|  |
| --- |
| CS5785 |
| Applied Machine Learning |
| Homework 0 Report |

|  |
| --- |
| Hao Geng, hg457  9-3-2021 |

**General summary**

In this homework, following the instruction, I

* set up python environment with Anaconda,
* download and parse the iris.data file and store them into array using
  + python function open(), list.append(), string.split(),
  + numpy
* plot 2-dimensional scatter plots from all pairs of two attributes using
  + matplotlib. pyplot
  + numpy array transpose .T

The generated figure looks like the sample figure from Wikipedia.

**Code repo**

<https://github.com/Jamie-Jay/AML-2021Fall/tree/main/hw0>

**Questions**

The iris.names file describes the structure of the dataset.

1. **How many features/attributes are there per sample?**

4

Attribute Information:

1. sepal length in cm

2. sepal width in cm

3. petal length in cm

4. petal width in cm

1. **How many different species are there?**

3

-- Iris Setosa

-- Iris Versicolour

-- Iris Virginica

1. **How many samples of each species did Anderson record?**

50 in each of three classes

**Result**

The parsed data are printed in the console. Please run the code to see the result.

The generated figure is as the following, which is also saved as *plot.png*:

