

Đồ án 3: Linear regression

Nội dung đồ án

File "wine.csv" là cơ sở dữ liệu đánh giá chất lượng của 1200 chai rượu vang theo thang điểm 1 - 10 dựa trên 11 tính chất khác nhau.

```
In [35]: import pandas as pd
df = pd.read_csv('D:/Data/wine.csv', sep=';')
```

```
In [36]: df
```

Out[36]:

| | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | density | pH | sulphates | alcoh |
|------|------------------|---------------------|----------------|-------------------|-----------|---------------------------|----------------------------|---------|------|-----------|-------|
| 0 | 7.4 | 0.700 | 0.00 | 1.9 | 0.076 | 11.0 | 34 | 0.99780 | 3.51 | 0.56 | 9 |
| 1 | 7.8 | 0.880 | 0.00 | 2.6 | 0.098 | 25.0 | 67 | 0.99680 | 3.20 | 0.68 | 9 |
| 2 | 7.8 | 0.760 | 0.04 | 2.3 | 0.092 | 15.0 | 54 | 0.99700 | 3.26 | 0.65 | 9 |
| 3 | 11.2 | 0.280 | 0.56 | 1.9 | 0.075 | 17.0 | 60 | 0.99800 | 3.16 | 0.58 | 9 |
| 4 | 7.4 | 0.700 | 0.00 | 1.9 | 0.076 | 11.0 | 34 | 0.99780 | 3.51 | 0.56 | 9 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1194 | 7.0 | 0.745 | 0.12 | 1.8 | 0.114 | 15.0 | 64 | 0.99588 | 3.22 | 0.59 | 9 |
| 1195 | 6.2 | 0.430 | 0.22 | 1.8 | 0.078 | 21.0 | 56 | 0.99633 | 3.52 | 0.60 | 9 |
| 1196 | 7.9 | 0.580 | 0.23 | 2.3 | 0.076 | 23.0 | 94 | 0.99686 | 3.21 | 0.58 | 9 |
| 1197 | 7.7 | 0.570 | 0.21 | 1.5 | 0.069 | 4.0 | 9 | 0.99458 | 3.16 | 0.54 | 9 |
| 1198 | 7.7 | 0.260 | 0.26 | 2.0 | 0.052 | 19.0 | 77 | 0.99510 | 3.15 | 0.79 | 10 |

1199 rows × 12 columns

Xây dựng mô hình đánh giá chất lượng rượu sử dụng phương pháp hồi quy tuyến tính.

- Sử dụng toàn bộ 11 đặc trưng đề bài cung cấp.
- Sử dụng duy nhất 1 đặc trưng cho kết quả tốt nhất.
- Xây dựng một mô hình của riêng bạn cho kết quả tốt nhất.

```
In [39]: df.quality
```

```
Out[39]: 0      5
          1      5
          2      5
          3      6
          4      5
          ..
        1194     6
        1195     6
        1196     6
        1197     6
        1198     6
        Name: quality, Length: 1199, dtype: int64
```

```
In [40]: df.alcohol
```

```
Out[40]: 0      9.4
          1      9.8
          2      9.8
          3      9.8
          4      9.4
          ...
        1194     9.5
        1195     9.5
        1196     9.5
        1197     9.8
        1198    10.9
        Name: alcohol, Length: 1199, dtype: float64
```

```
In [41]: import numpy as np
         a=np.array(df.alcohol)
```

```
In [42]: a
```

```
Out[42]: array([ 9.4,  9.8,  9.8, ...,  9.5,  9.8, 10.9])
```

Quy định bài nộp

- Thực hiện toàn bộ bài làm trên 1 tập tin Jupyter Notebook (.ipynb) hoặc Python (.py)
- File bài nộp chứa các tập tin sau:
 1. Báo cáo toàn bộ bài làm: MSSV.pdf
 2. Mã nguồn: MSSV.ipynb hoặc MSSV.py
- Trong đó, nội dung tập tin báo cáo gồm có:
 - Thông tin cá nhân: họ và tên, MSSV
 - Liệt kê các chức năng đã hoàn thành
 - Ý tưởng thực hiện, mô tả các hàm chức năng
 - Hình ảnh kết quả với từng chức năng