CS 5785 Applied Machine Learning

Homework 0

Team Member: Queenie Liu, Xueqi Wei

1. How many features/attributes are there per sample? How many different species are there, and how many samples of each species did Anderson record?

   According to iris.names,

- There are 4 numeric attributes and the class per sample.
- There are 3 different species.
- Anderson recorded 50 in each of three classes, which were 150 samples in total.

2.

   For problem2, we use read_csv in pandas to read the original data file, get rid of the last column and turn the dataframe to an array of 150*4. Additionally, we use the last column to create a 150-dimensional vector containing each sample's label. Below is what we got:

```
import pandas as pd
iris = pd.read_csv('iris.data', delimiter = ',',header=None)
print(iris)
iris1 = pd.read_csv('iris.data', delimiter = ',', usecols = [0,1,2,3],header=None)
d = iris1.values.tolist()
d1 = d[:-1]
print (d1)
iris2 = pd.read_csv('iris.data', delimiter = ',',usecols = [4], header = None)
data = iris2.values.tolist()
data1 = data[:-1]
print(data1)
```

N*p array:

```
[150 rows x 5 columns]
[[5.1, 3.5, 1.4, 0.2], [4.9, 3.0, 1.4, 0.2], [4.7, 3.2, 1.3, 0.2], [4.6, 3.1, 1.5, 0.2], [5.0, 3.6, 1.4, 0.2], [5.4, 3.9, 1.7, 0.4], [4.6,
3.4, 1.4, 0.3], [5.0, 3.4, 1.5, 0.2], [4.4, 2.9, 1.4, 0.2], [4.9, 3.1, 1.5, 0.1], [5.4, 3.7, 1.5, 0.2], [4.8, 3.4, 1.6, 0.2], [4.8, 3.0,
1.4, 0.1], [4.3, 3.0, 1.1, 0.1], [5.8, 4.0, 1.2, 0.2], [5.7, 4.4, 1.5, 0.4], [5.4, 3.9, 1.3, 0.4], [5.1, 3.5, 1.4, 0.3], [5.7, 3.8, 1.7,
0.3], [5.1, 3.8, 1.5, 0.3], [5.4, 3.4, 1.7, 0.2], [5.1, 3.7, 1.5, 0.4], [4.6, 3.6, 1.0, 0.2], [5.1, 3.3, 1.7, 0.5], [4.8, 3.4, 1.9, 0.2],
[5.0, 3.0, 1.6, 0.2], [5.0, 3.4, 1.6, 0.4], [5.2, 3.5, 1.5, 0.2], [5.2, 3.4, 1.4, 0.2], [4.7, 3.2, 1.6, 0.2], [4.8, 3.1, 1.6, 0.2], [5.4,
3.4, 1.5, 0.4], [5.2, 4.1, 1.5, 0.1], [5.5, 4.2, 1.4, 0.2], [4.9, 3.1, 1.5, 0.1], [5.0, 3.2, 1.2, 0.2], [5.5, 3.5, 1.3, 0.2], [4.9, 3.1,
1.5, 0.1], [4.4, 3.0, 1.3, 0.2], [5.1, 3.4, 1.5, 0.2], [5.0, 3.5, 1.3, 0.3], [4.5, 2.3, 1.3, 0.3], [4.4, 3.2, 1.3, 0.2], [5.0, 3.5, 1.6,
0.6], [5.1, 3.8, 1.9, 0.4], [4.8, 3.0, 1.4, 0.3], [5.1, 3.8, 1.6, 0.2], [4.6, 3.2, 1.4, 0.2], [5.3, 3.7, 1.5, 0.2], [5.0, 3.3, 1.4, 0.2],
[7.0, 3.2, 4.7, 1.4], [6.4, 3.2, 4.5, 1.5], [6.9, 3.1, 4.9, 1.5], [5.5, 2.3, 4.0, 1.3], [6.5, 2.8, 4.6, 1.5], [5.7, 2.8, 4.5, 1.3], [6.3,
3.3, 4.7, 1.6], [4.9, 2.4, 3.3, 1.0], [6.6, 2.9, 4.6, 1.3], [5.2, 2.7, 3.9, 1.4], [5.0, 2.0, 3.5, 1.0], [5.9, 3.0, 4.2, 1.5], [6.0, 2.2,
4.0, 1.0], [6.1, 2.9, 4.7, 1.4], [5.6, 2.9, 3.6, 1.3], [6.7, 3.1, 4.4, 1.4], [5.6, 3.0, 4.5, 1.5], [5.8, 2.7, 4.1, 1.0], [6.2, 2.2, 4.5,
1.5], [5.6, 2.5, 3.9, 1.1], [5.9, 3.2, 4.8, 1.8], [6.1, 2.8, 4.0, 1.3], [6.3, 2.5, 4.9, 1.5], [6.1, 2.8, 4.7, 1.2], [6.4, 2.9, 4.3, 1.3],
[6.6, 3.0, 4.4, 1.4], [6.8, 2.8, 4.8, 1.4], [6.7, 3.0, 5.0, 1.7], [6.0, 2.9, 4.5, 1.5], [5.7, 2.6, 3.5, 1.0], [5.5, 2.4, 3.8, 1.1], [5.5,
2.4, 3.7, 1.0], [5.8, 2.7, 3.9, 1.2], [6.0, 2.7, 5.1, 1.6], [5.4, 3.0, 4.5, 1.5], [6.0, 3.4, 4.5, 1.6], [6.7, 3.1, 4.7, 1.5], [6.3, 2.3,
4.4, 1.3], [5.6, 3.0, 4.1, 1.3], [5.5, 2.5, 4.0, 1.3], [5.5, 2.6, 4.4, 1.2], [6.1, 3.0, 4.6, 1.4], [5.8, 2.6, 4.0, 1.2], [5.0, 2.3, 3.3,
1.0], [5.6, 2.7, 4.2, 1.3], [5.7, 3.0, 4.2, 1.2], [5.7, 2.9, 4.2, 1.3], [6.2, 2.9, 4.3, 1.3], [5.1, 2.5, 3.0, 1.1], [5.7, 2.8, 4.1, 1.3],
[6.3, 3.3, 6.0, 2.5], [5.8, 2.7, 5.1, 1.9], [7.1, 3.0, 5.9, 2.1], [6.3, 2.9, 5.6, 1.8], [6.5, 3.0, 5.8, 2.2], [7.6, 3.0, 6.6, 2.1], [4.9,
2.5, 4.5, 1.7], [7.3, 2.9, 6.3, 1.8], [6.7, 2.5, 5.8, 1.8], [7.2, 3.6, 6.1, 2.5], [6.5, 3.2, 5.1, 2.0], [6.4, 2.7, 5.3, 1.9], [6.8, 3.0,
5.5, 2.1], [5.7, 2.5, 5.0, 2.0], [5.8, 2.8, 5.1, 2.4], [6.4, 3.2, 5.3, 2.3], [6.5, 3.0, 5.5, 1.8], [7.7, 3.8, 6.7, 2.2], [7.7, 2.6, 6.9,
2.3], [6.0, 2.2, 5.0, 1.5], [6.9, 3.2, 5.7, 2.3], [5.6, 2.8, 4.9, 2.0], [7.7, 2.8, 6.7, 2.0], [6.3, 2.7, 4.9, 1.8], [6.7, 3.3, 5.7, 2.1],
[7.2, 3.2, 6.0, 1.8], [6.2, 2.8, 4.8, 1.8], [6.1, 3.0, 4.9, 1.8], [6.4, 2.8, 5.6, 2.1], [7.2, 3.0, 5.8, 1.6], [7.4, 2.8, 6.1, 1.9], [7.9,
3.8, 6.4, 2.0], [6.4, 2.8, 5.6, 2.2], [6.3, 2.8, 5.1, 1.5], [6.1, 2.6, 5.6, 1.4], [7.7, 3.0, 6.1, 2.3], [6.3, 3.4, 5.6, 2.4], [6.4, 3.1,
5.5, 1.8], [6.0, 3.0, 4.8, 1.8], [6.9, 3.1, 5.4, 2.1], [6.7, 3.1, 5.6, 2.4], [6.9, 3.1, 5.1, 2.3], [5.8, 2.7, 5.1, 1.9], [6.8, 3.2, 5.9,
2.3], [6.7, 3.3, 5.7, 2.5], [6.7, 3.0, 5.2, 2.3], [6.3, 2.5, 5.0, 1.9], [6.5, 3.0, 5.2, 2.0], [6.2, 3.4, 5.4, 2.3]]
```

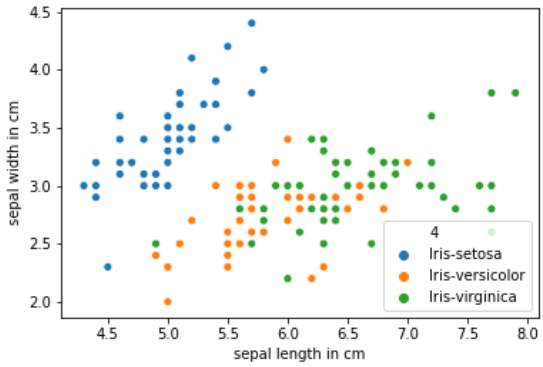
N-dimensional vector:

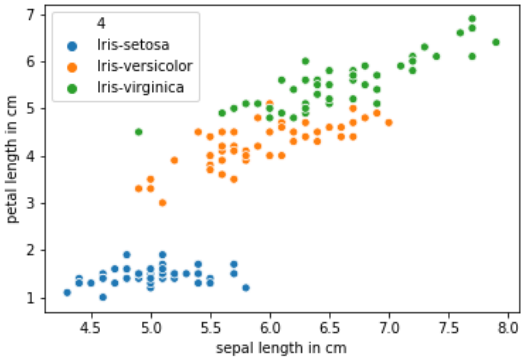
```
[['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'],
['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'],
['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'],
['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'],
['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'],
['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'], ['Iris-setosa'],
['Iris-setosa'], ['Iris-setosa'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'],
['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-
versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-
versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-
versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-
versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-
versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-
versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-
versicolor'], ['Iris-versicolor'], ['Iris-versicolor'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'],
['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-
virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'],
['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-
virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'],
['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-
virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'],
['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica'], ['Iris-virginica']]
```

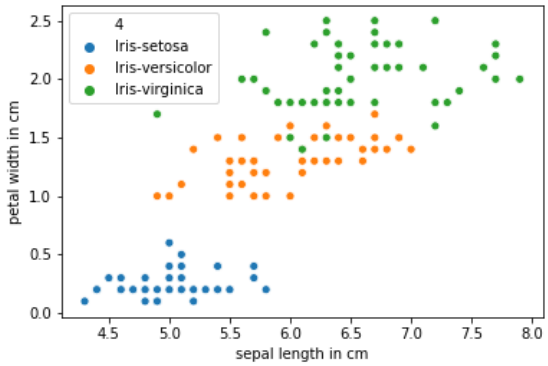
3.

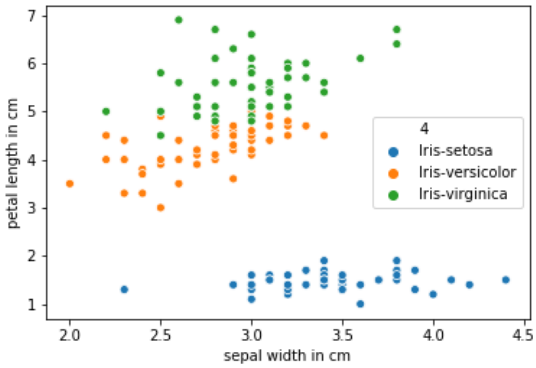
We use seaborn to build the following 6 scatterplots. We found that Iris-setosa is very different from the other two species because the blue dots which stand for Iris-setosa are distributed away from the green ones and the oranges ones while the other two species have similar distribution according to the plots.

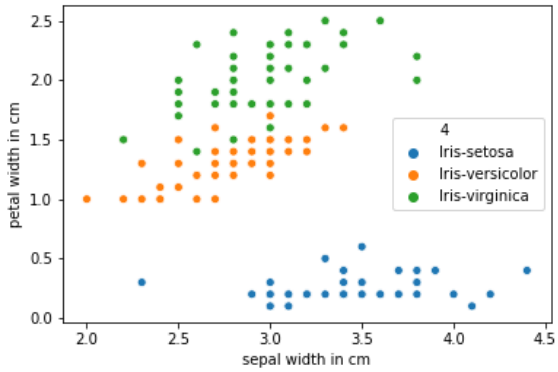
```
import seaborn as sns
import matplotlib.pyplot as plt
plt.figure()
sns.scatterplot(x=iris[0], y = iris[1], hue = iris[4])
plt.xlabel('sepal length in cm')
plt.ylabel('sepal width in cm')
plt.savefig("plot1.png")

plt.figure()
sns.scatterplot(x=iris[0], y = iris[2], hue = iris[4])
plt.xlabel('sepal length in cm')
plt.ylabel('petal length in cm')
plt.savefig("plot2.png")

plt.figure()
sns.scatterplot(x=iris[0], y = iris[3], hue = iris[4])
plt.xlabel('sepal length in cm')
plt.ylabel('petal width in cm')
plt.savefig("plot3.png")

plt.figure()
sns.scatterplot(x=iris[1], y = iris[2], hue = iris[4])
plt.xlabel('sepal width in cm')
plt.ylabel('petal length in cm')
plt.savefig("plot4.png")

plt.figure()
sns.scatterplot(x=iris[1], y = iris[3], hue = iris[4])
plt.xlabel('sepal width in cm')
plt.ylabel('petal width in cm')
plt.savefig("plot5.png")

plt.figure()
sns.scatterplot(x=iris[2], y = iris[3], hue = iris[4])
plt.xlabel('petal length in cm')
plt.ylabel('petal width in cm')
plt.savefig("plot6.png")
```

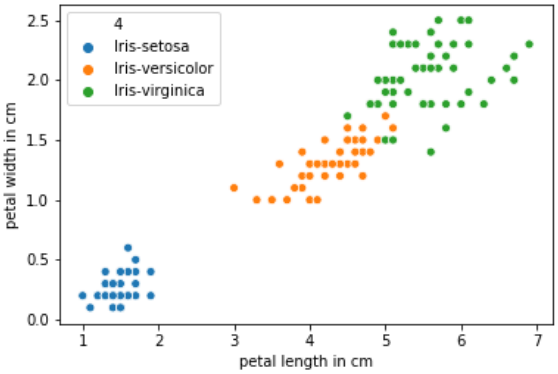









petal width in cm
sepal width in cm
4
Iris-setosa
Iris-versicolor
Iris-virginica


petal width in cm
petal length in cm
4
Iris-setosa
Iris-versicolor
Iris-virginica