bug_categorical_numbers

October 13, 2020

[1]:
```
import sys
sys.version
```

[1]:
```
'3.7.7 (default, Mar 26 2020, 15:48:22) \n[GCC 7.3.0]'
```

[2]:
```
import matplotlib
import matplotlib.pyplot as plt
matplotlib.__version__
```

[2]:
```
'3.3.1'
```

[3]:
```
import numpy as np
np.__version__
```

[3]:
```
'1.19.1'
```

[4]:
```
import pandas as pd
pd.__version__
```

[4]:
```
'1.1.1'
```

[5]:
```
import seaborn as sns
sns.__version__
```

[5]:
```
'0.11.0'
```

[6]:
```
n_points = 50

df = pd.DataFrame({"value": np.random.rand(n_points),
                   "class": np.random.randint(0, 2, n_points),
                   })
df.head()
```

[6]:

	value	class
0	0.081116	1
1	0.092313	0
2	0.329998	1
3	0.752526	0

```
4  0.187332      0
```

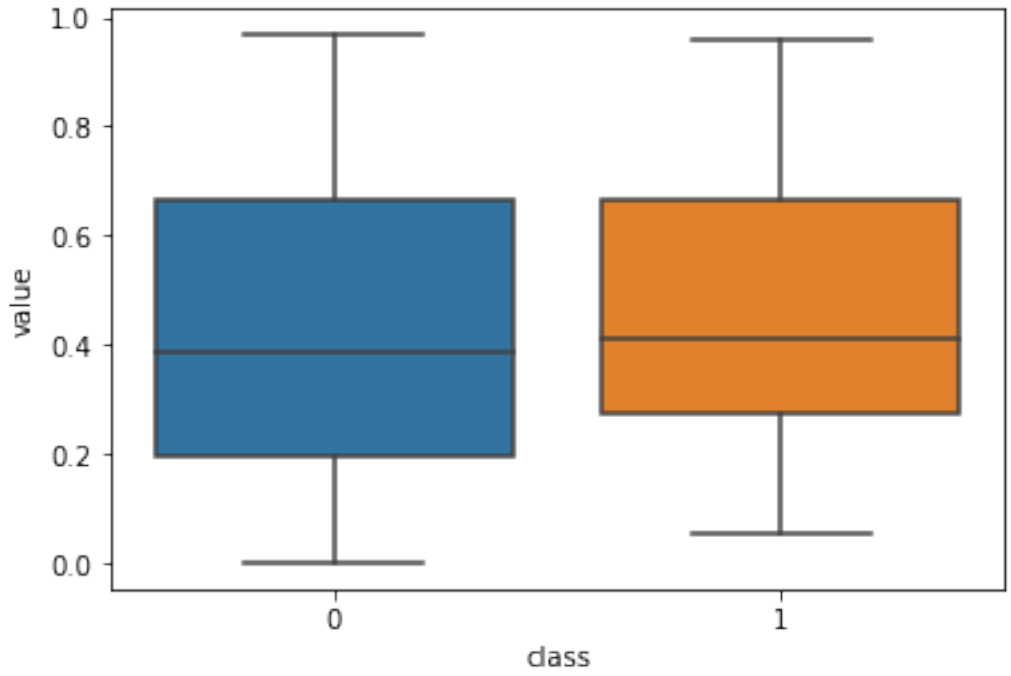
Explicitly set `class` column to categorical data type.

[7]:
```
df["class"] = df["class"].astype("category")
df["class"].dtype
```

[7]:
```
CategoricalDtype(categories=[0, 1], ordered=False)
```

If `class` is passed to `x` argument of a seaborn categorical plotting function, it is correctly treated as a categorical variable with two classes, 0 and 1.

[8]:
```
sns.boxplot(data=df,
            x="class",
            y="value",
            );
```

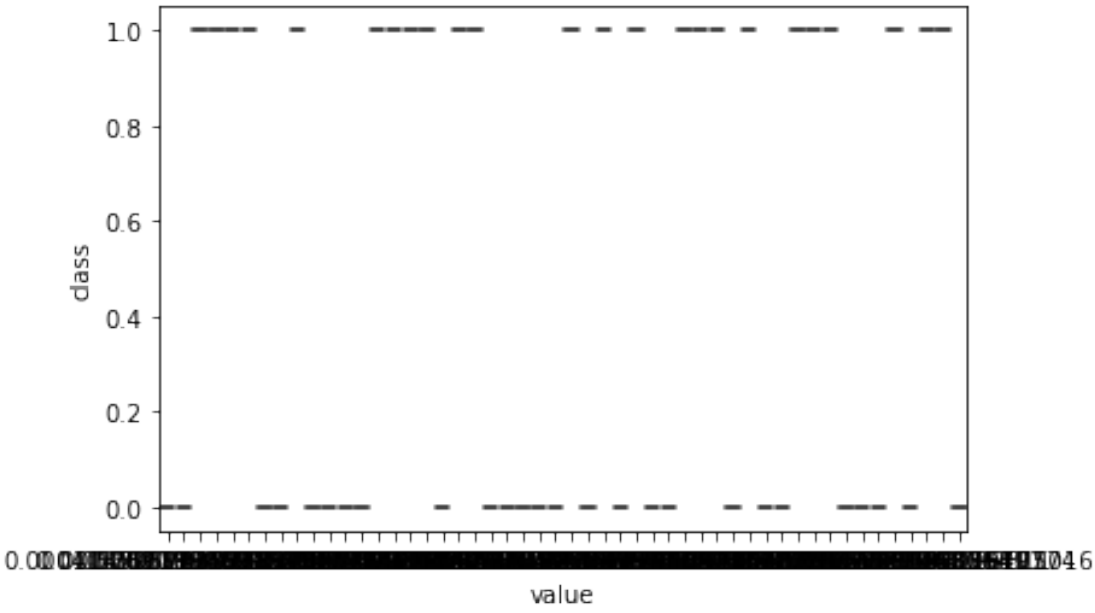


0.0.1 Bug:

If `class` is passed to `y` argument of a seaborn categorical plotting function, it is treated as numerical, and the other variable `value` is treated as categorical!

[9]:
```
sns.boxplot(data=df,
            x="value",
```

```
            y="class",
            );
```



[ ]: