

Phân tích rượu vang

Sử dụng NumPy để phân tích dữ liệu về chất lượng rượu vang. Dữ liệu chứa thông tin về các thuộc tính khác nhau của rượu vang, chẳng hạn như độ pH và độ axit cố định, cùng với điểm chất lượng từ 0 đến 10 cho mỗi loại rượu. Điểm chất lượng là điểm trung bình của ít nhất 3 người kiểm tra (bằng cách nếm rượu)

- "fixed acidity";"volatile acidity";"citric acid";"residual sugar";"chlorides";"free sulfur dioxide";
- "total sulfur dioxide";"density";"pH";"sulphates";"alcohol";"quality"

Yêu cầu:

1. Đọc tập tin "winequality-red.csv" ra mảng wines
2. Xem thông tin sơ bộ như: shape, xem 5 dòng đầu, xem 5 dòng cuối
3. In 5 dòng đầu, 5 dòng cuối gồm các cột "fixed acidity", "pH" và "quality"
4. Tạo wines_4 là mảng con trích ra từ mảng wines gồm 4 cột: "fixed acidity", "pH", "alcohol" và "quality"
5. Tính trung bình các cột trong mảng wines_4
6. Tính trung bình, min và max của quality
7. Cho biết các rượu có điểm > 5
8. Cho biết 5 rượu đầu tiên có điểm lớn nhất và có bao nhiêu rượu như vậy?
9. Cho biết 3 rượu đầu tiên có điểm >= 7
10. Cho biết các loại rượu có điểm > 7 và độ cồn > 10. Chỉ in ra 2 cột: "alcohol" và "quality"
11. Đọc tập tin "winequality-white.csv" ra mảng white_wines
12. Kết hợp wines và white_wines thành mảng all_wines
13. In ra 5 dòng đầu, 5 dòng cuối của mảng all_wines

```
In [1]: import numpy as np
```

Đọc dữ liệu

```
In [2]: # 1. Đọc tập tin rượu vang đỏ
wines = np.genfromtxt("data/winequality-red.csv", delimiter=";", skip_header=1)
```

Xem thông tin


```
In [3]: # 2. Xem thông tin sơ bộ như: shape, xem 5 dòng đầu, xem 5 dòng cuối
wines.shape

# "fixed acidity"; "volatile acidity"; "citric acid"; "residual sugar";
# "chlorides"; "free sulfur dioxide"; "total sulfur dioxide"; "density"; "pH"; "sulphates";
# "alcohol"; "quality"

# Lưu ý: các cột "fixed acidity", "pH", "alcohol" và "quality"
# (độ chua cố định, độ PH, độ cồn và điểm đánh giá)
```

Out[3]: (1599, 12)

```
In [4]: wines[:5].tolist()
```

```
Out[4]: [[7.4, 0.7, 0.0, 1.9, 0.076, 11.0, 34.0, 0.9978, 3.51, 0.56, 9.4, 5.0],
 [7.8, 0.88, 0.0, 2.6, 0.098, 25.0, 67.0, 0.9968, 3.2, 0.68, 9.8, 5.0],
 [7.8, 0.76, 0.04, 2.3, 0.092, 15.0, 54.0, 0.997, 3.26, 0.65, 9.8, 5.0],
 [11.2, 0.28, 0.56, 1.9, 0.075, 17.0, 60.0, 0.998, 3.16, 0.58, 9.8, 6.0],
 [7.4, 0.7, 0.0, 1.9, 0.076, 11.0, 34.0, 0.9978, 3.51, 0.56, 9.4, 5.0]]
```

```
In [5]: wines[-5:].tolist()
```

```
Out[5]: [[6.2, 0.6, 0.08, 2.0, 0.09, 32.0, 44.0, 0.9949, 3.45, 0.58, 10.5, 5.0],
 [5.9, 0.55, 0.1, 2.2, 0.062, 39.0, 51.0, 0.99512, 3.52, 0.76, 11.2, 6.0],
 [6.3, 0.51, 0.13, 2.3, 0.076, 29.0, 40.0, 0.99574, 3.42, 0.75, 11.0, 6.0],
 [5.9, 0.645, 0.12, 2.0, 0.075, 32.0, 44.0, 0.99547, 3.57, 0.71, 10.2, 5.0],
 [6.0, 0.31, 0.47, 3.6, 0.067, 18.0, 42.0, 0.99549, 3.39, 0.66, 11.0, 6.0]]
```

```
In [6]: # 3. In 5 dòng đầu, 5 dòng cuối gồm các cột "fixed acidity", "pH" và "quality"
# In 5 dòng đầu
wines[:5, [0,8,11]]
```

```
Out[6]: array([[ 7.4 ,  3.51,  5.  ],
 [ 7.8 ,  3.2 ,  5.  ],
 [ 7.8 ,  3.26,  5.  ],
 [11.2 ,  3.16,  6.  ],
 [ 7.4 ,  3.51,  5.  ]])
```

```
In [7]: # In 5 dòng cuối
# code here
```

Nhấn vào đây để xem kết quả!

```
array([[6.2 ,  3.45,  5.  ],
 [5.9 ,  3.52,  6.  ],
 [6.3 ,  3.42,  6.  ],
 [5.9 ,  3.57,  5.  ],
 [6.  ,  3.39,  6.  ]])
```

```
In [8]: # 4. Tạo wines_4 là mảng trích ra từ mảng wines gồm 4 cột:
# "fixed acidity", "pH", "alcohol" và "quality"
wines_4 = wines[:, [0,8,10,11]]
```



```
In [9]: wines_4
```

```
Out[9]: array([[ 7.4 ,  3.51,  9.4 ,  5.  ],
               [ 7.8 ,  3.2 ,  9.8 ,  5.  ],
               [ 7.8 ,  3.26,  9.8 ,  5.  ],
               ...,
               [ 6.3 ,  3.42, 11.  ,  6.  ],
               [ 5.9 ,  3.57, 10.2 ,  5.  ],
               [ 6.  ,  3.39, 11.  ,  6.  ]])
```

Tính toán thống kê

```
In [10]: # 5. Tính trung bình các cột trong mảng wines_4
wines_4.mean(axis=0)
```

```
Out[10]: array([ 8.31963727,  3.3111132 , 10.42298311,  5.63602251])
```

```
In [11]: # 6. Tính trung bình, min và max của quality
# code here
```

Nhấn vào đây để xem kết quả!

Điểm trung bình là: 5.64

Điểm nhỏ nhất là: 3.0

Điểm lớn nhất là: 8.0

Rút trích theo điều kiện

```
In [12]: # 7. Cho biết các rượu có điểm > 5
wines[wines[:, -1] > 5]
```

```
Out[12]: array([[11.2 ,  0.28,  0.56, ...,  0.58,  9.8 ,  6.  ],
               [ 7.3 ,  0.65,  0.  , ...,  0.47, 10.  ,  7.  ],
               [ 7.8 ,  0.58,  0.02, ...,  0.57,  9.5 ,  7.  ],
               ...,
               [ 5.9 ,  0.55,  0.1 , ...,  0.76, 11.2 ,  6.  ],
               [ 6.3 ,  0.51,  0.13, ...,  0.75, 11.  ,  6.  ],
               [ 6.  ,  0.31,  0.47, ...,  0.66, 11.  ,  6.  ]])
```

```
In [13]: # 8. Cho biết 5 rượu đầu tiên có điểm lớn nhất và có bao nhiêu rượu như vậy?
# code here
```

Nhấn vào đây để xem kết quả!

```
[[7.9, 0.35, 0.46, 3.6, 0.078, 15.0, 37.0, 0.9973, 3.35, 0.86, 12.8, 8.0],
 [10.3, 0.32, 0.45, 6.4, 0.073, 5.0, 13.0, 0.9976, 3.23, 0.82, 12.6, 8.0],
 [5.6, 0.85, 0.05, 1.4, 0.045, 12.0, 88.0, 0.9924, 3.56, 0.82, 12.9, 8.0],
 [12.6, 0.31, 0.72, 2.2, 0.072, 6.0, 29.0, 0.9987, 2.88, 0.82, 9.8, 8.0],
 [11.3, 0.62, 0.67, 5.2, 0.086, 6.0, 19.0, 0.9988, 3.22, 0.69, 13.4, 8.0]]
```



```
In [14]: wines[wines[:, -1] == wines[:, -1].max()].shape[0]
```

```
Out[14]: 18
```

```
In [15]: # 9. Cho biết 3 rượu đầu tiên có điểm >= 7  
# code here
```

Nhấn vào đây để xem kết quả!

```
[[7.3, 0.65, 0.0, 1.2, 0.065, 15.0, 21.0, 0.9946, 3.39, 0.47, 10.0, 7.0],  
 [7.8, 0.58, 0.02, 2.0, 0.073, 9.0, 18.0, 0.9968, 3.36, 0.57, 9.5, 7.0],  
 [8.5, 0.28, 0.56, 1.8, 0.092, 35.0, 103.0, 0.9969, 3.3, 0.75, 10.5, 7.0]]
```

```
In [16]: # 10. Cho biết các loại rượu có điểm > 7 và độ cồn > 10.  
# In ra 2 cột: "alcohol" và "quality"  
# code here
```

Nhấn vào đây để xem kết quả!

```
array([[12.8, 8. ],  
       [12.6, 8. ],  
       [12.9, 8. ],  
       [13.4, 8. ],  
       [11.7, 8. ],  
       [11. , 8. ],  
       [11. , 8. ],  
       [14. , 8. ],  
       [12.7, 8. ],  
       [12.5, 8. ],  
       [11.8, 8. ],  
       [13.1, 8. ],  
       [11.7, 8. ],  
       [14. , 8. ],  
       [11.3, 8. ],  
       [11.4, 8. ]])
```

Kết hợp

```
In [17]: # 11. Đọc tập tin rượu vang trắng  
white_wines = np.genfromtxt("data/winequality-white.csv", delimiter=";",  
                             skip_header=1)
```

```
In [18]: wines.shape
```

```
Out[18]: (1599, 12)
```



```
In [19]: # 12. Kết hợp wines và white_wines
all_wines = np.concatenate([wines, white_wines])
all_wines.shape
```

```
Out[19]: (6497, 12)
```

```
In [20]: # 13. In ra 5 dòng đầu, 5 dòng cuối của mảng all_wines
all_wines[:5].tolist()
```

```
Out[20]: [[7.4, 0.7, 0.0, 1.9, 0.076, 11.0, 34.0, 0.9978, 3.51, 0.56, 9.4, 5.0],
 [7.8, 0.88, 0.0, 2.6, 0.098, 25.0, 67.0, 0.9968, 3.2, 0.68, 9.8, 5.0],
 [7.8, 0.76, 0.04, 2.3, 0.092, 15.0, 54.0, 0.997, 3.26, 0.65, 9.8, 5.0],
 [11.2, 0.28, 0.56, 1.9, 0.075, 17.0, 60.0, 0.998, 3.16, 0.58, 9.8, 6.0],
 [7.4, 0.7, 0.0, 1.9, 0.076, 11.0, 34.0, 0.9978, 3.51, 0.56, 9.4, 5.0]]
```

```
In [21]: all_wines[-5:].tolist()
```

```
Out[21]: [[6.2, 0.21, 0.29, 1.6, 0.039, 24.0, 92.0, 0.99114, 3.27, 0.5, 11.2, 6.0],
 [6.6, 0.32, 0.36, 8.0, 0.047, 57.0, 168.0, 0.9949, 3.15, 0.46, 9.6, 5.0],
 [6.5, 0.24, 0.19, 1.2, 0.041, 30.0, 111.0, 0.99254, 2.99, 0.46, 9.4, 6.0],
 [5.5, 0.29, 0.3, 1.1, 0.022, 20.0, 110.0, 0.98869, 3.34, 0.38, 12.8, 7.0],
 [6.0, 0.21, 0.38, 0.8, 0.02, 22.0, 98.0, 0.98941, 3.26, 0.32, 11.8, 6.0]]
```

```
In [ ]:
```

