Worksheet 2

This worksheet is due Tuesday of Week 2, by 11:59pm.

You are encouraged to work in a group of up to 3 total students (working alone is also okay). Everyone should make a submission on Canvas, but it's fine for all group members to submit the same thing.

See Worksheet 1 for instructions on how to make a Education workspace (which will provide higher-priority access to Deepnote resources).

Loading a NumPy array from a pickle file

- Import `numpy` using the usual abbreviation `np`, and also import the Python `pickle` module (no abbreviation).
- Notice that there is an attached file `wkst2-starter.pickle`. Load the data from that file and store it in the variable `arr` using the following code.

```
with open("wkst2-starter.pickle", "rb") as f:
    arr = pickle.load(f)
```

- Check the type of this object using the function `type` and check its dimensions by computing the `shape` attribute (no parentheses after `shape`). The variable `arr` should represent an 8000 by 7 NumPy array.

In [1]:
```
import numpy as np
import pickle

with open("../Data/wkst2-starter.pickle", "rb") as f:
    arr = pickle.load(f)

print(type(arr))
print(arr.shape)

print(arr[:,4])
```

```
<class 'numpy.ndarray'>
(8000, 7)
[29 54 11 ... 70  6  3]
```

Practice with Boolean indexing

- Define the variable `x` to be the column at index `4`.
- Check that `x` is a length-8000 NumPy array.

- How many values in `x` are equal to `10`?
- Define `arrsub` to be the sub-array of `arr` containing all the rows for which the column at index `4` is equal to the integer `10`. Use Boolean indexing with `arr` and `x`.
- What is the shape of `arrsub`? Use the `shape` attribute again. (It should be `7` columns, but how many rows?)
- Define a new variable, `a`, to be equal to the number of rows of `arrsub`.

In [2]:
```
x = arr[:,4]
print(len(x))
print(np.sum(x == 10))

bool_arr = x == 10
arrsub = arr[bool_arr, :]
print(arrsub.shape)

a = arrsub.shape[0]
print(a)
```

```
8000
77
(77, 7)
77
```

Counting rows satisfying a condition using NumPy

- How many rows in `arr` contain the number `40`? You can use `arr == 40` to create a Boolean array. From the resulting Boolean array, you can use the `sum` method, in particular, `.sum(axis=1)`, to count for each row, how many times `40` appears. We want to find all the rows where `40` occurs at least once. Once you have a length-8000 array of `True` and `False` values (with `True` if `40` occurs in the row and with `False` otherwise). You want to count the number of times `True` occurs. Save the result with the variable name `b`. (Reality check: the answer should be between 500 and 700.)

In [3]:
```
bool_arr = (arr == 40).sum(axis=1) > 0
b = bool_arr.sum()
print(b)
```

```
590
```

Submission

- Store these answers, as well as the names of everyone in your group, as a Python dictionary

  ```
  ansdict = {"names": [name1, name2, name3], "a": a, "b": b}
  ```

The `"names"` value must be a list, with length the number of group members, and with entries the full names of your group members. The names of your group members should be strings. For example, `["Chris Davis", "Jinghao Chen"]`.

- Save the resulting dictionary in a pickle file using the following code.

```
with open("wkst2-ans.pickle", "wb") as f:
    pickle.dump(ansdict, f)
```

- Submit this pickle file on Canvas.

In [4]:
```
ansdict = {"names": ["Ilyas"], "a": a, "b": b}
with open("../Data/wkst2-ans.pickle", "wb") as f:
    pickle.dump(ansdict, f)
```