Lab 2. Hello World AI!

Yongjae Yoo, Ph. D.
Assistant Professor
Department of Artificial Intelligence
Hanyang University ERICA

한양대학교 ERICA
Education Research Industry Cluster @ Ansan

Today’s Agenda

- Operator, calculation, and Playing with the data.
- CSV basics.

- AI “Hello World!”
  - How computers understand, organize, and present data
  - Math and computing – basics!

Setting Up Data

- So-called Dataset
  - A data set is any collection of data.
    - An array
    - A complete database
    - A few tuples,
    - Etc.

ID	Date	Category	Item name	Price level
01	07/01/07	Milk	Cream Milk	Moderate
01	07/01/07	Milk	Cream Milk	Moderate
01	07/01/07	Bread	French Toast	Low
01	17/01/07	Milk	Cream Milk	Moderate
01	17/01/07	Milk	Cream Milk	Moderate
01	17/01/07	Bread	French Toast	Low
01	17/01/07	Bread	French Toast	Low
02	05/01/07	Milk	Skim Milk	Moderate
02	05/01/07	Milk	Skim Milk	Moderate
02	05/01/07	Milk	Skim Milk	Moderate
02	05/01/07	Bread	French Toast	Low

Id	SepalLeng	SepalWid	PetalLeng	PetalWidt	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5	3.4	1.5	0.2	Iris-setosa
9	4.4	2.9	1.4	0.2	Iris-setosa
10	4.9	3.1	1.5	0.1	Iris-setosa
11	5.4	3.7	1.5	0.2	Iris-setosa
12	4.8	3.4	1.6	0.2	Iris-setosa
13	4.8	3	1.4	0.1	Iris-setosa

Let’s Build An Array

- By put in the data manually :)

- Let’s put in
  - X = [3, 2, -1, 2, 4]
  - Y = [1,-1, 2, 2, 1]

- Find the multiplication of odd indexed elements (1^{st}, 3^{rd})
  - – note that the index starts 0^{th}.

Let’s Build An Array

- Merge X and Y to make a 2D array “xy”.

- And, draw a scatter plot by trying the following code
  - We will practice it more intensively tomorrow.

```
# Importing the required modules
import matplotlib.pyplot as plt
import numpy as np
# Creating a new figure
plt.figure(dpi=100)
# Creating array from list
plt.scatter(xy)
# Adding details to the plot
plt.title('Scatter Plot of a NumPy Array’)
plt.xlabel('x-axis’)
plt.ylabel('y-axis’)
# Displaying the plot
plt.grid()
plt.show()
```

Playing With The Dataset

- Often the dataset is provided by CSV format.
  - (Comma separated variables)
  - You can also take a look by Excel.

- Not to put in tons of data manually.

Loading A CSV File

- First, let's load the data and print it to the screen

```
f = open("FILENAME.csv", "r")
reader = csv.reader(f)
for row in reader:
  print(row)
```

CSV to List

- We may store each lines.

```
import csv
f = open("data.csv", "r")
reader = csv.reader(f)
# for row in reader:
# print(row)
data = list(reader)
# print(data)
```

- (Of course, you can also use Pandas library if you are familiar with it.)

Then, Reshape!

- The data what I gave you contains:
  - 1st row: header and variable name.
  - 2nd to end row: type, sepal length and width (sl, sw), petal length and width (pl, pw).

- You can extract each feature by using [index] in each line in the list.
  - Hint: use iteration. (for)

Type	sl	sw	pl	pw
Setosa	5	3	1.6	0.2
Setosa	5	3.4	1.6	0.4
Setosa	5.2	3.5	1.5	0.2
Setosa	5.2	3.4	1.4	0.2
Setosa	4.7	3.2	1.6	0.2
Setosa	4.8	3.1	1.6	0.2
Setosa	5.4	3.4	1.5	0.4
Setosa	5.2	4.1	1.5	0.1
Setosa	5.5	4.2	1.4	0.2
Setosa	4.9	3.1	1.5	0.1
Setosa	5	3.2	1.2	0.2
Setosa	5.5	3.5	1.3	0.2
Setosa	4.9	3.1	1.5	0.1
Setosa	4.4	3	1.3	0.2
Setosa	5.1	3.4	1.5	0.2
Setosa	5	3.5	1.3	0.3
Setosa	4.5	2.3	1.3	0.3
Setosa	4.4	3.2	1.3	0.2
Setosa	5	3.5	1.6	0.6
Setosa	5.1	3.8	1.9	0.4
Setosa	4.8	3	1.4	0.3
Setosa	5.1	3.8	1.6	0.2
Setosa	4.6	3.2	1.4	0.2
Setosa	5.3	3.7	1.5	0.2
Setosa	5	3.3	1.4	0.2
Versicolor	6.6	3	4.4	1.4
Versicolor	6.8	2.8	4.8	1.4
Versicolor	6.7	3	5	1.7
Versicolor	6	2.9	4.5	1.5
Versicolor	5.7	2.6	3.5	1
Versicolor	5.5	2.4	3.8	1.1
Versicolor	5.5	2.4	3.7	1
Versicolor	5.8	2.7	3.9	1.2
Versicolor	6	2.7	5.1	1.6
Versicolor	5.4	3	4.5	1.5
Versicolor	6	3.4	4.5	1.6
Versicolor	6.7	3.1	4.7	1.5
Versicolor	6.3	2.3	4.4	1.3
Versicolor	5.6	3	4.1	1.3
Versicolor	5.5	2.5	4	1.3
Versicolor	5.5	2.6	4.4	1.2
Versicolor	6.1	3	4.6	1.4
Versicolor	5.8	2.6	4	1.2
Versicolor	5	2.3	3.3	1
Versicolor	5.6	2.7	4.2	1.3
Versicolor	5.7	3	4.2	1.2
Versicolor	5.7	2.9	4.2	1.3

Lab Exercise

- Load the IRIS dataset (IRIS.csv)

- Extract each species' data, by reshape the array by 25 x 4 x 3
  - (or 3x4x25. Up to you).
  - Data should be put in the depth array in 3D in species-wise.
  - Four variables (sl, sw, pl, pw) organized in columns.
  - 25 observation data should be in the rows.

Lab Exercise

- Then,
  - Find the average of sl, sw, pl, pw for each species.
  - Find the maximum and minimum of sl, sw, pl, pw for each.

  - Draw conclusions based on the basic observations.

Further Exercise – Advanced Students

- Consider “how” we treat the data again. We will treat a little bit larger data this time.

- Load the Pizza.csv example. You can see tons of data more.
  - Do the same things to the IRIS
    - Reshaping to organize the data variable(feature)-wise.
    - Find the average, maximum, minimum of variables for each pizza types.
    - Draw scatter plots for each pizza’s each variables.
    - Draw conclusions from your observation.