读取白酒的数据

In [ ]:
```
import csv

#读取csv文件
f=open('white_wine.csv')
reader=csv.reader(f)
data=[]
for i in reader:
    data.append(i)
f.close()

#展示前5行
print('前5行的数据是')
for i in range(5):
    print(data[i])
```

```
前5行的数据是
['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar', 'chloride
s', 'free sulfur dioxide', 'total sulfur dioxide', 'density', 'pH', 'sulphates',
'alcohol', 'quality']
['7', '0.27', '0.36', '20.7', '0.045', '45', '170', '1.001', '3', '0.45', '8.8',
'6']
['8.1', '0.28', '0.4', '6.9', '0.05', '30', '97', '0.9951', '3.26', '0.44', '10.
1', '6']
['7.2', '0.23', '0.32', '8.5', '0.058', '47', '186', '0.9956', '3.19', '0.4', '9.
9', '6']
['7.2', '0.23', '0.32', '8.5', '0.058', '47', '186', '0.9956', '3.19', '0.4', '9.
9', '6']
```

处理数据

查看白酒有几种品质等级

In [ ]:
```
#就是读取最后一列，看有几种不重复的数

quality_allliat=[]
for i in data[1:]:
    quality_allliat.append(i[-1])
quality_list=set(quality_allliat)

print("白酒品质等级有%d种，分别是%r"%(len(quality_list),quality_list))
```

```
白酒品质等级有7种，分别是{'4', '5', '8', '6', '3', '7', '9'}
```

统计种品质等级的数量

In [ ]:
```
quality_nums={}
for i in quality_list:
    quality_nums[str(i)]=0

for i in data[1:]:
    quality_nums[str(i[-1])]+=1
print(quality_nums)
```

```
{'4': 115, '5': 1020, '8': 123, '6': 1539, '3': 14, '7': 616, '9': 4}
```

计算每个数据集中fixed acidity的均值

In [ ]:
```
fixed_acidity_total={}
for i in quality_list:
    fixed_acidity_total[str(i)]=0
for i in data[1:]:
    fixed_acidity_total[str(i[-1])]+=float(i[0])
fixed_acidity_average={}
for i in fixed_acidity_total:
    fixed_acidity_average[i]=fixed_acidity_total[i]/quality_nums[i]
print(fixed_acidity_average)
```

```
{'4': 7.052173913043476, '5': 6.907843137254891, '8': 6.708130081300811, '6': 6.81
2085769980511, '3': 7.535714285714286, '7': 6.755844155844158, '9': 7.5}
```

总结

通过本次实验，学习了如何使用csv库读取csv格式的文件，同时也学会了python的循环结构。 学会了使用append向数组中添加元素，也学会了用字典来记录不同类型的数据的值。 学会了使用set方法来去除元素中重复的数值。 学会了基本的数据统计方法，通过循环遍历数组，来得到我们所需要的求和数据。