

# Discretize Continuous Attributes

Arnold Jiadong Yu

March 7, 2019

Before we can do approach 1 in part 3, we need to discretize all continuous attributes in the dataset.

1. Age  $\Rightarrow$  Age\_cat with equal-width 10
2. Fnlwgt  $\Rightarrow$  Fnlwgt\_cat with equal-width 100000
3. CaptialG  $\Rightarrow$  CaptialG\_cat with equal-width 9000
4. CaptialL  $\Rightarrow$  CaptialL\_cat with equal-width 450
5. Hours  $\Rightarrow$  Hours\_cat with equal-width 10

We take the input dataset and output 4 modified dataset. They are

1. train\_removeMissing\_cat.csv
2. test\_removeMissing\_cat.csv
3. train\_replaceMissing\_cat.csv
4. test\_replaceMissing\_cat.csv

Instructions on compiling and running program:

IDE: Anaconda-Spyder with python 3.6.

1. Open conToCate.py.
2. Set to current directory (where the conToCate.py is).
3. Execute all the code.