



DAT Class 10

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



Machine Learning

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

Machine Learning

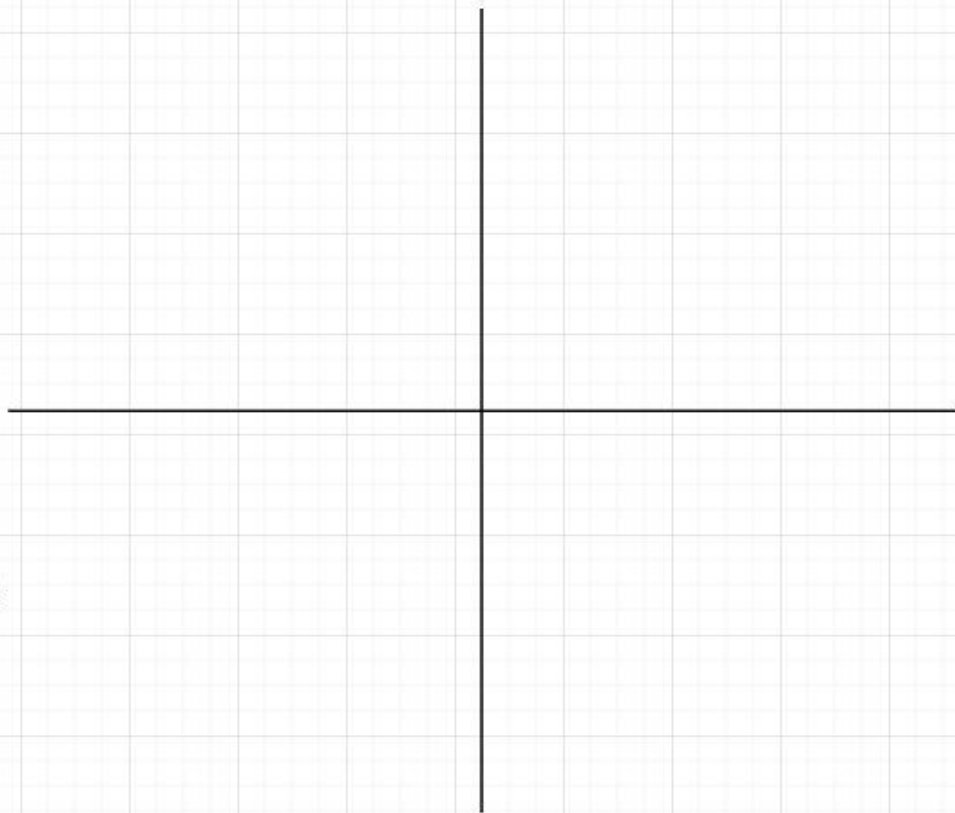
age	sex	bmi	children	smoker	region	charges	Prediction
19	female	27.900	0	yes	southwest	16884.92400	19270.248429
18	male	33.770	1	no	southeast	1725.55230	3715.049033
28	male	33.000	3	no	southeast	4449.46200	6492.553938
33	male	22.705	0	no	northwest	21984.47061	5686.393331
32	male	28.880	0	no	northwest	3866.85520	3744.188992

Supervised

Unsupervised

Regression

Classification





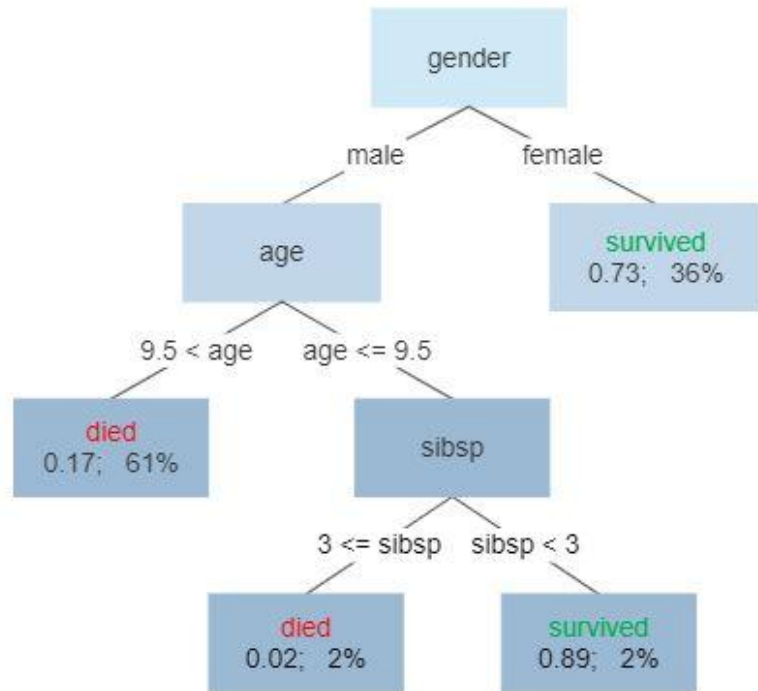
Decision Trees

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

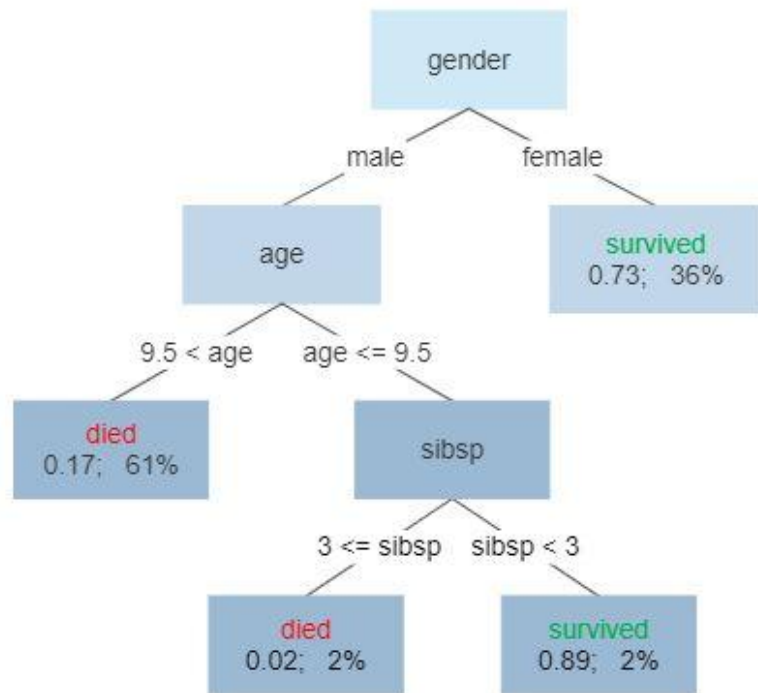
Survival of passengers on the Titanic



Decision Trees

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

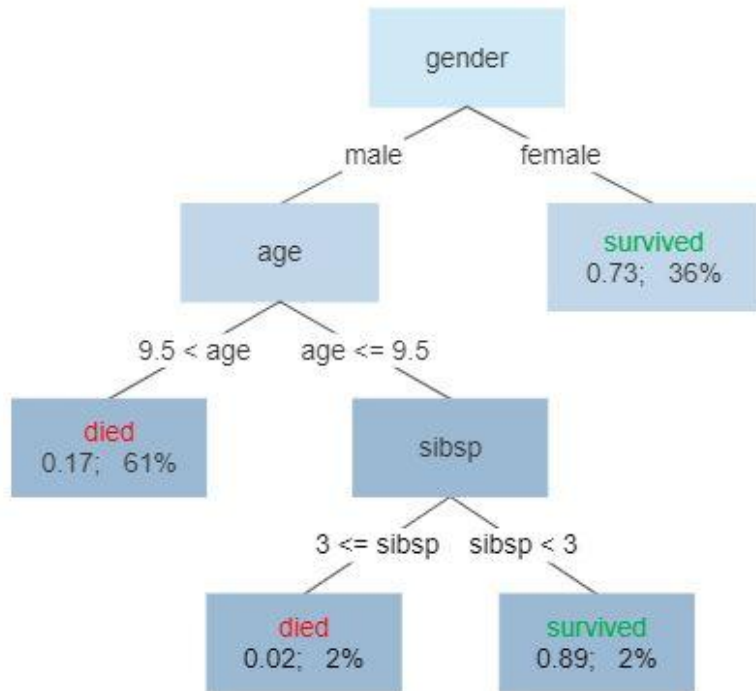
Survival of passengers on the Titanic



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



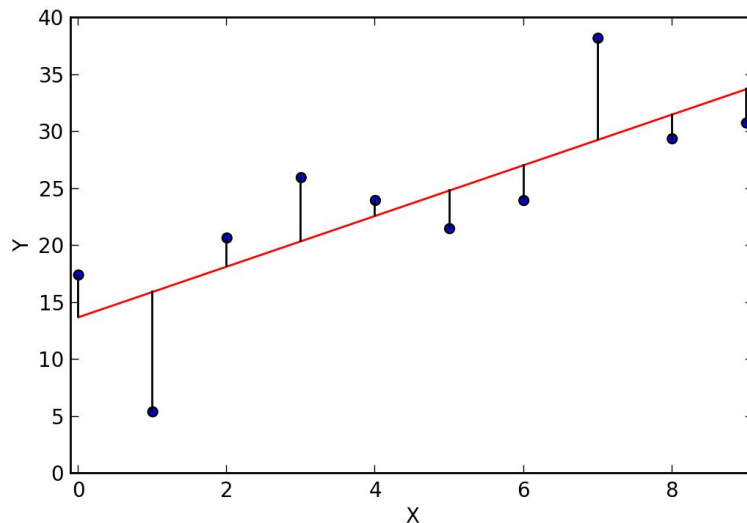


Scikit Learn

Machine Learning With Python

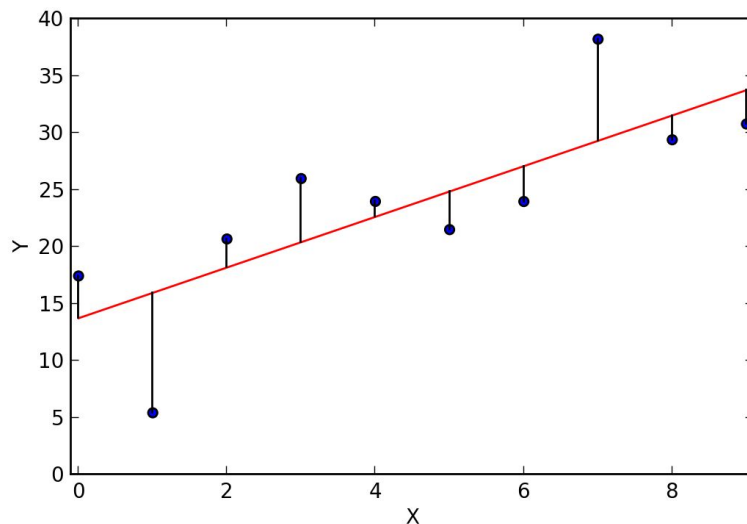
Scikit Learn

- 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





Model Interpretation

Machine Learning With Python

Model Interpretation - Common Metrics

R^2 Value

- How much of the change in y is described by the change in X
- Captures how much better your model is than the simplest alternative -- predicting the average

