CS5785 Applied Machine Learning

HW0 Writeup

Andrea Qiu (zq64@cornell.edu),

Roger Wang (rw575@cornell.edu)

## Procedures

1. Download the dataset from <https://archive.ics.uci.edu/ml/machine-learning-databases/iris/> The dataset has 150 lines representing 150 samples. There are three species of flowers: setosa, versicolor, and virginica. Each flower has 50 samples. Each line of data is separated by comma. The first column is sepal length in cm. The second column is sepal width in cm. The third column is petal length in cm. The fourth column is petal width in cm. The fifth column is the flower’s specie.
2. The program “hw0.py” opens and reads from the file “iris.data”. Splits each line by comma and stores the data into a 150 × 4 list. 150 represents 150 samples. 4 represents 4 attributes. The specie name is stored into a separate list.
3. Read lots of documentations and tutorials for using “numpy” and “pyplot”.

https://matplotlib.org/tutorials/index.html

<http://cs231n.github.io/python-numpy-tutorial/>

1. Draw six graphs and save them separately as six files. Add legend, x-axis, and y-axis. Add a specific color for each specie. (red for setosa, blue for versicolor, and green for virginica)

## Results

## Insights

* Among all three species, setosa’s petal width and length is the smallest among three species, while virginica’s petal width and length is the largest in average.
* Among all three species, setosa’s sepal width is the largest in average, while virginica’s sepal width is the smallest in average. Setosa’s sepal length is the shortest in average, while virginica’s sepal width is the largest in average.
* Among all three species, versicolor the largest range in sepal length and petal width. Setosa has the smallest range in petal width, petal length and sepal length.
* For each specie, there is a strong positive linear relationship between sepal width and sepal length.
* For versicolor and virginica, there is a weak positive linear relationship between petal length and sepal length. For setosa, there is no obvious relationship between petal length and sepal length.
* For each specie, there is no obvious relationship between petal width and sepal length.
* For versicolor and virginica, there is a weak positive linear relationship between petal length and sepal width. For setosa, there is no obvious relationship between petal length and sepal width.
* For versicolor and virginica, there is a weak positive linear relationship between petal width and sepal length. For setosa, there is no obvious relationship between petal width and sepal length.
* For each specie, there is a strong positive linear relationship between petal width and petal length.