

## - Pandas: 导入数据

```
1 import pandas as pd
2
3 names = ['mpg', 'cylinders', 'displacement', 'horsepower',
4          'weight', 'acceleration', 'model year', 'origin', 'car name']
5
6 # header=None 是否第一行为列名
7 # delim_whitespace=True 数据用空格分隔
8 # na_values='?' ? 表示NaN
9 df = pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-databases/'+
10                  'auto-mpg/auto-mpg.data',
11                  header=None, delim_whitespace=True, names=names, na_values='?')
12
13 # 丢弃无效行
14 df = df.drop('Unnamed: 0', axis=1)
15 # 查看前六行
16 df.head(6)
17
18 # 一些属性
19 # 返回 (行数, 列数)
20 df.shape
21 # 返回列名
22 df.columns
23 df.columns.tolist()
24 # 返回行数索引
25 df.index
26 # 返回每行的数据
27 df.values
28
29 # 选取其中某些列
30 df2 = df[['cylinders', 'horsepower']]
31 # 转化成数组
32 y = np.array(df['mpg'])
33 # 转化成多维数组, 注意这里要用两层 [[]] 包起来! 不知道为啥
34 y = np.array(df[['mpg', 'xxx']])
35
36 # 将列名转化成数组
37 names = df.columns.tolist()
38 names_x = names[0:7]
39 name_y = names[8]
```