

IMD0033 - Probabilidade

Aula 08 - Introdução a Numpy II

Ivanovitch Silva
Agosto, 2018



Agenda

- Reading CSV files with NumPy
- Boolean Arrays
- Boolean Index
- Assigning values
- Challenge



nyc_taxis.csv



Atualizar o repositório

```
git clone https://github.com/ivanovitchm/imd0033_2018_2.git
```

Ou

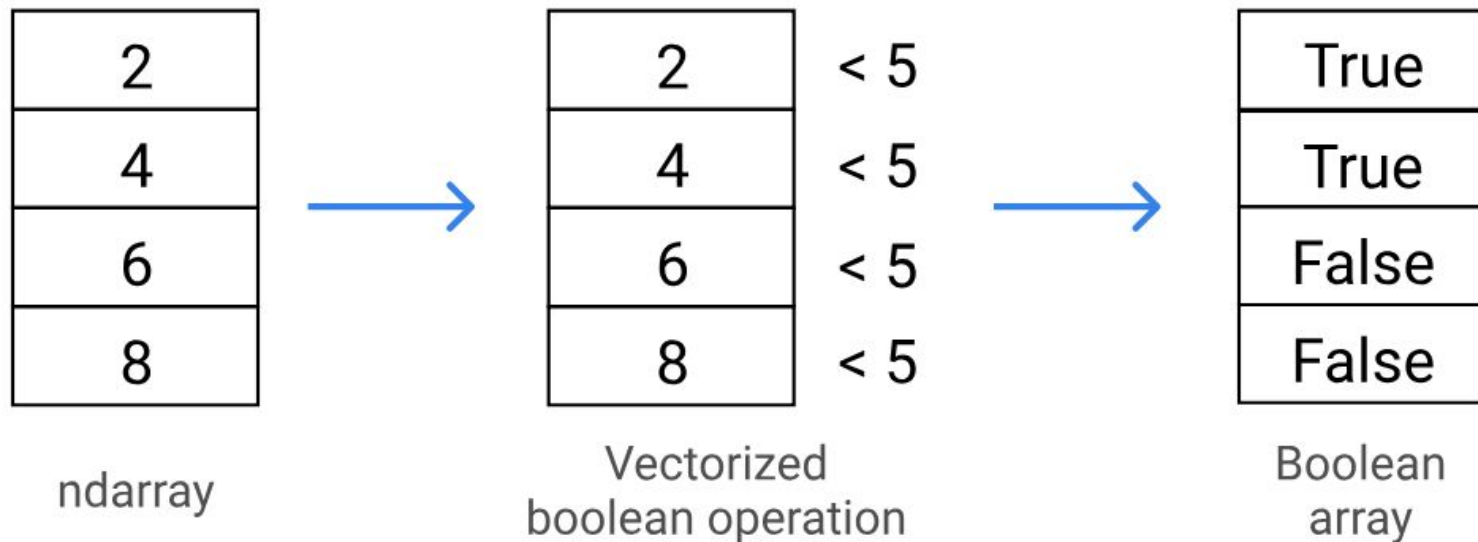
```
git pull
```

Reading CSV files from Numpy

```
taxi = np.genfromtxt('nyc_taxis.csv', delimiter=',')  
print(taxi)
```

```
[[ nan      nan      nan ...,      nan      nan      nan]  
 [ 2016         1         1 ..., 11.65    69.99         1]  
 [ 2016         1         1 ...,      8     54.3         1]  
 ...,  
 [ 2016         6        30 ...,      5    63.34         1]  
 [ 2016         6        30 ...,  8.95    44.75         1]  
 [ 2016         6        30 ...,      0    54.84         2]]
```

Slicing from boolean arrays



Boolean indexing with 1D ndarrays

```
c = np.array([80.0, 103.4,  
              96.9, 200.3])
```

```
c_bool = c > 100
```

80.0
103.4
96.6
200.3

c

False
True
False
True

c_bool

Boolean indexing with 1D ndarrays

```
result = c[c_bool]
```

False		80.0	
True	→	103.4	→ 103.4
False		96.6	↗ 200.3
True	→	200.3	↘

result

Code

```
arr = np.array([
    [ 1,  2,  3],
    [ 4,  5,  6],
    [ 7,  8,  9],
    [10, 11, 12]
])

print(arr)
```

Visualization

1	2	3
4	5	6
7	8	9
10	11	12

Explanation

The original array

```
bool_1 = [True, False,
          True, True]
print(arr[bool_1])
```

1	2	3
4	5	6
7	8	9
10	11	12

`bool_1`'s shape (4) is the same as the shape of `arr`'s first axis (4), so this selects the 1st, 3rd, and 4th rows.

```
print(arr[:,bool_1])
```

1	2	3
4	5	6
7	8	9
10	11	12

`bool_1`'s shape (4) is not the same as the shape of `arr`'s second axis (3), so it can't be used to index and produces an **error**

```
bool_2 = [False, True, True]
print(arr[:,bool_2])
```

1	2	3
4	5	6
7	8	9
10	11	12

`bool_2`'s shape (3) is the same as the shape of `arr`'s second axis (3), so this selects the 2nd and 3rd columns.

Boolean Indexing with 2D ndarrays

Assigning values in 1D ndarray

```
a = np.array(['red', 'blue', 'black', 'blue', 'purple'])  
a[0] = 'orange'  
print(a)
```

```
['orange', 'blue', 'black', 'blue', 'purple']
```

```
a[3:] = 'pink'  
print(a)
```

```
['orange', 'blue', 'black', 'pink', 'pink']
```

Assigning values in 2D ndarray

```
ones = np.array([[1, 1, 1, 1, 1],  
                 [1, 1, 1, 1, 1],  
                 [1, 1, 1, 1, 1]])
```

```
ones[1,2] = 99
```

```
print(ones)
```

```
[[ 1,  1,  1,  1,  1],  
 [ 1,  1, 99,  1,  1],  
 [ 1,  1,  1,  1,  1]]
```

```
ones[0] = 42
```

```
print(ones)
```

```
[[42, 42, 42, 42, 42],  
 [ 1,  1, 99,  1,  1],  
 [ 1,  1,  1,  1,  1]]
```

Assignment Using Boolean Arrays

```
a = np.array([1, 2, 3, 4, 5])
```

1
2
3
4
5

a

```
a[a > 2] = 99
```

False		1		1
False		2		2
True	→	3	→	99
True	→	4	→	99
True	→	5	→	99

a

Assignment Using Boolean Arrays

```
b = np.array([[1, 2, 3],  
              [4, 5, 6],  
              [7, 8, 9]])
```

1	2	3
4	5	6
7	8	9

b

```
b[b > 4] = 99
```

F	F	F		1	2	3		1	2	3
F	T	T	→	4	5	6	→	4	99	99
T	T	T	→	7	8	9	→	99	99	99

b

Assignment Using Boolean Arrays

```
c = np.array([[1, 2, 3],  
              [4, 5, 6],  
              [7, 8, 9]])
```

1	2	3
4	5	6
7	8	9

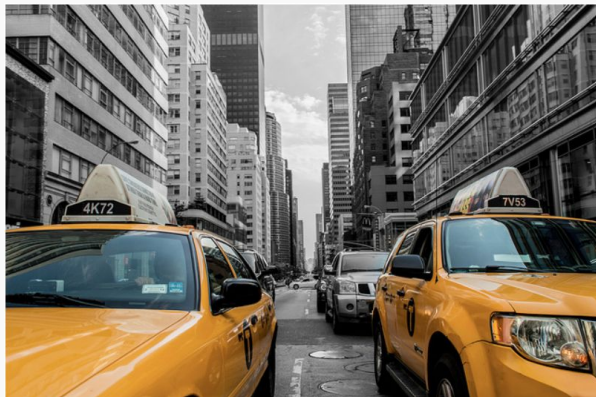
c

```
c[c[:, 1] > 2, 1] = 99
```

	F		→	1	2	3	→	1	2	3
	T		→	4	5	6	→	4	99	6
	T		→	7	8	9	→	7	99	9

c

Challenges



Which is the most popular airport?
Calculating statistics for trip?

Lesson_08_Introduction_to_numpy.ipynb

