



Phase 3 Expectations

Machine Learning





Agenda

1. **Phase 3 Overview**
2. **Week 1: Math & Classification**
3. **Week 2: More Models!**
4. **Gating / Assessments**
5. **What Will Phase 3 Feel Like?**

Overall Phase Timeline

Statistics

PHASE 2
Weeks 4 - 6

Advanced Topics

PHASE 4
Weeks 10 - 12

PHASE 1
Weeks 1 - 3

Data Engineering

PHASE 3
Weeks 7 - 9

Machine Learning

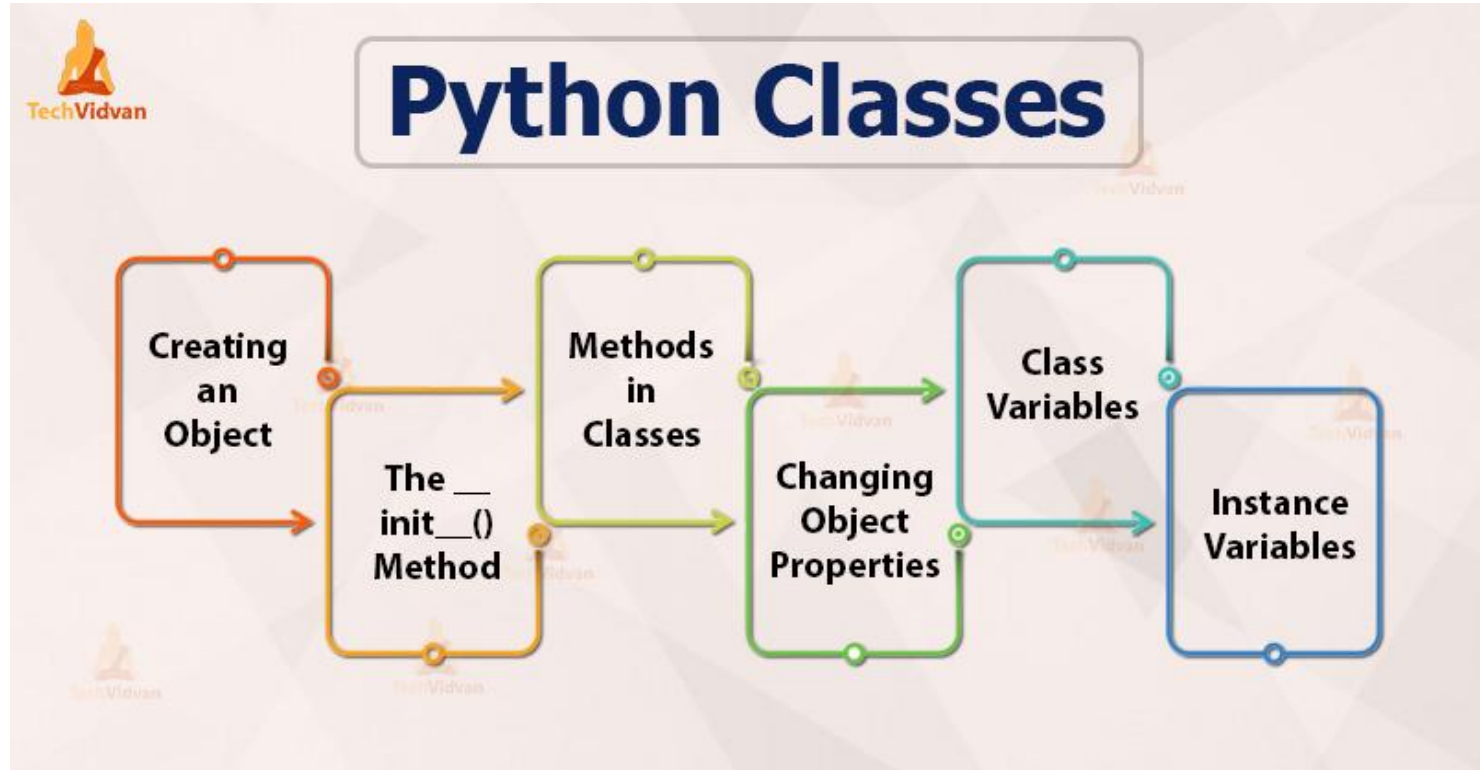
PHASE 5
Weeks 13 - 15

Capstone

Week 1

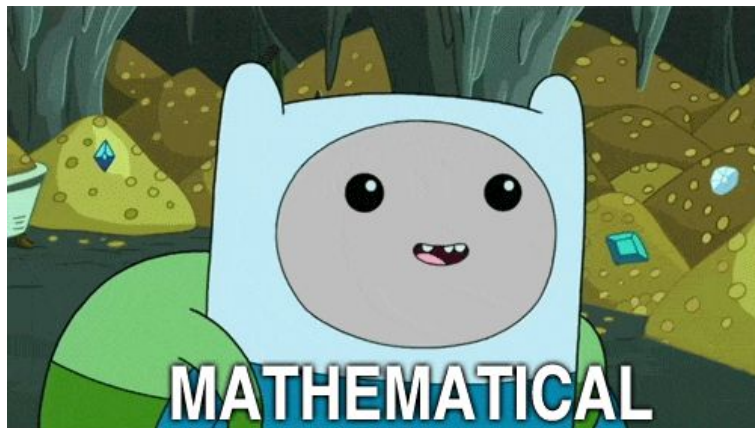
**Programming, Math,
Introduction to Classification**

Object-Oriented Programming



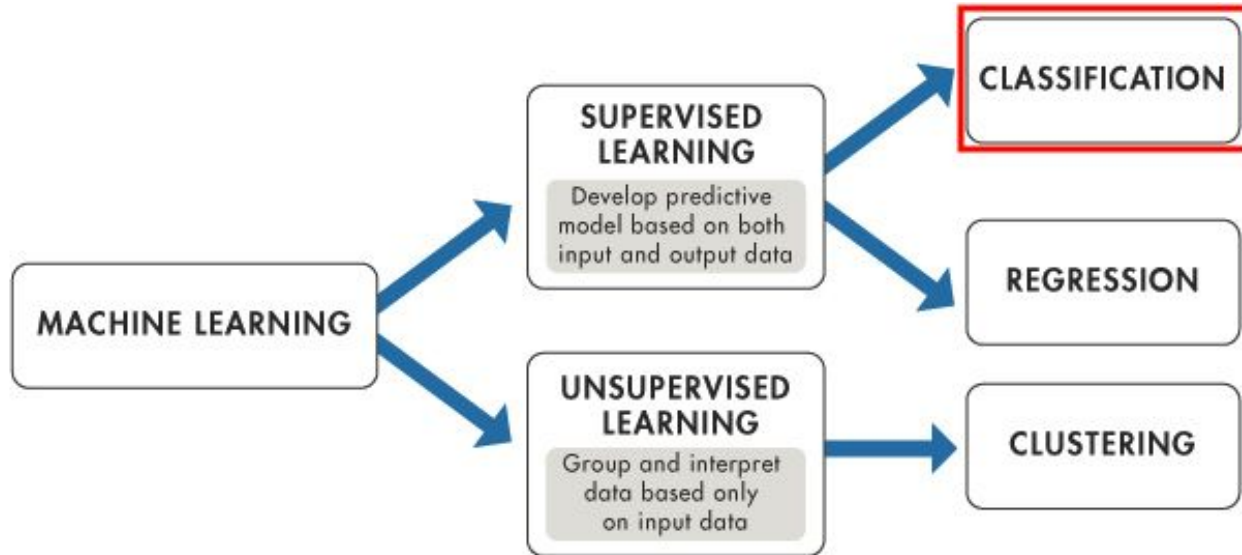
The Math Behind Data Science

- Linear Algebra
- Calculus, Cost Functions, Gradient Descent
- Focus on concepts and application!



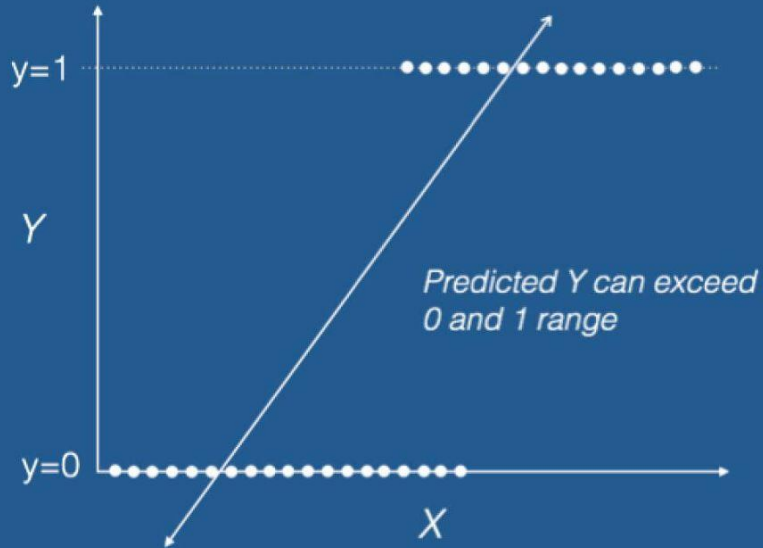
Classification

- Asking: Is it ____ or not?
- From continuous target to categorical target

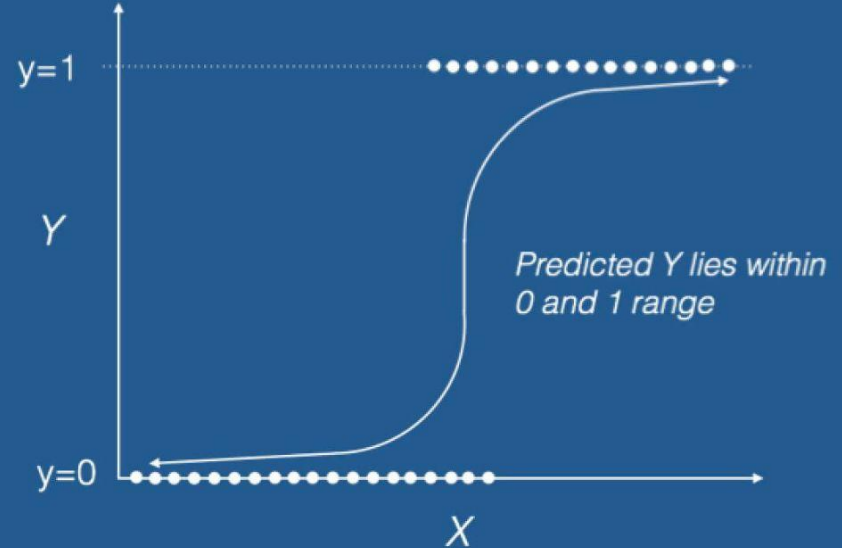


Logistic Regression

Linear Regression



Logistic Regression



Week 2

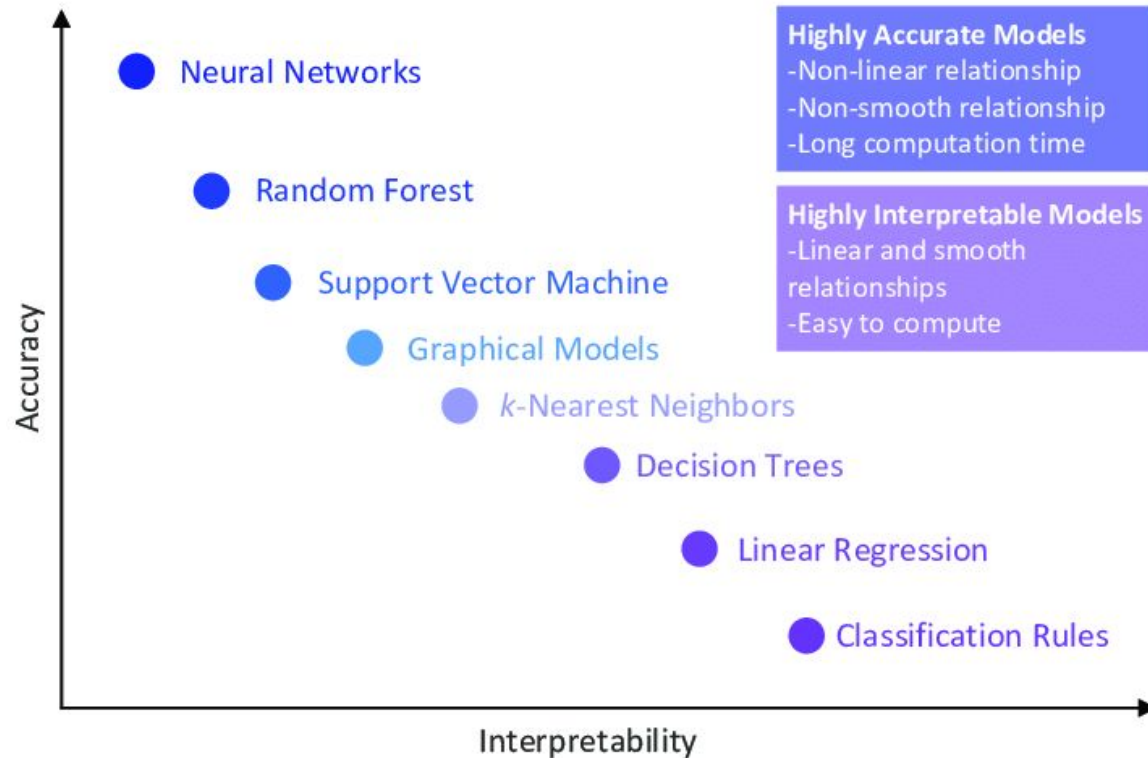
**Classification Evaluation,
More Models, Model Tuning**

Classification Metrics

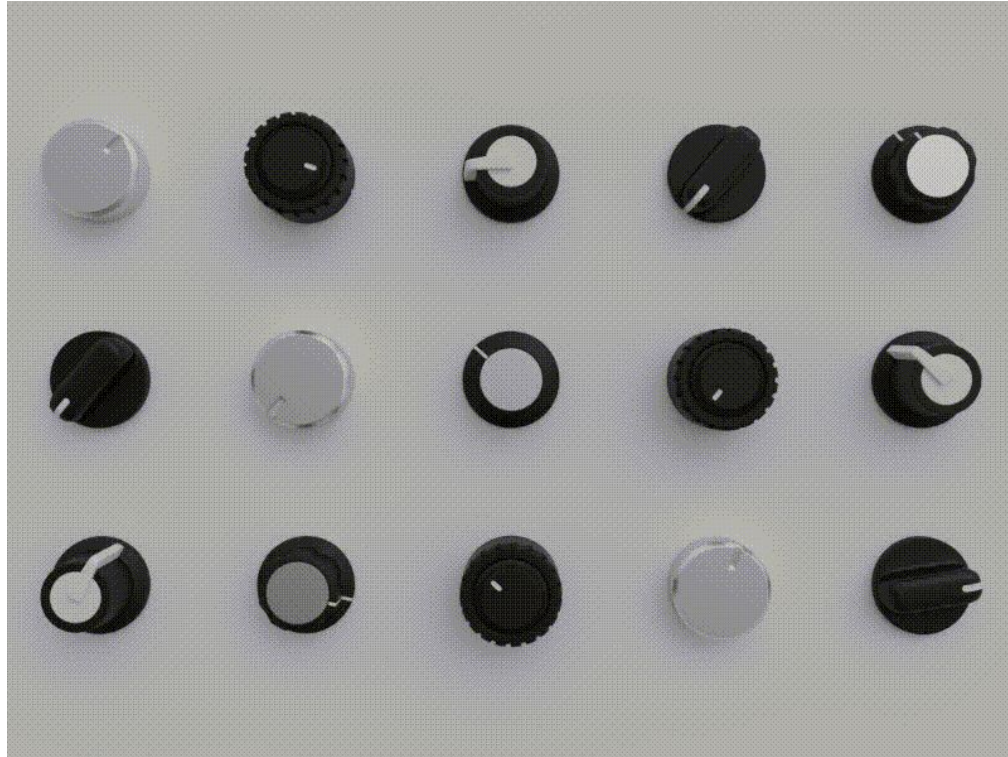
- New type of target - so, new metrics!

		Predicted		
		Negative	Positive	
Actual	Negative	True Negative (TN) 	False Positive (FP) Type I Error	Specificity $\frac{TN}{(TN + FP)}$
	Positive	False Negative (FN) Type II Error	True Positive (TP)	Recall / Sensitivity $\frac{TP}{(TP + FN)}$
		Negative Predicted Value $\frac{TN}{(TN + FN)}$	Precision $\frac{TP}{(TP + FP)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

More Models!



Model Tuning



Measuring Student Progress

1. Blog Post

- a. Due the **Tuesday** of the Second Week (Oct 18)
- b. Topic: Write a Tutorial (See next slide)

2. Canvas & Code Challenge

- a. Checkpoints
 - i. 4/4 : OOP
 - ii. 4/6: ML Fundamentals
 - iii. 4/10: Logistic Regression
 - iv. 4/12: Decision Trees
- b. Code Challenge: 4/14

3. Project

- a. Solo Project - Classification Problem
 - i. Can use a provided dataset or find your own

Measuring Student Progress: Blog #3

Write a tutorial (with a data set and code sample) on something that you think might be interesting to other people taking the course. It can be a topic we didn't cover at all, or can just go deeper into a topic that we did cover. Your tutorial can use the same tool/library from the previous blog post, or you can choose something new. Potential elements to include:

- I. An introduction explaining why a data scientist would want to do what your tutorial does
- II. A section explaining the data set you're working with (what are the features? if there is a target, what is it?)
- III. A section explaining the libraries you're working with (ideally with links to documentation and/or tutorials for someone just getting started)
- IV. Well-commented code
- V. A conclusion section that recaps what you did and why

What can you expect to **feel**
when going through Phase 3?

Feelings at the Start of Phase 3



Excited:
Let's tackle Machine Learning!

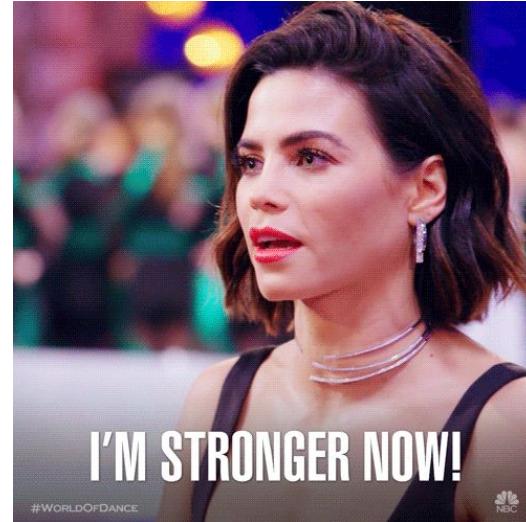


Overwhelmed:
Many new DS tools to learn

Feelings at the End of Phase 3



Stressed:
Starting to think about
capstone and end of course



Powerful:
Armed with modeling tools
to attack many different
problems

Capstone Project



- Sneaking up!
- What can you do now?
 - Explore data!
 - [UCI Machine Learning Datasets Repository](#)
 - [Kaggle Datasets](#)
 - [Awesome Datasets Repo on Github](#)
 - [New York City Open Data Portal](#)
 - [Inside AirBNB](#)
 - [Data is Plural](#)

**LET'S
DO THIS!**



KUNG FU PANDA © 2019 DREAMWORKS ANIMATION LLC. ALL RIGHTS RESERVED.