Machine Learning with Scikit-Learn

About Me

- PhD Student at Georgia Tech
- What I have done:

Understanding Fitbit Data

Stanford Continuing Studies

Machine Learning with Python Instructor

LinkedIn Learning

LinkedIn Learning Instructor

What you should know for this talk

- Python strings
- Integers and floats
- Conditionals and booleans-if, else, and elif
- For loops
- Lists, tuples
- Dictionaries

What is Machine Learning?

“field of study that gives computers the ability to learn from data without being explicitly programmed”

- Arthur Samuel

What is Machine Learning?

“‘Learning’ enters when we give these models tunable parameters that can be adapted to observed data; in this way the program can be considered to be ‘learning’ from the data."

- Jake Vanderplas

What is Machine Learning?

"Once models have been fit to previously seen data, they can be used to predict and understand aspects of newly observed data"

- Jake Vanderplas

Two Basic Types of Machine Learning

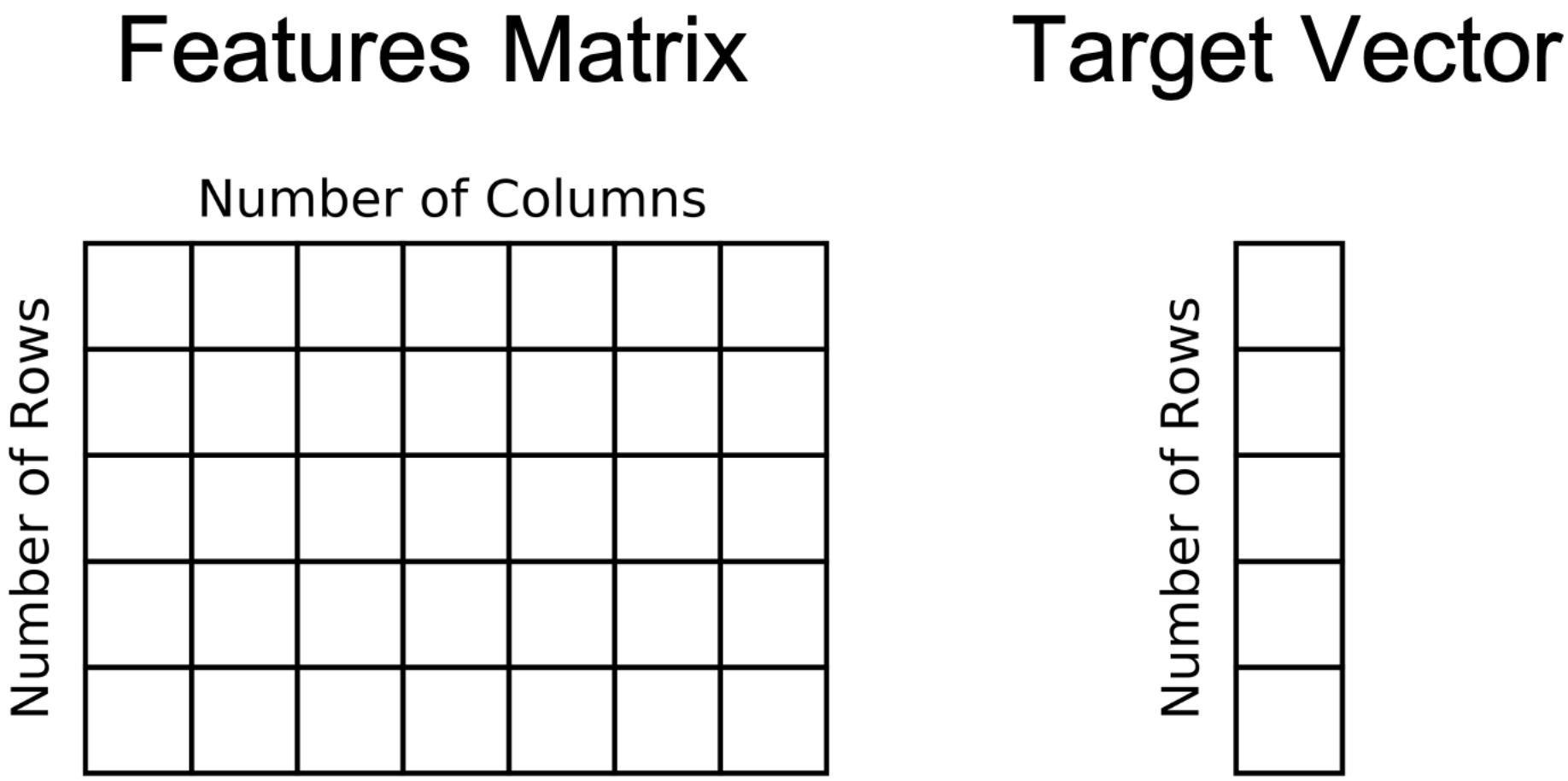
Supervised learning

Modeling the relationship between measured features of data and some target (label/value). When target is a label, it is classification. When target is a continuous value, it is regression.

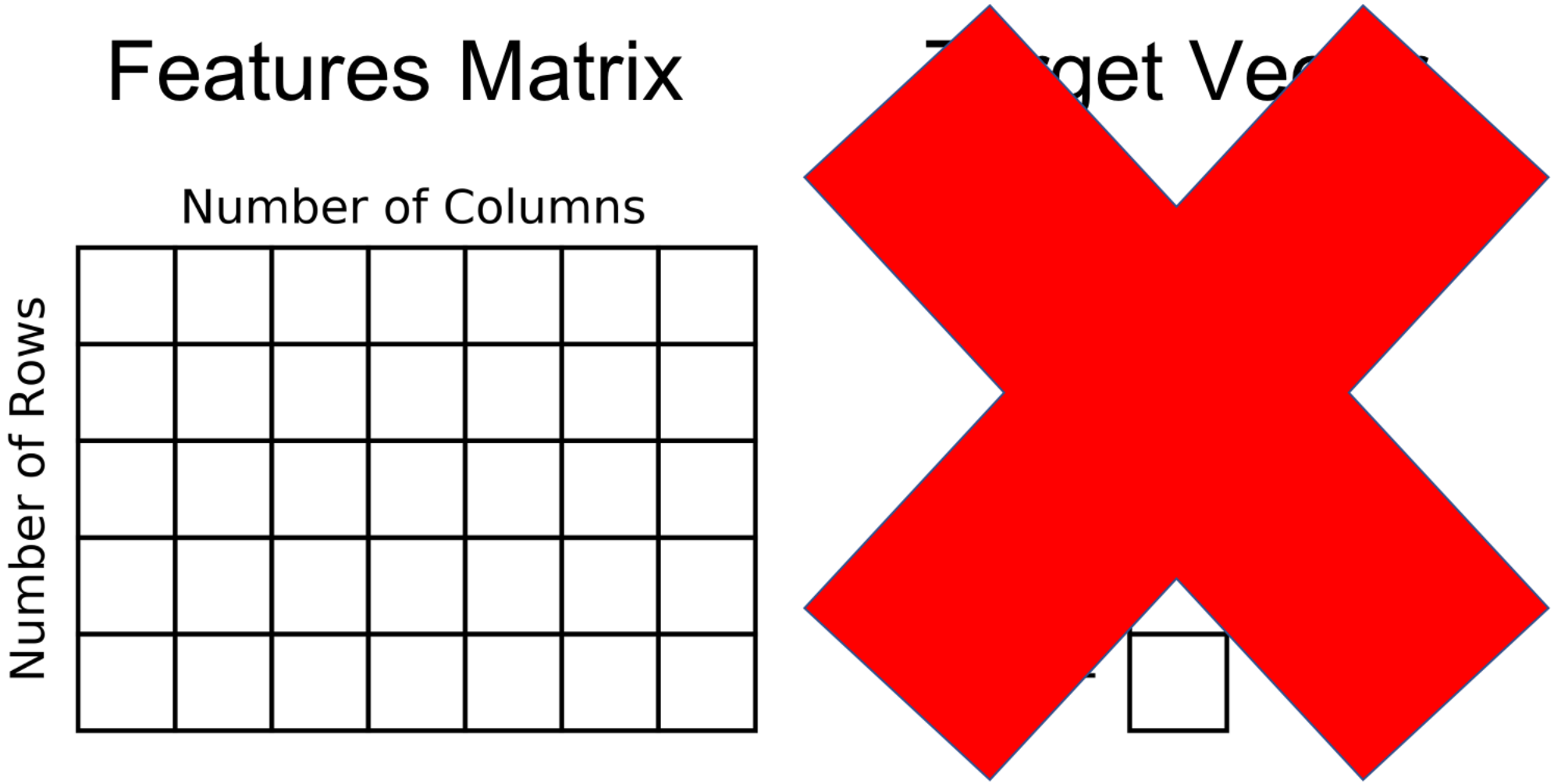
Unsupervised learning

Modeling the measured features of data without some target (label/value)

Supervised learning



Unsupervised learning



Why use Scikit-Learn for Machine Learning?

- Programming machine learning algorithms from scratch is no easy task.
- Scikit-learn provides a large amount of well programmed algorithms
- Syntax allows you to try multiple algorithms quickly
- Library works well with other python libraries like NumPy, pandas, and matplotlib
- Library is popular. There are always more scikit-learn tutorials being written. This means that you can get your questions answered more easily.

Why is this talk using Scikit-Learn?

- The goal of this talk is to try out multiple machine learning algorithms quickly and learn each algorithms strengths and weaknesses.
- In the process, learn concepts such as bias, variance, overfitting, and underfitting