Homework 0

CS 5785: Applied Machine Learning

Ryan Farr

rlf238

August 30th, 2018

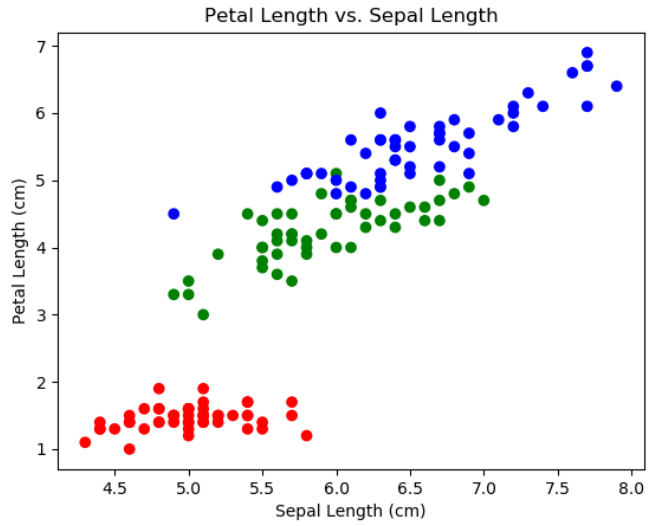
Writeup

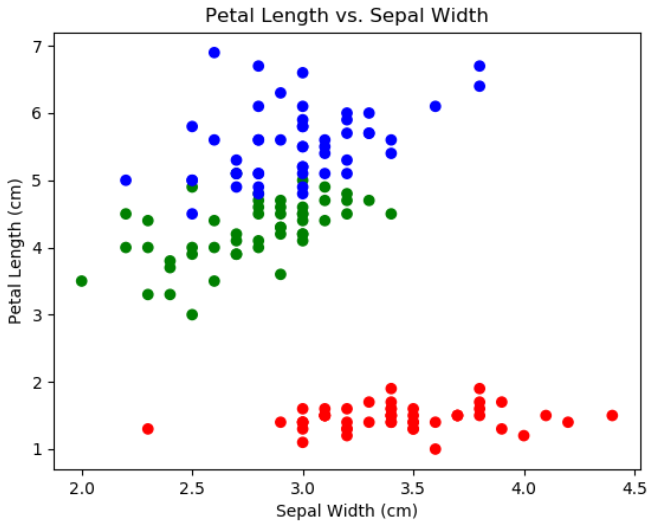
This assignment required that we create a program that reads the Iris dataset, then displays this data as a set of six plots. The data contains 150 samples, each with four attributes (sepal length, sepal width, petal length, petal width) and a label (the species). There are three species (each with 50 samples): Iris Setosa, Iris Versicolor, and Iris Virginica.

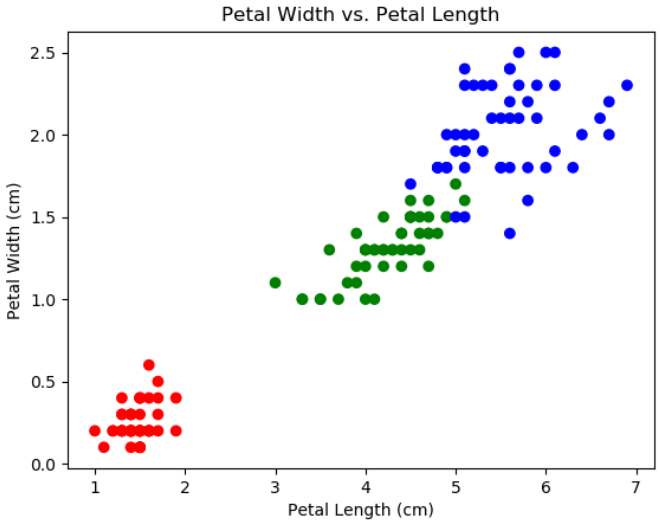
It's almost immediately obvious from the plots that each species, particularly Setosa, are fairly unique in nearly every feature. It's easy to see how something like linear regression would be able to separate the data with relatively low error, though with K-nearest neighbors you would likely run into problems near the boundaries of Versicolor and Virginica.

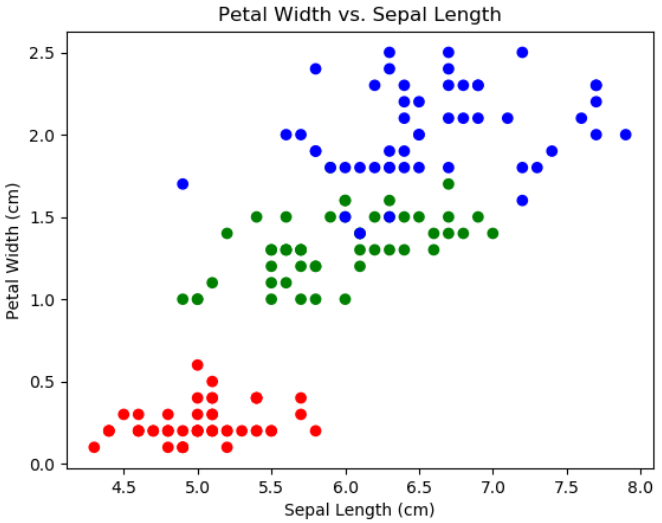
Figures

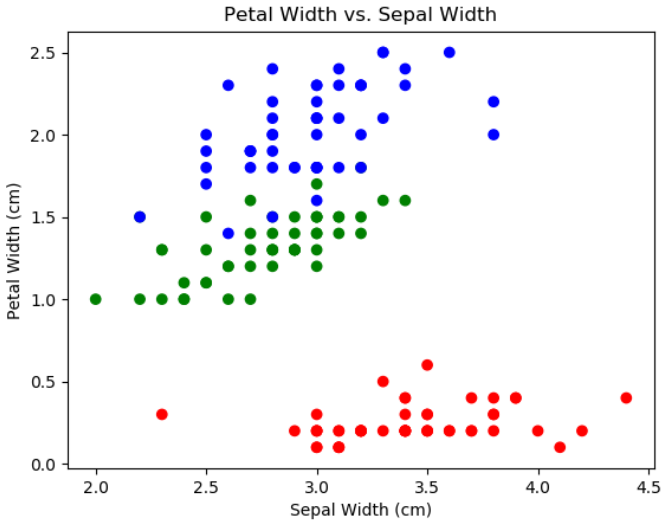
Figures were generated with generic code that can be found in *main.py*. In the plots below, red points represent Iris Setosa, green are Iris Versicolor, and blue are Iris Virginica.

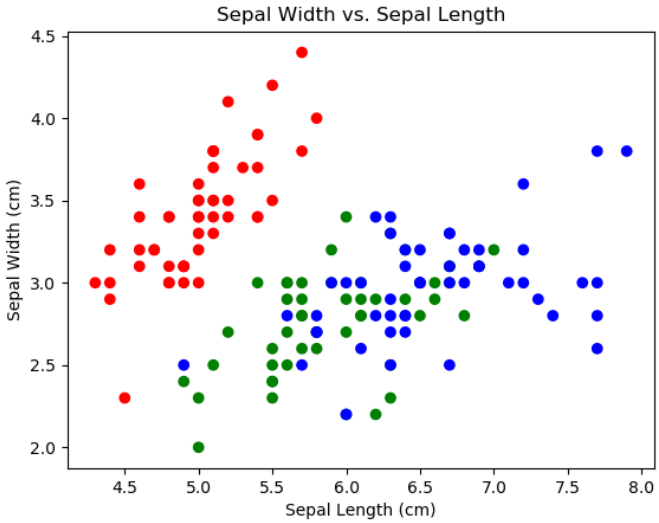









Petal Width vs. Sepal Width
Petal Width (cm)
Sepal Width (cm)
0.0
0.5
1.0
1.5
2.0
2.5
2.0
2.5
3.0
3.5
4.0
4.5


Sepal Width vs. Sepal Length
Sepal Width (cm)
Sepal Length (cm)
2.0
2.5
3.0
3.5
4.0
4.5
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0