Welcome to Weedle's Cave

Pokemon Analysis

STAT 445
Justin Lew

Pokemon is a franchise based on a game for Game Boy published by Nintendo

Ability to battle other Pokemon in order to catch them all

Datasets

Two datasets: pokemon and combats

X.	Name	Type.1	Type.2	HP	Attack	Defense	Sp..Atk	Sp..Def	Speed	Generation	Legendary
1	Bulbasaur	Grass	Poison	45	49	49	65	65	45	1	False
2	Ivysaur	Grass	Poison	60	62	63	80	80	60	1	False
3	Venusaur	Grass	Poison	80	82	83	100	100	80	1	False
4	Mega Venusaur	Grass	Poison	80	100	123	122	120	80	1	False
5	Charmander	Fire		39	52	43	60	50	65	1	False
6	Charmeleon	Fire		58	64	58	80	65	80	1	False
7	Charizard	Fire	Flying	78	84	78	109	85	100	1	False

First_pokemon	Second_pokemon	Winner
266	298	298
702	701	701
191	668	668
237	683	683
151	231	151

Combats

First_pokemon	Second_pokemon	Winner
266	298	298
702	701	701
191	668	668
237	683	683
151	231	151

Taken from
http://pokemonessentials.wikia.com/wiki/Battles

Data Manipulation

First_pokemon	Second_pokemon	Winner	HP	Attack	Defense	Sp..Atk	Sp..Def	Speed	Result
266	298	298	-20	-6	10	-15	10	-19	0
702	701	701	0	-39	-18	18	39	0	0
191	668	668	-20	-35	10	-45	10	0	0
237	683	683	-37	-80	-50	10	-50	-28	0
151	231	151	50	50	-105	105	-160	50	1

New columns correspond to the difference of all numerical stats

Ie. First_pokemon HP - Second_pokemon HP

Caveats

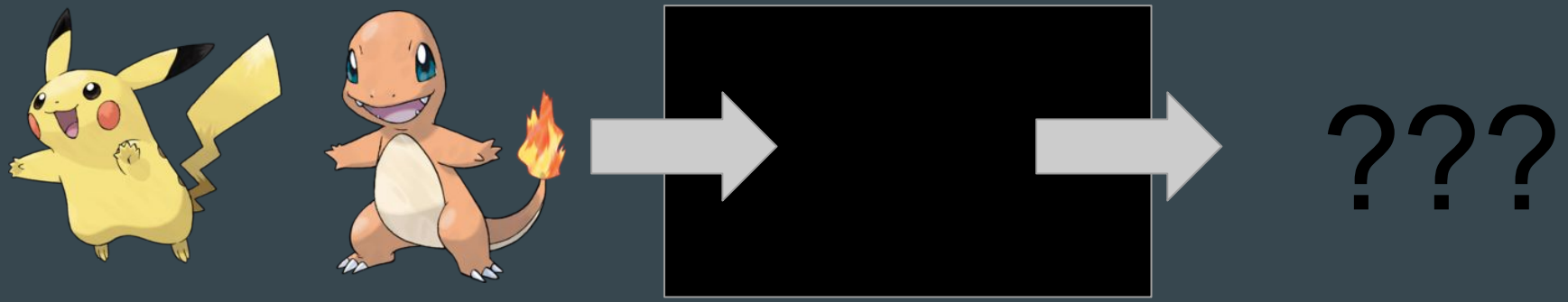
- Loss of information
  - Only know relative statistics between the two Pokemon
- Does not take into account the type of Pokemon
  - Ie. Fire versus water Pokemon
  - Water Pokemon has an advantage over fire Pokemon
- Pokemon in the “First_pokemon” column attacks first
  - Relationship between columns First_pokemon and Second_pokemon are not symmetric

Objective

Is there any way to predict the winners of future combats?

Classification!

We are going to go through Logistic regression and kNN classifiers.



Logistic Regression

Formula: **Result ~ HP + Attack + Defense + SP..Atk + SP..Def + Speed**

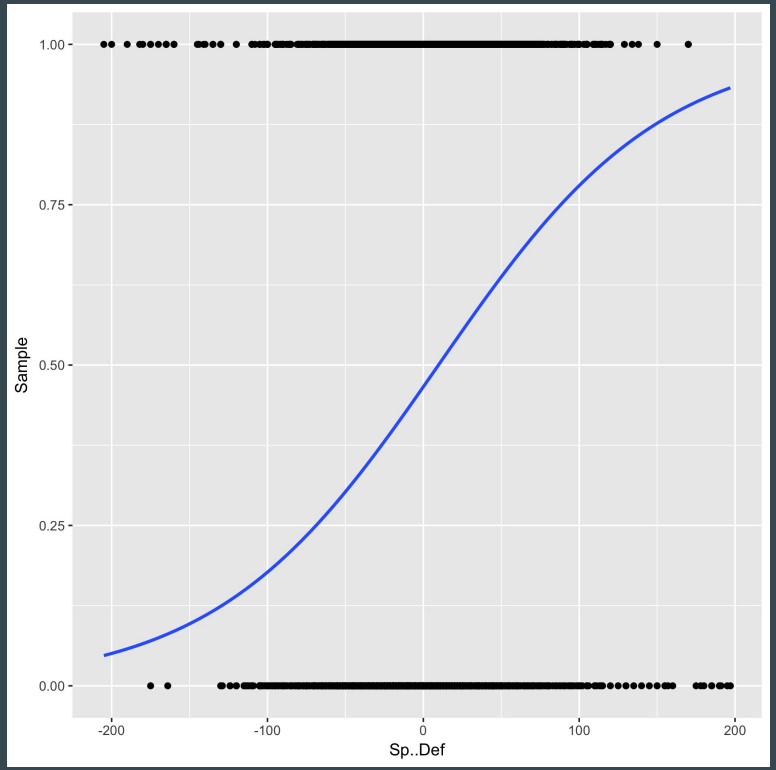
Result variable is binary (can take values 0 or 1)

Result = 1 if the first Pokemon wins (second Pokemon losses)

Result = 0 if the first Pokemon losses (second Pokemon wins)

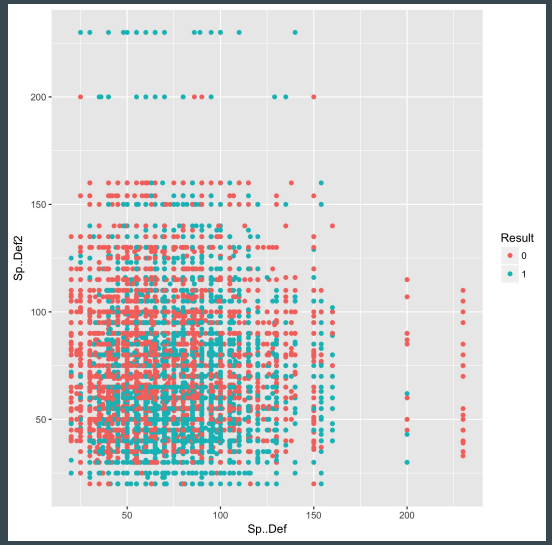
Conclusion: Can predict combats in the test dataset with an 89% accuracy!

Logistic Regression



k-Nearest Neighbors Algorithm

Set k = 2



Conclusion: Can predict test battles with an 86% accuracy!

Further Research

Include actual statistics of BOTH Pokemon rather than take its difference

Principle Component Analysis

Take into account the categorical variables (ie. Types of Pokemon and Legendary)

Fit a neural network classifier... and possibly SVMs