CORNELL TECH

Course: CS5785 Applied Machine Learning
Homework No.: 0
Name: Ray Su / Zhiyuan Jia
NetID: zs349 / zj74

Summary

The Iris Flowers dataset is a classic dataset that has been wildly referred by many papers. This homework makes it into graphs to get a visual insight of the patterns in the data.

Steps

1. Getting the data from http://archive.ics.uci.edu/ml/datasets.html as txt
2. Analyze the classes and attributes of flowers
3. Read the data into program
4. Convert numbers into diagrams
5. Obtain insight based on diagrams

Data Information

3 species:

1. Iris Setosa, 50 samples (Red spot in diagrams)
2. Iris Versicolour, 50 samples (Blue spot in diagrams)
3. Iris Virginica, 50 samples (Green spot in diagrams)

4 attributes:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm

Insights about Diagrams

We drew the following diagrams and obtained the insights below using the python code attached (also copied at the end of this write-up).

1. In all diagrams, spots representing Setosa almost locate very far away from green and blue spots, so it would be very easy using lazing learning method to locate if a new spot represents Setosa or not.
2. Three different color spots are rather compact in diagrams 4, diagram 5 and diagram 6, among which, diagram 4 and diagram 6 are more compact. Therefore, it would be better if we chose the training data from these to diagrams.
3. The boundaries of these three groups of spot are distinct in diagram 4, diagram 5 and diagram 6.
4. Finally, there are so few abnormal spots in diagram 6, so the comparison between petal length and petal width can be a great instruction for deciding to which class of flower the test data belongs.

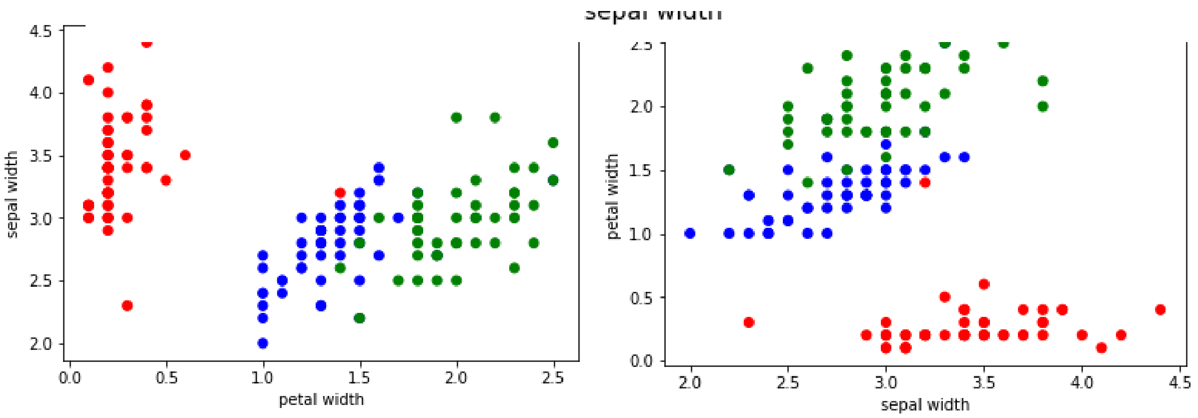

Diagram1 Correlation between sepal width and petal width

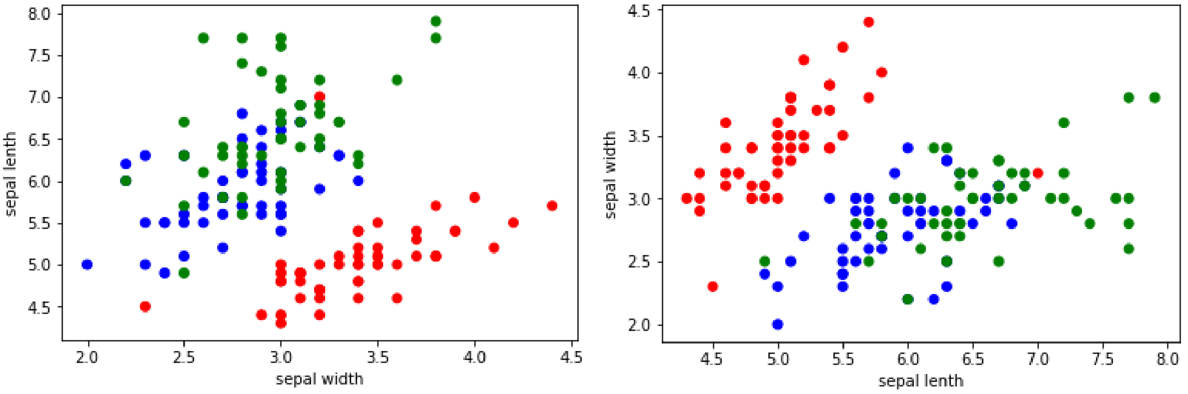

Diagram2 Correlation between sepal width and sepal length

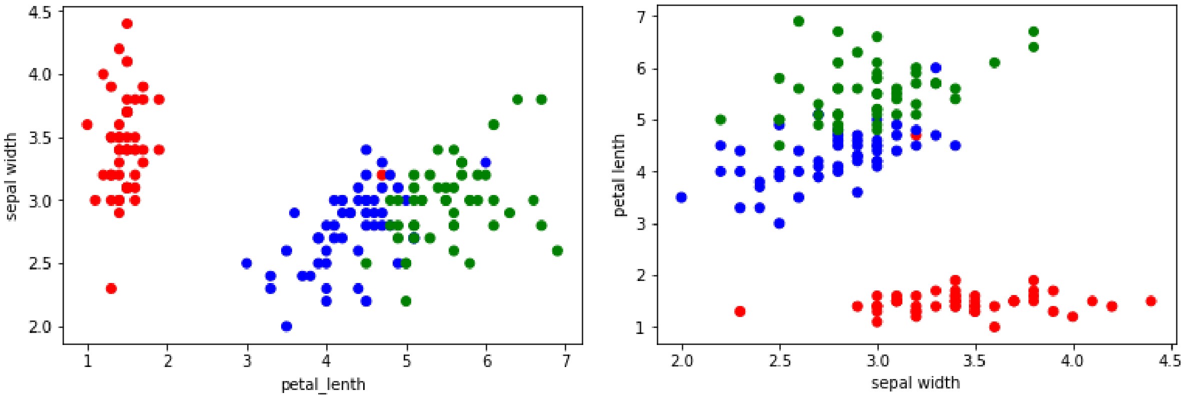

Diagram 3 Correlation between sepal width and petal length

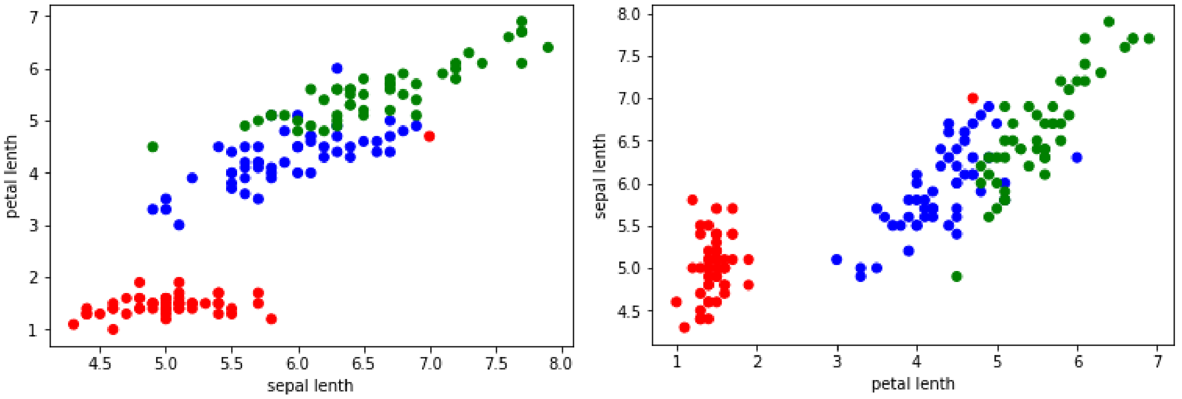

Diagram 4 Correlation between sepal length and petal length

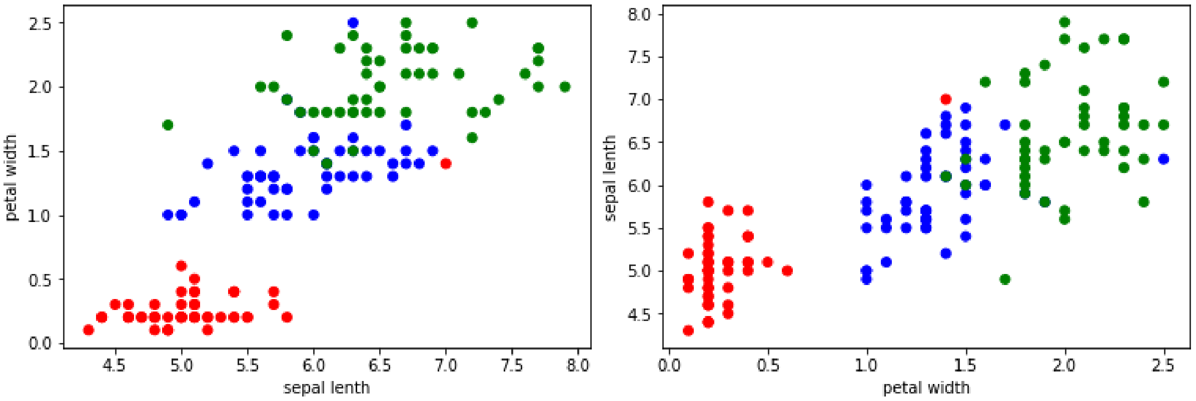

Diagram 5 Correlation between sepal length and petal width

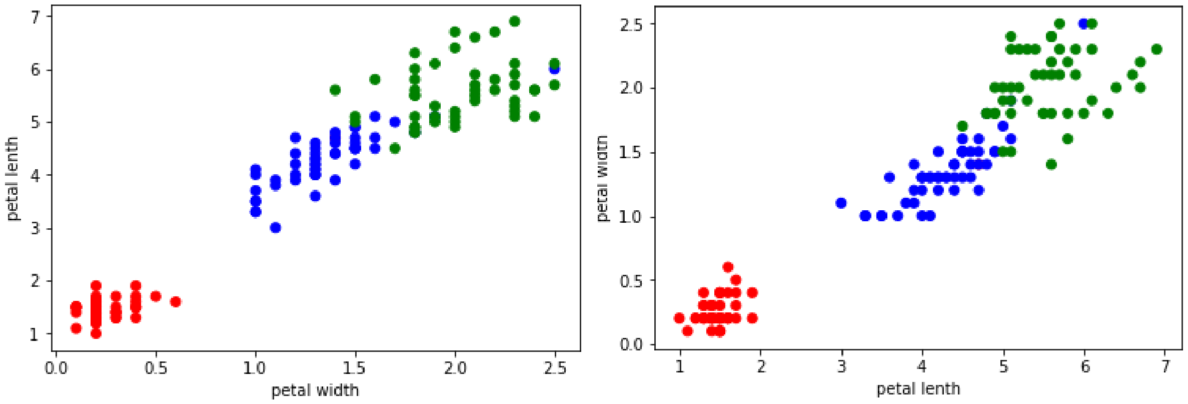

Diagram 6 Correlation between petal length and petal width

Python code:

```
from matplotlib import pyplot as plt
import sys,os
import pandas as pd
def cur_file_dir():
        path = sys.path[0]
        return os.path.abspath(path)
rpath = cur_file_dir()+'\\iris-data.txt'
lines = []

df = pd.read_csv(rpath, header=None)

i = 0
sepal_lenth = []
sepal_width = []
petal_lenth = []
petal_width = []
```

```
while i < 150:
        sepal_lenth.append(round(df[0][i],2))
        sepal_width.append(round(df[1][i],2))
        petal_lenth.append(round(df[2][i],2))
        petal_width.append(round(df[3][i],2))
        i = i+1


color    = []
for i in range(51):
        color.append('r')
for i in range(51):
        color.append('b')
for i in range(51):
        color.append('g')

plt.figure(1)
plt . scatter (sepal_width,sepal_lenth,c = color)
plt.xlabel('sepal width')
plt.ylabel('sepal lenth')
plt . savefig (" plot .png")

plt.figure(2)
plt . scatter (sepal_lenth,sepal_width,c = color)
plt.xlabel('sepal lenth')
plt.ylabel('sepal width')
plt . savefig (" plot .png")


plt.figure(3)
plt . scatter (sepal_width,petal_width,c = color)
plt.xlabel('sepal width')
plt.ylabel('petal width')
plt . savefig (" plot .png")

plt.figure(4)
plt . scatter (petal_width,sepal_width,c = color)
plt.xlabel('petal width')
plt.ylabel('sepal width')
plt . savefig (" plot .png")

plt.figure(5)
plt . scatter (sepal_width,petal_lenth,c = color)
plt.xlabel('sepal width')
plt.ylabel('petal lenth')
plt . savefig (" plot .png")
```

```
plt.figure(6)
plt . scatter (petal_lenth,sepal_width,c = color)
plt.xlabel('petal_lenth')
plt.ylabel('sepal width')
plt . savefig (" plot .png")


plt.figure(7)
plt . scatter (sepal_lenth,petal_lenth,c = color)
plt.xlabel('sepal lenth')
plt.ylabel('petal lenth')
plt . savefig (" plot .png")


plt.figure(8)
plt . scatter (petal_lenth,sepal_lenth,c = color)
plt.xlabel('petal lenth')
plt.ylabel('sepal lenth')
plt . savefig (" plot .png")

plt.figure(9)
plt . scatter (sepal_lenth,petal_width,c = color)
plt.xlabel('sepal lenth')
plt.ylabel('petal width')
plt . savefig (" plot .png")

plt.figure(10)
plt . scatter (petal_width,sepal_lenth,c = color)
plt.xlabel('petal width')
plt.ylabel('sepal lenth')
plt . savefig (" plot .png")

plt.figure(11)
plt . scatter (petal_lenth,petal_width,c = color)
plt.xlabel('petal lenth')
plt.ylabel('petal width')
plt . savefig (" plot .png")

plt.figure(12)
plt . scatter (petal_width,petal_lenth,c = color)
plt.xlabel('petal width')
plt.ylabel('petal lenth')
plt . savefig (" plot .png")
```