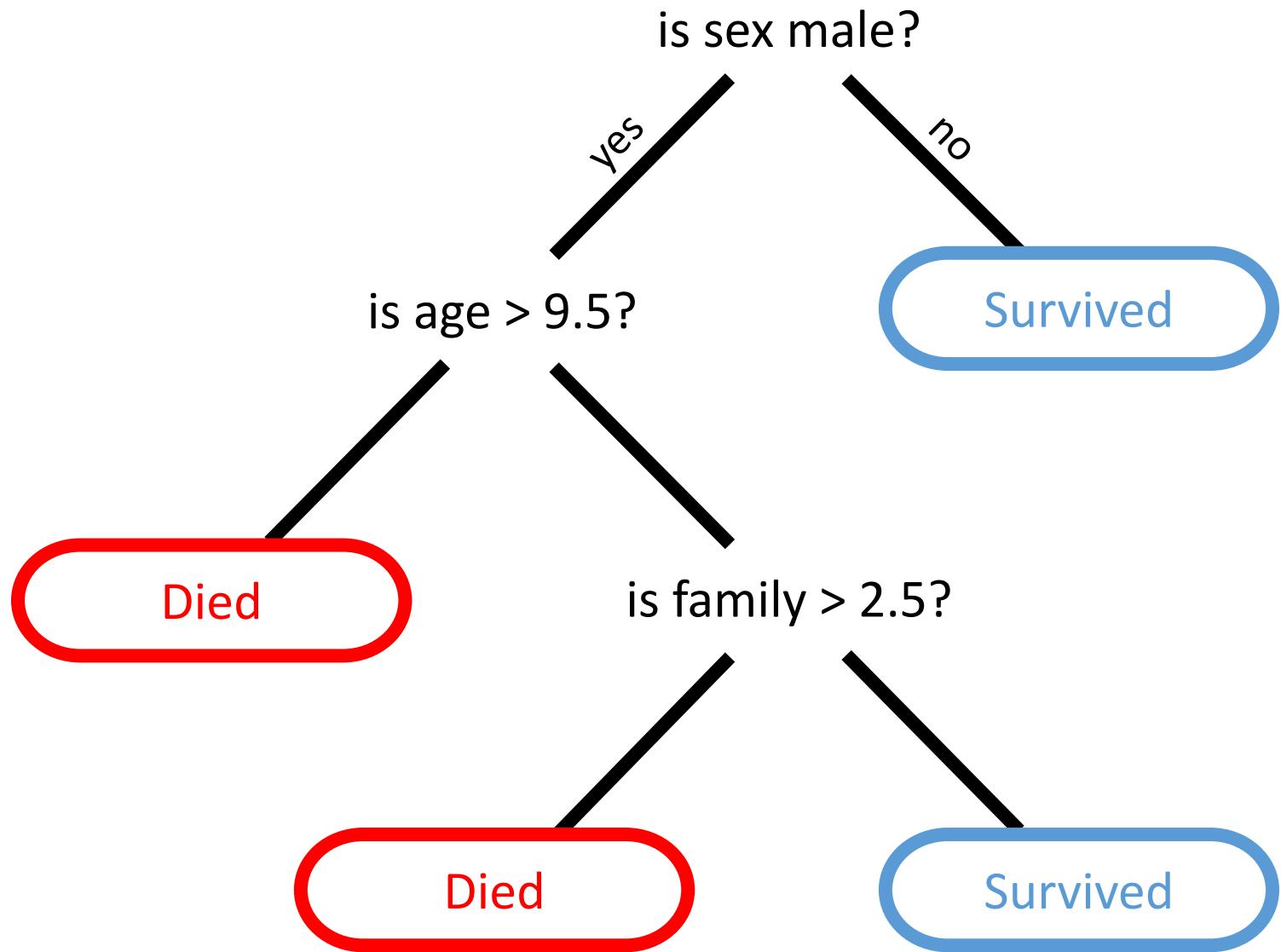
Supervised learning  
Tree of decisions  
Easy to understand



# Decision Tree Classifier

Supervised learning  
Tree of decisions  
Easy to understand  
Transparent





# Iris Data Set



Iris Setosa



Iris Versicolor



Iris Virginica

# 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
...		...	...	...

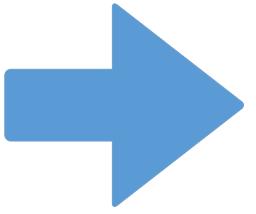
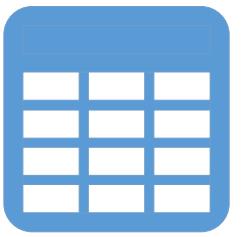
# Classification Demo

Goal: Predict species based on  
petal and sepal measurements

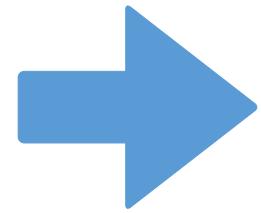
# Real-World Examples

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

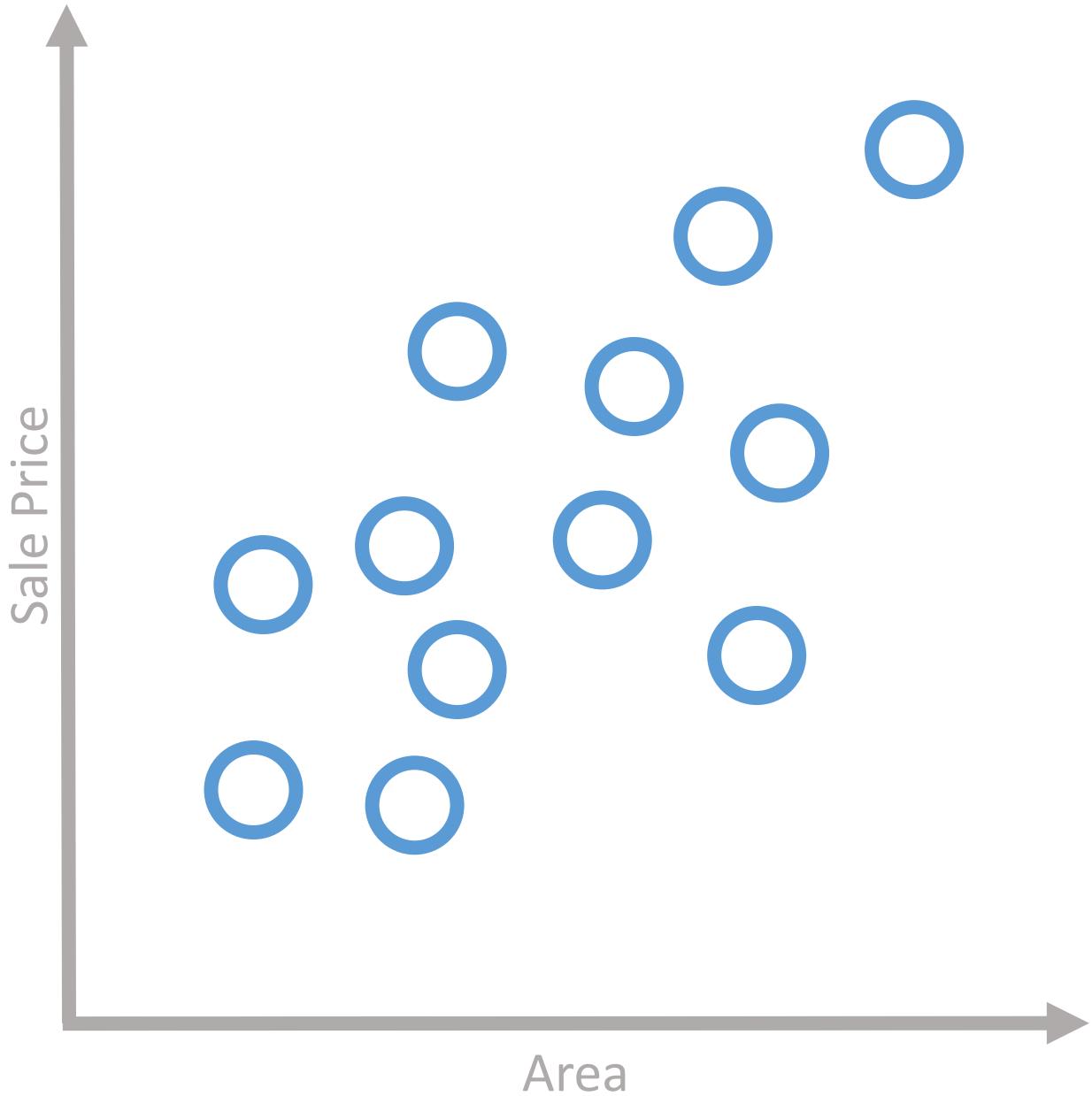
# Regression

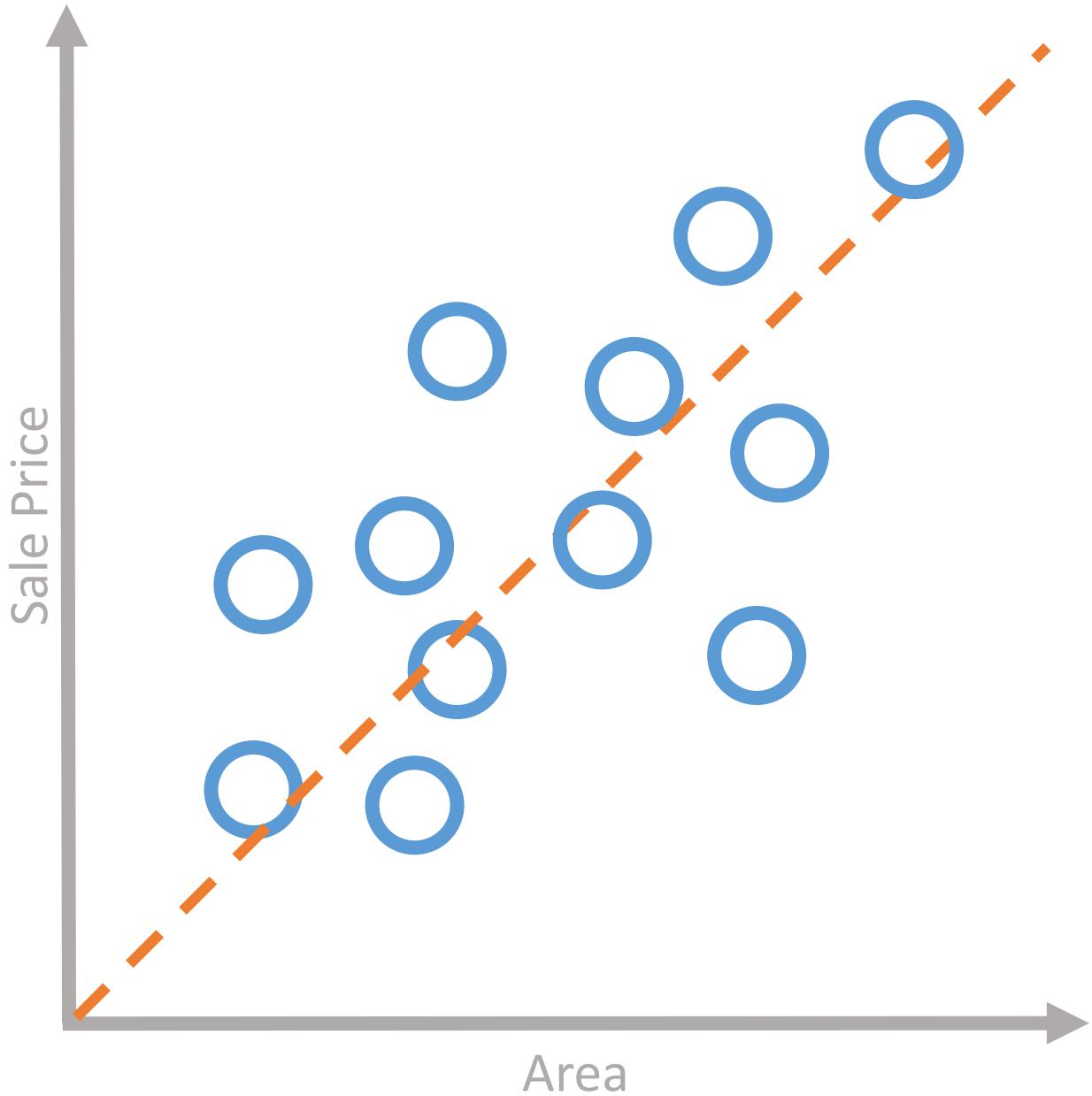


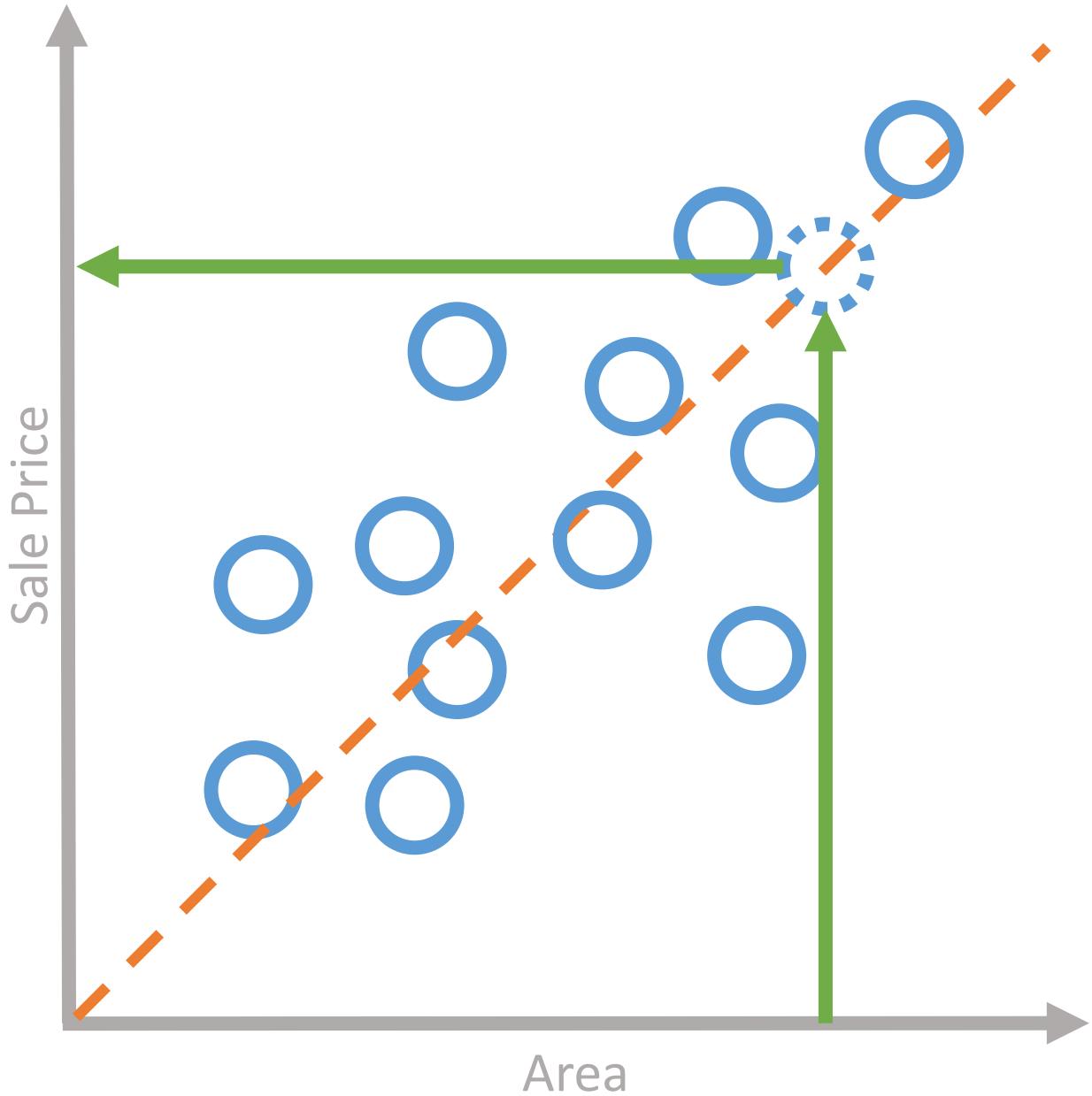
$f(x)$



1.23







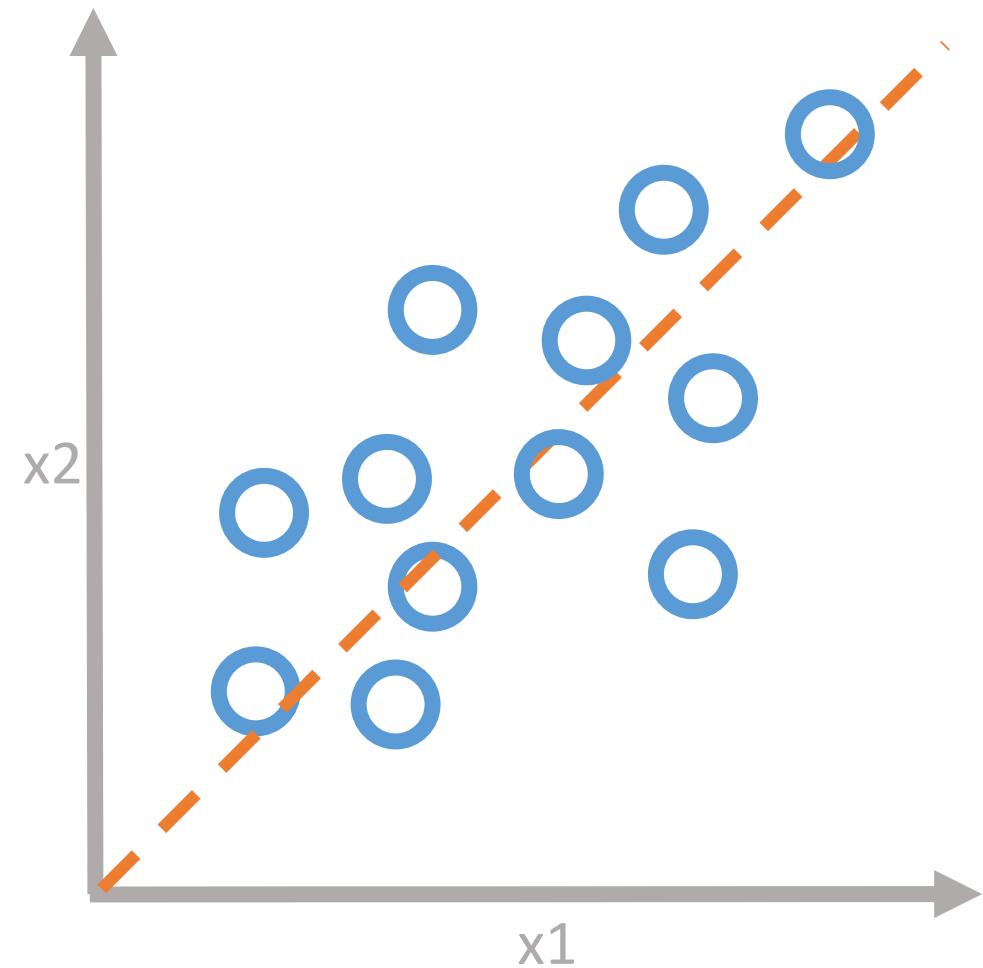
# Regression Algorithms

Linear Regression

Polynomial Regression

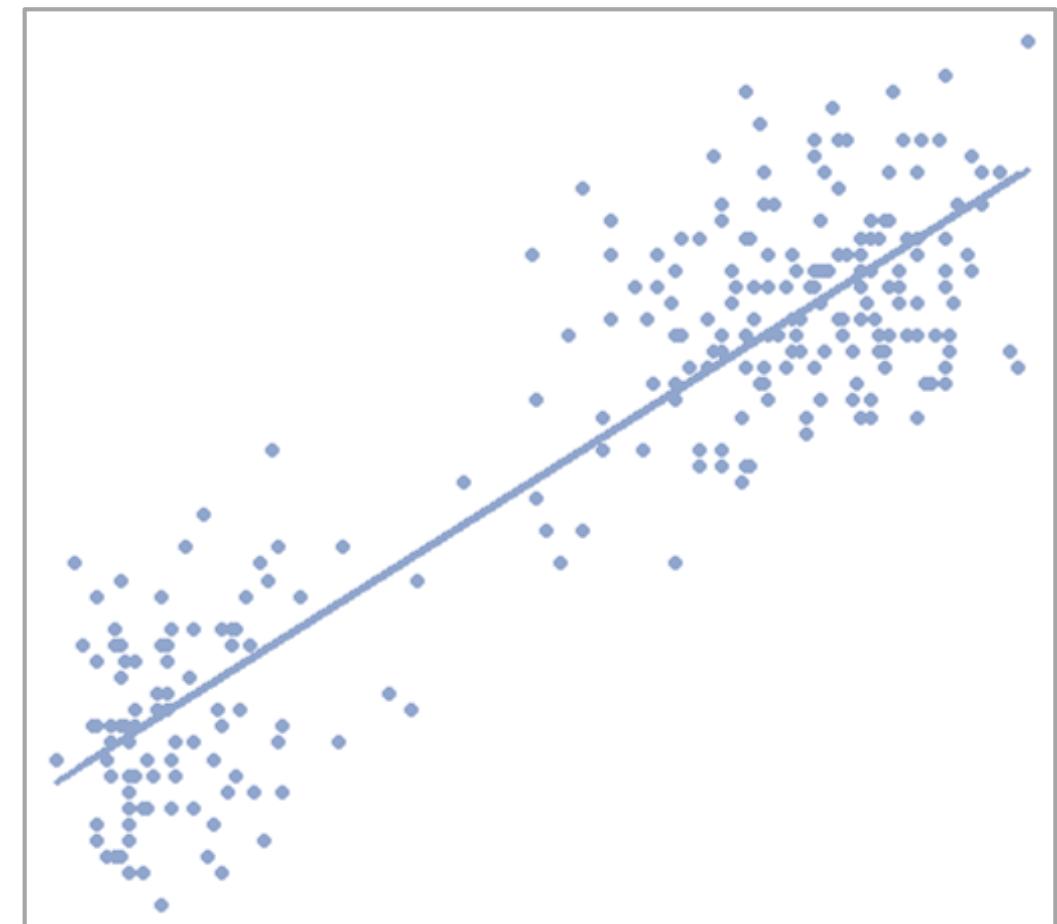
Lasso Regression

ElasticNet Regression



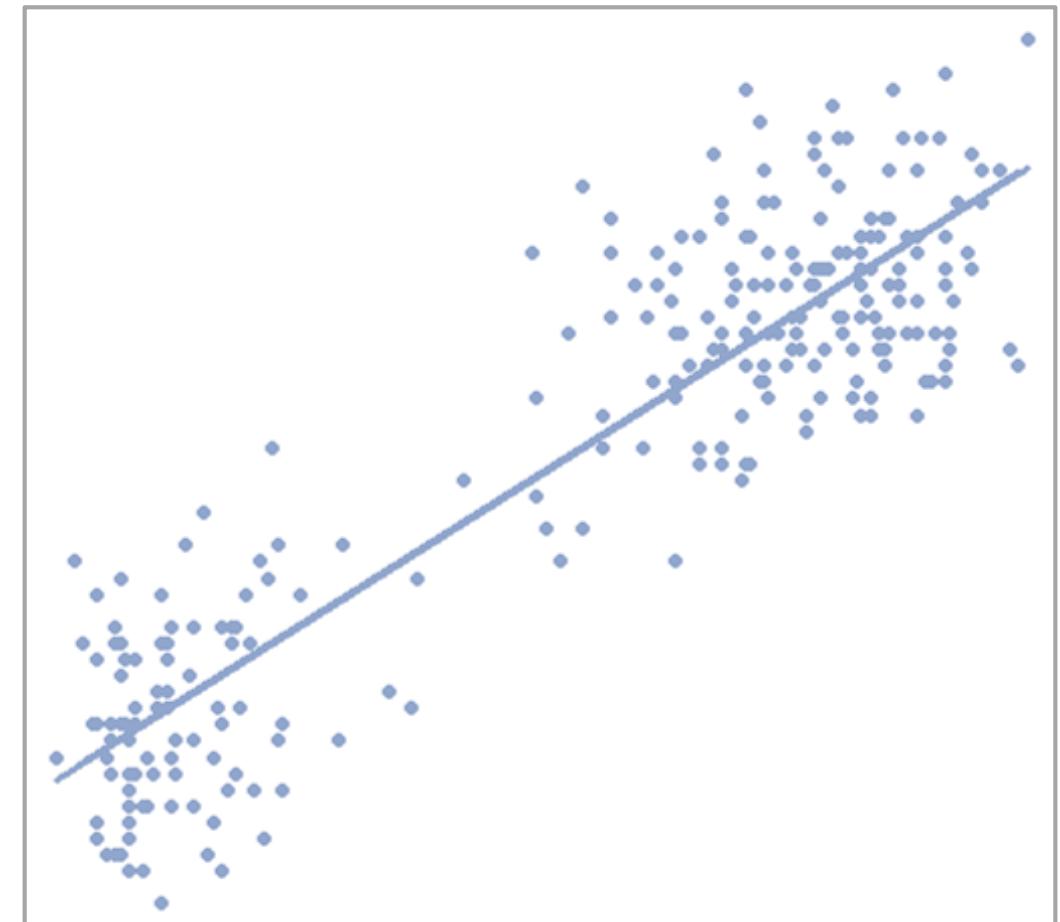
# Simple Linear Regression

Relationship



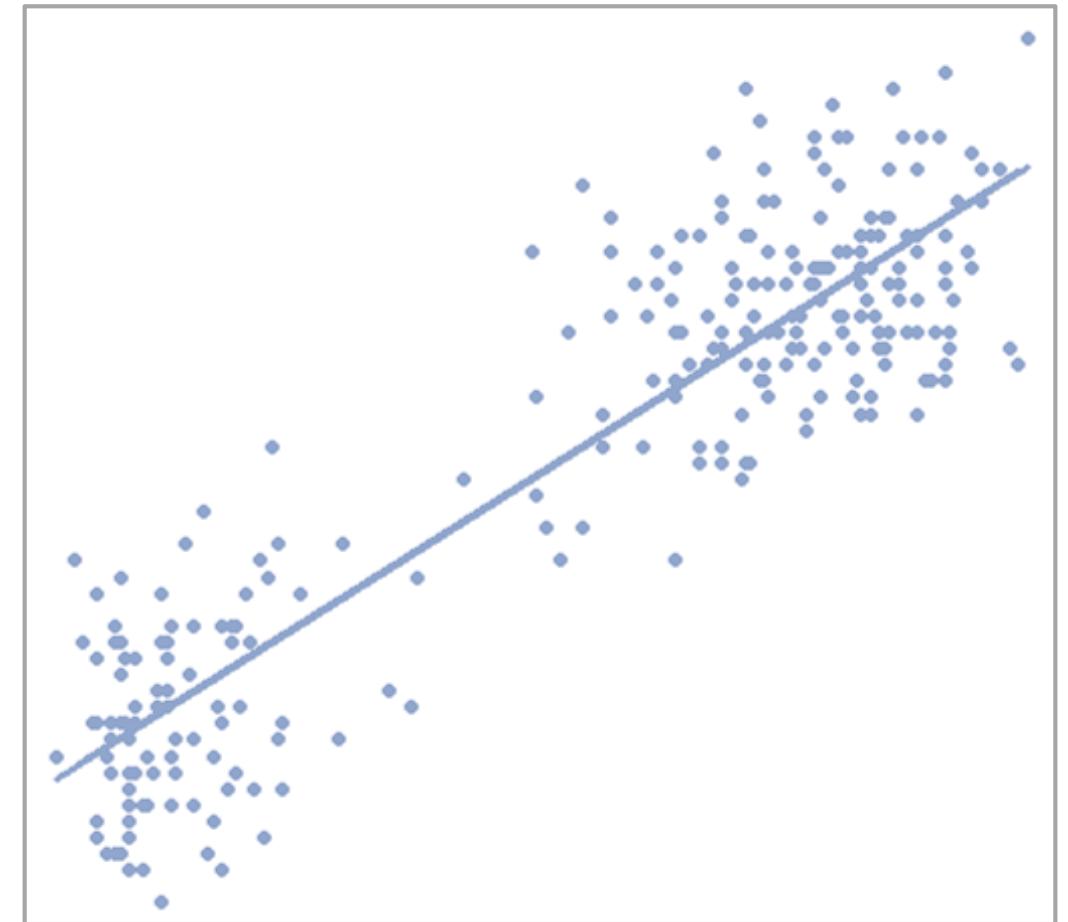
# Simple Linear Regression

Relationship  
Linear model



# Simple Linear Regression

Relationship  
Linear model  
Explanatory variable



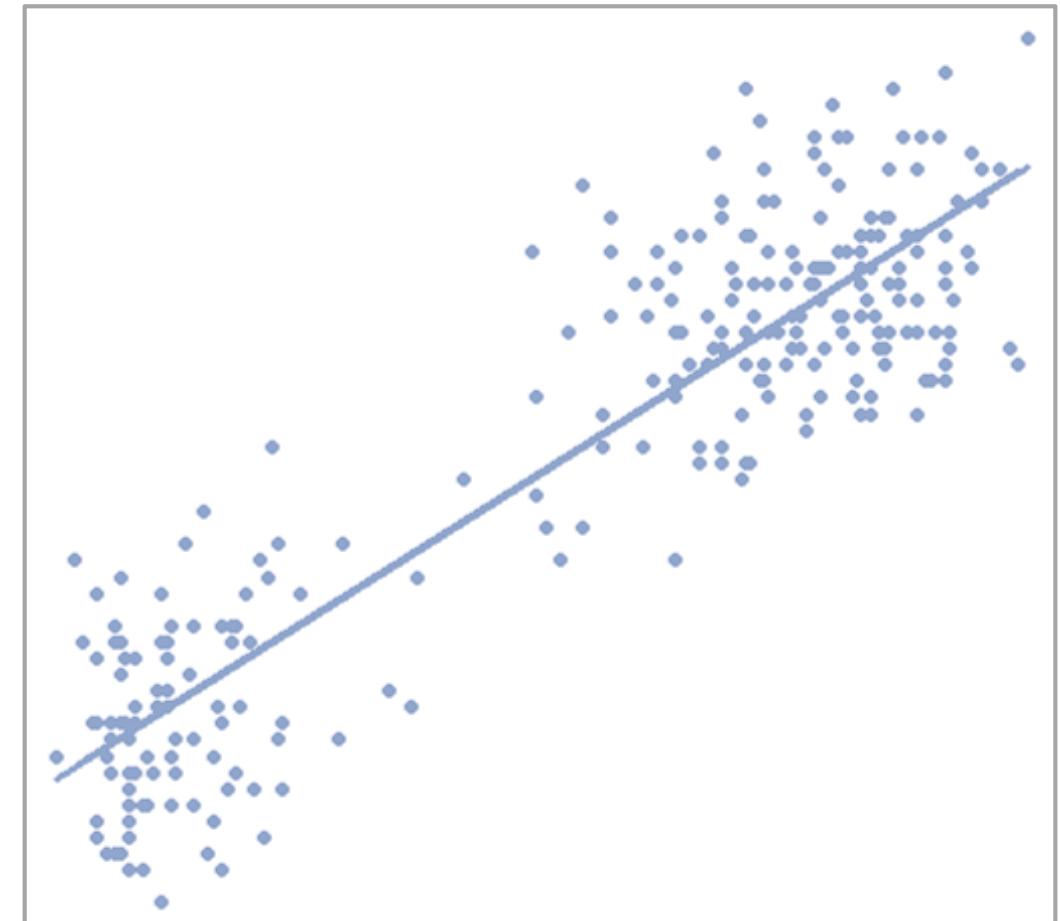
# Simple Linear Regression

Relationship

Linear model

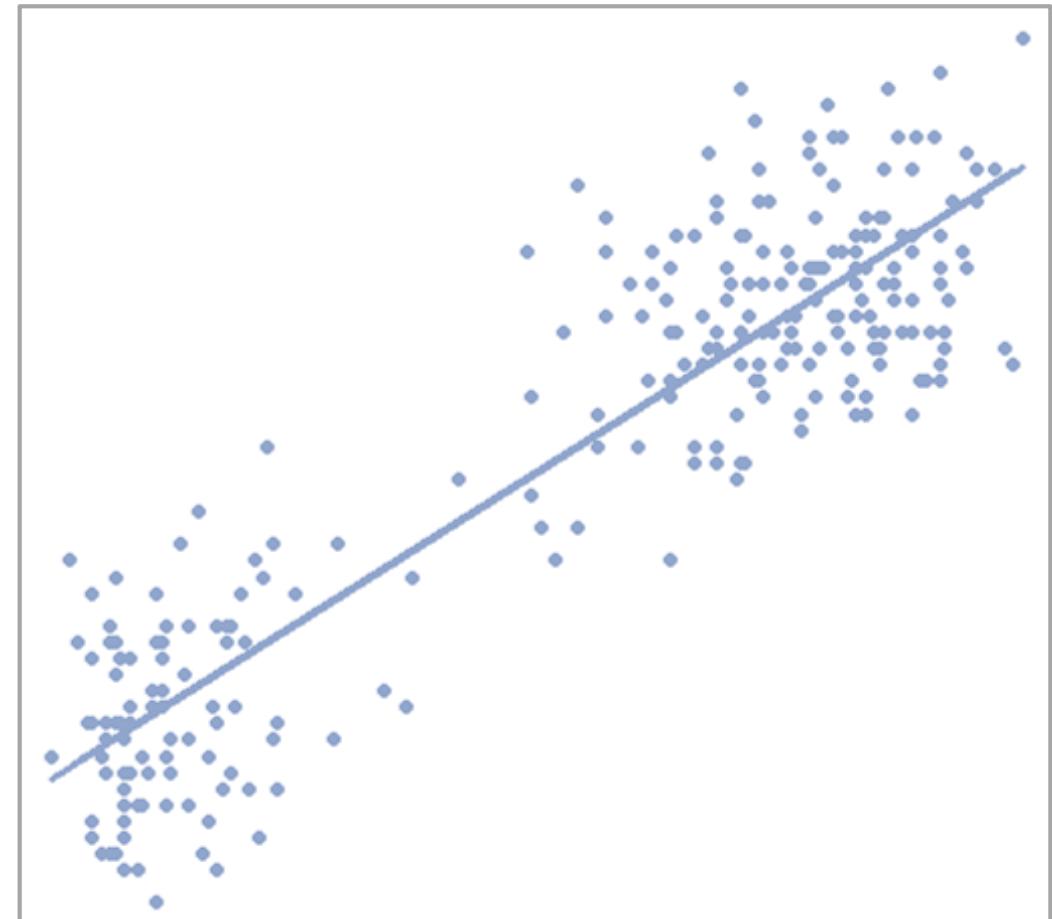
Explanatory variable

Outcome variable



# Simple Linear Regression

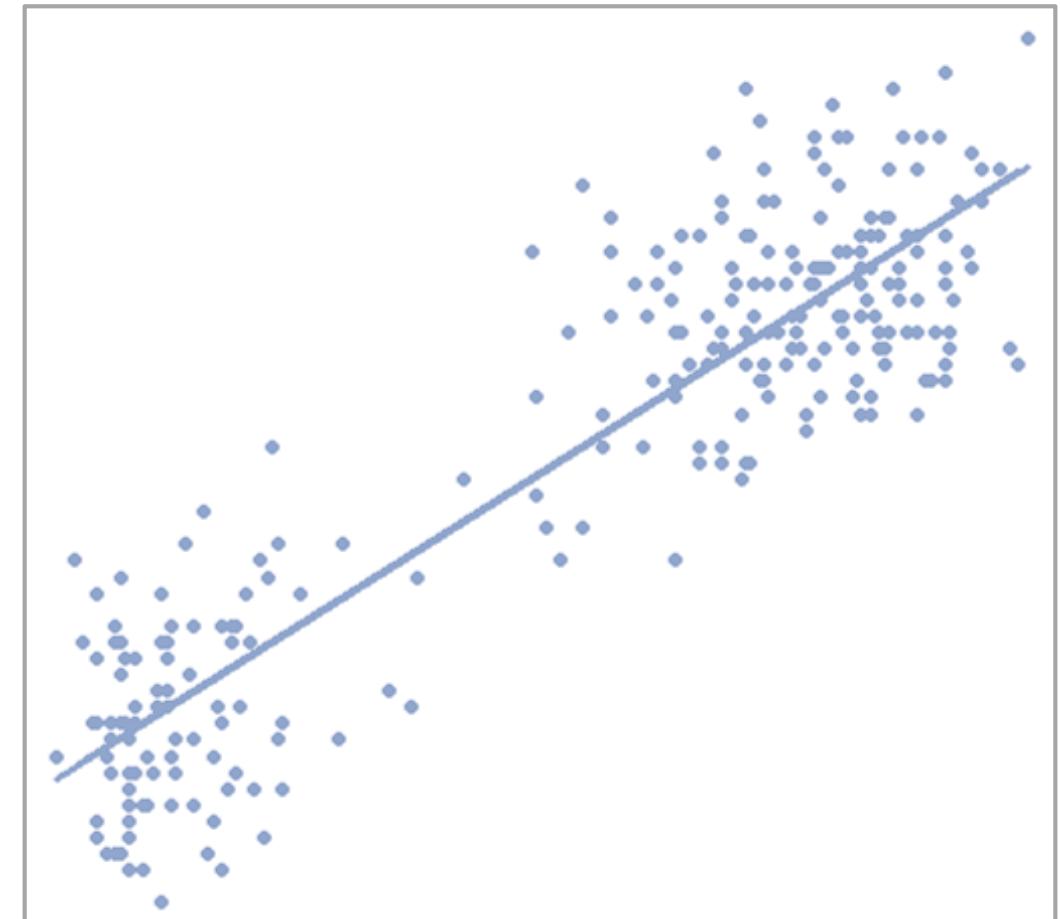
Linear predictor function



# Simple Linear Regression

Linear predictor function

$$y = m \cdot x + b$$

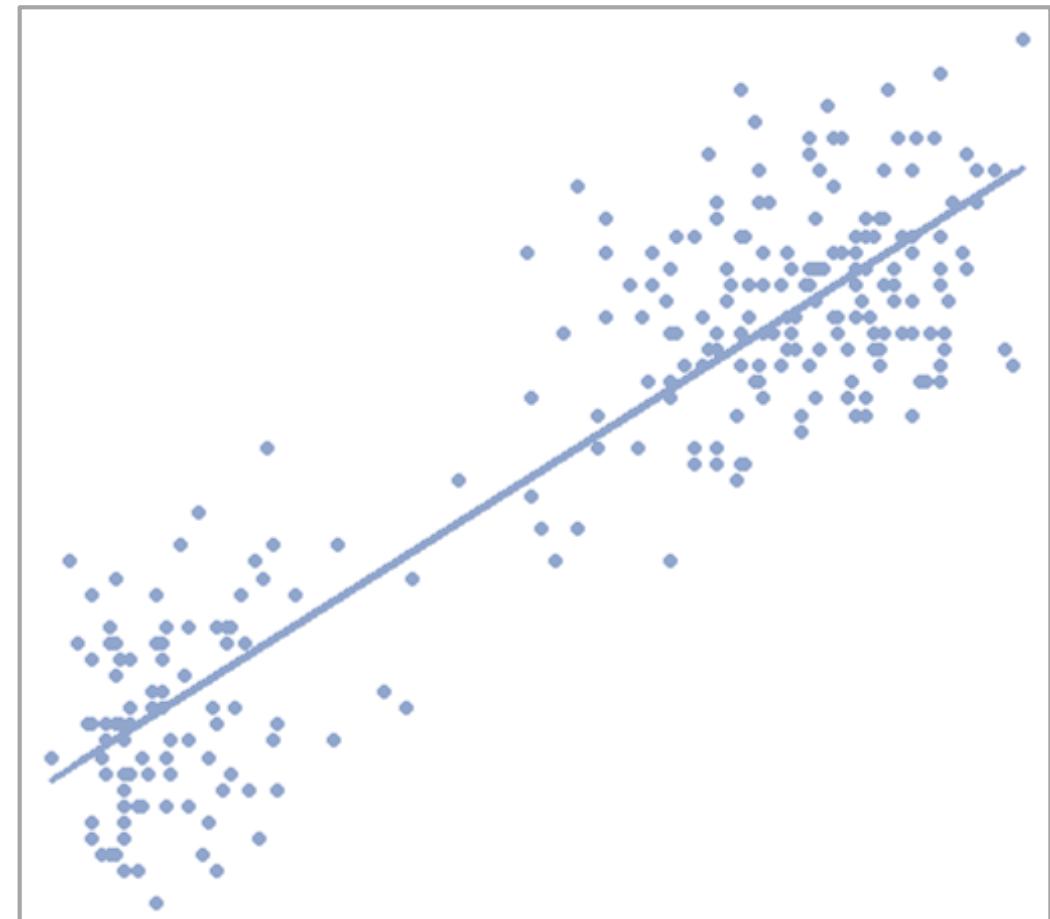


# Simple Linear Regression

Linear predictor function

$$y = m \cdot x + b$$

Parameters estimated



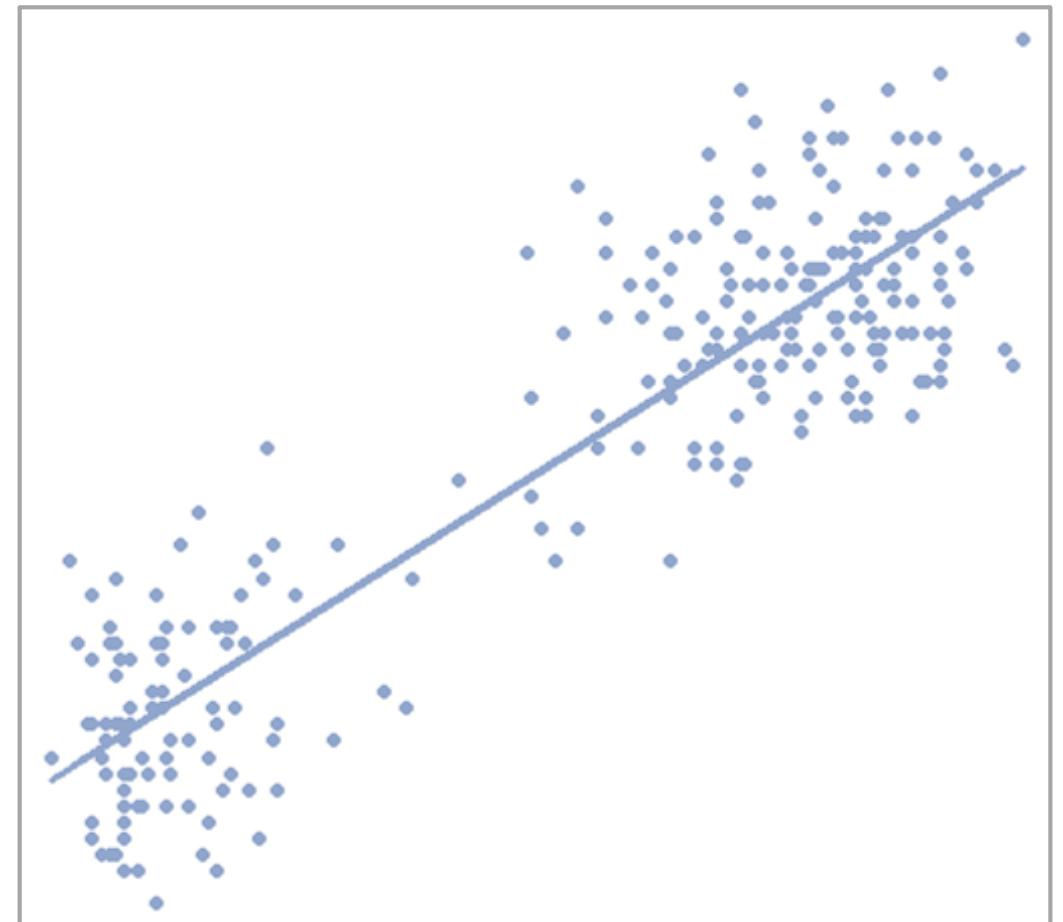
# Simple Linear Regression

Linear predictor function

$$y = m \cdot x + b$$

Parameters estimated

Relies on assumptions



# Regression Demo

Goal: Predict petal width  
based on petal length

# Real-World Examples

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

# Beyond the Basics



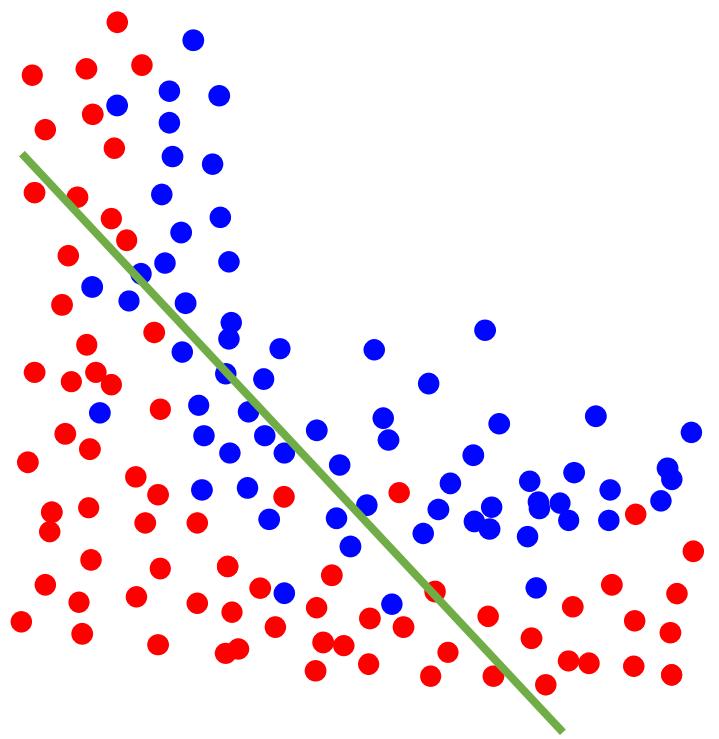
This is just the tip of the iceberg!

# Robust Models

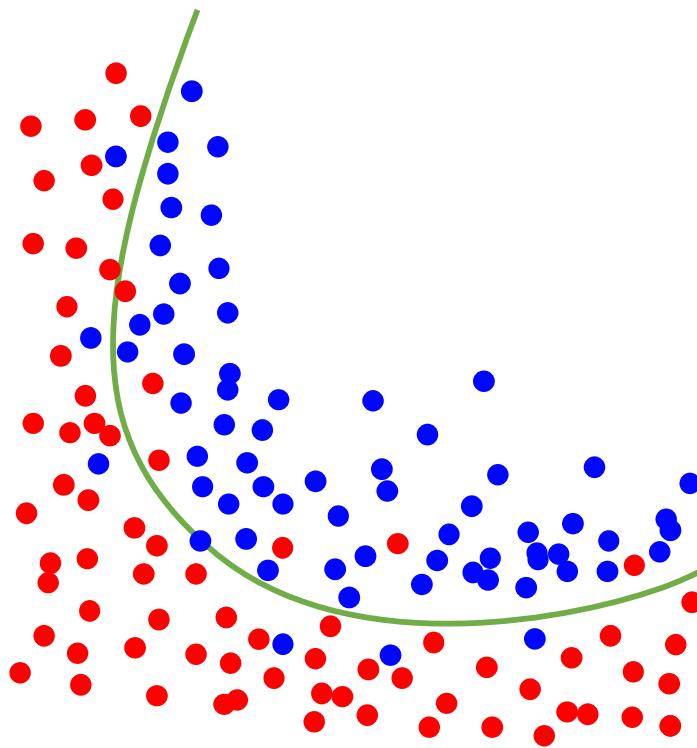
# Cleaning and Transforming Data

Data are messy  
80% of work  
R helps a lot  
Record all steps

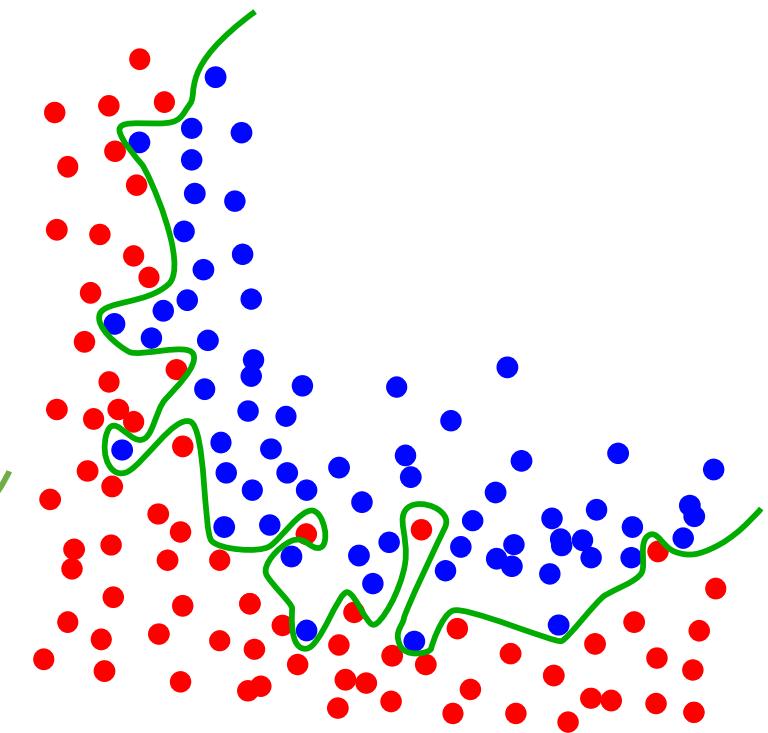




Underfit

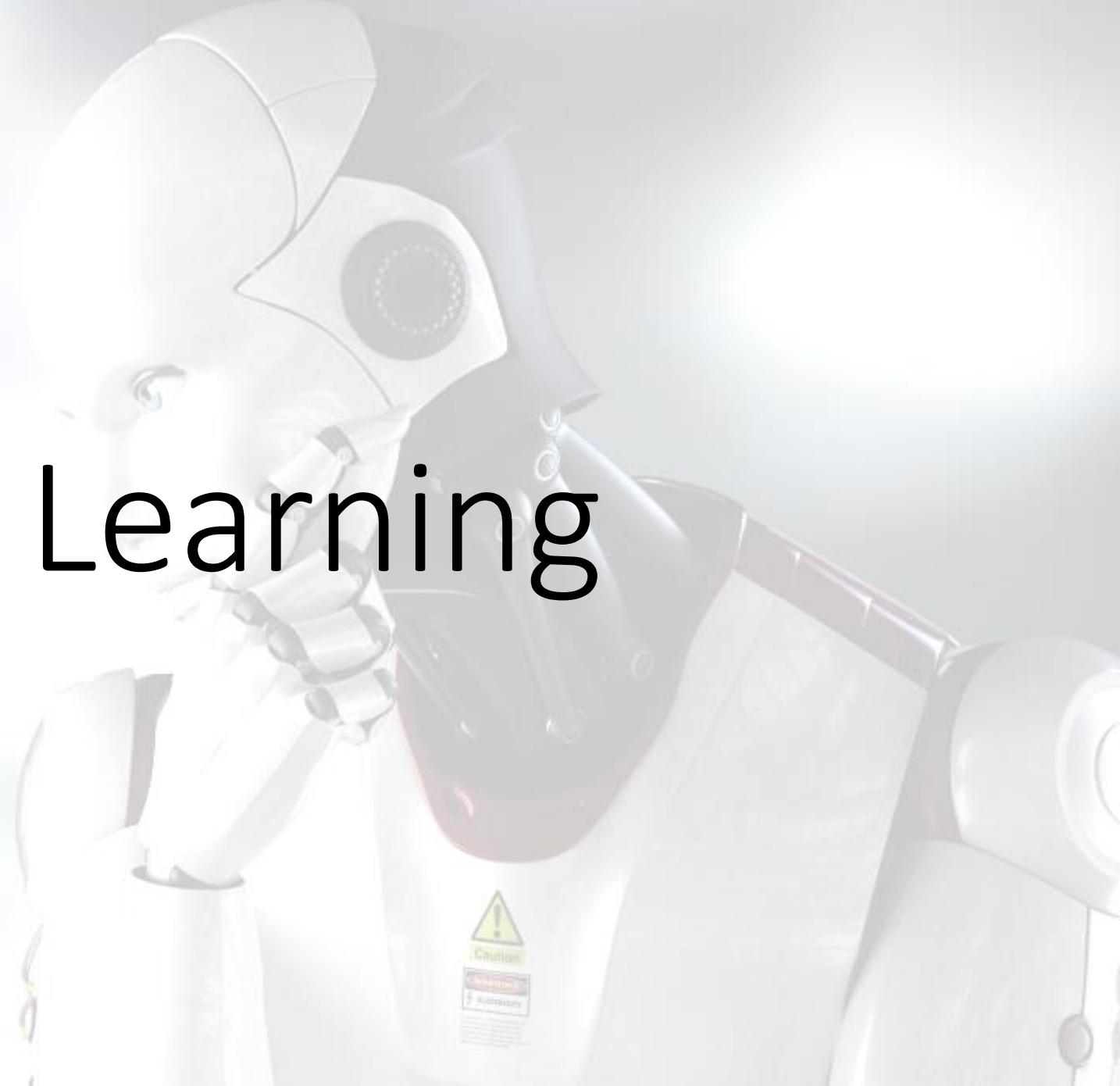


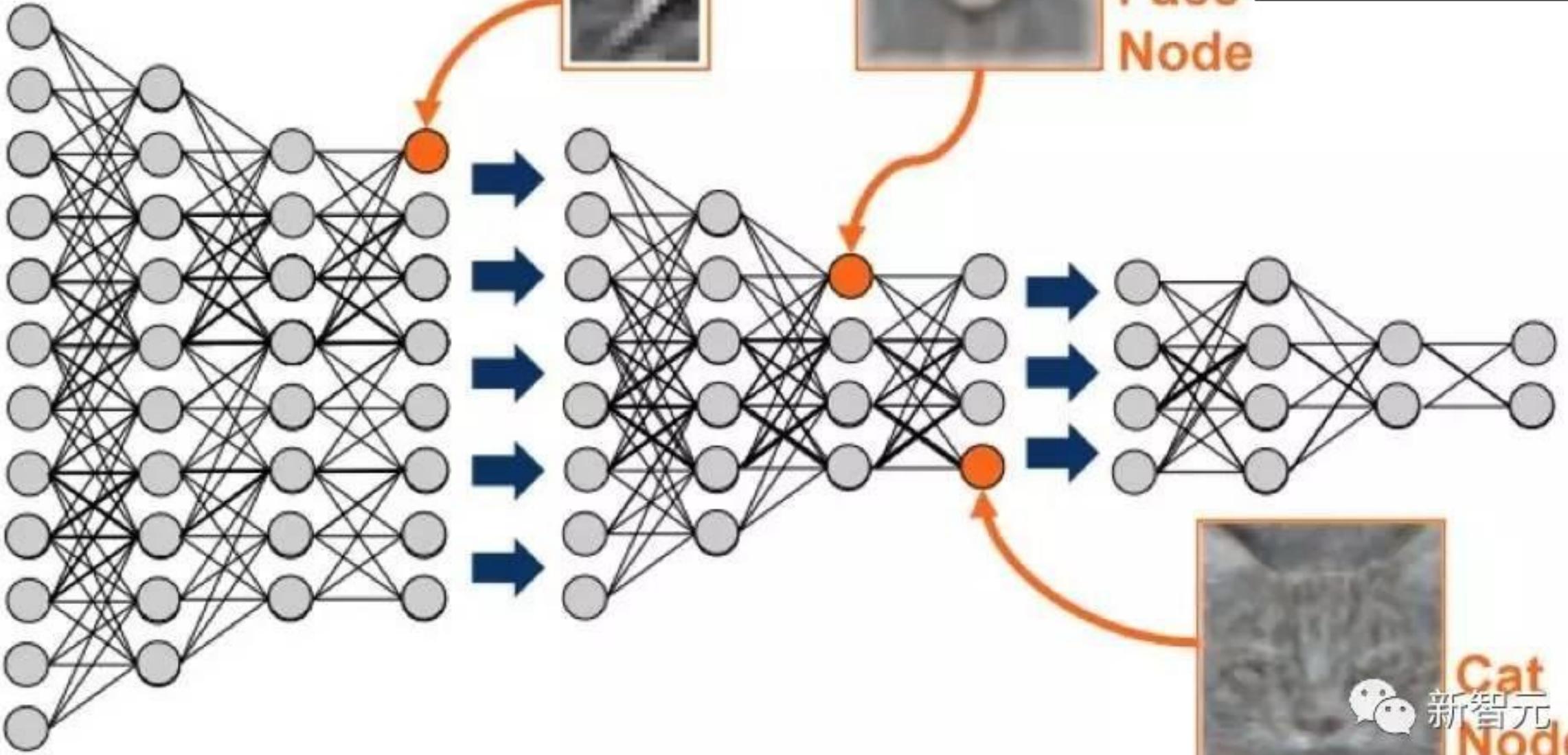
Good fit



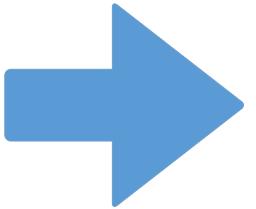
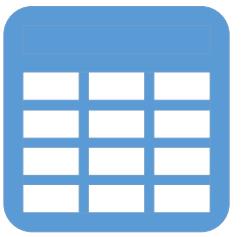
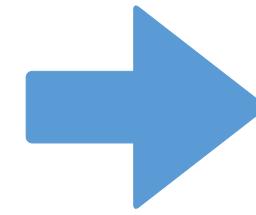
Overfit

# Deep Learning



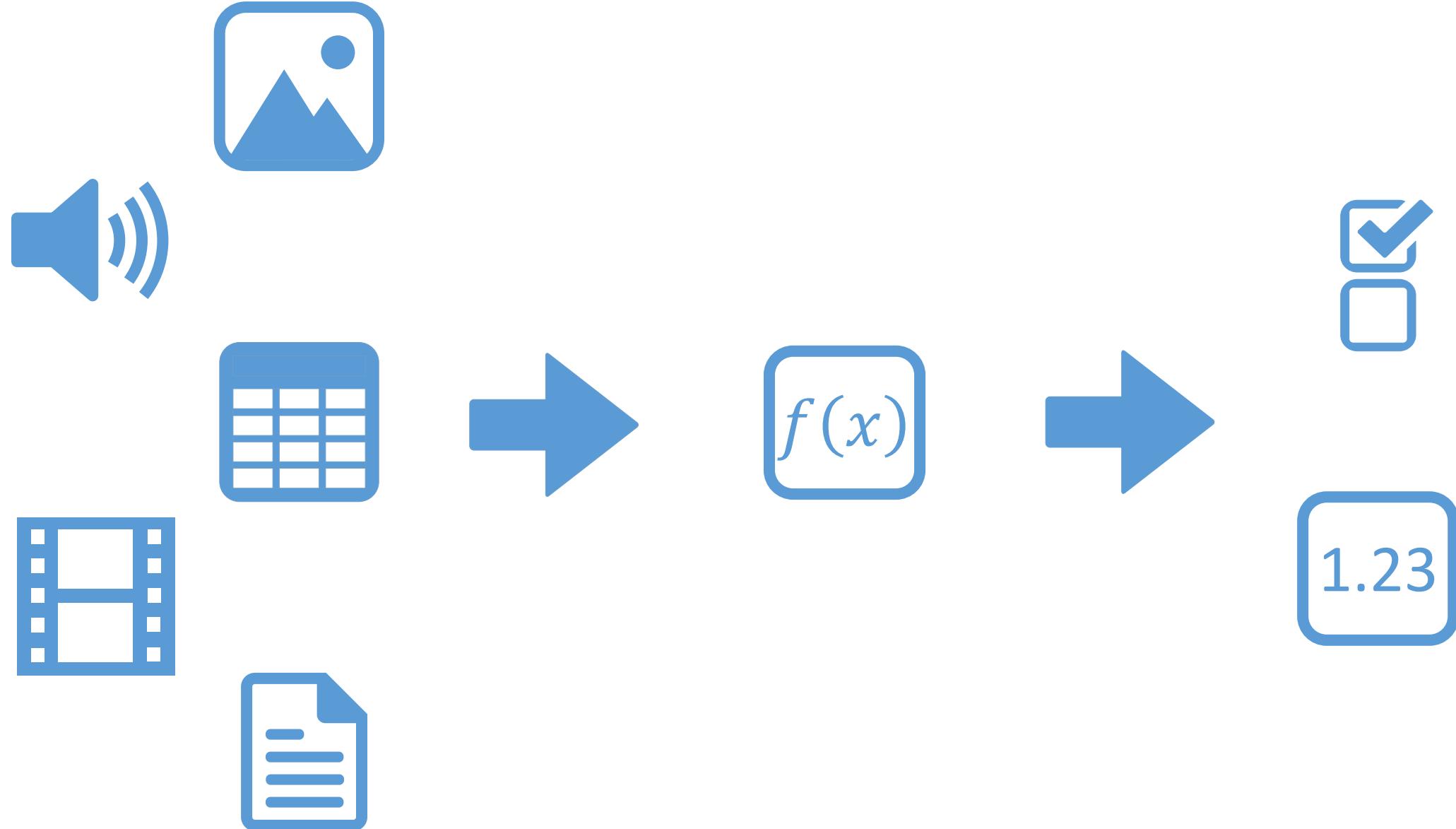


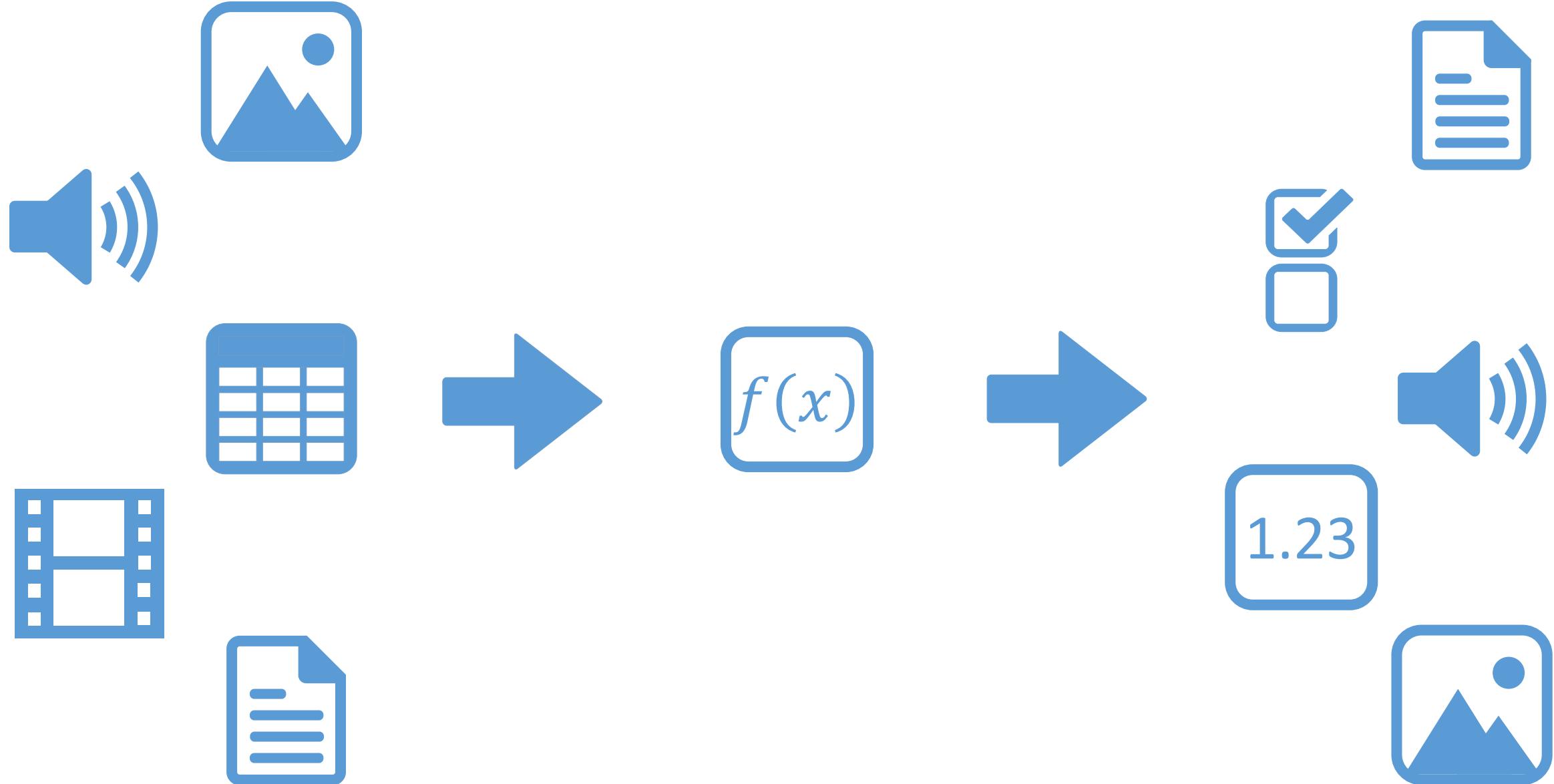


 $f(x)$ 

1.23









FaceApp

# Where to Go Next

Pluralsight: <https://www.pluralsight.com>

Coursera: <https://www.coursera.org>

Data Camp: <https://www.datacamp.com>

Tensorflow: <http://playground.tensorflow.org>

# My Website

Articles

Presentations

Source Code

Videos

Workshops

Matthew Renze

Home Articles Courses Presentations Software About Contact

## News

2016-07-11 - The Big Data Refinery

I wrote an article describing the Data Refinery pattern, which is a pattern for handing multiple consumers of Big Data. I learned about this pattern from my interactions with the Big Data Group at Microsoft.



2016-07-01 - Microsoft MVP Award

I received my first Microsoft MVP Award today. Very happy to be part of such an amazing group of people! In addition, I'm really looking forward to attending the Microsoft MVP Global Summit again in November.



Matthew is an independent software consultant, author for Pluralsight, international public speaker, a Microsoft MVP, ASPIndustry, and open-source software contributor.

2016-06-26 - JavaScript Air Interview

Kent Dodds invited me to be on his podcast JavaScript Air at KCDC. The video and audio of the podcast are now available online.



2016-06-25 - Lifelong Learning as a Developer

I participated in a discussion panel at KCDC on Lifelong Learning as a Software Developer. The video of the discussion panel is now available online. I thought all of the panelist did an excellent job.



[www.matthewrenze.com](http://www.matthewrenze.com)



PLURALSIGHT

# Data Science with R

★★★★★ By Matthew Renze

Data science is becoming more and more valuable to the workplace and to the global economy. Learn how to use the practice of data science and the programming language R to transform your data into actionable insight.

[Start free trial now](#)

[▶ Play course overview](#)

[www.pluralsight.com/authors/matthew-renze](http://www.pluralsight.com/authors/matthew-renze)

# Conclusion

# Conclusion

1. Introduction to ML
2. Introduction to R
3. Classification
4. Regression
5. Beyond the Basics



# Feedback

Very important to me!

What did you like?

What could I improve?



# Contact Info

Matthew Renze

Data Science Consultant

Renze Consulting

Twitter: [@matthewrenze](https://twitter.com/matthewrenze)

Email: [matthew@matthewrenze.com](mailto:matthew@matthewrenze.com)

Website: [www.matthewrenze.com](http://www.matthewrenze.com)



Thank You! : )

## TITANIUM SPONSORS



## Platinum Sponsors



## Gold Sponsors

