



# Machine Learning

# What Are Your Working Definitions of Machine Learning?



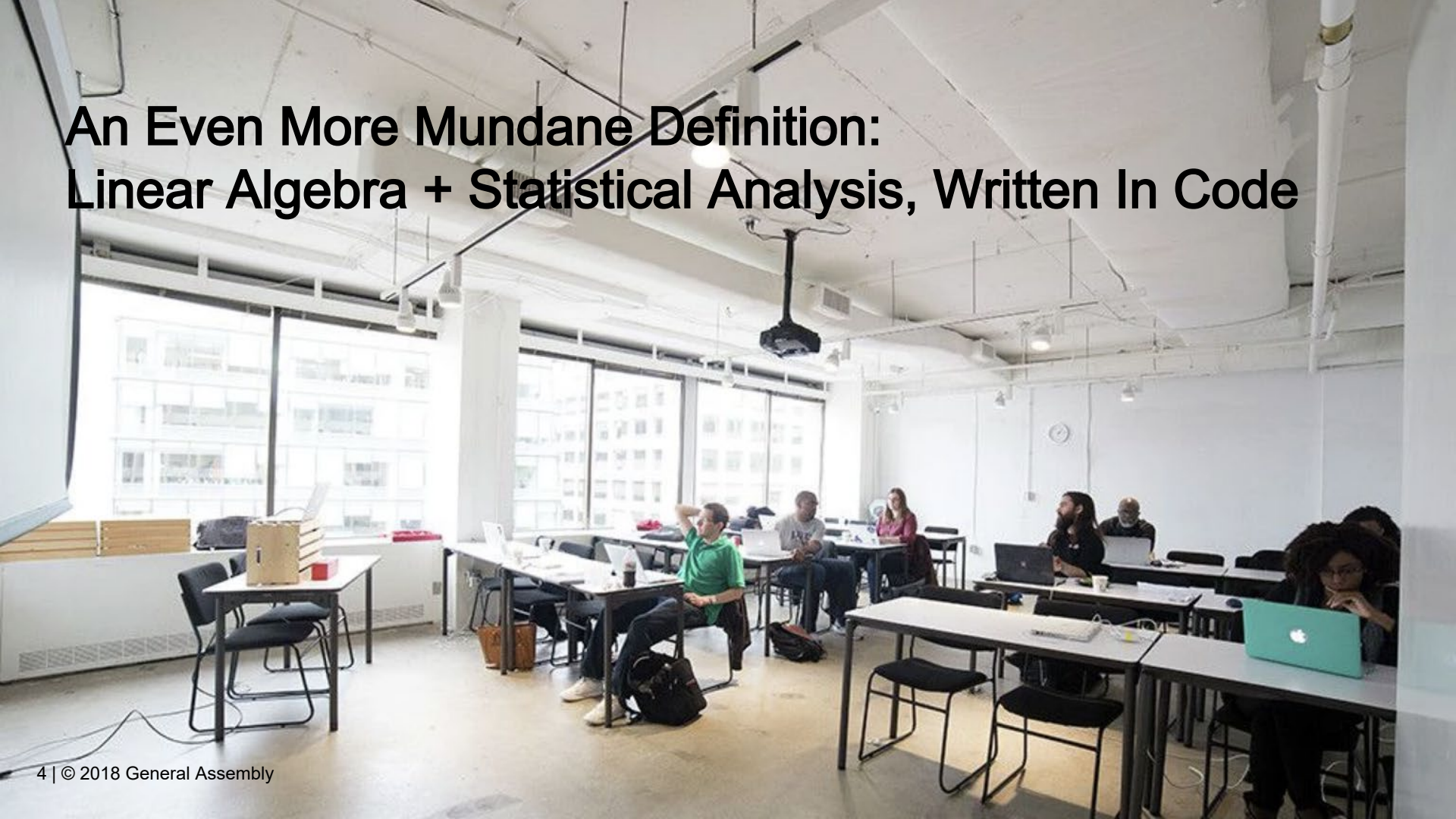
“

# The Broader Notion of Building Statistical Artifacts That Become More Accurate Over Time Based on Experience

Charles Isbell, Professor, Georgia Tech



# An Even More Mundane Definition: Linear Algebra + Statistical Analysis, Written In Code



# Machine Learning

$$\textit{cost} = \sum (\textit{answer} - \textit{guess})^2$$



# Machine Learning

AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	PRICE
65.2	4.0900	1.0	296.0	15.3	396.90	4.98	24.0
78.9	4.9671	2.0	242.0	17.8	396.90	9.14	21.6
61.1	4.9671	2.0	242.0	17.8	392.83	4.03	34.7
45.8	6.0622	3.0	222.0	18.7	394.63	2.94	33.4
54.2	6.0622	3.0	222.0	18.7	396.90	5.33	36.2

# Machine Learning

AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	PRICE	PREDICTION
65.2	4.0900	1.0	296.0	15.3	396.90	4.98	24.0	30.008213
78.9	4.9671	2.0	242.0	17.8	396.90	9.14	21.6	25.029861
61.1	4.9671	2.0	242.0	17.8	392.83	4.03	34.7	30.570232
45.8	6.0622	3.0	222.0	18.7	394.63	2.94	33.4	28.608141
54.2	6.0622	3.0	222.0	18.7	396.90	5.33	36.2	27.942882

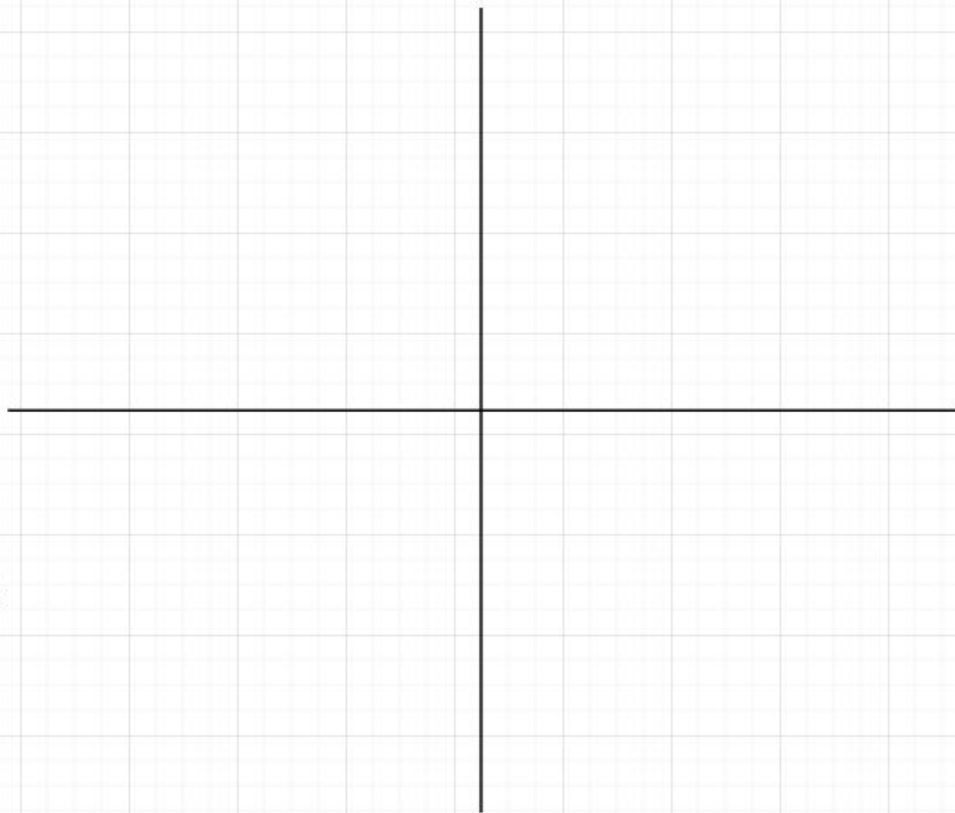


Supervised

Unsupervised

Regression

Classification







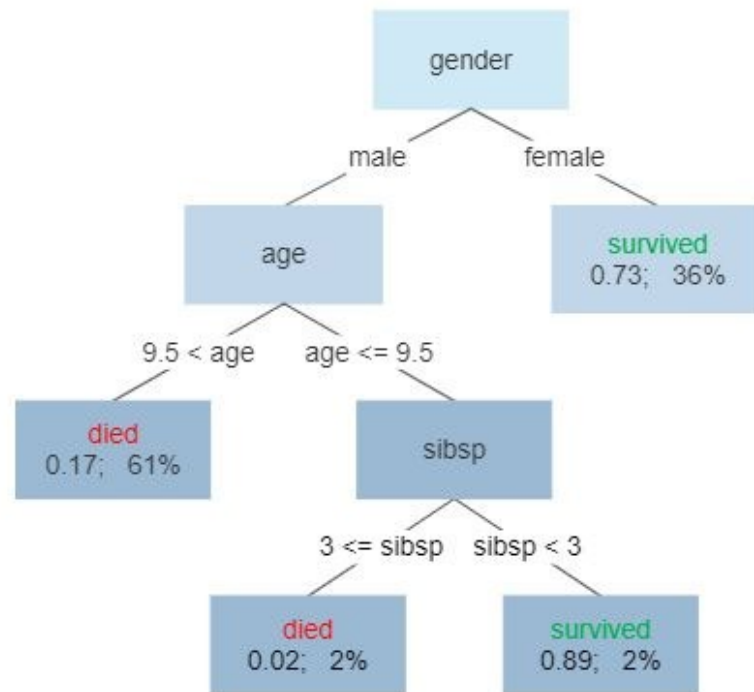
# Decision Trees

Machine Learning With Python

# Decision Trees

- Non-linear machine learning algorithm that continually partitions your data into smaller and smaller subsets
- Works by continually going through your dataset and finding the value that creates the largest weighted difference in accuracy on each side of the split
- Decision making process is very similar to how humans cycle through data to make decisions

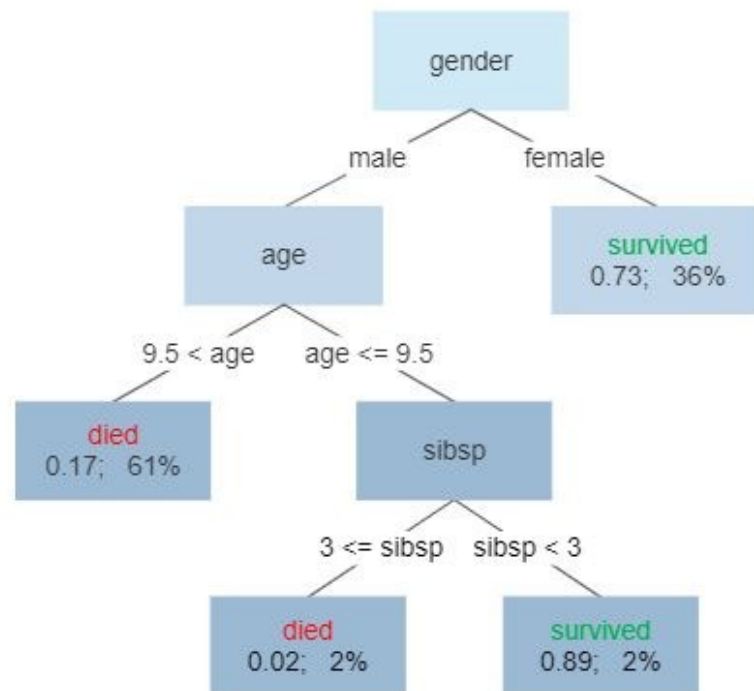
## Survival of passengers on the Titanic



# Decision Trees

- The good:
  - Non-linear
  - Require little to no data prep
  - Handle outliers very well
  - No statistical assumptions

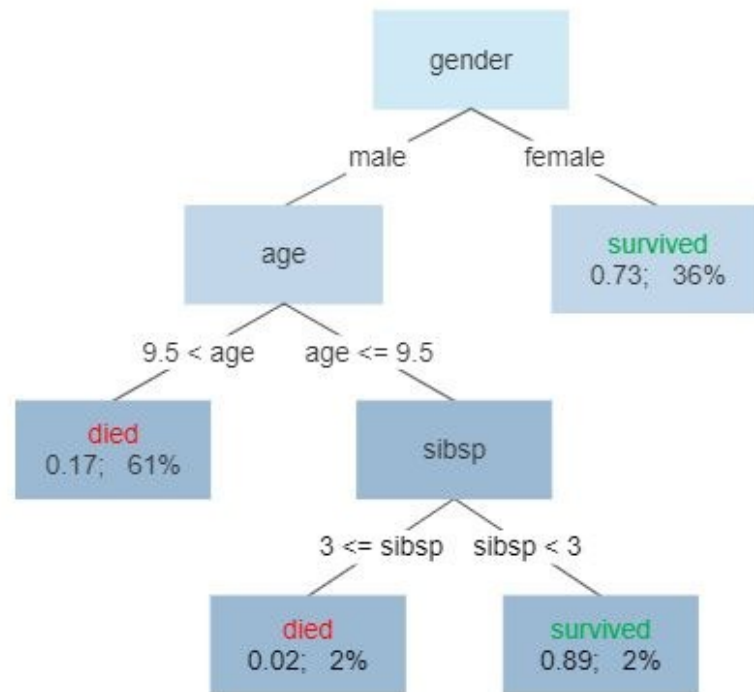
## Survival of passengers on the Titanic



# Decision Trees

- The bad:
  - Very unstable predictions
  - Memorize your data, more so than learn from it
  - Results can be very sensitive to tree order
    - Don't generalize well to out of sample data

## Survival of passengers on the Titanic



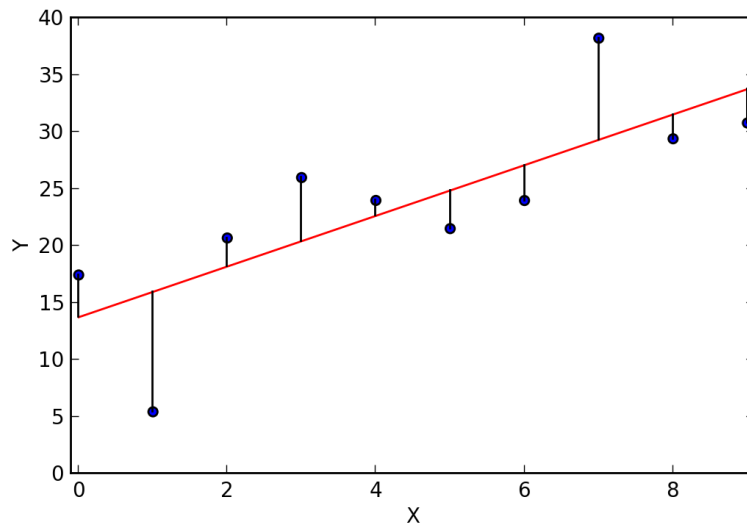


# Scikit Learn

Machine Learning With Python

# Scikit Learn

- The main library used to implement ML methods
- Jack-of-all trades, master of none
- Contains built-in techniques for most ML concepts
- Is primarily built to access your own computer's memory
- Runs on a CPU, but not a GPU



# Scikit Learn

- **fit()** - apply the algorithm to your data
- **score()** - evaluate your algorithm
- **predict()** - estimate answer based on new info
- **get\_params()** - access parameters of your algorithm
- **set\_params()** - change parameters of your algorithm

